COHORT ANALYSIS AND REPORTING FOR GRADUATE ATTRIBUTE ASSESSMENT

Aneta George and Liam Peyton
University of Ottawa, 800 King Edward, Ottawa, ON, Canada K1N 6N5
atraikov@uottawa.ca, lpeyton@uottawa.ca

Abstract - The Graduate Attribute Information Analysis system (GAIA) was developed at the University of Ottawa to support data collection and performance management of graduate attributes for engineering programs at the program level and at the course level [10]. This paper reports on our research to develop support for cohort analysis and reporting by providing a single consistent view of graduate attributes (GA) and performance indicators for groups of students who started and finished an engineering program at the same time. This is supported by two special purpose reports: Graduate Attribute Report per Cohort (GAR/C) and Course Progression Report per Cohort (CPR/C). The former shows average GA data per attribute, the latter tracks student achievement as students progress in their program. It also adds to the historic data trend analysis for a program. Furthermore, a COOP Progress Report per cohort (COOPR/C) is generated.

Keywords: Graduate attributes, cohort analysis, CEAB, accreditation.

1. INTRODUCTION

The Graduate Attribute Information Analysis system (GAIA) was developed at the University of Ottawa to support data collection and performance management of graduate attributes for engineering programs at the program level and at the course level [10]. This paper reports on our research to develop support for cohort analysis and reporting. Cohort analysis and reporting entails providing a single consistent view of graduate attributes and performance indicators for a particular group of students who started and finished an engineering program at the same time (e.g. “class of 2018”).

In the School of Electrical Engineering and Computer Science we have collected data from three engineering programs (electrical, computer and software) into a single graduate attribute (GA) database since 2013. Implementing GAIA involved a review of existing data. The data was restructured and sorted into categories simplifying reporting for accreditation and allowing historic trend analyses per year, per graduate attribute, per program and per course. GAIA was extended to process and report COOP data in 2015, and in 2016 it was revised to support common key performance indicators and a common continuous improvement process across all three engineering programs. More recently, a comprehensive review of existing algorithms leads to the development of data transformations to support reporting and comparison per cohort as well. To minimize the stress accompanying implementation of the new feature, we accommodated the UK Engineering Council principal “to make every effort to align accreditation requirements with those already in place for internal management and quality assurance” [9, p.5].

1.1. The Problem

In its draft version of “A Guide to Outcomes-based Criteria”, the CEAB refers to cohort as a synonym for group [3]. The document specifies that a program has to demonstrate “measured student performance (as a group or cohort) in respect of each attribute” [3, p.14]. In the Accreditation Criteria and Procedures 2017 document, CEAB steps away from the analogy and uses cohort for interpretive statements [4, p.80]. Furthermore, it adds the requirement for curriculum maps to be supplemented by “indication of sizes of the primary student cohort comprising the graduate class” [4, p.81].

Looking for practical implementation of the supplementary cohort data requirement, we explored related experiences of Canadian engineering schools undergoing accreditation [5, 7]. We performed a secondary search within countries members of Washington Accord [11, 12, 13] and CETA [9], the most recent economic and trade agreement between Canada and EU which involves a framework for mutual recognition of professional engineering qualifications [8]. Here is a list of successful implementations of cohort-based analysis on GA assessment data:

- We identified several Canadian studies reporting on the importance of cohort
reports for program improvement [5, 7].

- Cohort based GA analysis and their use on a larger scale, for providing data on different aspects of the important transition from the university to the labor market, is reported to be successfully implemented in ABET accredited universities in the Middle East [1, 2].
- Several Australian researchers pointed to successful use of data analysis per cohort for development of a systematic approach towards reporting on GAs [12, 13].
- International experience indicated that the use of a cross-country GA database for GA achievement comparison between programs per cohort and between cohorts within a program provides an opportunity to look at the GAs through the “lens of employability besides addressing concerns about unemployment” [2].

Despite the positive outcomes, implementation of long term cohort analysis remains limited. We were able to identify that the main reasons behind the low popularity of cohort analysis are mainly related to data complexity, volume, management and accessibility. Issues related to the actual process of collecting assessment data per cohort were repeatedly reported as accompanied by low interoperability of existing management systems in terms of access and compatibility [5, 7, 11].

