DESIGNING AND USING AUTHENTIC ASSESSMENT OF STUDENT CDIO SKILLS

Anastassis Kozanitis
Clément Fortin
Lina Forest
École Polytechnique Montreal

Rick Sellens
Queen's University at Kingston

Paul Hermon
Queen's University of Belfast

ABSTRACT

This paper presents a rationale for learner outcome assessment and an overview of the methods, instruments, and issues that instructors should consider when they design and use assessment of student CDIO skills. It argues the majority of actual assessment practices should be revised to satisfy learner needs, and to improve student learning and performance. To increase their performance, students need educative assessment, which has at least two essential qualities: firstly it is anchored in authentic tasks; secondly it provides students and teachers with feedback and opportunities they can readily use to revise their performance on these or similar tasks (Wiggins, 1998). This is known as authentic assessment and is intended as a richer and more stimulating method for assessing student learning. It provides students with complex and real world situations to be solved.

Authentic assessment requires students to act upon open-ended situations like a professional, by mobilizing their knowledge as well as their broad CDIO skills and attitudes. Only minor adjustments are made to fit any resources constrains (budget, time, and space). Also, continuous faculty feedback plays an essential role, since it allows students to improve before the final demonstration of their skills and competencies. In this sense, assessment is part of the learning process. Hands-on projects offer a natural setting for authentic assessment in engineering education. However, there are some issues one should consider before turning to this method. Specifically, authentic assessment requires well organized planning to help design and outline adequate assessment instruments, foresee effective feedback mechanisms, with the intention to foster student learning and success.

Hence, implementing authentic assessment requires: 1) formulating clear achievement targets; 2) choosing a real-world task; 3) writing a complex task; 4) giving clear instructions; 5) preparing a grading rubric; 6) designing scoring keys. This paper offers helpful information, practical advice, and handy examples to help instructors improve their methods for assessing student learning, particularly within the CDIO curricula.

KEYWORDS

Authentic assessment, rubrics, outcomes, competencies, team-work

INTRODUCTION

One very interesting and highly pertinent decision taken by ABET (the US Accreditation Board for Engineering and Technology), was to shift their accreditation criteria on what is learned rather than what is taught (ABET, 2006) [1]. This decision, seen as no less than revolutionary when it was introduced in 1997, is not a trivial matter, and for faculty and programmes to assume the transition is quite challenging. This implies changing from a teaching paradigm toward a learning paradigm, putting the focus on what students should know and what they should be able to do with their knowledge. The paradigmatic shift changes the question from "how will I teach this" to "how will students learn how to use this, and know when to use this".

This new approach has inspired many other accreditation boards worldwide to go about similar changes. For instance, Canada's accreditation board is contemplating adopting an equivalent approach as its southern neighbour. This new approach has also directly or indirectly motivated initiatives to transform the way engineering schools think of training their students. Therefore, many institutions and engineering programmes realized that sticking to old methods wasn't going to help them achieve the new ABET criteria. Indeed, this new educational paradigm required some innovative thinking and a different way of doing things.

The international CDIO initiative is certainly a convincing example of this evolution. Not only does it fit with the new criteria, but it also addresses the need for engineering curricula to pay more attention to broader skills such as communication and team work, to include active pedagogies that allow experimental hands-on approach, and to give the chance to develop effective life long learning.

Although ABET provides the criteria, many efforts are still needed to define what constitutes good practice for assessment. This article's aim is to provide guidance for faculty on engineering assessment practices. It addresses the need to use authentic assessment which appears to be close to ABET's approach to focus on outcomes. The first part of the paper describes what constitutes authentic assessment, and explains the knots and bolts of how it can be applied in a course. The second part of the paper presents two situations where authentic assessment has been implemented, and offers some solutions for the supervision of team work within project-based courses. The first situation is from the Product Design and Development (PDD) degree at Queen's University Belfast. It offers a solution for communicating with students and supervising the team work in a project. The second situation addresses the challenges instructors must face with supervising large groups in the Mechanical and Materials Engineering programme at Queen's University at Kingston.

What is authentic assessment?

Authentic assessment proposes students realistic situations, where tasks and problems refer to the reality of the professional life or to everyday life, and where they must combine knowledge, skills and attitudes in order to solve the problems. For Tardif (1997) [2], authentic assessment tasks are contextualised and do not refer to an artificial or made up situation. They require judgment and innovation, so students must use their knowledge and their skills with caution and effectiveness in order to solve the tasks properly or to come up with an elaborate answer. This answer can require a personal approach, a certain amount of creativity, or some form of innovation when compared to a rather standardized answer generally awaited through traditional assessment. Students should not simply produce an answer "*in abstracto*", try to guess the right answer, or replicate a learned by heart single right answer. Instead they should take into account several situational, circumstanced or conditional aspects as they exist in the working world, like rules and regulations, ethical aspects, a limited budget, an imposed data-processing environment, and uncontrollable

human behaviour like resistance to change. These conditions can be made contiguous to the required answer, and can affect the nature of the answer or work to be produced.

In this approach, students are evaluated in a constructive and iterative manner, resembling real life situations or in a professional context. Wiggins (1993) [3] lists several characteristics for authentic assessment: It must be realistic, it requires judgment and innovation, it asks students to do instead of to say, to repeat or reproduce, it simulates workplace situations, and allows consultations of resources, feedback and improvement.

