Towards a scale and tool for the appraisal of CEAB attributes – Progress report on a field test

Guy M. Cloutier, ing., Doc.
Dept of Mechanical Eng., École Polytechnique de Montréal
guy.cloutier@polymtl.ca

Pierre Savard, ing., Ph.D.
École Polytechnique
pierre.savard@polymtl.ca

Yves Boudreault, Ph.D.
École Polytechnique
yves.boudreault@polymtl.ca

Abstract

CEAB 2014 requires the appraisal of ‘attributes’. Quasi-competencies are not easy to ‘measure’. Results risk having a ‘local’ meaning with little transferability between institutions. From its interest into CDIO, and in parallel to its partaking in the DOCET project, École Polytechnique developed a 7-level scale, field-tested by nearly 100 appraisals pre-graduation work experiences. The paper summarizes the rationale behind the 7-level scale when compared to the 5-level CDIO scale, the possible mapping onto the EQF, and reports on the appraisals of students after 16 weeks internships. Supervisors appear to use the tool as a relative Likert-type scale, and will have to undergo a learning curve. Attributes appraisal tools need to be engineered rather than to be determined by consensus.

1 Introduction

Engineering programs must soon demonstrate that: i - graduates individually possess specific attributes, ii - outcomes are assessed in their context, iii - results are used to improve the program. This trend is shared by the International Engineering Alliance (IEA) [1], the Accreditation Board for Engineering and Technology (ABET) [2], the Canadian Engineering Accreditation Board (CEAB) [3], the European Accreditation of Engineering Programmes (EUR-ACE) [4], and the Conceive – Design – Implement – Operate (CDIO) organisation [5]. In embracing this trend, Engineering programs safeguard the mobility of their graduates. Experience with ABET shows the accreditation visit and review to be greatly simplified when external assessments are added [6].

CEEA 2010 introduced an information flow model — from stakeholders’ surveys to assessments — to satisfy the new CEAB requirements (see Figure 2, and [7]). It discussed how the CDIO Syllabus addresses the needs of the stakeholders in a well codified breakdown of expectations. The CDIO Syllabus covers 1. Engineering knowledge, 2. Personal and professional attributes, 3. Interpersonal skills, and 4. Conceiving, Designing, Implementing & Operating systems in the enterprise, societal and environmental context & innovation. The CDIO Survey shows stakeholders worldwide have a common understanding of what a current engineer is. The Syllabus has successfully satisfied outcomes-based conditions set by ABET (see [5]). A topical correlation shows the Syllabus spans the CEAB graduate attributes [8].

This paper presents the field test results of a 7-level scale and appraisal tool developed with the help of industrial advisors and a focus group.

The paper summarises aspects of curricula improvement at École Polytechnique de Montréal (EPM). It discusses the absolute scale referenced to the preparedness of an engineer about to become a “full” P. Eng. and relates it to the CDIO 5-level absolute scale and the European Qualification Framework (EQF) 8-level absolute scale [9]. Results as they stand are presented and interpreted, followed by challenges and developments.

2 Some aspects of curricula improvement at École Polytechnique de Montréal

EPM graduates some 1,000 students per year across 12 engineering programs. Committed to the evolving needs of its labour market, EPM has held a major redesign of its curricula in 2005. Few programs adhere to the letter of CDIO; all somewhat embrace its spirit with one mandatory project module every year making 12.5 % the 120 credits (Quebec students end college after K12 + 1). Students undergo a mandatory 16 weeks pre-graduation work experience, a prerequisite to the 6 credits 4th year capstone project that spans the fall and the winter semesters.

All programs maintain an update of expectations from employers through regular meetings with
specific consultative committees (COCEP). The 35 industrial advisors of the seven COCEPs ensure discipline-based distinctions. A central committee maintains consistency across disciplines.

Some programs deepened their understanding of expectations from the industry by conducting formal surveys. A 2006 survey took a very structured form based on the CDIO Syllabus [10], if only because of the commodity to maintain comparisons with a large international body of engineering faculties.

In 2010, a “CEAB attributes workgroup” was created to keep track of and prepare for the 2014 standard. Members have checked the perception of adequacy of the CDIO Syllabus within their departments. The faculty members found it to be an acceptable basis on which to build the ramifications of the CEAB attributes to a manageable level of details, without necessarily adhering to the CDIO Standards.

