OMAR CHEIDDE CHAIM

**Desenvolvimento de competências através de avaliação individual no
contexto de Educação em Engenharia: Pensamento Crítico e Criatividade**

São Carlos

2015

OMAR CHEIDDE CHAIM

**Developing Competences through Individual Assessments in an
Engineering Education Context: Creativity and Critical Thinking**

Master's dissertation submitted to the
Escola de Engenharia de São Carlos,
Universidade de São Paulo (São Carlos
School of Engineering, University of São
Paulo) for obtaining the degree of Master in
Science.

São Carlos

2015

OMAR CHEIDDE CHAIM

**Developing Competences through Individual Assessments in an
Engineering Education Context: Creativity and Critical Thinking**

Master's dissertation submitted to the
Escola de Engenharia de São Carlos,
Universidade de São Paulo (São Carlos
School of Engineering, University of São
Paulo) for obtaining the degree of Master in
Science.

Research field: Engineering Education

Tutor: Edson Walmir Cazarini

São Carlos

2015

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Chaim, Omar Cheidde

C434d Desenvolvimento de competências através de avaliação individual no contexto de educação em engenharia: Pensamento crítico e criatividade / Omar Cheidde Chaim; orientador Edson Walimir Cazarini. São Carlos, 2016.

Dissertação (Mestrado) - Programa de Pós-Graduação em Engenharia de Produção e Área de Concentração em Processos e Gestão de Operações -- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2016.

1. Educação em Engenharia. 2. Avaliação Individual.
3. Desenvolvimento de Competências. I. Título.

FOLHA DE JULGAMENTO

Candidato: Engenheiro **OMAR CHEIDDE CHAIM.**

Título da dissertação: "Desenvolvimento de competências através da avaliação individual no contexto de educação em engenharia: pensamento crítico e criatividade".

Data da defesa: 08/11/2016.

Comissão Julgadora:

Resultado:

Prof. Dr. **Edson Walmir Cazarini**
(Orientador)
(Escola de Engenharia de São Carlos/EESC)

Aprovado

Profa. Dra. **Ana Teresa Colenci Trevelin**
(Faculdade de Tecnologia de São Carlos/FATEC)

Aprovado

Prof. Dr. **Edilson Carlos Caritá**
(Universidade de Ribeirão Preto/UNAERP)

Aprovado

Coordenadora do Programa de Pós-Graduação em Engenharia de Produção:

Profa. Associada **Daisy Aparecida do Nascimento Rebelatto**

Presidente da Comissão de Pós-Graduação:
Prof. Associado **Luis Fernando Costa Alberto**

CONTENTS

1	INTRODUCTION.....	1
1.1	RESEARCH QUESTIONS.....	3
1.2	OBJECTIVES.....	5
1.3	DISSERTATION STRUCTURE.....	5
2	LITERATURE REVIEW.....	6
2.1	ENGINEERS AND ENGINEERING COMPETENCES.....	7
2.1.1	Defining Engineer.....	7
2.1.2	Engineering Competences.....	10
2.2	DEFINING EDUCATION IN ENGINEERING.....	14
2.3	COMPETENCE DEVELOPMENT.....	16
2.3.1	Revised Bloom Taxonomy.....	20
2.3.2	Critical Thinking.....	24
2.3.3	Creativity.....	27
2.3.4	Motivation.....	31
2.4	OBJECTIVES AND LEARNING OUTCOMES.....	34
2.5	ASSESSMENT THEORY.....	36
2.5.1	Learning Frameworks on assessment.....	37
2.6	ASSESSMENT TYPES.....	38
2.6.1	Diagnostic Assessment.....	38
2.6.2	Summative Assessment.....	39
2.6.3	Formative Assessment.....	39
2.7	ASSESSMENT FREQUENCY.....	40
2.8	ASSESSMENT TASKS.....	42
2.8.1	Structured Tasks.....	42
2.8.2	Unstructured tasks.....	52
2.9	ASSESSING INDIVIDUALS IN GROUP WORK.....	63
2.10	GRADING.....	65
2.11	FEEDBACK.....	67
2.12	RUBRICS.....	69
2.13	COMPUTATIONALLY ASSISTED ASSESSMENT.....	72
2.14	DIRECTIVE VERBS.....	73
3	USING INDIVIDUAL ASSESSMENT TO DEVELOP CRITICAL THINKING AND CREATIVITY IN THE ENGINEERING SUBJECTS	80

3.1	CONNECTING THE PIECES	80
3.1.1	Assessment and Motivation.....	80
3.1.2	Type of tasks, assessment and grading:	81
3.1.3	Critical Thinking, Creativity Skills and Competence in Engineering.....	82
3.2	ASSESSMENT PRACTICES.....	82
4	INTEGRATION IN ENGINEERING CONTEXT.....	89
4.1	DEFINING OUTCOMES.....	89
4.2	DEVISING STRATEGIES.....	90
4.2.1	Choosing the tasks	92
4.2.2	The possibility of repetition	95
5	CONCLUSION	99
5.1	LIMITATIONS.....	100
5.2	FUTURE RESEARCH	101
	BIBLIOGRAPHY	102

FIGURE LIST

FIGURE 1 - DISSERTATION STRUCTURE	7
FIGURE 2 - A ROADMAP TO 21ST CENTURY ENGINEERING	11
FIGURE 3 - THE CDIO SYLLABUS V. 2.0. AT THE SECOND LEVEL OF DETAIL .	12
FIGURE 4 - COMPETENCE INTEGRATION PROCESS AND TRANSFER	17
FIGURE 5 - MOTIVATION FRAMEWORK.....	33
FIGURE 6 - STATES IN DEVELOPING COMPETENCES.....	68
FIGURE 7 - RUBRIC GRID.....	70
FIGURE 8 - DIRECTIVE VERBS AND TYPES OF TASK	79

TABLE LIST

TABLE 1 - ORIGINAL AND REVISED BLOOM STRUCTURES.....	23
TABLE 2 - ASPECTS OF CREATIVITY ACCORDING TO THE FELLOWSHIP PROFESSORS	27
TABLE 3 - DIRECTIVE VERBS	74

RESUMO

As demandas da sociedade de conhecimento sobre os profissionais de engenharia são bastante diferentes daquelas presentes no século passado, para atender-las a educação em engenharia deve se adaptar. A acelerada evolução tecnológica e rápida mudança em contextos sociais e econômicos reforça a importância de duas competências que são chave nas profissões de engenharia, pensamento crítico e criatividade. O objetivo deste trabalho é promover através da compreensão do papel e da utilização de técnicas e ferramentas de avaliações individuais o desenvolvimento de ambas competências no contexto de engenharia. Para promover práticas de avaliação individual que permitam tal desenvolvimento, este trabalho se iniciou com revisões bibliográficas dos principais conceitos relacionados às teorias de educação, desenvolvimento de competências, motivação, demandas de engenharia e teorias de avaliação. Com uma compreensão mais profunda sobre estes conceitos inicia-se uma discussão das principais práticas que possam agregar valor à educação em engenharia e ao desenvolvimento de pensamento crítico e criatividade neste contexto. Como resultado de pesquisa são propostas estratégias para incorporar práticas em cursos de engenharia, considerando tanto os objetivos do processo educacional como limitações como tempo para preparação e aplicação de atividades, tamanhos de sala de aula, infraestrutura e disponibilidade de tarefas.

Palavras-chave: Educação em Engenharia, Avaliações individuais, Desenvolvimento de competências.

ABSTRACT

The demands on engineering professional in the knowledge society do not correspond to those of the last century and to meet them, the engineering education practices must adapt. The accelerated technological evolution and fast change on social and economic contexts corroborate the importance of two key competences in the engineering professions, critical thinking and creativity. The objective of this work is to promote through the comprehension of individual assessment role, techniques and tools the development of both competences in the context of engineering. To promote individual assessment practices that allow such development, this work began with the bibliographic review of education theories, competence development, motivation, engineering demand and assessment theories. With a deeper understanding of these concepts, a discussion on the main practices that can add value to engineering education and the development of critical thinking and creativity in this context. As a research result strategies are proposed to incorporate these practices in engineering courses, taking into account both the objectives of education processes as well as constraints such as time limitations for preparations and activities execution, the size of classrooms, infrastructure and task availability.

Key words: Engineering Education, Individual Assessment, Competence Development.

1 INTRODUCTION

Recent years have seen a great change in preconceptions on learning outcomes importance, especially in higher education where the content or curricula importance as the guide for university courses has been decreasing rapidly. While different fields of study can show varying degrees of change, the rapid increase in the body of knowledge combined with the increasing ease of access to information are driving a widespread change.

Regarding the engineering education field, attitudes, competences, skills and behaviors are the primary concern of the industry, academic community and students (Crawley & Lucas, 2011). Despite the huge recent technical advances of engineering fields, technical knowledge has lost its central role as the goal for educating engineers to the development of this integrated array of abilities. Much research has been conducted to elucidate matters on how to develop these (Baartman & De Bruijn, 2011; Drejer & Riis, 1999; Drejer, 2001; Grunert & Hildebrandt, 2004; Paulsson, Ivergård, & Hunt, 2005). Different teaching methodologies, styles (R. Felder & Silverman, 1988), environments, learning precepts, digital tools are some of the most explored topics in the field. This effort is an ongoing attempt to adapt to this new reality, and while many changes can be seen, some areas, such as student assessment, are falling behind.

Even when recurring to educational research, assessment is a much controversial topic (Fernandes, 2006). Since the goals and the purpose of education themselves have more than one definition following much distinct lines of reasoning, the practices advocated are also distinct, sometimes even opposite. They range, for instance, from traditional perspectives of mass learning and one teacher preaching to many students that should be absorbing, to more individualized approaches where each student should be encouraged to achieve his or her own capabilities.

To understand the importance of individual assessment, its uses must be made explicit: assessment of learning, assessment for learning and diagnostic assessment (Freitas, Da Costa, & De Miranda, 2014; Helena & Santiago, 2015; Trevelin & Neiva, 2011).

The first type, assessment of learning, is used throughout the world as a mean of measuring the successful outcome of the educational process, especially regarding knowledge acquisition. It is the type of evaluation responsible for attributing grades,

classifying students, comparing different institutes teaching efficiency and awarding diplomas (Harlen & James, 1997). Consisting many times of arbitrary questions featuring a wide range of difficulty levels for the same topics and goals (even more so when comparing tests given in different institutions), the validity of these tests as comparison parameters has been challenged, even when well designed (Fernandes, 2006).

The second deals with challenges and activities proposed during the course in order to improve learning by different means. The feedback gotten by the professor can be used to infer the development of learning outcomes enabling changes in the process to improve their result, while the feedback given to the students can nurture positive outcomes and highlight improvement areas (Race, 2015). While taking formative assessments, learners practice, given the proper learning outcomes, and develop their knowledge and skills. This kind of assessment can also be used as preparation tools by stimulating student to keep up with given material, classroom discussions and proposed problems (Baleni, 2013; Harlen & James, 1997).

The last cited type is diagnostic assessment. The focus of this approach is to identify students' previous knowledge, individual or group characteristics as well as learning issues (Trevelin & Neiva, 2011). In the beginning of a course, tests to determine how well prepared are the students fall into this category. This type of assessment can be used to infer important yet not directly connected to the study subject information such as different learning preferences, role preferences, expectations on the subject, to determine whether an specific class is more intuitive or deductive in their learning approach among others.

Evaluation and grading of individual characteristics serves many purposes in today's society, from accreditation to classification. In this context, a proper assessment should be able to tell how competent or knowledgeable a respondent is in an specific content or regarding a specific skill or competence.

Even when working in group activities, individual assessment has its place in academic environment. While teamwork and group problem solving are invaluable as learning and assessment tools, especially regarding attitudes and behaviors, their results alone cannot evaluate and provide accurate learning feedback on a personal basis. The issue is that each member should be able to perform independently and if assessed only for their group performance, there is little basis for that learner's accreditation.

The focus of this work is not in developing methods or discussing the theory that provides validation for measuring students capabilities, but rather practices that promote change, focusing on reaching learning outcomes rather than classifying people for achieved results. In this regard, as much as possible the punitive aspect is avoided.

1.1 RESEARCH QUESTIONS

The development of a research question for this work was not linear. The motivation for working with competence development came from different two sources: the observation of the students graduated at the University of São Paulo, in the courses of Mechatronics Engineering, Electrical Engineering, Production Engineering and Materials Engineering; the work experience in a multinational company where many workers came from this and other high-end Brazilian Universities.

Many students seemed to be limiting their reasoning to achieve conformity, hurting their creativity as well as their competence to think critically, affecting their learning development and the quality of presented tasks. In the working environment, the same type of behavior was observed not as an exception, but as a common occurrence limiting in this case the quality and efficiency of work tasks. People seemed to be more worried about working by the method they were taught than they were about the results, reliability, effectiveness or even delightfulness the process.

In the work environment, when questioned about why they were performing some task in a seemingly inefficient way, many times people figured out or knew how to improve their working process, but did not do so before being asked. Some of the tasks observed in the professional environment were performed akin to methods taught by courses in the very company. A lack of competence in using the acquired knowledge, in criticizing the current work practice and in devising new practices was perceived.

The next step was to look in the existing literature to check validity of the observation, the other shortcomings of Engineering Education, the spread of these issues and the proposed solutions. The aforementioned observations were found valid and are present in many different locations and contexts, but the exact cause of these issues and the solution for each can vary significantly, thus they were divided and focus was given to those more closely related to the education process. This review led to a first

and excessively broad research question: How to develop relevant competences in the context of Engineering Education?

At this point another round of preliminary research was made in order to better understand the “relevant competences”, this time both in literature review and by looking at the Brazilian engineering job market.

In order to further narrow the scope of the question the “How” was broken apart. The intention was to find relevant measures to influence learners’ outcome that were, as much as possible, within the reach of Professors and that affected as many students as possible.