What's more authentic assessment proposes students relatively complex tasks, where team work and a number of operations are often necessary to succeed. Solving these complex tasks requires appropriate integration of material from various courses and some iteration before attaining the anticipated result. Feedback from the professor is needed to improve performance. Thus assessment becomes an integral part of teaching. Many times during the semester, students will request feedback from the professor (or colleagues) for the improvement of their final performance. Thus a key ingredient to authentic assessment is allowing students to make errors and learn from them. Feedback is done in a context of transparency, not of secrecy, and based on clearly stated public criteria. It is continuous and provides students with a clear vision of what excellent work is like in order to help them improve future performance.

In contrast, traditional assessment, generally carried out in the form of an exam (mid-term and final), does not comprise this characteristic of realism and complexity, nor does it offer any regular instructor feedback intended for the improvement of performance. Indeed, mid-term and final exams are usually one-way, no turning back summative tests, where students are very often asked to reproduce a certain amount of information they have learned by heart or to solve a rather narrow, unique-right-answer problem. The students have very little information on what will be on the exam. The instructor has very little information on how well the student will perform. Traditional assessment focuses on information recollection, and mobilises little if any analysing, judging and critical thinking skills.

How to use authentic assessment process of competencies?

We propose a five step process for designing effective authentic assessment of competencies. This requires: 1) organizing and planning the process; 2) conceiving and producing the instruments; 3) foreseeing feedback mechanisms; all this in order to 4) allot a grade to each student; and finally trying to 5) continuously improve the process.

Organizing and planning the assessment process

The first step of the process is a preparatory stage, when considering all the factors of the teaching situation and analyzing their influence on how you project to assess the students. This is done by answering the following questions: Where does the course stand within the programme (first year, first semester or last year, last semester?). Is it an elective or compulsory course? Which students and how many students usually sign up for the course? What are the contents of the course, and what is their degree of difficulty? Which specific professional actions or behaviours is the course preparing the students for? Most importantly, what are the competencies or the skills that are needed to be assessed?

The answer to these questions will have an impact on the decisions regarding student assessment. For example, the complexity of a task will be different depending if the course is intended for freshmen, sophomores or upperclassman. The number of students in the classroom will directly impact the extent of the assignment, the number and the size of the teams. After taking into account all of this information, it is possible to choose the most

adapted combination of instruments which will be used to authentically assess students' competencies and skills.

Choosing an assessment instrument

The second step is conceiving and producing the assessment instruments. Assessment instruments can be classified according to their level of authenticity. A lower level of authenticity category includes examinations that call only upon memory: generally multiple choice exams, short answer exams, cookbook type laboratories, homework, drill and practice exercises, and theoretical papers. In our opinion, these types of examinations should no longer exist in universities because of their artificial nature, and their poor capacity to develop competencies. A higher level of authenticity replicates how people actually confront problems and performance challenges in the field. The following are examples of higher level authenticity instruments: team projects, professional work simulations, case studies, long answer exams within a professional setting, clinical evaluations, artistic creations, mock trials, internships, etc. These instruments are perfectly adapted to an assessment of competencies in authentic situations.

There are two elements to this step, the first is writing the performance task, the second is building the rubrics (whether informational or scoring rubric). Writing a relatively complex assessment task that provides students with a challenging ill-defined problem or situation similar to the ones professionals face at work, allows them to demonstrate what they have learned in their very own personal way. This means there is no one right answer and students must use higher end cognitive capacities like inquiring, reasoning and developing plausible hypotheses to solve the problem or situation they were presented. Authentic assessment should also allow them to make mistakes and try over, because that is how learning takes place. Instructors can help students expand and perfect their competencies by pointing toward more sophisticated and effective ways of doing things.