In view of the upcoming enforcement of the 2014 accreditation standards, the Mechanical, Biomedical, and Aerospace engineering programs have designed and field tested an appraisal tool that could cover i - the pre-graduation field appraisal of preparedness to practice (Loop 2 in Figure 2), ii - the field evaluation of graduates’ “competencies” (Loop 3 of Figure 2), as well as iii - the exposure of students to national standards and requirements early in the curriculum.

3 Absolute scales to express proficiency

The choice of a scale to express proficiency is a cornerstone in the design of a system that can demonstrate “possession of” and “program improvement against” the 12 CEAB attributes.

3.1 Characteristics of a “good” scale

A context and granularity insensitive, discipline free, field related, absolute scale is an invaluable element of the data fusion process shown in Figure 2, and would benefit all. As an added advantage, a scale with the potential of international recognition would contribute to the mobility of Canadian engineers.

Let the 12 attributes of CEAB define “coarse granularity”. Surveying the needs of stakeholders would require “intermediate granularity”, and the design of a course or project module would call for “fine topical granularity”. A level of proficiency need not be determined for every “granular element” at every stage, but every stakeholder needs to relate to the scale for the “granular elements” that is relevant to his involvement — either visiting a program for accreditation, expressing a need as a stakeholder, or designing a case study as an educator.

Consider the users of such a scale: employers (engineers in design, production, project coordination, research, consultants, ...), curriculum advisors, alumni, faculty, students, CEAB, and our respective provincial licensing bodies. The mere diversity of the list gives an appreciation for the difficulty in devising this scale. It is clear employers, alumni, the CEAB and provincial licensing bodies should be able to relate to the “preparedness to practice” the engineering profession, more than to some other indirect or academic definition of proficiency.

3.2 The CDIO five levels of proficiency

These “CDIO five” warrant a clarification:

1. To have experienced or been exposed to
2. To be able to participate in and contribute to
3. To be able to understand and explain
4. To be skilled in the practice of (P. Eng.)
5. To be able to lead or innovate in

The intention being “to resemble the progressive development of skills in a professional engineer, from those of an apprentice to those of a senior leader” [10]. The reference to a P.Eng. is strong, completed by a weaker link to Bloom’s taxonomy, “as a source of verb patterns”. Level 4 corresponds to the transfer from being a “trainee” to a “full” engineer, whatever the given local jurisdiction: an absolute reference.

A 5-level scale lacks the resolution to display i - cross differences between graduate attributes, ii - longitudinal differences along a curriculum, and iii - discriminating differences between students. A program could not demonstrate an improvement with such a coarse scale, a shift of only 1 level representing a strong departure from stakeholder’s expectations.

3.3 The DOCET project

The EQF provides a cross-discipline reference qualifications system applicable across Europe.

In 2009-2010, École Polytechnique and five other Engineering faculties/schools across six countries undertook the 18 months DOCET project. The objectives were i - to develop and test an EQF-CDIO equivalence model able to enhance the attractiveness of European Higher Engineering degrees and to improve their recognition worldwide; ii - to strengthen the links between the University and the labour market; and iii - to promote the use of the EQF for the recognition and improvement of engineering degrees.

In trying to bridge the Atlantic thanks to an EQF-CDIO similarity mapping, the DOCET is a strong thrust towards the requirement to converge and unite. Appendix A of the DOCET Report [11] provides a Syllabus-wide mapping of the CDIO 5-level scale to the EQF 8-level scale of competencies. Appendix B supplies examples on learning outcomes expressed for
many of the topics within the CDIO Syllabus. In its recommendations, the DOCET argues:

On one hand, allowing for explicit description of different levels of autonomy and responsibility and of different contexts and use of knowledge / skills / competencies acquired, appears to be much more than a formal exercise. This suggests the introduction or the strengthening of the above-mentioned two very important educational dimensions (autonomy and responsibility, context) which in turn are of high interest for the world of work.

An improvement over the CDIO Scale would thus be to incorporate some description of autonomy and responsibility, context, for international recognition.

3.4 A ‘working engineer’s perspective’

Independently from the DOCET, the Mechanical, the Biomedical, and the Aerospace engineering programs at ÉPM underwent the development of such a scale, with the perspective of a “working engineer”.

3.4.1 Ongoing development

The development of the scale and of the appraisal tool was done in three phases. The concerns of 12 industrial advisors over three engineering programs were answered by modifications of the first draft. Advisors remained aware of one another’s opinions inside groups, and unaware between groups. Fifteen field supervisors then had the opportunity to use and to criticise the scale and the tool independently.