Individual Assessments were the chosen topic as they fill both criteria, in many universities they are at least partially under the Professor’s control (especially regarding assessment for learning), and most, if not every, student is affected by them. With this in mind, the possibility of using Individual Assessments tasks as a mean to help developing different competences was analyzed and found feasible.

At this point, three questions were formulated:

How assessment influences engineering students’ outcomes beyond acquired knowledge, but also regarding attitudes, skills and behaviors? Can the extrinsic motivation to get good results lead to deeper changes and develop learners beyond the content learning level? Can the negative aspects of summative assessment be mitigated?

The second and third questions belong to the learning theory as a whole and while their discussions are relevant to this work, they do not answer the original question.

When choosing the competences many aspects were taken into consideration, what is their relevance, how much importance are they treated in the academic environment, and how could they be affected by individual assessments. Creativity and critical Thinking were chosen, for their relevance, perceived lack of stimulus in the engineering courses, their connection to the advancements in education theory and the possible impact of assessment tasks in their development.

Taking into account the chosen competences and trying to limit the scope of this work while providing an understandable framework of the implications of assessment in their development, the proposed research question became:

How different assessments practices can be integrated in engineering courses in order to promote the development of critical thinking and creativity in engineering students?

1.2 OBJECTIVES

In order to adapt assessment practices to reach more holistic learning outcomes, this study aims to increase the understanding of evaluation theory and competence development applied to engineering education, not by prescribing solutions, but by proposing practices and alternatives, discussing their implementation in different contexts and inferring their outcomes.

1.3 DISSERTATION STRUCTURE

In the first chapter an overview of the thesis is given regarding its motivation, the research question and objectives are presented.

The second chapter contains the literature review of the most relevant concepts in order to understand the requirements of engineering as a profession, the goals of education and the tools and tasks available to achieve these outcomes. This tools include both the understanding of assessment tasks as well as the theories on how to understand the levels of learning, as well motivation and special characteristics of developing both critical thinking and creativity.

The third chapter is discussion of how the concepts explored in chapter two connect in order to promote the development of competences as well as highlighting other connections that can affect designing assessments.

The fourth chapter includes the suggestions for implementation, the connections of the practice with the context of engineering education, the rational usage of evaluation resources and a discussion on what are the expects impacts of these proposals.

The conclusion sums up the main points, points the challenged in implementation, a contextualization of the proposal, its limitations, and how future works can contribute to the development of a more holistic and well-rounded assessment practice.

2 LITERATURE REVIEW

Before discussing the assessment process in engineering education, a clear understanding of the involved concepts is needed.

Despite being presented sequentially, the first two topics were developed in parallel, determining the goals of education engineering in an informed manner require both the comprehension of the objectives of education in general and those of engineering. The definition of engineer and the demanded competences was selected as the first to topic to allow directing the discussion of education to the context of engineering.

Having set the goals, the next step was to explore the competence development and motivation theory to determine if, and by which mechanisms the assessment practices would be able to influence these competences outcomes. Both critical thinking and creativity definitions and development practices are discussed.

The revised bloom taxonomy is used in order to organize and provide a clearer framework as a basis for argumentation.

In order to use assessment as a tool to promote change, the comprehension of the different characteristics, tasks and practices involved is important. These tasks are further divided in structured and unstructured tasks. The frequency, feedback and reward system of assessment are also reviewed. Figure 1 shows the proposed structure.

Figure 1 - Dissertation Structure

Source: Author's creation

2.1 ENGINEERS AND ENGINEERING COMPETENCES

2.1.1 Defining Engineer

What is an engineer? While it is possible to make a regression of the term and find engineers throughout history, the origins of modern usage of the term are quite recent and related to the 18th century European industrial revolution (Merriam Webster Dictionary, 2015). The word "Engineer" nowadays carries many different definitions. The Merriam Webster Dictionary has three meanings for engineer, the definition that this work relates to is: "a person who has scientific training and who designs and builds complicated products, machines, systems, or structures: a person who specializes in a branch of engineering". While providing a general idea, this definition is of little help

to understand what type of professional should be achieved; the main concept provided being scientific training, complexity and performing engineering.

In his book “Shigley’s Mechanical Engineering Design” (Shigley, Mischke, & Budynas, 2002), the author provides an insight on what he perceives as a defining characteristic of engineering in comparison with science.

[...] Thus, while science has not yet completely explained the complete mechanism of fatigue, the engineer must still design things that will not fail. Engineers use science to solve their problems if the science is available. But available or not, the problem must be solved, and whatever form the solution takes under these conditions is called engineering. (Shigley et al., 2002)

From this definition comes that engineers need to be able to solve problems, if possible, based on science, if not in experimentation, logic or in any other tool that can make the solution viable.

The United Nations International Standard Classification of Occupation provides in the document ISCO-08 (2008) defines the work of Engineering professionals according to the following extract:

Engineering professionals (excluding electrotechnology) design, plan and organize the testing, construction, installation and maintenance of structures, machines and their components, and production systems and plants; and plan production schedules and work procedures to ensure that engineering projects are undertaken safely, efficiently and in a cost-effective manner. (ISCO, 2008, p. 63)

This definition uses “Engineering projects” as means of defining what this type of professional does and is one of the many examples where, in one way the other, engineers are defined as those who work with engineering.

Michael Davis (1996) has written an article on defining “Engineer”, and the importance of the definition, that adds much to discussion by evaluating other definitions, recurring to engineering history and context and comparing it to other professions in order to better understand what the definition should be, some of the relevant points extracted from this article:

- The definition of Engineer is that accepted by those in the field:
 - Mathematics and sciences are central to engineers;

- It is related to physical objects and physical systems (in contrast with “money” for accountants, “rules” for lawyers and “people” as in management);
- Engineers do not seek to understand the world by itself, but to achieve a goal or create solutions.
- There are three mistakes in the common definitions of engineering:
 - to equate engineering to technology;
 - to define it as the knowledge of the profession;
 - to attach to it the concept of moral and ethics as intrinsic.

The author gives in his conclusion the following explanation about the definition of the occupation:

Nonetheless, we can, I think, see that, as engineers became clearer about what engineers were (or, at least, should be), they tended to shift from granting membership in their associations ("at a professional level") based on connection with technical projects, practical invention, or other technical achievements to granting membership based on two more demanding requirements. One specific knowledge (whatever its connection with what engineers actually do)-is occupational. This requirement is now typically identified with a degree in engineering. The other requirement commitment to use that knowledge in certain ways (that is, according to engineering's code of ethics)-is professional. While many professions (law and medicine, especially) make a commitment to the profession's code of ethics a formal requirement for admission, engineering has not (except for licensed P.E.'s). Instead, the expectation of commitment reveals itself when an engineer is found to have violated the code of ethics. The defense, "I'm an engineer but I didn't promise to follow the code and therefore did nothing wrong", is never accepted. The profession answers, "You committed yourself to the code when you claimed to be an engineer. (Davis, 1996, p. 101)

While these arguments can help one understand better who can or cannot be called an engineer, however attached to an association of engineers and the accreditation of courses, it directs the definition towards the understanding of engineering held by the professionals of the area.

Considering the discussion, some key points are to be taken into account: Engineers are people that solve problems basing, whenever possible, their solutions on scientific theories while dealing with resources and conditions constraints and belong to formal Engineering classes. To overextend the discussion of what an engineer is would not contribute as much to the development of the engineering education field as understanding what they should be able to do. Many published papers deal with the engineer's role and engineering competences (Berry, DiPiazza, & Sauer, 2003; Duderstadt, 2008; Nguyen, 1998; Rugarcia et al., 2000) .

2.1.2 Engineering Competences

Many authors and entities discuss the competences and activities that belong to the Engineering scope (Crawley & Lucas, 2011; Goldberg & Somerville, 2015; ISCO, 2008; Stevens Kevin; Garrison, Lari; Jocuns, Andrew; Amo, Daniel M., 2008).

One of the oldest documents on the matter, a 1918 report of engineering education in the United States, presents a surprisingly current discussion on the matter. It is important to remind that it was written before the implementation of the Engineering Science policy in the 1950s. The report (Mann, 1918) contains the result of two relevant surveys on the engineering profession at the time (both from the year of 1916). First, they mailed the following written question: *"What are the most important factors in determining probable success or failure in engineering?"* Analyzing fifteen hundred answers, they published that personal qualities were mentioned seven times more than did knowledge of engineering science or practice techniques. The second survey consisted in a letter asking respondents from four major engineering societies to grade the importance of the following characteristics from one to six: *"Character, Judgment, Efficiency, Understanding of men, Knowledge, and Technique"*. Of the over seven thousand respondents, 94.5% choose Character as the most important, and an equally decisive majority put Technique as the last. The most important contributions to this topic might be the authors' statement that motivation and interrelation are some of the most important topics for research in the field and the conclusion that future engineering schools could not afford to ignore students' personality. By personality, the author exemplified integrity, initiative and common sense. The understanding of

the limitations of a curricula or content driven education is expressed in many points in the report, with actual recommendations to provide a broader range of capabilities. Another contributive aspect described by Duderstadt (2008) regards the range of requirements for engineers. Current engineers have a wide variety of attributions, starting from routines engineering services, such as sales or maintenance, up to designers and engineering scientists. According to the paper, the differences in requirements for different jobs can, and it is implied that it should, translate into formal requirement of different levels, such as bachelor for conventional jobs and master or Ph.D. for the highest end applications. An important contribution is a framework of the current characteristics and future needs of engineering as shown in Figure 2. The article also stresses the importance of the increase in engineering recognition and attractiveness for students and points out that engineers should change from an occupational profession towards the paradigm of a learned profession. The author states that when law and medicine professionals (examples of learned professions) are asked about their activities they respond with their specialization (cardiologist, surgeon, corporate law, litigation...) while engineers place the company they work for first.

Figure 2 - A Roadmap to 21st Century Engineering

Source: (Duderstadt, 2008)

According to the CDIO (Conceiving, Designing, Implementing, and Operating Systems in the Enterprise, Societal and Environmental Context) Syllabus v. 2.0 (Crawley & Lucas, 2011), conceived to create a set of goals for educating engineers that allows implementation, the expectation is that: *“Graduating engineers should be able to: conceive-design-implement-operate complex value-added engineering systems in a modern team-based environment.”* The same document states this is another way of saying that engineers engineer, or *“[...] they build systems and products for the betterment of humanity.”* With this set of definitions the document reviewed a Syllabus proposed ten years before to address the new challenges of the field and also incorporate feedback. In order to assert its validity, the authors compared their results with other high level framework from other sources, such as the UNESCO taxonomy, professional fields of engineering and also accreditation agencies. The five-Es framework, as described in the same paper (Johan DeGraeve, 2008 apud Crawley & Lucas, 2011) provides an action oriented engineering education with five terms around which their program of educating integral engineers is built: Engineering – making things, Enterprising – getting things done, Educating – developing oneself and others, Environmenting – embracing all elements, Ensembling – transcending and including. The framework inspired changes for CDIO’s revision. The version 2.0 of the Syllabus for engineering learning is presented in Figure 3 in two hierarchical levels.

Figure 3 - The CDIO Syllabus V. 2.0. At the Second Level of Detail

1 DISCIPLINARY KNOWLEDGE AND REASONING 1.1 KNOWLEDGE OF UNDERLYING <u>MATHEMATICS AND SCIENCE</u> 1.2 CORE FUNDAMENTAL KNOWLEDGE OF ENGINEERING 1.3 ADVANCED ENGINEERING FUNDAMENTAL KNOWLEDGE, <u>METHODS AND TOOLS</u>	3 INTERPERSONAL SKILLS: TEAMWORK AND COMMUNICATION 3.1 TEAMWORK 3.2 COMMUNICATIONS 3.3 COMMUNICATIONS IN FOREIGN LANGUAGES
2 PERSONAL AND PROFESSIONAL SKILLS AND ATTRIBUTES 2.1 <u>ANALYTICAL</u> REASONING AND PROBLEM SOLVING 2.2 EXPERIMENTATION, <u>INVESTIGATION</u> AND KNOWLEDGE DISCOVERY 2.3 SYSTEM THINKING 2.4 <u>ATTITUDES, THOUGHT AND LEARNING</u> 2.5 <u>ETHICS, EQUITY AND OTHER RESPONSIBILITIES</u>	4 CONCEIVING, DESIGNING, IMPLEMENTING, AND OPERATING SYSTEMS IN THE ENTERPRISE, SOCIETAL AND <u>ENVIRONMENTAL</u> CONTEXT 4.1 EXTERNAL, SOCIETAL AND <u>ENVIRONMENTAL</u> CONTEXT 4.2 ENTERPRISE AND BUSINESS CONTEXT 4.3 <u>CONCEIVING, SYSTEMS ENGINEERING AND MANAGEMENT</u> 4.4 DESIGNING 4.5 IMPLEMENTING 4.6 OPERATING

Source: (Crawley & Lucas, 2011)

Many other works also provide some insight on the matter. Nguyen (1998), concludes that even though students, academics and the industrial sector agree on the importance of “technical knowledge and skills” and attitudes, the industry regards “personal and professional attitudes” as more important than those in the academic environment. Gibuena et al (2015) conclude that even though technical skills are essential, professional skills are of equal importance but professor and students do not see them as such. The same paper points out risks of deviating to far from practice since if students feel that what is learned is not applicable, they might end up believing that they will learn everything in industry. This kind of reasoning leads to disregard for formal education.

In this regard it is possible to understand that technical skills are, more often than not, prerequisites to perform a task or solve a problem, while many of the other competences have a more scaled contribution related to work. Customer relation, rapport with colleagues, decision making abilities and perceived value of a service create a grayscale that is directly related to engineering performance.