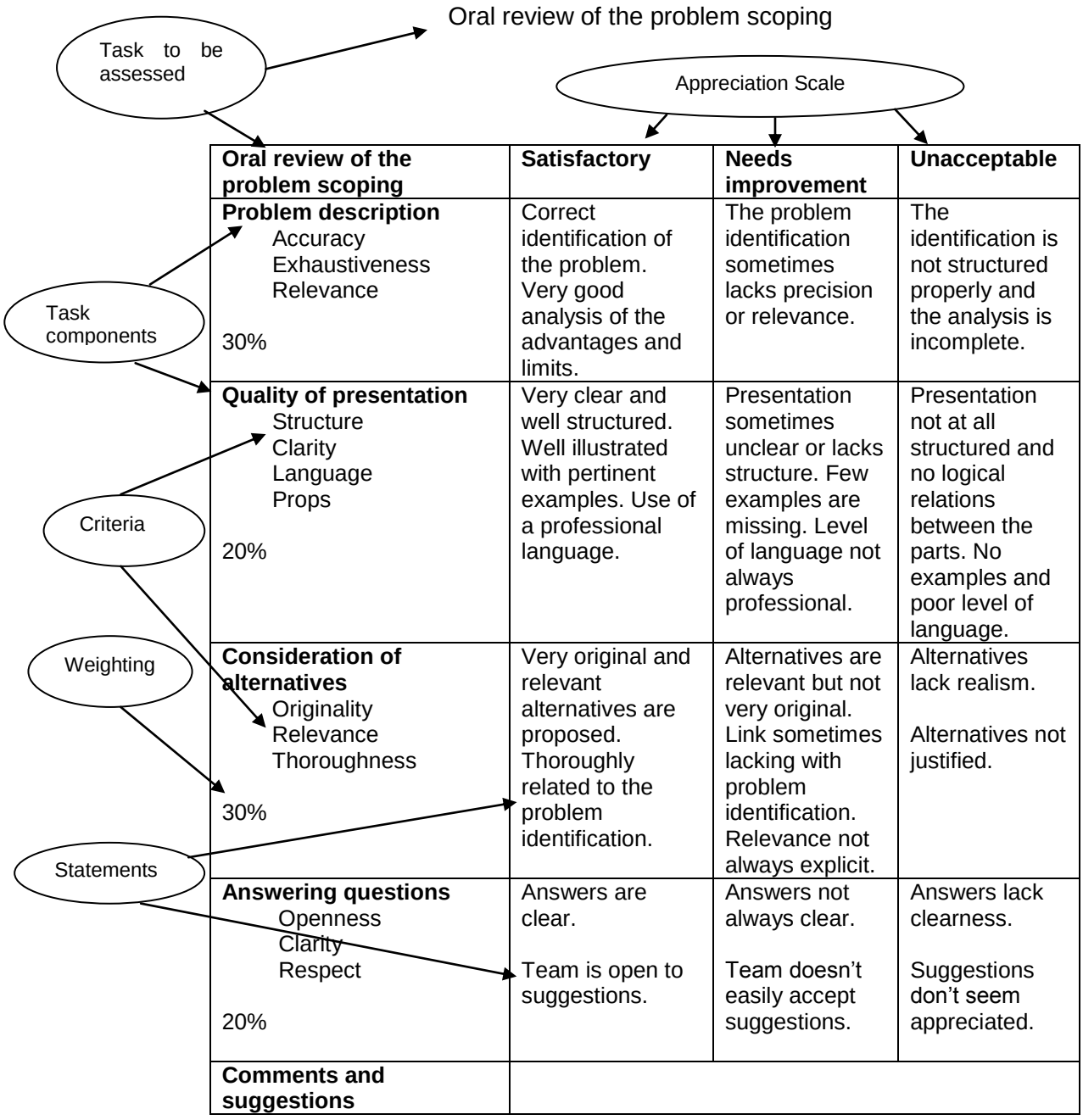
Using an ill-defined problem does not mean students should be left in total absence of direction. Some guidance will be needed in order for them to know what is expected from them. Hence, identifying specific responsibilities and tasks to be accomplished will help them organize the work and increase productivity. For example, defining the steps to follow, the operations to realize, the outcomes to produce, and setting a schedule to respect will definitely contribute to their success. Furthermore, writing clear instructions and communicating them is crucial for quality work. The more precise and clear are the instructions, the better chances will the work handed in meet up with your expectations. This will also facilitate supervision of student work as well as its correction.

The instructions can define team size, how teams will be made up, sharing responsibilities, requirements for intermediate and final reports, requirements for the oral presentation, the bill book and the handover dates, how to get consultations with the instructor, and the performance criteria that come in the assessment rubric.

Assessing student work is always a delicate exercise because of the possible bias du to subjectivity of human judgment. However, this is only a problem when judgment is based upon fuzzy or undisclosed criteria. In fact, assessing student performance using well-specified criteria enhances the reliability of the judgment. That is where rubrics come in handy, allowing explicit listing of criteria important to assess performance. What's more, rubrics can help distinguish between levels of performance for each criterion. In this sense an assessment rubric gives clear guidance of what constitutes successful performance, especially for anything prone to subjective grading. In the following sections we will describe the different elements that compose a rubric.

A rubric usually contains information allowing students to know what is expected from them, and providing descriptive feedback that helps them improve performance. A rubric contains the criteria and the levels of performance, but it should also specify the task that is being assessed, the task components, the weighting allotted to each component, an appreciation scale, and statements describing the performance levels. There should also be space for constructive written comments. Figure 1 gives an example of a rubric and its elements (Prégent, Bernard, & Kozanitis, 2009) [4].

Figure 1. Example of an analytical rubric and its different elements



Oral review of the problem scoping	**Satisfactory**	**Needs improvement**	**Unacceptable**
Problem description Accuracy Exhaustiveness Relevance 30%	Correct identification of the problem. Very good analysis of the advantages and limits.	The problem identification sometimes lacks precision or relevance.	The identification is not structured properly and the analysis is incomplete.
Quality of presentation Structure Clarity Language Props 20%	Very clear and well structured. Well illustrated with pertinent examples. Use of a professional language.	Presentation sometimes unclear or lacks structure. Few examples are missing. Level of language not always professional.	Presentation not at all structured and no logical relations between the parts. No examples and poor level of language.
Consideration of alternatives Originality Relevance Thoroughness 30%	Very original and relevant alternatives are proposed. Thoroughly related to the problem identification.	Alternatives are relevant but not very original. Link sometimes lacking with problem identification. Relevance not always explicit.	Alternatives lack realism. Alternatives not justified.
Answering questions Openness Clarity Respect 20%	Answers are clear. Team is open to suggestions.	Answers not always clear. Team doesn't easily accept suggestions.	Answers lack clearness. Suggestions don't seem appreciated.
Comments and suggestions			

Task definition corresponds to the professional action that needs to be accomplished or to the skill to be developed. In the figure 1 example, two professional skills are being assessed; the capability of giving a clear and well structured oral presentation as well as the capability of analysing a problem and suggesting plausible alternative solutions.

The task components are the more detailed professional tasks required during the oral presentation of the problem scoping. In the example, the components are: 1) problem description; 2) quality of presentation; 3) consideration of alternatives; 4) answering questions.

An assessment criterion is an indication, an attribute, or a benchmark that allows qualifying every component of the task. The number of criteria per component may vary from one to three. In the example, the criteria are: accuracy, exhaustiveness, relevance, structure, clarity, etc.

The weighting is the numerical value, generally given in percentage, allotted to each component of the task. The percentage indicates the weight of each component on the final grade. This valued information helps students adjust their work. In the example, 60% of the weight goes to technical skills, and 40% of it goes to communication skills.