Addressing those gaps, we performed a comprehensive review of our existing extract, transform and load (ETL) processes for data management and pre-processing. This led to the development of new transformations to support reporting and comparison per cohort in GAIA. The techniques applied by Catalanian researchers to compare cross-cohort performance of a particular component (GA) against several programs [6, p.114] are similar to the ones used by us in our new transformations to support cohort analysis. The GAIA implementation enables cohort-based comparison of a particular GA measurement using both cumulative mean and standard deviation.

1.2. Literature Review

In 2008 the Australian Council of Engineering Deans performed a consultative review of the engineering education system and reported that demand for engineering graduates was continuously exceeding engineering graduate supply [12]. It acknowledged the role of accreditation and resulting curriculum improvements and reforms, and it examined “the ability of the Australian engineering education system to meet future challenges” referring to issues like students “attrition from engineering programs” and “uneven exposure” to contemporary practice [12]. The issue of tracking graduate outcomes along the course of study is raised.

One of the first solutions to support cohorts, the online system ‘ReView’, is described by Thompson et al [13]. Although the study explores the tool’s application in Australian business education, the system follows a generic process applicable to any university program. The main advantage of ‘ReView’ is its ability to cluster assessment criteria into GAs making explicit alignment of GA development across courses and program levels for a cohort performance.

A mechanism behind the development and use of cohort assessment analysis for two entry years is discussed by Godfrey, Aubrey and King [11]. The study highlights the importance of the ability to access and extract data from different learning management systems and the difficulties that lead to data entry errors in the process of gathering the required database.

At home, overcoming the challenges in adopting existing assessment tools is described by Chong and Romkey [5]. The authors use GA performance per cohort and identify the balance between respective GA components, multiple evaluators, maintaining rubric usability, and validity as the most challenging tasks.

Dupuis and St-Pierre explore the use of built-in data analysis algorithms to report on expected and actual cohort performance on six graduate attributes by the end of a 4-year engineering program. The authors identify the ability for the scoring rubric to capture students’ performance as learners evolve along the program as its main advantage. Furthermore, the study highlights the importance of capturing data for processing in developing an efficient feedback tool [7].

In ABET-accredited universities from outside the U.S., Aoudia, Marji and AlQiwhsi present a new approach assessing continuous quality improvement based on cohort analysis [1]. They view an engineering program as a closed-loop control system where inputs (entry year students), processes (training), outputs (end-year students/graduates) and feedbacks (statistical data gathered by employers acting as external customers, and instructors as internal customers) provide cohort-based assessment data that support analysis and the use of reports to inform program improvement [1]. Informing on compatible tools of assessment is listed as an advantage of the approach. The authors view students as the “real subject of assessment” and identify the lack of support for students as its major flaw.

In recent research Belwal, Priyadarshi and Al Fazari present a model for measuring GA and employability skills based on cohort analysis implemented in Oman. In their research the authors use an online model of data collected for three cohorts of graduates [2].
1.3. Motivation

In 2010, Corominas, Saurina and Villar perform a study on higher education and graduate employment using a very large sample of three cohorts involving 30000 graduates in total, 10000 per cohort. The study takes place in Catalonia and involves data provided by the Catalan public universities. It is one of the largest databases of this kind in Europe and supported various studies mainly related to graduate employment [6]. The results are differentiated into measuring achievement for the following components: theoretical learning, practical learning, written expression, oral expression, teamwork, leadership, problem solving, decision-making, critical thinking, creativity, administration, documentation skills, languages and computing skills.

Analyses are performed at different levels and categories. For example, reports against mean (\(\bar{x}\)) and standard deviation (\(SD\)) for the acquired level of competence-based learning and Student \(t\)-test for comparison according to job match [6, p.108]; correlation between competence-based learning achievement and job usefulness according to field of study (theoretical learning, practical learning and written expression); rating of competency/job usefulness (based on achievement against oral expression, teamwork and leadership, problem solving, decision making and critical thinking components). We found this approach to cross-cohort data analysis inspiring and looked to adapt it for cohort-analysis of engineering programs in terms of the 12 CEAB GAs.