Another important source for defining the competences that belong to this field is that of the Accreditation Board for Engineering and Technology (ABET) (2016). According to the commission, the learning outcomes of engineering courses should include, at least the following items:

- (a) an ability to apply knowledge of mathematics, science, and engineering*
- (b) an ability to design and conduct experiments, as well as to analyze and interpret data*
- (c) an ability to design a system, component, or process to meet desired needs within realistic constraints such as economic, environmental, social, political, ethical, health and safety, manufacturability, and sustainability*
- (d) an ability to function on multidisciplinary teams*
- (e) an ability to identify, formulate, and solve engineering problems*
- (f) an understanding of professional and ethical responsibility*
- (g) an ability to communicate effectively*
- (h) the broad education necessary to understand the impact of engineering solutions in a global, economic, environmental, and societal context*
- (i) a recognition of the need for, and an ability to engage in life-long learning*
- (j) a knowledge of contemporary issues*

(k) an ability to use the techniques, skills, and modern engineering tools necessary for engineering practice. (Commission, 2016, p.3)

Despite the different form of presentation, ABET and CDIO converge in most of their propositions on the engineering education outcomes.

Regarding the engineering competences, it is clear that the complementary skills needed in the profession are as important as core engineering knowledge. Despite the debates over the exact skills, some of them appear consistently on different sources: critical thinking, creativity, communication skills, leadership, problem solving skills, common sense and core knowledge.

2.2 DEFINING EDUCATION IN ENGINEERING

Whereas it is common to find Engineering Education papers skipping definition of educational goals, their relevance for the outcomes of assessments makes them an essential discussion topic. If these goals are well designed and implemented accordingly, they can direct the change in engineering education.

At least three sets of goals are identifiable: the industry, the university and the students (D. Jonassen, Strobel, & Lee, 2006; Nguyen, 1998). While their definition on actual abilities and competences might differ, they all revolve around preparing students to perform as engineers in the global market.

In the field of pedagogy there is a huge discussion on the purpose of education and while it takes into account different conditions than that of educating engineers, this discussion adds to the comprehension of the issue.

The very definition of the first element, the student varies. There is a clear displacement between the traditional and non-traditional standpoints regarding the amplitude and complexity of learners. The simplified student element becomes integrated, and as such more active and dependent on the environment. The concept of a uniform mass of students to be educated turns into the individual student, with unique characteristics (Ally, 2004).

In her work “Ensino, as abordagens do processo” (Teaching, the approaches of the process), Mizukami (1986) points out the approach of different frameworks regarding the purpose of and what is education. The traditional approach treats students as knowledge receptors that should update their knowledge up the point where he or she

is able to perform according to values that are prescribed. The behaviorist approach sees people as the byproduct of environmental influences and has as goals the achievement of basic abilities in order to manipulate and control the world/environment, also according to a source external to the learner. The humanist approach considers people as individuals and preaches the freedom from external pressures and direction, the objectives in this context should come from the student herself or himself. The Cognitivist approach believes that knowledge is a result of the interaction of the human and his or her world and it aims to prepare students to observe and extract the knowledge from this relation. The socio-cultural approach aims to create situations of self-education based heavily in critical dialog, based on the individual experiences and context.

One of the key developers of the socio-cultural school, Paulo Freire (1968) proposed concept of education considers taking into account not only the educatee, but also his or her entire condition. The goal of this approach is education for liberation. According to its precepts, the paradigm of the master-student relation, the first transferring his or her expertise to the second, should be changed to reciprocal relation of educators-educatees. This would lead to an education system where the basis for learning would be dialog. Education is necessarily biased by politics, and the critical understanding of the bias together with the individual development of critical thinking have a central role in starting a liberation process. By engaging in this process, the educators-educatees are able to act as transformers of their own reality, in a continued process of unalienating themselves (Freire, 1968).

The aforementioned frameworks deal with fundamental school and in order to understand their impact in the engineering education context, some differences should be highlighted.

While each of these theories brings possible contributions to the field of engineering education, some of them can be more directly connected to the practice (among others). From the humanist framework directing the field to the individualization of learning, within engineering boundaries, leading to higher degree of freedom to pursue one's interest. The socio-cultural brings up the importance of critical dialogue and deep discussion within the students reality, both as an individual and assuming the future role of engineers, and can lead to professionals more connected to the needs of society and aware of their impacts on it. The cognitivist approach, despite engineering

education dealing almost exclusively with adults (the discussion of learning phases does not directly apply), adds to the importance of preparing engineers to learn from their experiences and exchanges with the environment. Behaviorist tools can be used in order to direct students to develop attitudes and skills that are deemed important. The objectives preached by the humanist, cognitivist and socio-cultural frameworks have contributed to the choice of critical thinking and creativity as the key competences of this study.

Trying to follow throughout a framework that is not either traditional or behaviorist, if possible, would require a disruptive change in engineering education. While the validity of this change is subject to discussion, one would be hard-pressed to present arguments defending changes starting in engineering education instead of basic education. This is not to say that incrementally applying these concepts and considering their discussions in a higher education environment is undesirable, but that this is not the central goal of engineering education.

However, a change from the traditional perspectives to more holistic and individualized ones is necessary in order to adapt to the innovations and changes of the knowledge era.

By defending the development of characteristics deemed important to their role as engineers, this work is more aligned with the precepts of Behaviorist education. The emphasis on critical thinking is a step towards, even limited as it is compared to the full scope of this framework, towards the socio-cultural approach. The development of creativity has many of the cognitive precepts built in, such as free exploration of ideas and experimentation against reality (or models of reality) itself. When limiting this discussion to the practices of individual assessment, the impact remains relevant, however, it is reduced.

2.3 COMPETENCE DEVELOPMENT

Many theories have been developed and experiences have been conducted in order to assess how competences can be developed (Baartman & De Bruijn, 2011; Drejer & Riis, 1999; Grunert & Hildebrandt, 2004; Paulsson et al., 2005). Before discussing development, it is necessary to understand better competence itself.

Competence can be defined from, at least, two different point of views. By itself it consists of integrated pieces of knowledge, skills and attitudes (Lizzio & Wilson, 2004). From the professional stand point it can be assumed as a prerequisite for an adequate performance on the job (Eraut, 1994; Hager, Gonczi, & Athanasou, 1994). The individual consideration of the proposed components of competence, Figure 4, can help its understanding.

While at schools ill-defined problems are normally rare, tending to inexistence, this is not the reality in many working environments (D. Jonassen et al., 2006). From clients that do not know well the parameters of their problems or even which solutions fits them best, through design variables that cannot be properly measured and also market theories that might fail to properly predict demands, there are many ill-defined problems in industry. Learning to work with them in a controlled environment might give the student a competitive edge and help informed decisions.

Figure 4 - Competence Integration Process and Transfer

Source: (Baartman & De Bruijn, 2011)

Regarding knowledge, there is the classical differentiation between tacit and explicit knowledge. While a beginner might know every step of melding parts and is even able to describe the steps in different levels of detail, this does not mean that he or she is able to do it properly. The explicit part deals with direct instructions and can be learned verbally while the tacit part requires the internalized knowledge of the activity, in the forementioned example, an experienced melder that can simply meld the parts properly (Baartman & De Bruijn, 2011).

Skills are interwoven with knowledge in more than way. Skills require both the knowledge and the psychomotor domain (Morrison, Ross, & Kemp, 2001 apud Baartman & De Bruijn, 2011). Skilled behavior depends on a goal oriented use of sensory feedback and response movements, thus involving both cognition and motor skills (Baartman & De Bruijn, 2011).

The aforementioned work describes three manners by which skill and knowledge relate to one another.

The first one, called the low road, refers to skills trained to the point of automation. After the explicit knowledge is practiced new patterns and relations start to be understood. The next step includes transforming explicit knowledge into procedural knowledge, what culminates in automation.

When referring to high road, the author describes a process of reflection on the process. This level implies redesigning and adapting previously automated skills into a new solution, the automation in this case ceases. The new process might require a new motor skill development or rethinking the process in order to adapt. The title high road denotes a increase in difficulty and competence required to use this process.

The transformational road of knowledge-skill integration includes not only a reflection on the process itself, but also on the professional precepts on which the process is based. To illustrate this principle, the author describe the case of a plumber that is asked to install a solar heated distribution and decides he does not have the competence for doing so. For this decision the plumber had to access his existing capabilities and evaluate critically his capability. This road allows for transformation of precepts regarding both skill and knowledge. To develop such a level of competence learning, flexibility is essential.

The third component of competence, attitude, also has a diverse theoretical background. There are discussions regarding the term's definition, the generality level of its objects, wheter there is a separation between implicit and explicit types, and also if they stable or context dependent. Attitudes can influence behavior, and act differently on the three discussed "roads".

In the low road attitudes work as autonomous habits, are well-trained and many times with limited or no thinking.

When moving to the high road concept, attitudes have a deeper connection to the thought process. Facing a problem that requires adapting raises some questions: Is it

worth it? What are the rewards? How much effort is needed? Whether a risk-taker, laid-back or aggressive attitude is chosen, makes a great difference in the outcome (Krosnick, Judd, & Wittenbrink, 2005).

The third road, transformational, also has its effects on attitude. To have an attitude of self-criticism plays a major role since without it the reevaluation of one's own knowledge and skill hardly happens. This redesign can many times lead to the need of exposure to other's judgment, further illustrating the importance of attitudes for this behavior (Mezirow, 1990).

Since the three aspects of competence can be developed through practice (Paulsson et al., 2005), why not leave them to be learned in the work environment?

The main issue would be in the quality of the acquired ability and the damages that may be caused in the learning process. While many companies have outstanding training programs for a wide range of competences (Hinkin & Tracey, 2010), sometimes programs far more advanced than those found on schools, this is not always the case. The classroom environment is in its essence a place of learning, designed for improvement and that assumes lack of mastery on the students' part, be it on knowledge, skill or attitude. As such, doubts and mistakes not only can happen, but actually should happen. In a professional environment, errors and mistakes while working are seldom welcome, such as long training times.

In the processes described in order to develop competence, both critical thinking and creativity are required e.g. to perceive the need to improve, critical thinking is required, to develop alternatives to the current process, creativity. The application of both abilities in the field are competences that require the three characteristics, the engineering knowledge to base them, the skill and the attitude of reflection and will to promote change.

To further the discussion, the following sections deal with the specifics of developing both creativity and critical thinking. In order to provide a clear framework to discuss these competences the revised Bloom's taxonomy is presented. The last subtopic of this chapter is a theoretical motivation framework, that is used both for proposing practices to promote change and to base the discussion on assessment and grade impacts on motivation.

2.3.1 Revised Bloom Taxonomy

The original and revised Bloom taxonomy are widely used in the education practice and research fields in order to provide a framework for understanding the level of cognitive depth of a given practice. In Engineering education this is also one of the most widespread frameworks (Villanueva & Taxonomy, 2015).

The complete Bloom taxonomy consists of three different models divided in domains, the cognitive, the affective and psychomotor (Bloom, Englehard, Furst, Hill, & Krathwohl, 1956).

The psychomotor domain is directly related to the learner's ability to manipulate tools or instruments. In the original work (Bloom et al., 1956), this domain did not have subcategories, but in 1966 Simpson (1966) proposed a division in the following lines:

- Perception: dealing with the ability to perceive sensory cues to guide the task.
- Set: how ready one is to act.
- Guided Response: is composed by the early stages of psychomotor learning. It includes learning, trial and error.
- Mechanism: the intermediate step, where the learner is able to perform the task with some confidence and proficiency.
- Complex overt Response: in this level the practitioner has reached a level of proficiency where the task is skillfully done with accuracy and minimal effort and hesitation.
- Adaptation: Skills are developed enough to adapt to meet different requirements.
- Origination: Developing new movement patterns in order to solve a specific problem or address a different situation.

While this division can help the understanding of psychomotor development, its use is limited in the majority of the challenges presented in Engineering Education, especially taking into account Individual Assessments and the development of critical thinking and creativity. However, the two last levels do require both skills in order to be performed successfully.

The affective domain is related to the emotional responses and is further divided in to five different levels:

- Receiving – this level is a prerequisite to learning and involve openness to the topic and attention.
- Responding – Building on the receiving level, this level includes some sort of reaction in the learning process, be it mental (as a reflection or indignation) or physical (such as expressing an opinion or writing a response).
- Valuing – In this level the learner attaches or changes the value given to an idea, phenomena or piece of information.
- Organizing – Where the learner integrates the valued object in his or her own framework, incorporating it with the preexistent knowledge.
- Characterizing – The actual process of building abstract knowledge.

The affective domain also provides some insights regarding the Individual assessment practices, question and other tasks require the usage of at least the second level of this taxonomy. However, if a task calls on the experience of a student or successfully challenges his or her preconception it is possible to work in even higher levels, reaching even the last, characterizing.

The cognitive domain describes the different levels of cognitive depth and for the purposes of this work is the most relevant. This is due to its relation with the development of creative and critical thinking, the cognitive framework is the most suitable for this discussion as many author use the last three levels to describe both skills.

There are two version of the cognitive domain taxonomy, in this work the revised taxonomy will be used. Originally this domain was composed by, in increasing order of complexity: Knowledge, Comprehension, Application, Analysis, Synthesis and Evaluation. The revised framework, as proposed by Anderson and Krathwohl (2001) swaps the last two levels and renames Synthesis as Creating, and change the levels' names to verbs. Both frameworks cognitive domain levels are described in Table 1.

One of the changes that can be seen in the revised taxonomy regards the addition of one extra dimension to the framework. According to Anderson and Krathwohl (2001), the description of learning outcomes are composed by two parts, some subject or matter (described by a noun or noun phrase) and the description of what is to be done with it (described by a verb). In the original taxonomy, the knowledge level contained both elements of the subject and the usage. In the revised version a structure of the knowledge dimension is proposed and divided in the following categories:

A. Factual Knowledge - The basic elements that students must know to be acquainted with a discipline or solve problems in it.