The appreciation scale is a qualitative continuum containing levels of achievement used for judging the quality of the work; this for every task component and its corresponding criteria. The appreciation scale continuum spans from exceptional all the way to an unacceptable performance.

The number of levels is a personal choice, and depends on what better suits the situation. We suggest using a qualitative only scale to assess student's work in progress. A three level scale seems perfectly adequate for this matter. It tells students if their work so far is satisfactory or not, and gives them hints for improving, without judging the level of excellence. Indeed, a student may not be willing to continue his efforts or will not look to improve his work if the instructor considers it excellent. This way, a work considered as satisfactory will make him understand that he has well begun, without however having reached the excellent or exceptional level.

The statements describing the performance levels clearly indicate the quality of the performance for each task component. The rubric should have well defined statements that best describe the quality of the performance for each criterion and level of excellence. The statements should be short, explicit and unambiguous. This is what will probably take the most time during the conception and the writing of the rubric. Therefore they should only be written once all the task components, the criteria and the levels of the appreciation scale are chosen. For example, a satisfactory presentation will be judged as so when it is "very clear and well structured, well illustrated with pertinent examples, and using a professional level language."

Information of the level of performance for each component wouldn't be complete without additional written comments intended to support students' progression. Constructive comments are always welcomed and have a positive effect on motivation. For any shortcomings, they will appreciate concrete suggestions for improvement, useful documentary resources, or consulting an expert that can assist them with better understanding the aspects to be corrected.

Advantages for using rubrics

Using rubrics has at least four advantages: objectivity, equity, transparency, and standard setting within the programme. Rubrics make assessment and grading more objective, consistent, and defendable. Turning to explicit criteria avoids judgement that is subjective, emotional, and random. It also avoids judgement based on the comparison of students' work. This reinforces the equity of judgement; all student work is assessed equally, using the same components, criteria, and appreciation scale. Rubrics support transparency in the communication between the instructor and the students. By making the rubrics publicly

available, students are aware of what is expected from them right from the beginning of the semester. The transparency also helps reduce disputes concerning the grade. Finally, when all the instructors within a department or a programme share their rubrics with one another, it makes it easier to identify some common standards, and gives a clear idea of the quality of the work that is expected throughout the entire programme. This information can be very useful to heads of department, instructors, and any stakeholder interested by student assessment.

Types of rubrics

There are many different types of rubrics, we will however only cover two of them in this paper; the analytical rubric which is used mostly for feedback purposes, and the holistic rubric used for certification and grading at the end of the semester.

An analytical rubric usually holds a higher number of task components. However, an instructor can choose the degree of detail of the appreciation scale by adding sub components that give a more precise description. The goal is to guide the students' work in the most precise manner. An analytical rubric is generally used with any new complex tasks. Instructors will use it for giving feedback to students, but it can also serve as a self-assessment or a peer-assessment instrument.

The holistic rubric has a more global or general character. In this rubric, the task components are not as detailed as in the analytical rubric. The statements describing the levels of performance are more succinct. This type of rubric is generally used for summative assessment of final work, since students have been given the opportunity to improve their work with previous feedback from an analytical rubric.

Here are a few tips for conceiving and writing an assessment rubric. Firstly, if there are many tasks to assess, it is preferable to write more than one rubric. For example, during a project students may be asked to orally communicate during a formal design review, to write a complete report, and to test the prototype they built. In this case, the instructor should write three assessment rubrics, one for each of these tasks. However, writing a rubric is time consuming, it can take up to two hours, and the first draft may not be completely satisfactory. A few iterations and some help from peers or from an assessment expert will give a more satisfying product. Furthermore, this effort invested will be very profitable when comes the time to give student feedback or during the final correction.

Foreseeing feedback mechanisms

Supervision of student work in progress is an important part of the assessment process, mostly because students can come away with insights about their progress in learning. Indeed, instructor feedback is what allows students to improve their skills and develop their competencies, keeping in mind that adequate feedback should help, not hurt. Students need continuous feedback in order to improve the quality of the work. With this intention, here is some advice that can help enhance the quality of the feedback. First of all, feedback remarks should be phrased in a detailed, precise, and constructive manner. By referring to the instructions given at the beginning of the semester, to the intended learning outcomes, and to the competencies, feedback information provides direction for learners and reveals if they are on the right track.

Second, Huba and Freed (2000) [5] refer to feedback as intermediate when it is given during the development of a project, as opposed to final feedback. The different types of rubrics mentioned earlier are very useful to judge the quality of work at various stages of development. Rubrics help students understand the standards by which they will be evaluated, and focus on the characteristics that will be graded. Feedback discussion

sessions should be scheduled in a timely manner, during, as well as after assessment. Feedback information should be delivered orally or in written format, and look at sharing information and exploring alternatives, not giving advice or imposing solutions. Helpful feedback focuses on specific characteristics rather than generalities, and is directed toward observed student work and behavior, not those you infer.

Finally, questioning students is a powerful tool to help them develop an understanding of what they have learned so far, and what they still need to work on. Open-ended questions help develop their critical thinking, as well as their metacognitive abilities. Metacognition is one of the most important ingredients in the learning process. Some examples of open-ended questions are: Why did you choose x instead of y? How do you know that this is the right way of doing x? What was the hardest thing you had to overcome in order to do x? What are the implications of your conclusions to x? How can you improve the x process by using the y model? Etc.

Allotting the final grade

Each institution has its own marking system, some are literal (for example: A, B, C, D) and some are numeric (for example: 65%, 75%, 85%, 95%). When a literal marking system is in place, there is usually a numeric value associated with each letter used to calculate the grade point average. For allotting grades, instructors will use the scale provided by their department. Table 1 presents an example of equivalence between a literal marking system and its numeric equivalent.