The third phase called for field supervisors to appraise 98 internships (96 after filtering for validity). No observations taken to see if the interns influenced the way the field supervisors interpreted the scale.

As an ongoing phase, the first industrial advisors are now being met with the results of these appraisals.

3.4.2 Advantages

Table 3 abridges the 7 levels of the scale as it stands. It is non-linear, to improve resolution in the central portion of its range while remaining compact. It remains compatible with the “CDIO five”, as confirmed by the founders of CDIO. It expresses autonomy and responsibility, and thus addresses major conclusions of the DOCET report.

The focus group reported good understanding. Good reproducibility is expected by using paraphrases. These would avoid the intrinsic confusion from single terms carrying multiple definitions or large synonymic domains. No formal R&R study was performed.

It easily maps onto the levels of EQF by transitivity through the levels of CDIO, and the appendices of the DOCET report. No initiative was yet taken to gauge the extent to which this would favour the international recognition of the scale.

3.4.3 Robustness

The abridged texts of Table 3 are thought to convey the interpretation of the levels adequately. Neither the original nor the abridged versions were tested for robustness: interpretations could be culture, corporate, or context sensitive.

A ‘best’ scale would display robustness against cultural and individual bias to ensure the desirable repeatability. Parent to validations in psychometrics, this verification has to be repeated for translations, validity depending on such concepts as the subjective synonymic space of the terms that express the levels.

To our knowledge, no accreditation body had the budget nor took the time to ‘engineer’ robustness into their statements and/or scales. As engineers, scholars and educators, we have to question ourselves about the steps and the precautions we should take — and ask CEAB to take — before we entrust the future of our students to an appraisal system based on consensus. What else must we do, not to breach our own code of ethics by stretching ourselves beyond the limits of our competencies? Consensus is not engineering.

4 Results and interpretation

4.1 Global results

The appraisal tool asks supervisors to indicate their expectations and their observations. Expectations should be conveyed at the start of the internship, and in the presence of the student intern. It was not possible to implement this precaution at the time.

4.1.1 Expectations

Expectations are displayed in Table 1 (N = 76 for Mechanical Engineering, N = 20 for Biomedical Engineering, filtered for credit points above 55). Both programs taken together or separately, expected levels are essentially flat across attributes, with a small dip at 3.1.9 and increase at 3.1.10, together with a larger tally count of irrelevancy for 3.1.4, .9, .10 and .11. In the context of a 16 weeks internship, social and environmental repercussions could weigh less than problem analysis or investigation in the eyes of an employer. In addition, it is conceivable that involving a student in a project likely to handle sensitive or confidential information stresses the need for professionalism (3.1.8) and ethics (3.1.10).

However relevant these explanations may seem, the fact remains this nearly flat profile of expectations does not match the numerous surveys of stakeholders available to us through the CDIO organisation. One reason could stem from the numerous surveys of stakeholders available to us through the CDIO organisation. One reason could stem from the numerous surveys of stakeholders available to us through the CDIO organisation.
into single attributes would make the latter indistinctive. Another explanation could be that the scale is misused, and does not supply the desired ‘absolute’ reference. For instance, level 7 of the scale: ‘can lead or innovate’ (unshown) was expected by some responders in both programs. Section 4.2 further explores the possibility of misusing the scale.

4.1.2 Observations

As seen from Table 2, observed levels display similar patterns, and support similar possibilities (level 7 unshown). The reassuring difference being global observations lie above global expectations.

A closer look at observations reveals surprisingly high levels for engineering tools, teamwork, and communication: as high as or above the reference level of a P.Eng. Either i - these were attributed to exceptional students and hopefully correlate with academic evaluations, ii - field supervisors were biased by comparisons with their own workforce, or iii - supervisors, for some reason, do not treat level 6 in particular, and the scale in general, as an absolute reference.

4.1.3 Individual differences

Collective patterns would not demonstrate distinct students possess the wanted attributes. In addition, a plot of differences between individual observations and expectations could indicate the scale is practically used as a subjective Likert-type scale, and not as an absolute and referenced scale.

Figure 1 displays individual differences for the N = 76 internships of the Mechanical engineering program (69 to 71 appraisals left after filtering for credit points). Not only are there very few negative differences (observations lower than expectations), but these match students having received lower than average evaluations on a separate questionnaire about this pre-graduation working experience.