Aa. Knowledge of terminology

Ab. Knowledge of specific details and elements

B. Conceptual Knowledge - The interrelationships among the basic elements within a larger structure that enable them to function together.

Ba. Knowledge of classifications and categories

Bb. Knowledge of principles and generalizations

Bc. Knowledge of theories, models, and structures

C. Procedural Knowledge - How to do something; methods of inquiry, and criteria for using skills, algorithms, techniques, and methods.

Ca. Knowledge of subject-specific skills and algorithms

Cb. Knowledge of subject-specific techniques and methods

Cc. Knowledge of criteria for determining when to use appropriate procedures

D. Metacognitive Knowledge - Knowledge of cognition in general as well as awareness and knowledge of one's own cognition.

Da. Strategic knowledge

Db. Knowledge about cognitive tasks, including appropriate contextual and conditional knowledge

Dc. Self-knowledge. (L. W. Anderson & Krathwohl, 2001, p. 3)

The objective of this taxonomy is providing a model to fit educational objectives, goals and even standards (L. W. Anderson & Krathwohl, 2001). There are many examples on educational literature that uses this division to propose practices, formulate curricula and evaluate outcomes (Aljarallah, 2011; Hoeij, Haarhuis, Wierstra, & Beukelen, 2004; Palmer & Devitt, 2007; Whiteley, 2006).

This framework can be pivotal in understanding whether a task is requiring the usage of higher order thinking skill or not, both for assessing these levels and for promote development through practice.

Table 1 - Original and Revised Bloom structures

Structure of the Original Taxonomy	Structure of the Cognitive Process Dimension of the Revised Taxonomy
1.0 Knowledge	1.0 Remember - Retrieving relevant knowledge from long-term memory.
1.10 Knowledge of specifics	1.1 Recognizing
1.11 Knowledge of terminology	1.2 Recalling
1.12 Knowledge of specific facts	
1.20 Knowledge of ways and means of dealing with specifics	
1.21 Knowledge of conventions	
1.22 Knowledge of trends and sequences	
1.23 Knowledge of classifications and categories	
1.24 Knowledge of criteria	
1.25 Knowledge of methodology	
1.30 Knowledge of universals and abstractions in a field	
1.31 Knowledge of principles and generalizations	
1.32 Knowledge of theories and structures	
2.0 Comprehension	2.0 Understand - Determining the meaning of instructional messages, including oral, written, and graphic communication.
2.1 Translation	2.1 Interpreting
2.2 Interpretation	2.2 Exemplifying
2.3 Extrapolation	2.3 Classifying
	2.4 Summarizing
	2.5 Inferring
	2.6 Comparing
	2.7 Explaining
3.0 Application	3.0 Apply - Carrying out or using a procedure in a given situation.
	3.1 Executing
	3.2 Implementing
4.0 Analysis	4.0 Analyze - Breaking material into its constituent parts and detecting how the parts relate to one another and to an overall structure or purpose.
4.1 Analysis of elements	4.1 Differentiating
4.2 Analysis of relationships	4.2 Organizing
4.3 Analysis of organizational principles	4.3 Attributing
5.0 Synthesis	5.0 Evaluate - Making judgments based on criteria and standards.
5.1 Production of a unique communication	5.1 Checking
5.2 Production of a plan, or proposed set of operations	5.2 Critiquing
5.3 Derivation of a set of abstract relations	
6.0 Evaluation	6.0 Create - Putting elements together to form a novel, coherent whole or make an original product.
6.1 Evaluation in terms of internal evidence	6.1 Generating
6.2 Judgments in terms of external criteria	6.2 Planning
	6.3 Producing

Source: Adapted from (L. W. Anderson & Krathwohl, 2001)

2.3.2 Critical Thinking

Critical thinking is a recurrent topic in the discussions about education and in order to discuss it systematically, a definition is needed.

The understanding and study of critical thinking is more than a century old, and one of the most relevant definitions for the understanding of the current use of the term was given by Dewey not directly on critical thinking, but rather on reflective thought:

Active, persistent, and careful consideration of a belief or supposed form of knowledge in the light of the grounds which support it and the further conclusions to which it tends. (Dewey, 1909, p.6)

The main aspects of critical thought can be seen in this early definition, namely the active reflection on thought itself, the evaluation of its foundations and expected consequences.

Extrapolating the “grounds which support it” to an evaluation of the context in which this type of thinking is inserted, this definition takes a form more closely related to those used in modern articles (Barnett & Francis, 2011; Niewoehner, 2006; Whiteley, 2006). One of the most relevant definition of critical thinking for the scope of this work is that given by Paul and Elder (Paul & Elder, 2008): *“Critical thinking is the art of analyzing and evaluating thinking with a view to improving it.”* Despite its simplicity this definition takes into account the main aspects of the process of critical thinking, being corroborated by many authors. Its direct relation to the Bloom’s taxonomy, relating to the fourth and sixth level of the original taxonomy and the fourth and fifth level of the revised, increases this definition usefulness.

In a similar direction goes the definition that is given by Ennis (1993) *“Critical Thinking is a reasonable and reflective thinking focused on what to believe or to do”*.

In his literary review of critical thinking (Petress, 2004), the authors gives different sets of definitions, varying according to area, level of specificity and length. Despite some differences, the characteristics of questioning, contextualizing, analyzing, evaluating and reflection are present in the majority of the definitions. These characteristics agree with and complement those mentioned before.

2.3.2.1 Why Critical Thinking

The development of critical thinking serves many purposes not only in the process of empowering engineers as professionals, but also as people.

The role of critical thinking in problem solving, evaluating sources of information, understanding a person's impact on his or her surroundings and directing efforts in a more meaningful way are general reasons for any professional to develop this competence, in engineering however there are even more reasons. The lack of space for discussion and open debate in strongly technical engineering subjects is another issue. While professors place the development of critical thinking as high priority, few can clearly delimit their teaching practices to promote this ability (Niewoehner, 2006). In the engineering competences discussion, the models presented consolidate the importance of critical thinking in the profession. According to Duderstadt (2008) some of the desirable characteristics of the profession are having global understanding of their work, working in increasingly multidisciplinary projects, providing end-to-end solution while keeping the commitment to the role of an engineer, and the understanding of the ethical implications and societal impact of their work. The CDIO Syllabus regards critical thinking in items 2.5 – ethics, equity and other responsibilities, 4.1 – external, societal and environmental context and 4.2 – Enterprise and business context. ABET objectives requires this competence in items (b) - an ability to design and conduct experiments, as well as to analyze and interpret data, (e) - an ability to identify, formulate, and solve engineering problems, (f) - an understanding of professional and ethical responsibility and specially (h) the broad education necessary to understand the impact of engineering solutions in a global, economic, environmental, and societal context.

In their work about the future of engineering education, despite its age providing a relevant discussion, Rugarcia, Felder, Woods and Stice have select “problem solving, critical thinking and creative thinking skills” as one of the key development points (Rugarcia et al., 2000).

2.3.2.2 Developing Critical Thinking

Researchers have been worried about fostering critical thinking skills for more than a century, and many studies have been made in order to comprehend how thinking skills can be developed. Some of the main answers proposed is skillfully asking questions, inquiring, debating and reflecting on the thought process itself (Cottrell, 2005; Ennis, 1993; Llano, 2015; Wendland, Robinson, & Williams, 2015).

Bailin and Battersby (2015) have written an article proposing a new way of teaching critical thinking, by teaching inquiry principles and promoting its practice. Their initial proposal is to either include an extra subject to develop such competence or integrate inquiry in other subjects, a proposal that they see as more feasible.

Since a discussion about bloom taxonomy was already started, it is relevant to understand the proposed link between this taxonomy and development of this competency. The relationship between the high-order thinking skills and the three upper levels of either the revised or original Bloom taxonomy is stated by many authors (Barnett & Francis, 2011; Kong, 2014; Whiteley, 2006). According to Ennis however (Ennis, 1993) the concepts presented by the taxonomy are not enough to provide a clear understanding of how to assess critical thinking.

By using this taxonomy in order to determine which question require or not critical thinking, it is possible to direct assessment practice towards activities that demand this kind level of reasoning, and according to aforementioned authors, fostering its development.

Due to the role of questioning in the development of critical thinking, while some authors defend that multiple choice questions can be used to assess critical thinking skills, the very definition of it makes this set of abilities closer to what is expected from essay questions. The aforementioned abilities of reflecting upon the credibility of information, understanding and contextualizing relations and the scope of statements and solutions, assessing the very process of thinking (Ennis, 1993) are discursive by nature and as such closer in format to unstructured questions.

2.3.3 Creativity

The classical definition of creativity has two key elements, creativity requires both originality and effectiveness (Runco & Jaeger, 2012). In order to better understand this concept it is helpful to look the characteristics of works and results deemed creative where according to Margaret et.al there is a significant level of consensus (Edwards, McGoldrick, & Oliver, 2006).

Despite the many theories and frameworks, Jackson and Shaw agree on the consensus in the actual characteristics of creativity, summarized by the author as being imaginative, original, exploring for the purpose of discovery, using and combining thinking skills and creative communication (Jackson & Shaw, 2006).

In her survey with fellow professors from the National Teaching Fellows, Fryer (2006) asked about which of these aspects pertain to creativity. The results presented in Table 2 are similar to those proposed by Jackson and Shaw.

Table 2 - Aspects of Creativity According to the Fellowship Professors

Aspect of Creativity	%
Imagination	90.0
Seeing unusual connections	86.7
Combining ideas	80.0
Original ideas	80.0
Innovation	76.7
Thinking processes	72.2
Discovery	66.7
Invention	61.1
Generative thinking	53.3
Self-expression	52.2
Valuable ideas	52.2
Sudden Inspiration	51.1
Analytical thinking	44.4
Awareness of beauty	25.6
Aesthetic products	21.1
Unconscious activities	21.1
Tangible products	18.9
Mysterious processes	14.1
Other	14.3

Source: (Fryer, 2006)

In his lecture Ken Robinson states that one of the prerequisites of creativity is divergent thinking, the capacity of generating alternatives and combining things in novel ways independent from being useful or not, and creativity as the process of having original that have value (Robinson, 2009).

Being creative is, summarizing the aforementioned definitions, to be able to contribute in an original or innovative way, so to develop creativity skills is to increase one's ability to create ideas of value. Developing creativity competence would require connecting this skill with engineering knowledge and the attitude of applying for its problems.

2.3.3.1 Why creativity

Creativity has been chosen as one of the studies focus in this work for three main reasons, its importance in the engineering context, lack of attention in engineering education environment and impact of assessment in creativity development.

Regarding importance, the increasing level of change in society associated with the change from a manufacturing to a knowledge centered economy makes the development of creativity increasingly important. In the specific competence frameworks discussed the importance of creativity is stated in: According to Duderstadt in the characteristic of the profession as Innovative and directly in the section "New R&D Paradigms" under "Stress on creativity/innovation" (Duderstadt, 2008); the CDIO syllabus items 2.1 (problem solving), 4.3 (conceiving) and 4.4 (designing) (Crawley & Lucas, 2011); in ABET in items b and c due to design requirement and e regarding real world problem solving (Commission, 2016).

While there have been movements from teaching to learning in later years, many of the current practices in education in general are being accused of diminishing creative thinking (Beghetto, 2005; Min-sik, 2015; Robinson, 2009). Understanding better if, how and why this takes places can enable changes, be it to decrease the negative impact, eliminate it or, hopefully, nurture the development of this characteristics.

2.3.3.2 Developing Creativity

Creativity and its development process are still a matter of much debate. There is little basis of understanding on how to foster and what are the workings of creativity. As

with many other activities however, many researchers believe that the practice of creative freedom and an environment that foster this characteristic lead to positive outcomes.

In her article about Creativity in Schools, Anna Craft (Craft, 2006) reviews a collection of works from different lines of thought as well as different time frames to compile a list of factors that can help developing creativity in children. For this work and in the context of higher education the most relevant are: the study of disciplines in depth, beyond the immediate experiences and observations; involving students in the creation of new routines; provide an environment where students can go beyond what is expected and reward them for doing so; provides means for students to find personal relevance in the activities; incorporate alternatives when imparting information; given time to incorporate ideas.

In proposing a framework for the conditions of emergence of creativity in higher in education Paul Tosey (2006) proposes the usage of Richard Seel (2005) conditions for the emergence of creativity in higher education, these conditions were originally devised for organizations:

- Connectivity – emergence is unlikely to occur in a fragmented world.
- Diversity – if there is not enough diversity in the system, it will be hard for different patterns to emerge.
- Rate of information flow – either information overload or too little information flow can make emergence unlikely.
- Lack of inhibitors – e.g. inappropriate power differentials, too much anxiety or threats to identity can all inhibit emergence.
- Good constraints to action – effective boundaries can enhance the possibility of emergence.
- Positive intention – a clear sense of purpose can influence the chances of emergence occurring.
- Watchful anticipation – while a clear sense of purpose can influence the chances of emergence occurring, a period of expectant waiting is often necessary to facilitate emergence.

These conditions lead to many changes in the way higher education institutions should work. The separation in subjects, classes mainly organized with people from similar

areas, power differences, lack of sense of purpose and deadline managements are topics that should be reviewed to facilitate the emergence of creativity.

In his research on Students' experiences of creativity, Martin Oliver et.al. (2006) state that one of the perception regarding creativity and assessment is that objective exams have are more problematic than essays tests.

In an article reviewing creativity based on the views of the members of the "National Teaching Fellows", Marilyn Fryer (2006) points out the following as approaches as practiced for developing creativity:

- Providing stimuli for imaginative thinking or heuristic strategies;
- Learning in a particular context, as a real situation, or providing a context for creative thinking;
- Showing examples of creative thinking and solutions;
- Providing supportive factors, such as in the relationship between tutor and students;
- Fostering personality characteristics as risk taking and building self-confidence;
- Teaching skills for use in creative work;
- Setting tasks which require creativity;
- Developing students' motivation.