Table 1. Equivalence between a literal marking system and a numeric grading system

Excellent	**Very good**	**Good**	**Insufficient**
A+ = 4.3	B+ = 3.3	C+ = 2.3	D+ = 1.3
A = 4.0	B = 3.0	C = 2.0	D = 1.0
A - = 3.7	B - = 2.7	C - = 1.7	E = 0

In reality, even though the marking system is literal, many instructors will begin by allotting a numeric score to each student, and will then convert it into a literal mark. The probable reason for this behaviour is that it isn't easy to add letters (B + C- + A =?) and come up with an average. Nevertheless, we suggest using a literal marking system for all grades to be allotted. The example in tables 2 and 3 illustrates how to assign a grade by using a literal system. The final grade was allotted using rubrics for three homeworks (HW) and one project for student X, student Y, and student Z. The steps to allotting a grade are the same for one assignment or for various assignments. For every assignment corresponds a weighting established by the instructor. In this case HW1 is worth 15%, HW2 and HW3 are each worth 20%, and the project weighs in at 45%, for a total of 100%. Taking student Z as an example, let us see how we came up with B- as the final mark. Table 3 helps understand the process.

Table 2. Example for allotting the final mark

Students Weighting	**HW1** 15%	**HW2** 20%	**HW3** 20%	**Project** 45%	**Total**
Student X	B-	B	A-	A	**A-**
Student Y	C	B-	B	B+	**B**
Student Z	C	C	B-	B	**B-**

Table 3. Example for calculating the final literal mark

Results – student Z	Weighting	Grades
HW1 = C	15%	**C =** (2.0) X 15% = 30
HW2 = C	20%	**C** = (2.0) X 20% = 40
HW3 = B-	20%	**B- =** (2.7) X 20% = 54
Project = B	45%	**A** = (3.0) X 45% = 135
Total	**100%**	**Total** = **259 /100 = 2.59 = B-**

1. In the left column appear the literal grades for student z, as well as the weighting for each means of assessment.

   In the right column, the literal grade is transposed in its numeral equivalent based upon the values of table 1. (Ex.: **C=2.0**).

2. It is then multiplied by the weighting value (Ex.: **C = 2.0 X 15%)**

3. We get the results in the right column.
   (Ex.: **C = 2.0 X 15% = 30**)

4. We then add up all the results from the right column:
   30 + 40 + 54 + 135 = 259

5. We divide that total by 100 (the total weighting)
   (Ex. 259 /100 = 2,59)

6. By looking up the literal equivalent of **2.59**, in table 1.

7. The final operation is to transform the numerical result into its literal equivalent, in this case a B-
   (Ex. 2.59 = B- = (2.7)

In this example, student's Z final numerical score is 2.59, putting him between the 2.7 mark for a B- and the 2.3 mark for a C+. Can we allot him a B-? We believe the answer should be yes, because it would be unfair to allot him a C+, since 2.3 is under his 2.59 mark. However, this decision belongs entirely to the instructor, and depends on the knowledge and appreciation he has of each student. For example, a student that has shown constant efforts for improvement can merit a better grade.

Continuously improve the process

At the end of the course we suggest instructors take some time to assess the entire assessment process and look for areas of improvement. The purpose is to ensure the validity and fidelity of the student assessment process. There are several methods instructors can use to analyse and reflect upon the situation. It is important to review all the steps of the assessment process by asking oneself relevant questions. Table 4 presents some open-ended questions that can help with the reflexion and improvement process. We also suggest three other very simple and easy to use methods. The first one is to analyse students' results to the various assessments and try to identify any difficulties they encounter. These difficulties can be related to some eventual weaknesses in the assessment process. The second method is to analyse the results of students' course and instructor evaluation. Quantitative results as well as written comments can provide relevant and worthy information.

The third method consists of meeting a few students, after the final results are handed in, and discuss with them desirable improvements. This discussion will point to very concrete and operational aspects drawn from the student experience.

Table 4. Examples of open-ended questions for continuous improvement.

Planning and preparing the assessment *Taking into account your teaching context,* ▪ Were the competencies to be assessed well targeted? ▪ Do they have to be reformulated? Is it necessary to add some or remove some? ▪ Were the most important contents targeted for the assessment? ▪ Do the students refer to them sufficiently at the time of realizing their work? ▪ Do the selected instruments allow adequate assessment of the targeted competencies?
Improving the instruments ▪ Are the tasks to be realized clearly presented? ▪ Are the responsibilities and the instructions clearly stated? ▪ Are the components and the criteria sufficiently explicit? ▪ Is the rubric clear, univocal, and complete? ▪ Is the appreciation scale suitable?
Improving feedback ▪ Does the assessment process as a whole allow feedback and the continuous improvement of work by the students? ▪ Do students receive sufficient support? ▪ Did the students find the feedback sufficiently detailed, precise and constructive? ▪ Did the discussions take place in a reciprocal open-to-criticism atmosphere? ▪ Are your personal availability, your appointment system and your assistants' availability enough for supporting students? ▪ Is the balance between the collective supervisions, the team supervisions and the individual supervisions adequate? ▪ Is the meetings schedule appropriate for all parties?

The next section deals with two issues instructors might have to address with authentic assessment situations. The first issue brings up the challenge of implementing and assessing project based courses with large enrolment groups. The second issue relates to the challenge of properly communicating with all team members during the development of a project.