4.2 Stratified results

The number of credits completed stratified the Mechanical Engineering data (sample sizes being too small in Biomedical Engineering to do the same) The rationale is that, on an absolute scale, interns with fewer credits should populate levels lower than students closer to graduation and thus closer to performing the duties of an Engineer in training.

To summarize, the sample sizes of N = 6, 31, 18, and 21 respectively for students in the sub-ranges of [27, 51], [55, 63], [67, 87], and [89, 110] credits were insufficient to show significance in the slight longitudinal trends along the curriculum. Some attributes even displaying slight negative trends.

Table 1. Global expectations, levels 1-6; internships after 55 credits of the 120 credits curriculum

<table>
<thead>
<tr>
<th>Scale level</th>
<th>CEAB attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 0 0 0 0 0 1 2 2 2 1</td>
</tr>
<tr>
<td>2</td>
<td>10 14 9 7 6 2 7 9 7 7 9</td>
</tr>
<tr>
<td>3</td>
<td>42 36 35 32 35 27 34 31 25 20 30 35</td>
</tr>
<tr>
<td>4</td>
<td>26 24 25 22 28 37 35 23 12 16 15 30</td>
</tr>
<tr>
<td>5</td>
<td>5 8 14 6 8 14 12 18 9 20 8 14</td>
</tr>
<tr>
<td>6</td>
<td>1 1 1 3 2 2 0 3 1 2</td>
</tr>
<tr>
<td>99</td>
<td>6 7 5 19 7 2 0 11 32 20 24 1</td>
</tr>
<tr>
<td>Total</td>
<td>91 89 90 91 90 91 91 88 90 70 71 88</td>
</tr>
</tbody>
</table>

Table 2. Global observations, levels 1-6; internships after 55 credits of the of the 120 credits curriculum

<table>
<thead>
<tr>
<th>Scale level</th>
<th>CEAB attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 0 0 0 0 0 1 0 2 1 1 1</td>
</tr>
<tr>
<td>2</td>
<td>3 6 3 5 3 3 2 0 2 4 4 0</td>
</tr>
<tr>
<td>3</td>
<td>22 19 14 19 14 8 16 15 16 10 15 13</td>
</tr>
<tr>
<td>4</td>
<td>32 27 30 22 28 31 34 22 14 24 19 29</td>
</tr>
<tr>
<td>5</td>
<td>22 23 26 18 26 29 21 27 15 25 15 31</td>
</tr>
<tr>
<td>6</td>
<td>5 6 9 6 9 13 13 9 4 7 9 11</td>
</tr>
<tr>
<td>99</td>
<td>6 8 5 19 6 2 0 12 35 19 25 2</td>
</tr>
<tr>
<td>Total</td>
<td>91 90 90 91 91 91 91 89 89 90 91</td>
</tr>
</tbody>
</table>

Figure 1. Differences: observations – expectations (Mechanical engineering program).
Not only are there no indications students with more credits would stand more prepared to practice engineering, but observations sometimes diminish with increases in credit points. This would support the assumption field engineers practically tighten their subjective criteria when they supervise an intern about to complete his curriculum, regardless of the scale provided with the appraisal tool.

5 Developments and questions

As it stands, the findings point to important questions, challenges, and areas of developments.

An obvious undertaking is to increase the sample size of the current tool, unchanged, and provide substantiated answers for the assumptions and/or speculative interpretations of the patterns reported.

For employers to keep appraising attributes, over and above attending their core business, institutions will need to share a common scale and survey tool. Consensus not being engineering, by what process can we build robustness and lack of bias into such tools, to ensure the common scale and external appraisal tool we end up using possess the required characteristics?

Psychometrics showing the various translations of a survey tool all need to be distinctly validated (a major endeavour), what provincial/national bodies will perform the required validations? From what budgets?

By asking the assessment of “attributes”, we may be asking for an intellectual synthesis that is not easily mapped onto an absolute scale. Could the solution to robustness not lie in a breakdown of attributes into indicators, and for the field supervisors to undergo an effortless “box ticking” exercise, leaving the synthesis into attributes to schools and faculties as an aftermath exercise? As a contrary argument, how then to devise and test indicators that stand a chance of being shared between specialties, schools and faculties?