Another work that proposes practices to foster and help teaching creativity is that of Edwards (Edwards et al., 2006)

- Developing critical thinking
- Encouraging lateral thinking
- Move between the university and outside – as proposed tactics the author suggests case studies, working with external teachers, visiting specialists, among others.
- Give space for group work
- Increase student confidence
- Have fun – where the authors describe manners of making the learning environment more fun.

Developing critical thinking, encouraging lateral thinking, encouraging students to take risks, breaking divisions of area, knowledge or subject, giving time for the

development of answers and rewarding creativity are the key factors taken from the literature in order to help the development of this skill.

2.3.4 Motivation

One of the recurrent themes in advancing education in general has been how to improve students' motivation. In their article, Williams and Williams compare theorists working towards the understanding motivation to blind people trying to describe an elephant. Even though each may be right in their description of the phenomena, one that compares it to a spear has a similar hold of the truth as the one who compares to a broom (fangs and tail), the problem is complex enough to make explanations seem completely unrelated (Williams & Williams, 2011). With this in mind, it is still possible to use each part description attempting to improve education.

Perhaps the most relevant article to the understanding of motivation in this document context is delivered by Ryan and Deci (Ryan & Deci, 2000). Motivation is described as falling in one of three categories, the extremes intrinsic and amotivation, and the intermediate level of extrinsic motivation. The extrinsic motivation is further divided into four categories.

Intrinsic motivation regards the condition of coincidence of the goal and the activity or objective. In this case, the task would be performed by its own sake. As an example, studying calculus for the pleasure of it or for the improvement of student abilities in the subject would be regarded as being intrinsically motivated activities. Skinner (Skinner), according to whom everything is motivated by rewards, explained the same phenomena by stating that the reward in this case would be the activity itself.

Some activities called intrinsically motivated, but they are only named thus due to the amount of people who find an end in itself (Ryan & Deci, 2000). Even though there are no universally intrinsically motivated activities, internal motivation is more common to some.

The amotivation case regards the simple lack of motivation, meaning that no effort will be made in order to perform the activity. At this stage the activity itself is considered to the amotivated observer as irrelevant to his or her objectives (Ryan & Deci, 2000).

Regarding the assessment influence on the students, extrinsic motivation holds the highest value. It is further divided in four sub-categories, from the most external to the

most internalized: External Regulation, Introjection, Identification, and Integration. The more internalized a motivation is for a person, the more he or she is self-determined, and thus less influenced by environmental variation in that specific activity or behavior. The most environmental susceptible of the subcategories is External Regulation. In this category, rewards and punishment are the real drivers of motivation. An example of such is a student whose driver for studying a given subject is the cost of failing or the grades given. In this case, there are several steps between the actions (studying the subject) to the motivation (not failing).

Introjection is a type of motivation based on psychological pressure, such as other's expectations or a sense of pride. Since it is based on an internal reward or punishment, one own feeling as motivator, the level internalization is greater than that of External Regulation. It can be strongly related to self-esteem, where the person copes with the activity in order to maintain or attain a sense of self-worth. In daily routine, this might appear as studying a course because of parental expectation, attributing one's sense of worth to how well you perform in the university or even believing your grades are a measure of intelligence. In these cases, the activity is performed to provide an internal reward.

Closer to the intrinsic motivation, there is Identification. When there is a clear understanding on how performing an activity contributes to inner goals, thus the importance of the task and the pressure to perform it comes from the person alone. A student that wants intrinsically to be an engineer and believes the mathematics subjects are a requirement for achieving this goal would have such a motivation, the motivation is attached to the goal itself. When an activity is in this level of motivation, the external influence is less significant and coping with difficult tasks becomes more natural.

The most intrinsic level of extrinsic motivation is the Integration level. This last level has assimilated the identified reasoning to the very self. According to the authors, the process by which this change occurs is *"through self-examination and bringing new regulations into congruence with one's other values and needs."* Even though it shares many characteristics with true intrinsic motivations, namely being autonomous and lacking conflict, it is considered extrinsic because the objective or valued outcome is not the behavior or activity itself (Ryan & Deci, 2000). Figure 5 shows a framework locating all described motivation levels.

Figure 5 - Motivation Framework

REGULATORY STYLES	Extrinsic Motivation					Intrinsic Motivation
	Amotivation	External Regulation	Introjection	Identification	Integration	
ASSOCIATED PROCESSES	Perceived non-contingency	Compliance/Reactance	Ego involvement	Conscious valuing of activity	Congruence	Interest/Enjoyment
	Low perceived competence	Salience of extrinsic rewards or punishments	Focus on approval from self or others	Self-endorsement of goals	Hierarchical synthesis of goals	Inherent satisfaction
	Nonrelevance					
	Nonintentionality					
PERCEIVED LOCUS OF CAUSALITY	Impersonal	External	Somewhat External	Somewhat Internal	Internal	Internal

Source: Adapted from (Ryan & Deci, 2000)

Some professors end up assuming that students must have intrinsic motivation and study for several hours for the sake of the topic by itself. While working only with intrinsically or integrally motivated students would be theoretically ideal, it would also be a waste of human capability. Even if to perform well as an engineer a broad range of abilities are required, this does not imply the obligation of thoroughly understanding and internalizing the importance of each concept. In the real classroom environment, many of these concepts should simply be learned.

A distinction that is closely related to this framework is the that of learning for mastery and learning for performance. The first relates to the practice of learning to develop oneself in the given field, the second the practice of learning to achieve a good performance as measured by grades and other activities. Learning for mastery is associated with better perceived results (Black & Wiliam, 1998). The first is clearly more intrinsic as the interest is in the subject itself, while the second practice is closely related to the External Regulation and Introjection levels.

Regarding assessments, there are at least two viable approaches dealing with motivation. First, it is possible to create more appealing challenges and tasks, thus trying to amplify their internalization levels (such as by using games). Secondly, by using the external rewards of assessments to promote the developing of more competence oriented tasks.

2.4 OBJECTIVES AND LEARNING OUTCOMES

To direct efforts in the classroom and be able to direct learning, having goals is essential. One of the most widely used tools for such practice is learning outcomes, but devising these to serve as guidelines useful for making decision and improving learning experience is another issue.

Learning outcomes can be understood in two different manners. The first as voluntary tool to guide the professional practice in order to assist the learning process and create learning experiences aligned with broader objectives and the second is as tool of accountability and assessment in order to validate and measure the success of different institutions, courses and professors (Bennett & Brady, 2014; Maher, 2004).

In their work about the revision of Bloom's taxonomy, Anderson and Krathwohl (2001) state usual the description of learning outcomes in two parts, the knowledge or subject that is to be learned, and a verb signaling what is to be done with or to that content. This division was a driving force in the proposition of a new dimension in the revised taxonomy and can serve either of the proposed descriptions.

While working with learning outcomes can be valid in understanding what is to be expected from a course, guide assessment practices or even defining prerequisites of learning, it can make the course rigid if followed too strictly and become a simple bureaucratic step when forcefully pushed on professors. Regarding the usage of institutionalized learning outcomes, Professor Frank Fureti (2012) point out four negative points:

- Disrupting the conduct of academic relationship between teacher and student – The professor states that considering the rigid characteristics of learning outcomes, in this text closely related to a promise of learning, can hinder the experience of experimentation and discovery when these are not following the outcome intended path. The professor also states that learning outcomes do not encourage dialogue but a one-sided teaching process.
- They foster a climate that inhibits students and teacher to deal with uncertainty – The argument is that the clarification promoted in the learning outcomes is a futile attempt of gaining certainty through process. He claims that what is gained is not clarity, but rigidity.

- Learning outcomes devalue the art of teaching – Professor Fureti states that learning outcomes reduce the need of professors to think by defining beforehand what should be reached. He continues by saying that “The art of teaching also requires a readiness to treat different students differently. The more adventurous students will flourish only if they are let off the leash and given space to experiment [...]”, which could be inhibited by the usage of learning outcomes.
- Breeds a culture of cynicism and irresponsibility – Fureti defends this argument by saying that the initial creation of learning outcomes was to hold institutions accountable for their teaching and that the focus becomes not the learning and experience of the student, but whether he or she has achieved the desired outcomes or not.

A second article presenting critiques to the usage of learning outcomes is that of professor Ian Scott (Scott, 2011). He points out three main issues with this tool.

- Find an adequate level of detail – The trade-off between accuracy and scope. Too broad learning outcomes do not deliver the promise, too narrow and they over limit the learning experience.
- They are not really student centered – Scott also states that to be student centered, the student should have voice regarding both what is taught and how. Previously defined learning outcomes do not meet this criterion.
- Issues with assessment – If achieving the learning outcomes is prerequisite to approval, should a student who achieved genius level in 80% of the intended outcomes but failed one be approved? By basing assessment exclusively on learning outcomes, the professor misses the opportunity to assess unintended, even though many times relevant, learning outcomes.
- Bennet and Brady (Bennett & Brady, 2014) have published a fierce critique on learning outcomes as assessment practices, however they make it clear critique is about their use not to foster learning, but to hold institutions and professors accountable.

It is noticeable that most of the critiques described are related to the usage of learning outcomes as a tool of accountability. The rigidity of the process, the promise of learning and the exclusion of the student from defining what are those learning outcomes can

be better dealt with when learning outcomes are regarded as guidance tools. The trade-off issue described by Scott (Scott, 2011) is a more complicated matter, but is also related with the need for precision, if these outcomes can be proposed or even discussed with students, this issue can be mitigated.

It is not the intent of this work to propose practices that accurately measure students abilities or to advocate the use of standardized measures of comparison, so in this context, in agreement with the conclusion proposed by Scott (Scott, 2011) learning outcomes should serve as guidelines of a desirable destination not as an strict travelling plan.

In her article Angela Maher (Maher, 2004) gives recommendations on using leaning outcomes, of special interest are: developing a broader conception of learning outcomes and making learning outcomes congruent with good learning. The broader conception recommendation deals with a more fluid concept of learning outcomes reached by active dialogue with the involved parties, and aiming at a level of specificity that is both responsive and flexible. The good learning and teaching practices deals mainly with managing the ambiguity that is inherent to a learning environment and maintaining the ability to fetch unplanned outcomes possibilities of learning.

2.5 ASSESSMENT THEORY

The relevance of assessment in the learning context is well documented (K. M. Scouller & Prosser, 1994). While some students are more interested in learning than in their grades, according to Fransson (1977) students tend to focus their learning on the abilities and knowledges that they believe they will be asked to demonstrate, “[...] prepare for what they expect to be the performance requirements [...]”. This view is consistent with the presented review on motivation, where the most extrinsic level is related to searching directly for what bring an external reward.

Due to the great importance of the goals of education in assessment, before discussing it, an introduction on the traditions of learning theory is presented aiming to clarify the different takes on the roles of assessment.

Following this discussion, the three main types of tests, summative, formative and diagnostic are presented.

2.5.1 Learning Frameworks on assessment

There are many different sets of frameworks, some of them theoretically fresh and rather untested, other well established but nonetheless problematic.

Behaviorist assessments are considered by many theorists as the most widespread type of evaluation in schools worldwide (Fernandes, 2006). They focus on the measurement of acquisition of desired knowledge, transferred from an expert (teacher) to a student. Fail in behaviorist assessments are by definition wrong behaviors that should be eliminated (Helena & Santiago, 2015). Fernandes also states that this leads to restrictive and limited assessments, reduced to punctual measurements of desired behaviors, even in their formative proposal.

The cognitivist theory is more focused on the process of learning and understanding mistakes as a natural part of the learning process. As such, learners should face new concepts interact with them and develop their own way through knowing. Assessments as means to measure up students achievements find little support in this context since knowledge and development are view as not measurable (Helena & Santiago, 2015). On the other hand, presenting challenges to create unbalance in order to promote learning opportunities is highly recommended.

In the humanist framework, challenges for development are also welcome, but standardized assessments are also disregarded. The main approach to evaluation of goal achievement is self-assessment (Helena & Santiago, 2015).

The critical learning school has another approach where the most capable evaluator of a learning process is the learner himself or herself followed by his or her peers, where the educator can be situated, (although in a minor degree). In this context, there is no use for a standard or comparative measure of learning. Again, challenges presented in integrated context are also welcome as opportunities for mutual learning. Summarizing, while there is a general acceptance of the importance of challenges for learning, the value and accuracy of knowledge measurement for comparison and classification is not accepted by most learning theories.

2.6 ASSESSMENT TYPES

Assessment can be categorized according to many different criteria. Tests can be high or low stakes, formal or informal, oral or written among many others. One division however stands out in terms of usage and clarity, the threefold division in Diagnostic, Summative and Formative. It is important to notice that this classification does not imply that these categories are mutually exclusive, meaning that a formative or summative test can be also used for diagnostic purposes, however assessment practices devised for different functions lead to different practices (McAlister, 2005). These three types of test are described below.

2.6.1 Diagnostic Assessment

Diagnostic assessments are, as the name suggests, used for diagnosing. They differ from the other two types because their objective is neither directly to promote learning, although they can (and should) be used as auxiliary practices for doing so, nor to assess how much has been achieved. Their intent is help understanding the current state of classroom. Although they can relate directly to the topic of the subject, many times it is not so.

One of the key aspects this type is to actually verify the existence or lack of a pre-required ability or competence in order to evaluate if the learners are ready to proceed to in the learning process or if some corrective measure should be taken (Freitas et al., 2014).

Another type of assessment that would fall into this category is that of individual and or group characteristics definition. Many factors not directly related to subject being taught are proposed to benefit the courses. Students favored group roles, their expectations on the subject, their main learning preferences, their psychological profile and many other types of characteristics evaluating tests are considered diagnostic assessments. Busato has developed studies to try to understand the impact on academic achievement based on a set of characteristics, the types of tests used are mostly diagnostic in nature (Busato, Prins, Elshout, & Hamaker, 1998, 2000).