An assessment perspective from Queen's University at Kingston

In Mechanical and Materials Engineering (MME) we have a broad consensus that our graduates should be well prepared to participate in conceiving, designing, implementing and operating real products, systems or processes as a central element of their professional education. This can be a challenge to implement with a typical enrolment of 140 to 160 students per year in MME. Each of our three options (General, Materials, Biomechanical) requires students take a core based on at least one authentic experience each year. Still, the majority of courses retain a traditional format with midterm(s) and final exam providing the bulk of the marks and assignments weighted just high enough to motivate students to complete them.

In first year students take APSC 100, a common core course across the entire engineering faculty that incorporates group projects aimed at solving practical problems. Groups of 6 to 8 students are directly supervised by a senior undergraduate student TA while working on a problem for a "client," usually a faculty member. Multiple different problems spread the load over multiple faculty members and allow some student choice according to personal areas on interest.

In second year students start a specifically MME curriculum, including an open ended design course (MECH 215) and an active learning measurement lab (MECH 215, Activating a Second Year Measurement Lab Sequence, 5th International CDIO Conference, Singapore Polytechnic, Singapore, June 7-10, 2009). The design course includes two group project segments, the first focused on competition among groups to each successfully complete the same (usually whimsical) task within a set of physical and time constraints. The second segment is a Client Based Design (CBD) project in which each group (typically 6 students) is assigned to a client with a real world design problem. All these clients are external to the course and include faculty members, local organizations and businesses, and industry, selected in part for their patience with novice designers. The groups are supervised and coached through the process by a team of Grad student TAs and (preferably more than one) faculty instructor.

In third year students complete a mechanical design course (MECH 323) closely focused on the traditional and important elements of machine design, gears, shafts, bearings, etc. that many people picture whenever they hear the word “design” in a mechanical engineering context. Student teams each work on their own design for a specific subsystem such as a gearbox, where the requirements are defined in detail and the results are quite similar, team to team. This is the most traditional of our design offerings, yet it also includes teams, potential for innovation, and multiple satisfactory solutions.

In fourth year students have a wide variety of options for design and research projects, however the program requires each student complete a conceive and design exercise, typically as part of a group of four students in MECH 460. Ideally, each MECH 460 group would have an individual client from industry with a practical problem requiring a solution, and the experience would closely replicate real world engineering practice. In actuality, we have a little over half of our teams working with industrial clients, about a quarter working on sub tasks for competition projects like SAE Baja or Formula cars, and the remainder working with faculty members. All projects must have a clear design requirement, as distinct from a research or literature survey focus. In addition to a client, each group has a faculty supervisor for guidance and evaluation.

Students have the option for much more authentic experience than the compulsory core through the design stream that starts in third year and culminates in a full year interdisciplinary project in APSC 480, or through the TEAM experience in APSC 400, both of which match student groups with industrial clients who commit to a significant level of cash sponsorship. These students self select for interest and ability and clients receive genuine value from the process.

All of these courses require a higher commitment of resources than traditional classroom courses. Some of this is direct cash costs for travel, communications, and materials to interact with industrial clients at a distance, however the vast majority is personnel time for technologists, machine shop, teaching assistants and faculty to work closely with students in small groups of 4 to 6 at a time rather than in lecture halls of 140 or more at a time. This resource commitment is diffused through the department by involving many faculty members as project supervisors. Spreading the load is a response to budget realities, but adds complications to the assessment approach.

Anecdotal Experience

Although not formally quantified, there is clear anecdotal evidence from observation of students working in projects in the different years that they do, as a group, advance substantially in CDIO curriculum areas over the course of their degree. It is also clear from comparisons with other groups of students that participation in our program correlates with that advancement. However, we lack the measurements we need to conclude that our

program is responsible for that improvement, let alone to identify which elements contribute most to reaching those objectives. That information is essential to scientifically improving our offerings and to scientifically allocating more of the resource pool to those elements that contribute most. Getting that information is not the focus of this work and there's evidence that efforts to measure actual outcomes (Bill Lucas, MIT, self efficacy ref) can be seriously confounded by any association with the assignment of grades.

Past accreditation based on teaching content has been very much a bean counting exercise, with individual faculty members deciding which category a particular teaching hour falls into and accreditation visitors doing due diligence by auditing actual course materials, assignments, and exams for content. As accreditation organizations move to outcome based criteria, it is likely that they will accept measures that focus on student learning.

Assessment Objectives in the MME Context

In the context of our authentic activities assessment has to meet objectives that are sometimes contradictory:

- Summative results for various learning objectives that:
    - Accurately rank order students on a scale
    - Show that individual students have met the objective for pass/fail
    - Show that groups of students have generally met some objective for accreditation of the program or budget justification
- Cost effectiveness
- Perceived fairness
- Motivate students
- Normative results that help students develop their abilities
- Authentic experience

Opinions of engineering education stakeholders differ substantially on the relative importance of these objectives, or even whether they should even be on the list, however, just about all of our stakeholders require summative results and sometimes place substantial weight on them. It is therefore a temptation to concentrate on getting the summative results right and to address the other objectives only where they don't interfere with that primary objective. This is certainly the traditional approach, still taken in most engineering science courses and student grades in those courses provide a clear indicator of purely technical abilities.

Within our design stream we have elected to focus first on the authenticity of the experience and delivering normative guidance, and then address the other goals where they don't interfere with those two objectives. In designing an authentic experience with authentic assessment we have replicated, as far as possible, some important elements of engineering practice within a consulting services model:

- Clients define the task, independent of the course instructor. The task is typically incompletely specified at the outset and the students define it in consultation with the client. Requirements may change over the course of the project. Different teams have different clients and different projects.
- Students work in teams, with a TA or faculty member in the role of engineering supervisor. The supervisor provides regular advice on technical questions, managing the project, and managing the client. It is crucial that any one supervisor not have too many project groups so that the advising process does not become formulaic.
- There are formal checkpoint assessments at proposal, detailed design, and fabrication stages, where clients and supervisors provide feedback before the project continues.