Meeting a group of industrial advisors with the results, one expressed that “Whatever the scale, bear in mind we consider it will take seven years to have a fully developed engineer.” Others simply nodded.

6 Conclusion

A 7-level scale and 12-attribute appraisal tool underwent a three-phase development. The scale is compatible with both the CDIO and the EQF levels of proficiency. Nearly 100 field supervisors tested this work on interns with from 27 to 110 credit points of a 120 credits curriculum. Global and individual results can show “(dis)satisfaction”. Neither the “expected” nor the “observed” levels seem to hide by the absolute nature of the scale.

The sample size is being increased. Care is being taken not to use procedures not maintained in the long run, and only temporarily improve reproducibility.

These raise important questions about the ways field supervisors practically assess interns, and about the unproven robustness of any appraisal tool. Precautions should be taken before we entrust the future of our students to a new appraisal system. We should be careful not to breach our own code of ethics by stretching ourselves beyond the limits of our competencies, being satisfied with a system that is consensual rather than engineered.

7 References


Table 3. A "working engineer’s perspective" of levels of proficiency (abridged).

“1. Has been exposed to...”: Heard about it. This may orient quest for information. No immediate insight or contribution.

“2. Participates and contributes to...”: Servile participation. Replicates long sequences or simple branching thoroughly enough. Entrusted with small tasks with several closely spaced verifications. (Not left on own with unforeseen conditions.)

“3. Executes, with signs of judgment...”: Repeats, but also judges the nuances or ramifications of decisions "well enough". Fewer verifications and milestones spaced-out. Premature to encourage initiative on own.

“4. Delivers and provides full explanations for...”: Relevant contribution. Can thoroughly explain reasons. Entrusted with more comprehensive tasks. Encouraged to demonstrate initiative (mainly to gauge potential) while seeking approval.

“5. Could exercise initiative without major risks...”: Answers ‘well enough’ to unforeseen events (not as well as a P.Eng.). Can be encouraged ‘to dare’, but within limits, as truly considered capable (and not only to gauge potential).

“6. Ready for the practice of engineering...” (threshold of P.Eng.): Ready to practice on own (but for legal restrictions). Approval of work is a mere formality. Verifications are less frequent (except on major topics requiring years of experience).

“7. Can lead or innovate in...”: Can be trusted with supervision of others or in situations never encountered (specific topics).

Table 4. ‘Essentials’ of comments from field supervisors and conclusions from a 4-level Likert scale questionnaire.

<table>
<thead>
<tr>
<th>Source of the comment</th>
<th>Mechanical Eng. (~ 1,000 students)</th>
<th>Civil Eng. (then ~ 700 students)</th>
<th>Biomedical Eng. (~ 100 students)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early comments from the COCEPs of large, medium and small programs</td>
<td>Keep it simple, and shorten the texts. We trust academia about the level of knowledge.</td>
<td>Our concern is attitude and transferable skills. They are not ‘soft’. They are ‘key’.</td>
<td>Teamwork (3.1.6), communication skills (3.1.7) and professionalism (3.1.8) are particularly important in biomedical engineering</td>
</tr>
<tr>
<td>Recurrent comments from the field [v. 1 of the form and scale, full CEAB descriptions of 12 attributes]</td>
<td>“Don’t change anything to the clarity of the scale; just shorten a bit.” (Done)</td>
<td>“Shorten the 12 descriptions of CEAB” (Done) “Regroup – e.g. 8, 9, 10 – into fewer.”(a)</td>
<td>“Maintain both ‘Expected’ and ‘Observed’ appraisals: interesting and easy to use.”</td>
</tr>
<tr>
<td>Comments from the field</td>
<td>“Shorten all you can.” “Keep it simple for us.”</td>
<td>“The whole process raises fundamental questions. You should ask about those.”</td>
<td>“‘Impossible’ levels, if only to trace questionable responses, to ensure validity.”</td>
</tr>
</tbody>
</table>

(a) In the eyes of the employer, it is easy to picture: 3.1.5 as embedded into 3.1.2, 3.1.3 and 3.1.4; to regroup 3.1.6, 3.1.7 and 3.1.12; and to regroup 3.1.8, 3.1.9, 3.1.10 and 3.1.11. Thus leaving us with a 6-point questionnaire 3.1.1, 3.1.2, 3.1.3, 3.1.4, 3.1.A, 3.1.B. If such is the strong perception of the employer, academia and CEAB will not “unblur” the picture by adding elements.