2.6.2 Summative Assessment

Summative assessments have as their main objective to grade and/or classify students based on their learning outcomes. While most of the modern learning theories disregard the validity of measuring knowledge and comparing grades among different learners, stopping to do so would represent a major disruptive change in higher education as it is. The vast majority of universities around the world use this type of assessment as the primary method for accountability and degree awarding.

Nowadays summative assessments have a large variety of attributions such as classifying students, determining capabilities, defining grades and even certifying professionals. While it can be dangerous not having some extrinsic level of certainty on a student capacity of performing his or her profession, simple summative assessments are not much better in inferring this (Harlen & Deakin Crick, 2002).

Motivation and summative assessments have a two-fold relation. On one hand the stakes of this type of test can, through mechanism of Introjection and External Regulation, motivate the student to prepare and work in alignment with the required skills to perform well. On the other there is a connection between the perceived importance, mainly grade-related, of the test and the student result, higher stakes correlate positively with better results (Outcomes, Education, & Role, 2013).

Final exams and written tests have several problems in inferring even knowledge by itself. They are more closely related to picture of current psychological and test preparation state than to real measure of the state of mastery.

The usage of learning outcomes in assessment theory is mainly criticized due to the summative aspect of the current practice more than by issues of the outcomes themselves (Bennett & Brady, 2014).

2.6.3 Formative Assessment

Formative assessment as a concept deals only with tests, activities and measurements used to improve the quality of learning. They are composed by challenges, questions, projects and observations that can provide feedback to be applied in the learning process.

The nature of the feedback can vary significantly. It can be used to facilitate understanding, to suggest alternatives, to reformulate previous concepts, to challenge conclusions, to help developing critical analysis, among many others.

This alternative is not used in the end of the learning process, but in the process itself. The objective is assessment for learning. The more frequently it occurs, the faster difficulties can be addressed, and according to Paul Black and Dylan Wiliam (1998) the better students learn.

It is important to note that tests can be both summative and formative in nature, but while the first is mainly attempting to categorize, the second is concerned with the opportunity of development given to the learners. These opportunities can happen during the test, due to its impact on their organization and motivation, to the feedback provided, or even to the changes in their course that their results can lead towards (an intersection with diagnostic tests).

2.7 ASSESSMENT FREQUENCY

The frequency of assessments is a relevant topic in order to promote learning. The impact of increasing the number tests has been studied for over fifty years, and some authors found relevant evidence that there is a correlation between the frequency and learning.

Many of the assumptions on assessment frequency possibilities have been made before the worldwide spread of the internet or even personal computers. While it would be an impractical task for a professor to personally evaluate and/or grade daily (or even more frequently), computer based tests have made it relatively easy.

One of the main summative assessment practices regarding frequency is based on a low number of high stakes tests, often no more than three. A key limitation of this model is the low level of adaptability that is provided since these assessments are normally used to close topics, losing significant values as a course feedback tool. Another limitation is the accumulation of topics and learning outcomes that can lead to an unfair selection of coverage. According to Paul and Dylan (Black & Wiliam, 1998) many studies have shown a positive correlation between increase in the frequency of assessment and learning.

When deciding on assessment frequency, it is also necessary to consider the size of the classes, the more students there are the more difficult it is for the professor to provide useful feedbacks and grade their activities. This issue is mitigated by tools such as computer aided assessments and peer and self-grading. Some techniques and instructions, such as rubrics, can diminish the gap between students grading and feedback and that of a professor.

Khan academy and Duolingo are two web-based courses that use an inverse approach to learning (Khan, 2011; Vesselinov & Grego, 2012). Learners should be able to perform tasks, and even though both initiatives provide content related material, how they learn is the student's concern. This is done by testing small steps in the process and if a student gets the correct answers with a sufficiently low error rate, they are able to proceed. Even if they are not approved, if they decide to do so, they can advance.

This theory can be applied with little adapting effort to higher education. As long as the learning outcome is reached, does it matter how many times the student had to redo the task? When students fail a course, a significant amount of effort, time and energy is wasted. The traditional process would be to try the subject again, and when the new score is given that is de facto measured level of knowledge. Since computers can repeat not only classes but even the smallest concept repeatedly, students are able to fail each part as many times as they need, learning from each attempt.

Although not a qualitative difference, smaller parts can allow a better understanding of the difficulties saving both time and effort. On the other hand, breaking down subjects can lead to fragmentation and consequently decrease the global understanding of the topic.

Some situations allow getting the best of both alternatives regarding general and specific focus. The just-in-time approach to learning, often used in inductive learning, assumes that when a piece of information, concept or skill is needed, it is learned. As such the student will be solving a general problem (global) obtaining the specific knowledge (local) that is required. In this case, the more specific the retrieved is information the better, since it can be quickly applied to a broader problem. Changing the assessment frequency in this context can also lead to a trade-off (assuming a well-dimensioned task), this time between the coverage of the problems (lower frequencies)

and the capacity to generalize, and perhaps understand, the concept (higher frequencies).

Whether using frequent or infrequent assessments, it is important to understand why such a choice was made. Why only one test a semester, or two? What are the impacts of starting every class with an assessment, or each module?

Different objectives and constraints can lead to different proposed frequencies, however making a conscious choice on this variable can be quite positive. The time constraint, both for professor and for students is a key issue in increasing the quality of assessment practices. Even in a condition where the number of feasible tests and activities is low, it is better to try and divided proposed concepts logically than simply splitting the semester in two or three parts.

One of the challenges of this approach, besides the designing costs, is the balance of the aforementioned trade-offs.

2.8 ASSESSMENT TASKS

2.8.1 Structured Tasks

Structured tasks are those that have a limited variety of response options. They can take many forms, true or false, multiple choice, multiple response, matching and fill in the blanks being the most used types of structured questions, but despite their spread, they are subject to many critics and heated debates.

Their limited set of answers makes them easier to grade and correct but can often change answering from a development to a simple recognition. While their grading can seem more just, given that a well-constructed answer should have clear arguments for making it the best alternative, they are more limited in measuring partial and lateral knowledge, and often higher cognitive levels. Being pre-defined, they limit the learner's expression to the selection of test designer provided options.

Comparing with open-ended and unstructured tasks however, this more recent development has a significant advantage, they can easily be computer graded. While some years ago this difference could be considered of small relevance, except perhaps

for national and other massive institutional tests, with the recent and increasing penetration of computers in classrooms it is now more relevant. The larger the class is the more difficult it gets to grade and review tests, hence the increasing importance of computer grading, especially when considering the huge sizes of Massive Open Online Courses (MOOC) classes. This advantage can lead to a qualitative difference in assessment techniques, adaptable assessments.

Another key aspect of structured tasks is that they are normally less time consuming for test takers than unstructured tasks. This allows, or at least facilitates, practices such as chapter quizzes, assessments every class, multiple repetition tests among other practices that would be too time consuming if manually graded (McAlister, 2005; Race, 2015).

There is consensus that despite their capabilities to assess high order thinking, the three higher levels of Bloom's taxonomy, to develop tasks to do so is time consuming and regarding these thinking levels, the expression and feedback characteristics of open ended tasks may support their usage.

In this section, each of the main types of structured section will be shortly described and then reviewed regarding its strong and weak points, general usage strategy and designing.

2.8.1.1 Type and Usage

The sheer variety of possible ways and forms to build structured tasks makes the description of every type unfeasible. The types selected to be described in this context are the most used, cited and widely used, be it to assess learning or for learning.

2.8.1.2 Multiple Choice Question

Multiple choice questions consist of primarily three parts, the stem, distracters and key. The stem is the description of task, where the conditions for choosing the alternative are given and, despite the name of multiple choice questions, many times it can be an affirmative or imperative task. The distracters are the part of alternatives designed to distract the test taker from the best suited, many times simply called correct, alternative. The key is the best suited alternative itself.

Developed as a testing tool in 1913, and first applied in large scale in 1917 for measuring intelligence in the army of the United States, multiple choice questions have caused much impact in assessments technics. After being applied to more than 1.7 million people, the results this first test were deemed inconclusive by the military, which did not stop its spread to the education field (Trewin, 2007)

They have some important advantages and are in the core of many teaching programs, especially online, but some of their limitations have put this kind of task in the spotlight of many heated debates (Min-sik, 2015; Rodríguez-Díez, Alegre, Díez, Arbea, & Ferrer, 2016; Seçmeliye et al., 2016).

The study of multiple choices is aided by a huge literature body, mainly because of its uses for standardized tests. Multiple choice questions are arguably the most studied type of task, partially due to the aforementioned reason, in forms of application, objective, techniques and delivering tactics. This variety, however, does not imply a similar variety of uses especially regarding the disputes in the cognitive levels assessed (Boud & Falchikov, 2007; K. Scouller, 1998).

2.8.1.2.1 Strengths

Clarity of given tasks – It is much easier to direct the student reasoning towards intended learning outcomes when using this type of task (Zimmaro, 2004).

Getting diagnostic information from incorrect alternatives – When properly designed for accounting for the most likely mistakes, the distracters can help understanding where the student has made a mistake. This ability is a strength compared to other structured tasks (Toolkit, 2008; Zimmaro, 2004).

Ease, Objectivity and Reliability of scores – Since scoring can be easily automated, it is easier when compared to unstructured tasks. For the same reason it is possible to state that it is more objective. When considering unstructured tasks, the scores are more reliable, since they do not depend on an evaluator to grade (McAlister, 2005; Toolkit, 2008).

Difficulty analysis – It is possible to easily analyze numerical data regarding percentage of correct answers, well and poorly designed distracters, poorly written stem, under discussed topics and how effectively students are meeting learning outcomes.

Comparing performances – This type of task enables comparisons among students from different years, classes, professors or whatever group test takers can be divided in with ease. To do the same with essay questions would require much more effort in guaranteeing the score reliability, which in this task is given. If aligned with proper learning outcomes, this can mean feedback on the progress of the class as well as on the fly reports of difficulties and obscure points (Zimmaro, 2004).

Test Coverage – Since a MCQ is normally much faster than unstructured tasks, a lot of material and learning outcome topics can be covered in a relatively short amount of time. This is so for both formative and summative tests, with the aid of computers or smart phones creating the possibility of even multiple tests per class (McAlister, 2005; Zimmaro, 2004).

Can be written for student to differentiate levels of correctness – This is a strength compared to other structured tasks. True or false, match the alternatives and fill in the blanks for example lack this advantage, on the other hand essay questions, for example, can ask student not only to differentiate, but also debate the level of validity of different statements (Zimmaro, 2004).

Recyclable questions – While it is time consuming to create any type of task, MCQs are especially reusable. Given that the outcomes are clearly defined, it is possible not only to search huge databases worldwide, but even generate different tests on the fly by recombining and using the tasks in a distinct context. The sheer amount of available MCQs allows the test designer to focus his or her efforts in the faster task of selecting and adapting tasks rather writing from scratch (Toolkit, 2008).

Can assess high order thinking – This is subject to much debate, but many authors defend that it can be done, but even in this case, the tasks are to be designed for such end. This questions however normally take more time for the students to answer (Toolkit, 2008; Zimmaro, 2004).

2.8.1.2.2 Limitations

The following limitations of multiple choice questions were found:

Time consumption of design a proper task – Not only must the stem and key be well written, but also the distracters. The process of creating multiple plausible non-scoring

answers is time-consuming since not only the content, but also the form of the alternatives must be well thought of to discourage student from taking an informed guess. There are answering techniques based on the size, repetition of arguments in alternatives, choosing one of opposing alternatives and many others. (McAlister, 2005; Toolkit, 2008; Zimmaro, 2004)

Difficulties in distracter design – In order to create plausible and effective distracters there are limitations regarding size, grammar forms, arguments repetition among others. These can lead to controversial practices such as “all of the above” and “none of the above”(Waterloo, 2014; Zimmaro, 2004).

Ineffective when measuring the ability to organize and express ideas and with some types of problem solving (Zimmaro, 2004) – These three factors can be attributed to lack of space for expression inherent to this type of task.

Lack of connection to real-world problem solving (Zimmaro, 2004) – While it is possible to argue that many real world problems are a matter of choosing the correct option, one would be hard-pressed to defend MCQs as good modeling tool for such situations, especially because in reality the alternatives normally present trade-offs, and not right-wrong or even best-option solution types.

Influence on scores by the test taker reading skills and the designer writing skills (Toolkit, 2008; Zimmaro, 2004) – One of the main biases of MCQs are their reliance, especially when assessing high-order thinking skills, on reading comprehension and a clearly written task. When comparing to more open-ended questions, this type of task is more sensible to poor writing, both for having more critical parts and for the rigidity of the format. Subtle differences in the stem can many times change the answer, and the reading skill gets measured as a by-product.

Lack of insight and feedback in the thought process – Having only the chosen alternative as a mean of inferring thought process, this feedback opportunity gets limited. It is still possible to infer common mistakes when the distracters where designed for such (Palmer & Devitt, 2007; Zimmaro, 2004).

Over reading – The level of depth in reading the question can surpass the task designer's intention, hindering the alternatives less effective (possibly both in providing the answer and distracting from it).

Failing to measure high order thinking skills – While it is possible for this type of question to assess and promote high-order thinking skills (Palmer & Devitt, 2007; K.

Scouller, 1998; Toolkit, 2008) it is more common in this format to assess only information recall and understanding (Toolkit, 2008; Zimmaro, 2004).

More than defensible correct answer – It is possible, especially when trying to design challenging distracters, to end up with alternatives that can overcome, at least in specific lines of reasoning, the key answer.

Does not measure writing skills – Written expression is not a part of this kind of task (Zimmaro, 2004).