- In addition to evaluation of the final product by the immediate supervisor, the results of the project are assessed by faculty who have not been directly involved in the project.
- Peer feedback is used both during the project to identify and address team dysfunction and at the end of the project for summative evaluation of individual contributions.

This authenticity brings many advantages. Students get regular, project specific feedback from both the client and the supervisor and must resolve potentially differing opinions, making it clear there's no single right answer. Multiple iterations are possible on the same design issue. Supervisors get personally invested in the success of the team and the individual team members thus the normative evaluations flow much as real world guidance from an immediate supervisor or senior colleague would. Final course marks are determined in ways similar to success in employment.

This authenticity also brings disadvantages bundled with the advantages. Supervisors and clients who are personally invested in the project cannot be expected to be fully objective in assessing its success, nor can anybody directly involved be expected to be fully objective on the relative contributions of the various team members. It is a virtual certainty that discrepancies in contributions will not be fully reflected in final course mark variations within teams, and that varying supervisor bias will colour the scores between teams.

Such individualized assessments are costly in the time of both the supervisors and the independent secondary evaluators.

Students often see the evaluation process as unfair and different students have different ideas of what would be fair.

With clients and projects varying among teams and from year to year it is difficult to assess the general quality of the learning experience, or the resulting learning outcomes as they are intentionally not uniform.

An Imperfect Solution

As in professional practice, no evaluation and reward system will provide unequivocally objective and correct results. The best we can aim for is useful feedback to students, a reasonable reflection of performance in their record, a perception that we sought to be fair, and opportunities that motivate students to improve their results. In MECH 212 and 460 we seem to be succeeding in all of these, with the possible exception of an absolute measure of performance for individuals.

In MECH 212 the first project is uniform across the class and multiple judges evaluate the results according to a rubric that is developed collaboratively with the students. Although the second project problem varies from team to team, a detailed rubric for evaluation of deliverables is available, along with exemplars of good results from previous projects. For students whose designs fail in testing, the "Bank Account Model" of course marking allows redemption. Although they don't get the chance to re-do a project, students can propose bonus projects to improve their marks, so the course mark is limited only by the amount of work students are willing to do. Failure analysis of a design and/or a design process provides an ideal bonus project for those most in need of them. Overall course marks are high, as is student satisfaction with the experience and the fairness. Useful feedback is evident in iterative improvement of work product.

In MECH 460 the rubrics for the deliverables are more open ended to reflect the breadth of project variations, however detailed marking guides and exemplars are provided for each

element. Multiple assessments by different faculty with feedback for improvement at the proposal stage, and in the final presentations in advance of report submission provide opportunities for improvement of work product. Subjective elements in the marking scheme and an end of term peer assessment allow the course coordinator to address fairness issues with a broad perspective. Overall course marks are high and the vast majority of teams have a positive experience with a high quality outcome. The diffuse structure of the course means that occasional teams achieve poor results from the combined dynamics of team members, client and supervisor, an unfortunate, but truly authentic outcome. Usually these situations can be caught by coordinator intervention following the early term peer assessments in which students do not hesitate to point out issues with their client or supervisor.

Although the final marks in either course do not appear to be a good indicator for selecting students for success in future authentic tasks, they do show participation in an authentic project course where most groups eventually produce high quality original work, as evidenced by their final reports. That quality shows that assessment structure does provide adequate motivation and perceived fairness to work hard on meeting objectives. These projects provide students with an opportunity to shine and with a concrete result they can show to potential employers as evidence. Collectively, those reports provide anecdotal evidence of outcomes for accreditation.

Operating project activities within an authentic framework requires that students participate in a series of processes that eventually lead to a product, with normative feedback along the way. Exposure to the activity provides an important learning opportunity that most students take good advantage of. By putting authenticity first we create the conditions necessary to reach many of the CDIO syllabus learning objectives.

Ideally we would have independent assessment of learning outcomes in individuals and that is a direction we are pursuing in courses that are part of our optional specialization in design, for inclusion in the general curriculum.

Example of authentic assessment from the Product Design and Development (PDD) degree at Queen's University Belfast

Project 3P (MEE3052) 24 week duration, groups of 4 or 5 typically.

This major project carried out in the 3rd year of the degree is an advanced team design build and test (DBT) project which combines technical and commercial elements typically found in industry. Students are required to develop a product from the identification of a problem / customer need through concept development, product design specification, detailed design, prototyping, testing and finally to a plan for manufacture. The teams also develop an associated marketing plan for their product in a concurrent module.

The groups are required to plan their activities and must schedule regular formal design review meetings with their supervisors. It the latter stages of the project these often become weekly events. Agendas must be distributed in advance along with minutes of the previous meeting, as is standard practice in industry. The teams can request that the supervisor chairs the meeting but are encouraged to take on this role, and that of secretary, on a rotating basis to gain experience. These formal meetings demonstrate professional practice in the context of the project and the supervisors can provide immediate feedback in the way a design manager might do in the workplace. Action points such as task assignments with deadlines are typical outputs from these meetings and help organize the progress of the project.