1.4. Methodology

The techniques applied by Catalonian researchers to compare cross-cohort performance of a particular component (GA) against several programs [6, p.114] are similar to the ones used by us in the development of new ETL transformations in GAIA to support cohort analysis. We developed an analysis framework allowing for horizontal comparison across the three programs within the School of Electrical Engineering and Computer Science: software engineering (SEG), electrical engineering (ELG) and computer engineering (CEG). Figure 1 illustrates the comparison of GA measurements on different bases as follows:

a) Comparison based on GA cumulative mean assessment data per cohort (\(\bar{x}\)):
- Vertical comparison per year (from entry to final year of the program) indicates the extent to which different cohorts acquire skills related to particular GA;
- Horizontal comparison of selected GA achievement across programs within a selected cohort is used to identify successes in acquiring a particular GA. It indicates comparative improvement achieved by a program and leads to shared methodology at faculty level. Alternatively, a report of this type can indicate an unexpected drop in particular GA achievement.

b) Comparison based on standard deviation data per cohort (\(SD\)). The use of \(SD\) as a measure of dispersion over inter-quartile range is due to the fact that \(SD\) takes into account every variable (student) from the data set.
- Vertical comparison per cohort within a program indicates the dispersion of GA data across the respective GA mean value thus relates to the consistency of acquiring particular skill denoted by that indicator;
- Horizontal comparison provides similar analysis between programs within a selected cohort;

c) Comparison based on enrollment data represented by a variable (n). (n) indicates the number of students in a cohort and allows for:
- Assigning weight of a respective GA across cohorts within a program (vertical analysis);
- Assigning GA weight across programs within a cohort (horizontal analysis).

![Fig. 1. Multiple GAs cohort analysis.](image)

A data transformation specifically developed to perform cohort data management within GAIA is supported by a mathematical model, shown in Fig. 2. It modifies related data entered by users into \(n\), \(\bar{x}\) (calculated using population mean) and \(SD\) (calculated using the standard deviation \(\sigma\)) as a measure of the spread/variability of the assessment scores per GA. The symbol ‘\(\mu\)’ represents the population mean. The symbol ‘\(\sum X_i\)’ represents the sum of all scores present in the population, in this case \(X_n\), where \(1 \leq i \leq n\). The symbol ‘\(n\)’ represents the total number of students for a given cohort in a particular program for a particular year.
of the program. The symbol Y represents the total number of courses for a given GA. The data transformation allows for generating informative reports per cohort showing clustery or dispersion of data per program measured against selected GA(s). Cohort data management transformation illustrated in Fig. 2 demonstrates the use of selected GA data from three consecutive years in the same program.

Figure 3 illustrates the ETL processes within GAIA’s data warehouse that support cohort analysis. Raw data is loaded from a variety of academic and non-academic sources into the GAA database. RDBMS and mathematical modulations illustrated in Fig. 2 support the transformation process and loading the data into the respective target data marts (RC1, RC2 and RC3).

Data behind the variable $n$ can be used to demonstrate the extent to which a particular program contributes to developing skills associated with allocated attribute in terms of percentage graduates for particular cohort. For example, applied vertically per cohort, $n$ indicates increase/decrease of applicants’ interest in particular program. Thus, setting enrollment limits or increasing promotion of a related engineering program can be correlated to the employment markets current demand.

The theoretical model described above is illustrated in Figure 4 for a single GA cohort analysis. It is implemented per cohort, for a 5-year program, over $P$ number of programs, reflecting achievement for GA 1. System algorithms support extension of the model across all twelve CEAB graduate attributes as illustrated in Figure 5 where $C$ represent the number of cohorts. GAIA’s database collates assessment data since 2013 and is programmed to accumulate additional data till 2020 when the next EECS accreditation is to take place.
Graduate Attribute Report per Cohort (GAR/C) is illustrated in Table 3 below. It displays average student achievement per graduate attribute and follows a complete cohort cycle per program. Meeting accreditation requirements is reported as per CEAB 60-80 scale where overall achievement below 60% indicates failing, data within the range 60%-80% indicates meeting and data above 80% indicates exceeding accreditation requirements.

Table 3: Graduate Attribute Report per Cohort (GAR/C).