Encourage guessing – Unless counter measures are taken, guessing is rewarded; more the less alternatives are given. (McAlister, 2005; Toolkit, 2008; Zimmaro, 2004)

Assessing trade-offs and grayscale topics – The existence of a single solution makes it more difficult to assess subjects where there are multiple criteria for answering. The lack of a best alternative or the existence of clear next best and so forth can difficult the usage of MCQs (Toolkit, 2008)

Anti-cheating practices are more necessary – From impersonators to direct cheating MCQs are more vulnerable than other forms of assessment, at least in their traditional form. Since there is less to no sample of handwriting impersonators may be a concern (Toolkit, 2008).

Encourage superficial studying practices – Due to the perceived need, by test-takers, to only recognize the proper answer and the natural tendency towards factual and comprehension level of assessing, this tasks have been shown to encourage superficial studying practices (McAlister, 2005; K. M. Scouller & Prosser, 1994).

2.8.1.2.3 Scenario Questions

One development over multiple choice questions are the scenario questions. In their form they are similar to multiple choice questions, but instead of delivering a direct question or proposition, before their stem, there is a case description. One of the main contexts where this kind of question is applied is the medical sciences, where a set of relevant (and in many good questions irrelevant) information is given, and the test-taker must decide how to decide in that given context.

Describing the scenario is one of the most utilized ways to facilitate the assessment of high order thinking in MDQs. The introduction of a context, a greater number of variables (compared with more traditional implementations of multiple choices), the

possibility of contradicting elements, and the opportunity of describing real cases are elements used by test makers to take the question up to the levels of analysis and even evaluation (American Society for Engineering Education, 1996; Mujeeb, M. L. Pardeshi, & Ghongane, 2010; Palmer & Devitt, 2007; Toolkit, 2008).

One of the disadvantages of writing the scenarios and modifications is the time cost. To be able to write a plausible scenario that actually allows thinking process with a similar depth of real world issues takes not only training but a fair amount of time. This issue, however, is mitigated since multiple questions can, and many times should, be placed under a single scenario description.

Comparing to other types of unstructured tasks, multiple-choice questions have by the largest literature body. Their extensive usage in standardized and large scale test have made them a subject of study of many studies.

2.8.1.3 Multiple Response Question

Multiple response questions are similar to multiple choice questions, the components are the same, but the number of distracters (non-scoring or wrong alternatives) and keys (scoring alternatives or correct alternatives) varies. This type of question is similar to multiple choice questions in many aspects, and despite allowing a larger set of possible answers (even 3 alternatives provide eight possible outcomes, 4 would mean sixteen and so forth) most of the limitations are kept.

This type of question however is not suitable for task containing a grayscale of better and worse, yet right, alternatives, they must be divided in those that fulfill the stem requirements and those that not. This type of dichotomy makes many of this questions more limited in assessing high order skills, similarly to what happens in true/false questions. It does not mean however that these questions cannot assess them; the analysis of coherence between the stem proposal and the alternatives without a limitation of a single correct response has a positive impact regarding guessing and the level of confidence required from the test taker.

According to Burton et.al (Burton, Sudweeks, Merrill, & Wood, 1991), scoring this type of question can also be problematic. The two most used methods for doing so are either scoring all-or-nothing or distributing points. In the first case the criterion is clear, if the student has chosen all the correct elements and none of the distracters, she or

he scores full, any mistake and it is zero. In the second a grade may be attributed for each correct selection be it either for a distracter or a key. A third relevant alternative, not described by the authors, is the usage of negative grades for wrong alternatives, as an example, in a question with five alternatives the student is correct about four, so he would score 3 out of 5. The last alternative can be further developed by the addition of a “do not know” state, implying in no score and no loss. The main issues with each of the possibilities are that in the first case any partial knowledge would be disregarded. In the second, failing to answer would represent a grade, diminishing the validity of the grade as a measurement device (since no knowledge would be graded as partial) and allowing students to use tricks to achieve better grades. The last method can lead to negative scores, and while using the third may discourage guessing, it can be a complicated implementation in many systems.

2.8.1.4 True or False

Another common type of structure questions is true/false or binary answer question. In this modality the learner reads a statement and simply selects whether it is true or false, yes/no, agree/disagree or any other type of self-exclusive determined answer. This kind of question has stronger limitations comparing to other types regarding the level of skill being assessed and guessing.

The author has failed to find any scientific works presenting clear arguments and/or examples that support the usage of this type of task to assess higher order cognitive processes. This does not imply that they do not exist, but evidences the common use of such activities to assess either recalling or understanding information. The guidelines of development of test of the universities of Illinois and the “Learning Management Corporation” state that it is hard to asses high order thinking skills with this type of test and provide no attempt to do so (Corporation, 2008; Illinois, 2016). The university of Wisconsin provides some guidelines with an example a true/false that would assess the upper levels of bloom’s taxonomy, the question is as follows:

“1 - In the equation, $E=mc^2$, when m increases E also increases. True False

Why is it GOOD? All the commandments are followed. A bonus is that it requires some higher order thinking.” (Wisconsin, 2011).

A basic understanding of the equation is enough to answer this question, and as such it does not relate to higher order thinking skills.

A common tactic in dealing with guessing is the grouping of statements in blocks where an incorrect answer invalidates the block and negative grades attributed to mistakes.

2.8.1.5 Matching

The matching type includes task where the objective is to order connections among a number of elements. This is a rather difficult type of activity to define in its entirety, especially when considering the many formats it can assume, but in the most traditional sense the task requires the selection of connection between two sets of statements.

In this case the best pairs should be selected and there is a correct and direct outcome.

Some advantages of this type of question are: the ease to address a large amount of content; easier to write (Ben, 2001). Paired with scenarios, this format can be used to schematize decision making, allocate resources and other tasks of higher complexity. Despite this possibility, this type of task is more often used to assess factual recognition.

In the “Archive of Learning Design Instructional Guides” from the Pennsylvania State University the following characteristics are attributed as relevant for the understanding of matching type questions:

- *Used for recognition of relationships and making associations*
- *Can be used for a wide range of subject matter*
- *Can be used to match*
 - *terms and definitions*
 - *symbols and names*
 - *questions with answers*
 - *cause with effect*
 - *parts with functions*
 - *procedures with operations*
 - *principles with situations in which they apply*
- *Good matching items can easily be converted to multiple-choice items.*
- *The premises and responses should be homogeneous.*

According to the same source, the following are the guidelines for writing this type of question:

- *Directions should provide a clear basis for matching.*
- *Have from 4 to 10 items in each list.*
- *Have about 3 extra items from which responses are chosen.*
- *Mix the order of the items in the columns.*
- *Use brief phrases and as few words as possible in each list.*
- *Longer statements should be in the premise (left) column and shorter statements in the response (right) column.*
- *Use a larger or smaller number of responses than premises. That is, avoid perfect matching which allows for elimination of responses when the test-taker is unsure of the correct answer.*
- *Format the entire item on the same page.*
- *All responses should be plausible solutions.*

2.8.1.6 Short Answer

Since the term can also refer to short essays, essays with relevant space limitations, short answer tasks are defined for the purposes of this work as those that require the student to answer a question with a single term or a short sentence. The key difference between this type of task and other structured tasks is the elimination of recognition as a means of answering and making it more difficult to guess. A common usage for short answer task is the answering of numerical exercises. Interviews with students evidence that this type of structured task is perceived as more difficult (Mcmillan & Turner, 2014). When considering computer aided tests, many of the tasks that require calculation can be answered by short answer. Within a given precision, the student is required to type in the answer, which is evaluated. This takes the grading burden of the professor at the cost of limiting both feedback possibilities as well as the understanding of the students' thought process. Coupled with randomly generated exercises, this can lead to an interesting tool for practice lists which is especially useful when considering the possibility of usual-mistake-based feedback. If the input result is wrong because of a common mistake (a usual mistake in selecting the right formula, for instance), a

feedback can be tailored to help the learner. This is however a more complex implementation.

Fill-in-the-blank or completion tasks are other types of short answer questions, these two types, however, frequently measure recalling abilities. One of the key issues when computer grading this type of question is the check the collection of answers, verifying if answers are not being considered incorrect because of blank spaces, synonyms, different phrasing and similar issues. Yet this verification is usually much easier than checking individual responses since the evaluator is only concerned about unique entries, which in a well-designed task should be limited in number.

2.8.2 Unstructured tasks

The second division of assessment tasks deals with the activities where the answer must be composed rather than selected. In this category there is a great variety of tasks, starting with the widely used essay questions, through the more limited version of short essays and holistic activities such as portfolios and projects.

One of their advantages is the high level of freedom and versatility that test-takers have when addressing these activities. In an essay question or presentation, for instance, the student has more freedom to express and defend his arguments than in any of the structured questions. This freedom is not only present in the answer itself, but also many times in the comprehension of the context and the task, there is versatility in this format to account for differences in point of views, values and even to soften the negative impact of misreading or poorly written questions.

On the other hand, a byproduct of this versatility is the need for evaluators to understand and account for all this diversity. The differences of opinions, values, rigor and emotional state, among others can make the grading and feedback of such activities less objective and more detached to the achievement of the actual learning outcomes. This type of activity has its reliability questioned much more often than structured tasks (American Society for Engineering Education, 1996; Boud, 2007).

A common issue when evaluating unstructured task is dealing with subjective answer when grading. There is no widely accepted method to do so, there is simply too much loss of value when moving from an answer towards a grade, no matter how many criteria were evaluated (Kohn, 2006).

Since grades do serve many purposes, they are likely to stand, and despite the critiques there is still no feasible way to eliminate grading sufficient accountability, given existing constraints. To apply other methods, such as no grades initiatives, would imply drastically reducing the scale of learning institutions.

For formative purposes, discussions based on either a holistic view or based on a set of criteria where there is dialog, whether written or spoken, in order to promote growth do not face the same critique.

Being more closely related to real life situations, this type of question has a broader set of uses, however these come with higher costs, especially for larger classes.

Following are the descriptions of the main types of tasks that are used in Engineering Education as it is or can be used to develop creativity and/or critical thinking competences.

2.8.2.1 Essay Questions

Assessing student's abilities with essay questions can be traced to at least the 13th century (Krishnamurthy & Subbarao, 2015), making it, even in the written form, one of the oldest forms of asking questions. They can vary from questions that demand simple listings to complete texts supporting an argument or explaining a concept. Since they can be open ended and qualitative, they are one of the most versatile types of question available.

While there is more than one definition of essay questions and some gray area between this concept and that of short essay question, there is a significant consensus on what is an essay question, to the point that some articles use the term without discussing or even defining it (Criswell & Criswell, 2003; Mujeeb et al., 2010; K. Scouller, 1998; Seçmeliye et al., 2016). The Merriam Webster Dictionary¹ defines essay questions as "an examination question that requires an answer in a sentence,

¹ Merriam Webster Dictionary online version available at <http://www.merriam-webster.com/dictionary/essay%20question>, accessed 15/3/2016.

paragraph, or short composition". One of the most relevant definition, that despite the age is commonly used, is that given by Stalnaker (1951), defining four conditions for a question to be an essay: Requires a written answer rather than selecting a response; Needs more than a single sentence to be answered; allows different patterns and original responses; Requires a competent specialist to judge the question.

In this work essay question are those in which there is interest in the entirety, or at least majority, of the written answer (even if to different extents) and some freedom is allowed in the composition, if not in the final result, at least in the method. Comparing to Stalnaker's proposal, the intention is to reinforce the focus on the body of the answer, since in the context of Engineering a significant portion of the essay questions require calculating or defining project parameters, if the main interest is in the answer itself with little or no regard to the reasoning, methods or arguments, it does not fit the proposed definition. The main objective of this proposal is to separate essay questions from the short answer type, where the interest lies in the answer itself instead of in the discussion or development, first because the grading of short answer questions is much easier (possibly even automatable) and second because of differences in the learning outcome discussions and implications for these two types of question. By this definition, however, the same prompt and description can (but should not) lead to either type of question. In a question with a numerical answer, for instance, if the interest is only in the final number, the question will be considered, for the purposes of this discussion, as a short answer.

2.8.2.1.1 Types and usage

A common division of essay type questions is regarding applying constraints in the prompt, such as number of words and definite spacing, for this reason essay questions are further divided in two different types: restricted (or closed-ended) and extended essay questions (or open ended) (Criswell & Criswell, 2003; Reiner, Bothell, & Sudweeks, 2002b). A third type, which can be a restricted or extended question, is called the Modified Essay Question where a scenario is set and question are asked based on it (Aljarallah, 2011; G I Feletti & Smith, 1986; Grahame I Feletti, 1980; Palmer & Devitt, 2007).

Restricted or closed ended essay questions have some kind of limitation on how the answer should be delivered. This might be limiting the number of words, defining the number of arguments (e.g. "provide **three** reasons for"), providing a table to be filled among many others.

The authors (Criswell & Criswell, 2003) define the restricted type as means of assessing student retention, comprehension and application outcomes, the restricted part meaning that there are constraints in the very question prompt. The example given by the authors of this type of question is: "Describe three ways in which communicable diseases might be spread.". The same authors state clearly that this type of question fails to access higher level thinking, being restricted to assessing up to the third level of bloom's taxonomy, application. Despite this claim, the authors point out three advantages in comparison to multiple choice questions, namely: The opportunity to explain their answers by writing, eliminating the possibility of guessing and assessing the ability not to simply recognize the answer in a set of alternatives, but actually recall the information (Abduljabbar & Omar, 2015).

While restricting the outcomes decreases the freedom in delivering the answer, this does not imply directly that the level of thinking used will be constrained to application. The argument that this limitation can make analyzing less profound and decrease the amplitude of tasks related to evaluation is understandable, however it can be argued that dealing with constraints can be in itself a way to promote high order thinking as the original Bloom taxonomy itself (R. M. Felder & Brent, 2004) regards synthesizing as the fifth level. To say that restricted essay questions do not normally assess higher level thinking is more reasonable than saying that are they unsuitable for.