These projects represent 2.5 of 6.0 modules in stage 3 of the degree program and even at the frequency of 1 design review meeting per week there is still a lot of activity which can take place between these formal meetings. It is not feasible for the supervisors to be with all

of the groups they are involved with all of the time. In order to provide a facility for supervisors to stay informed the web based “Google Groups” tool is used. Group access is by invitation only and is managed by the supervisor (a member of faculty). As well as a threaded discussion forum there is a shared file space and an area where “pages” can be developed collaboratively. Group members receive notification of activity by email and can chose between instant notification or to receive a digest email with activity over a 24 hour period collated into a single email. Supervisors generally receive the digest version and can provide near immediate feedback to the group. These digests conveniently provide a transcript of activity from which individual members contributions can be assessed both in terms of quantity and quality. A metric has been devised to convert these contributions into weekly grades which form part of the overall assessment.

Table 5: Grade Descriptors for Online Contributions to Google Groups

Grade	Description
0	no entry / absent from meeting
1 - 3	brief note of activity
4 - 6	detailed description of own work done
7 - 10	as 4 - 6 above plus evidence of engagement which has helped to progress the project as a whole

Each contribution is marked according to the scheme described in Table 5. Students are expected to make 5 such contributions per week. All contributions are averaged and 0.5 added for each contribution in excess of the expected 4. Similarly 0.5 would be subtracted from the average if only 3 contributions have been made. Students are made aware of the supervisor’s grading of their contribution at the midway point of the project along with a summary of a peer assessment exercise.

The 0809 stage 3 students were questioned about a number of aspects of the Project 3P module including their opinion of the use of Google Groups. It is clearly seen as an effective communication tool and has a marginally positive effect in terms of encouraging a consistent work rate. There is a slightly more mixed opinion however about whether it reduces the need for face to face meetings and it is best to consider it as supplementary to the design review meetings rather than as a replacement.

Table 6: Stage 3 QUB PDD Opinions Regarding the Use of Google Groups

	STRONGLY AGREE	AGREE	NEUTRAL	DISAGREE	STRONGLY DISAGREE
The Google Groups facility assists communication between team members	12	5	2	3	0
Google Groups reduces the number of face to face meetings necessary to manage the progress of the project	4	3	5	8	2
Knowing that my online contribution is being assessed on a weekly basis encourages me to work consistently throughout the semester	2	8	5	5	2

CONCLUSION

Authentic assessment of competencies proposes to move away from traditional practices which far too often require from students the repetition of an awaited good answer or learned "by heart" answers, as students say so themselves. Thus, authentic assessment requires conceiving instruments capable to evaluate students' competencies in realistic and complex situations. These situations simulate real life professional tasks. The production of the assessment instruments needs the identification of a complex task, accompanied with clear instructions in order to guide its achievement. Students are asked to apply their critical thinking, and a certain dose of originality. Furthermore, students are entitled to making mistakes and learning from them with proper instructor feedback. Writing a rubric and communicating clear and explicit criteria will help the assessment process as well as the students learning process. Appropriate use of authentic assessment, because of its ongoing formative nature, will eventually help instructors save time by accelerating evaluation of final student performance. Nonetheless, various constraints can make the instructors' work more complicated, like time constraints, high student enrolment, and adequate individual support. With these challenges that may lie ahead, instructors will need to reflect critically and design more effective assessment methods and instruments, within the authenticity mindset.

References

[1] ABET (2006). Engineering change: A study of the impact of EC 2000. Retrieved April 11, 2009, from http://www.abet.org/Linked%20Documents-UPDATE/White%20Papers/Engineering%20Change.pdf

[2] Tardif, J. (2006). L'évaluation des compétences. Documenter le parcours de développement. Montréal: Chenelière éducation.

[3] Wiggins, G. (2003). Assessing Student Performance: Exploring the Limits and Purpose of Testing. San Francisco: Jossey-Bass.

[4] Prégent, R., Bernard, H., & Kozanitis, A. (2009). Enseigner à l'université dans une approche-programme. Montréal: Presses internationales Polytechnique.

[5] Huba, M. & Freed, J. (2000). Learner-Centered Assessment on College Campuses. Shifting the Focus from Teaching to Learning. Boston: Allyn and Bacon.

Biographical Information

Anastassis Kozanitis is a Pedagogical Consultant for the faculty development centre at École Polytechnique Montreal (EPM), Canada, and has been collaborating in the international CDIO Initiative since 2005. He is also a lecturer at the University of Montreal, Canada. He has a PhD in educational psychology from the University of Montreal. His research interests are in higher education pedagogy, school motivation, and innovative teaching methods.

Clément Fortin is a Professor and Head of the Department of Mechanical Engineering at École Polytechnique in Montreal, Canada. His research background is in Design and Manufacturing, Advanced Manufacturing, and Aerospace, Aeronautical and Automotive Engineering. He has been collaborating in the international CDIO Initiative since 2003.

Lina Forest is a Pedagogical Consultant and Head of the faculty development centre at École Polytechnique Montreal, Canada. She holds a M.A. in education and is specialized in assessment and evaluation.

Rick Sellens is an Associate Professor in the Department of Mechanical and Materials Engineering at Queen's University. His research background is in Fluid Mechanics and more recently in Biomechanics and he has been actively involved in facilities development to support active learning, including the recently completed Integrated Learning Centre at Queen's.

Paul Hermon is a Teaching Fellow in the School of Mechanical and Aerospace Engineering at Queen's University Belfast. He holds a MEng Degree in Mechanical and Manufacturing Engineering (QUB 1987) and is Program Director for the Product Design and Development degree pathway. He has a Postgraduate Certificate in Higher Education Teaching (PGCHET) and is a Fellow of the Higher Education Academy.

Corresponding author
Dr. Anastassis Kozanitis
École Polytechnique Montreal
C.P. 6079, Succ. Centre-ville
Montreal, Québec, Canada, H3C3A7
1-514-340-4711 ext. 3229
anastassis.kozanitis@polymtl.ca