<table>
<thead>
<tr>
<th>Graduate Attribute</th>
<th>COURSES</th>
<th>Meeting Accreditation Requirements</th>
<th>GA Achievement Data in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA1</td>
<td>SEG1101</td>
<td>Exceeded</td>
<td>84.0</td>
</tr>
<tr>
<td>GA2</td>
<td>SEG3101 SEG4111</td>
<td>Met</td>
<td>63.1</td>
</tr>
<tr>
<td>GA3</td>
<td>SEG3103</td>
<td>Met</td>
<td>76.9</td>
</tr>
<tr>
<td>GA4</td>
<td>SEG4911</td>
<td>Met</td>
<td>55.1</td>
</tr>
<tr>
<td>GA5</td>
<td>SEG2105 SEG3105 SEG4111</td>
<td>Met</td>
<td>82.3</td>
</tr>
<tr>
<td>GA6</td>
<td>SEG4911 SEG2911</td>
<td>Exceeded</td>
<td>87.2</td>
</tr>
<tr>
<td>GA7</td>
<td>SEG4111 SEG1911</td>
<td>Exceeded</td>
<td>80.4</td>
</tr>
<tr>
<td>GA8</td>
<td>SEG2911</td>
<td>Met</td>
<td>60.0</td>
</tr>
<tr>
<td>GA9</td>
<td>SEG2911</td>
<td>Met</td>
<td>70.2</td>
</tr>
<tr>
<td>GA10</td>
<td>SEG4105 SEG1911</td>
<td>Exceeded</td>
<td>86.3</td>
</tr>
</tbody>
</table>

Course color codes indicate the relative level of the course within the program as follows:
- red shading is used for Year 5 courses
- grey shading is used for Year 4 courses
- peach shading is used for Year 3 courses
- green shading is used for Year 2 courses
- blue shading is used for Year 1 courses

Course Progression Report per Cohort (CPR/C) tracks student achievement as a student cohort progresses in their program. It uses a summative average GA achievement per year and allows for comparison of GAA against end of year average course assessment data. Information provided is used for follow-up on assessment data consistency and assessment techniques accuracy. A sample CPR/C report using sample data is illustrated in Table 4.

Table 4: Course Progression Report / Cohort (CPR/C).

<table>
<thead>
<tr>
<th>Cohort Progression (years)</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course Average</td>
<td>80.5</td>
<td>83.8</td>
<td>75.3</td>
<td>84.7</td>
<td>91.4</td>
</tr>
<tr>
<td>GA Average</td>
<td>80.4</td>
<td>72.6</td>
<td>68.0</td>
<td>82.4</td>
<td>84.6</td>
</tr>
<tr>
<td>Courses</td>
<td>SEG1911 SEG2105 SEG2106 SEG3101 SEG4111</td>
<td>SEG3103 SEG2911 SEG3102 SEG4105 SEG4145</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data from both cohort reports are used for comparison allowing for evaluation within a process of continuous improvement and providing a historic data trend for further curriculum development.

To illustrate the mechanism of data sorting behind the two types of cohort reports, we generated random data that is being used for demonstration purposes in this paper in all tables and graphs. To follow the course
progression within a cohort GAIA combines a historic trend data per course according to the course sequence of the specific engineering program (i.e. software engineering in this sample). To generate GAR/C GAIA uses algorithms that select and feed in the respective data into the required cell from the report template. A simplified procedure for generating a report for sample cohort Y requires combination of cohort data from \((Y - k)\) years, where \(k = 1, 2, 3, 4\) and their corresponding graduate attributes. A sample report developed from this data arrangement is shown in Table 5 below.

**Table 5: Use of historic trend data for generating cohort reports (GAR/C and CPR/C).**

<table>
<thead>
<tr>
<th>Cohort (Year)</th>
<th>Course</th>
<th>Graduate Attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA Data (Y)</td>
<td>SEG4911</td>
<td>2, 4, 6, 7, 8, 12</td>
</tr>
<tr>
<td></td>
<td>SEG4105</td>
<td>6, 11</td>
</tr>
<tr>
<td>GA Data (Y-1)</td>
<td>SEG3101</td>
<td>1, 2, 5</td>
</tr>
<tr>
<td></td>
<td>SEG3102</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>SEG4145</td>
<td>11</td>
</tr>
<tr>
<td>GA Data (Y-2)</td>
<td>SEG2106</td>
<td>xxx</td>
</tr>
<tr>
<td></td>
<td>SEG2105</td>
<td>7, 9, 10</td>
</tr>
<tr>
<td>GA Data (Y-3)</td>
<td>SEG105</td>
<td>5, 6</td>
</tr>
<tr>
<td></td>
<td>SEG3103</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>SEG3125</td>
<td>xxx</td>
</tr>
<tr>
<td>GA Data (Y-4)</td>
<td>SEG1911</td>
<td>8, 12</td>
</tr>
</tbody>
</table>

Generating this report GAIA used GA data as follows:
- Year (Y) data for SEG4911 and SEG4105;
- Year (Y-1) data for SEG3101, SEG3102 and SEG4145;
- Year (Y-2) data for SEG2106 and SEG2911;
- Year (Y-3) data for SEG2105, SEG3103, SEG3125;
- Year (Y-4) data for SEG1911.

*** notation indicates database allocation for a course which has not been assigned GA(s) measurement at the time when the report is being generated. GAIA accommodates addition of courses and assignments of GAs as part of continuous improvement and further program developments.

Data arrangement per year as students progress in their program allows for generating a **COOP Progress Report per cohort** (COOPR/C) as well. Such a report increases reliability of data analysis on employability and adds consistency to the professional skills assessment provided by COOP employers. Table 6 below illustrates a COOPR/C report, based on the SEG COOP program sequence. In all sample reports used in this paper we refer to SEG program course progression as approved by the program GA Committee for reporting on GAA.

**Table 6: COOP Progress Report per Cohort (COOPR/C).**

<table>
<thead>
<tr>
<th>Year</th>
<th>Year</th>
<th>Year</th>
<th>Year</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Measuring Accreditation Requirements</td>
<td>Exceeded</td>
<td>Not Met</td>
<td>Met</td>
<td>Exceeded</td>
</tr>
<tr>
<td>COOP Courses</td>
<td>SEG</td>
<td>COOP Placement</td>
<td>SEG</td>
<td>SEG</td>
</tr>
<tr>
<td>1911</td>
<td>2901</td>
<td>3901</td>
<td>3902</td>
<td></td>
</tr>
</tbody>
</table>

3. **SIGNIFICANCE OF THE RESULTS**

Our approach to cohort analysis and reporting for graduate attribute assessment, leverages the existing data collection mechanisms and GA database of GAIA. The process does not involve any extra effort from users. Data is populated automatically, processed simultaneously, and reports are being generated concurrently. This enables program committees to perform GA assessment on a cohort basis with both vertical and horizontal analysis and a sound basis in fundamentals of statistics.

4. **CONCLUSION**

GAIA allows for comparison between cohorts by generating two reports – Graduate Attribute Report per Cohort (GAR/C) and Course Progression Report per Cohort (CPR/C). The former arranges averaged GA data per attribute, while the latter tracks student achievement as they progress in their program. Data from both cohort reports are used for comparison within a continuous improvement process for our engineering programs and provide a historic data trend for further curriculum development. Furthermore, a COOP Progress Report per cohort (COOPR/C) is generated. It adds reliability in analyzing students’ employability and professional skills assessment provided by employers.

**Acknowledgments**

The work was supported by NSERC and the School of Electrical Engineering and Computer Science at the University of Ottawa.

**References**


[7] Dupuis D., St-Pierre C., “Assessing graduate attributes as described by CEAB: an exploratory study in a first-year design course”, in Proc. of CEEA Canadian Engineering Education Conf., CEEA13, (École Polytechnique de Montréal, Montréal, Québec: 17-20 June 2013), DOI: https://doi.org/10.24908/pceea.v0i0.4912


[10] George A., Lethbridge T., Peyton L., “Graduate Attribute Assessment in Software Engineering Program at University of Ottawa – Continual Improvement process”, in Proc. of CEEA Canadian Engineering Education Conf., CEEA16, (Dalhousie University, Halifax, NS, 19-22 June 2016), DOI: https://doi.org/10.24908/pceea.v0i0.6484