The extended or open ended essay questions allow the student to express themselves more freely. Some articles (Cashin, 1987; Palmer & Devitt, 2007) point out the usage of such questions for assessing more elevated cognitive functions. One of the key arguments is the possibility of directing such questions towards the high order function, which is defined as (Abduljabbar & Omar, 2015; Palmer & Devitt, 2007) or at least partially overlaps with the last three levels of either the original or the revised Blooms taxonomy (L. W. Anderson & Krathwohl, 2001). According to Criswell and Criswell, it is able to assess higher order of cognitive skills by the usage of verbs to direct the respond to a different level of discussion (Criswell & Criswell, 2003). Defend, create, analyze, differentiate and contrast and explain are examples of the verbs to be used

to increase the cognitive level required to answer the questions. The importance of directing the question through different verbs is addressed in many articles, webpages, books and university directives (Abduljabbar & Omar, 2015; Cashin, 1987; Criswell & Criswell, 2003; Hoeij et al., 2004; Palmer & Devitt, 2007; Reiner, Bothell, & Sudweeks, 2002a).

Another proposed practice would be to force decision making by using into in order to contextualize and generate abstract thinking (Criswell & Criswell, 2003).

2.8.2.1.2 Strengths

There are several reasons to use essay questions both to summative and formative assessments, some of these points are clarified below.

Testing complex or high order of thinking learning outcomes (Cashin, 1987; Criswell & Criswell, 2003; Reiner et al., 2002b) – One of the main advantages of this type of question, especially the extended variety, is the possibility of directing the questions to different levels of thinking. One of the most common feature in the studied literature are lists of verbs used to define what skills and cognitive abilities students should use in order to answer the questions. In the end of this chapter there is a table compiled from literature sources with the equivalent level of thinking expected to provide an adequate answer.

Can Test Thought Processes (Cashin, 1987; Criswell & Criswell, 2003; Reiner et al., 2002b) – A significant advantage of essay questions especially over more objective questions is the possibility to analyze thought processes. If a mistake is made while answering a multiple choice question it is more difficult, despite being possible to some extent, to identify the causes or where the reasoning went wrong. The ability to identify these causes not only helps the development of an individual student, but may also show some mistake patterns that can be addressed in class.

Require Students to Use Own Writing Skills (Cashin, 1987) – The need to write allows the student to choose his or her own words to deliver an answer, making explicit what are the points deemed relevant and developing the ability to express themselves by writing.

Pose a More Realistic Task (Cashin, 1987; Reiner et al., 2002a) – When comparing to some of the more objective tasks, essay questions provide a more accurate way to

depict real questions. It has a closer relation to real challenges and activities that pose similar questions not only in written form, but also orally especially outside the academic environment.

Cannot be Answered Correctly by Simply Recognizing the Correct Answer (Abduljabbar & Omar, 2015; Cashin, 1987) – The lowest cognitive level in both the revised and original Bloom taxonomy are recognizing, a practice that should not be possible in a well-designed essay question. Other authors (Becker & Johnston, 1999; K. Scouller, 1998) also point out the importance of eliminating the possibility of guessing answers, making the assessment more accurate.

Can be Constructed Relatively Quickly (Cashin, 1987) – When comparing to objective questions where both the question and the possible answer should be properly designed, essay questions hold the advantage of being easier to build.

{Insert other sources found}

2.8.2.1.3 Limitations

Some characteristics of this type of questions limit their uses, and may render other type of tasks more suitable depending on the conditions.

Only Limited Content Can Be Sampled – Both Reiner and Cashin (Cashin, 1987; Reiner et al., 2002a) argue that given the time consumed by the students to answer essays the amount of content that can be assessed, and consequently practiced, is limited. While it is possible, according to the same authors, to answer dozens of multiple choice or short answer questions in an hour, essay question demand much more time. While using them to assess and train high order levels of thinking, this time may be further increased by the need to reflect and create more complex and relevant answers. While it can be considered a well-spent time, this makes essay question more time consuming, and less broad in the content covered, given the same duration.

Yield Unreliable Scores (Cashin, 1987) – When talking about summative assessments, a difficulty arises in providing reliable and just scores. Correct and incorrect are not as clear as in more objective tests and even same evaluator gives different scores when correcting depending on mood, level of tiredness and sometimes reputation of the student being scored.

Can be influenced by the impression that the examiner has of the examinee – (Cashin, 1987) There is evidence of consistent differences in scores given by the evaluator caused by previous work, class participation and other tests. Professors can give better scores to students that are already performing well.

Score may be influenced by factors unrelated to those being tested– Poor handwriting, lack of organization in the answer, spelling and grammar mistakes, untidy answer among others.

Allows the practice of poor writing skills (Cashin, 1987) - Especially when considering tests with significant time constraints, the quality of writing can hinder the ability to provide a complete answer. As such, many students write poorly, leading to the practice of poor and underdeveloped writing.

Essay tests are time consuming to score and provide feedback (Cashin, 1987) – One of the biggest issues especially in large classes is scoring and providing feedback to the students. Reiner suggests that tests are scored in smaller batches to make it fairer by decreasing the level of stress and tiredness caused by the grading process (Reiner et al., 2002a). One of the most relevant advantages of using essay tests is the ability to detect difficulties and points to improve, without proper feedback these advantages may be wasted.

2.8.2.2 Reports

Reports can be used for assessing a wide variety of activities, be them technical visits, project descriptions, field work, laboratorial practices, literature research, among many others. It is one of the practices most directly related to the engineers working environment.

Among the advantages of reports, Race (Race, 2015) points out: high authenticity due to the importance on many work fields; usage of end products of useful learning activities; allow students to display their talent since they often have more time and control of the process in comparison with other type of tasks; can be developed incrementally; can be collaborative.

They are useful to assess students' level of both understanding on ideas and capacity to describe practices as well as reflections on challenges or require even propositions for improvement.

As disadvantages the same author points out: unwanted collaboration may be difficult to detect since it can be easy to students to acquire old reports or share reports among peers; takes much of the students' time; demands much time from instructors.

2.8.2.3 Oral Questions

Oral questions and interviews share many similarities with essay questions. These activities are open-ended, allow students to express themselves with some freedom and have the capacity to assess and promote the development of high order thinking skills. They differ regarding the form of expression, written versus spoken, formality and versatility.

Oral exams can be used as tool by themselves, where the test taker is asked questions by the evaluator, or as a complement to other forms of assessment such as presentations.

Some of the relevant advantages pointed out by Race (Race, 2015) are: have strong validity; are closely related to successful career skills; useful for checking ownership of evidence; can be useful when searching for particular outcome since this can be exploited just as much as needed. As limitations he proposes that: some candidates never show themselves well due to individual or cultural characteristics; Speed of response can be critical, harming slower and more reflective students; the questions' agenda may leak, meaning that students can share information providing advantages to later test-takers; the coverage is narrow in comparison to other methods, the author states that this type of task is not commonly used to measure how well students understand the syllabus; the lack of anonymity and its effects on bias. In his suggestions, this author stresses the importance of preparation, understanding of the goals of the activity, clarity and people skills to diminish the levels of stress on the student side, sensibility in delivering feedback and listening.

2.8.2.4 Portfolio

A portfolio is a learning journal of sorts, a student's own organization and production for the learning process. This tool based on companies' and professional portfolio that show the projects that were performed and the achieved results, but instead of focusing

on the results themselves, the focus is on the student's improvement over time. According to Race these type of activity is becoming increasingly common in the digital era, both in traditional and digital (or e-portfolios) forms(Race, 2015).

This type of task is more directly applied on subjects that have constant deliverables to be analyzed instead of a building up structure of concepts; as such it is commonly in subjects such as technical drawing, rendering and product development disciplines. There are, however, ways to adapt the usage of portfolios to different learning contexts, such as: journal entries and reflective writing; peer reviews on discussions and participation; artwork, diagrams, charts, and graphs; group reports; student notes and outlines; rough drafts and polished writing; electronic, video, and/or digital items (Ben, 2001).

In his book, Phil Race (Race, 2015) points out the following advantages on assessing portfolios: tell more about the students than the results of exams; can be of high validity - due to their more direct connection to the intended outcomes; can be strong regarding authenticity – due to their possible connection to work activities; can reflect development – for their continuous nature, the author argues that they can illustrate progress instead of a snapshot of the process; can reflect attitudes and values as well as skills and knowledge. Regarding limitations, the author points out: Time consumption; being much harder to mark objectively; are difficult to assess online – according to the author this is the difficulty in seeing the whole context of the portfolio; doubts in the ownership of evidence – especially in collaborative work it may be difficult to verify ownership of the presented work, the author suggests coupling this practice with oral assessment or interview.

In the last session of his description of portfolio practices the author gives 21 suggestions on how to assess and design portfolios, being the key points to clarify the objectives and expectations with the students (negotiating if possible), emphasize quality and care in the marking process.

Portfolios can bring assess students' development in a more holistic manner in comparison with written or objective exams depending on how much learning outcomes agree with this method. Subjects such as automotive design, factory planning, electrical installations or concrete bridges are more likely to benefit from this form of evaluation then less hands-on subjects, like for instance statistics quality control, electrical circuits or transport phenomena.

Regarding the revised Bloom taxonomy, portfolios deal mainly with the level of application and creation, depending on the proposed task. Analysis and evaluation can be developed either in the creating process or through related tasks.

2.8.2.5 Presentations

To assess students' characteristics while on group is a challenge faced by many professors. While regarding grading it is rather common for professor to just use a common score for every member, there is potential for assessing and promoting individual development, however not as directed as in other tasks.

The simple act of asking a question directly to a group can reveal which learners are more willing to answer, which can indicate different levels of preparation. To separate the members in different roles, can contribute by creating a space to observe how well they assume different points of view. One of the most widely used ways is to simply observe how well each of the students deliver the presentation, either in general or following a given set of criteria.

Some key points of Race (Race, 2015) description of presentations are: they are authentic tasks since virtually in any work environment the ability to deliver presentations well is useful; students take presentations seriously; they can provide for useful reflections when students are asked to evaluate their own presentation; opening for question can help oral expression skills. As limitations, the same author points out that: presentation rounds take a long time especially in large classes; some students find presentation traumatic; evidence is transient, if not recorded it is difficult to reevaluate if students appeal; they cannot be anonymous, making it difficult to eliminate grading bias; in a series of presentations there can be order bias since the last groups or individuals will have learned from the previous presentations.

One of the key difficulties, particularly for large classes is that most presentations are not individual, increasing the challenge of assessing individuals. Grading them maybe even more complex, adding the difficulties present in unstructured tasks. Some of the tools provided for facilitating this process are described in assessing individuals in group work.

2.8.2.6 Projects

Perhaps the most holistic type of task among those described in this work, projects are high-cost high-reward activities that are rather complicated to explore completely. This task is a key motivator of change in education through the development of project based learning.

Projects lack close-ended answers, meaning that the student is not guided towards an outcome, but is rather left free to explore a variety of solution possibilities. For their size and complexity in comparison with other tasks, they allow for much more nitpicking and argumentation, creating multiple opportunities for deepening the discussion and for the development of analysis and evaluation skills. They can include changing conditions, forcing adaptation. Projects can easily generate spin-offs such as presentations, debates, questioning sessions, proposal defenses, role plays and many others that can be incorporated for competence development. They can also be divided in university (purely in an academic environment), industry (involving real companies) and community (Chandrasekaran, Stojcevski, Littlefair, & Joordens, 2012). This type of task can be evaluated directly, by the observation of the development process, evaluation of the end result, by the analysis of written reports, class discussion, debates, presentations, technical or financial reports, simulations and a wide variety of associated assessment practices. Many times being a group activity, much of the discussion presented under Assessment Individuals in Group Work is relevant for evaluating projects (Race, 2015). Since it is closely related to many other assessment practices, many other sections are also relevant for this topic, including oral questions, essays, reports, presentations and rubrics.

As other unstructured tasks, assessing can be done following a set of criteria or a general view. Grading can use rubrics, compare to expectations, be based on decision making process, on achieved results and many others, but it still subject to the same considerations of grading in other unstructured tasks.

Projects require at least the level of application and allow related tasks that explore any of the other levels. As with portfolios, knowledge and understanding are prerequisites.

2.9 ASSESSING INDIVIDUALS IN GROUP WORK

Before assessing group works, it is interesting to decide whether they should be assessed by product or process. Product means that the quality will be based on the project deliveries while process implies that evaluation will take into account interaction and teamwork (Eberly, 2015; Waterloo, n.d.-b).

One of the key issues of group work is how to assess individual participating in them. Failing to consider each student in the assessment practice can lead to issues like lack of motivation, feeling of unfairness, inconsistent markings, rewarding egoistical behavior among many others.

The lack of motivation can occur both to dedicated learners that feel others are not doing their share, as well as less committed students that can feel as if they do not need to work in order to succeed (Chin & Overton, 2005; Eberly, 2015; Race, 2015). The feeling of unfairness can arise from some grades and recognition being attributed to very distinct levels of effort and achievement. Not assessing individual students can lead to inconsistent markings for all the main goals of grading, being them a measure of achieved outcomes, a tool to promote learning, recognition of achievement or a measure of work quality.

With this in mind, why would professors assess and grade groups with little regard to the individuals? First, the methods for assessing the individuals in this context have clear limitations. According to an article available in the webpage of the “Centre for teaching excellence” (Waterloo, n.d.-b), this method is the easiest to implement, may be appropriate if the mark is a minor part of the total course mark and enforce group responsibility. Being easy to implement and appropriate for low stake works are practical notes, however enforcing group responsibilities is an important outcome, particularly when considering work environments.

To mitigate these negative impacts, some tools and practices have been developed:

Peer assessment

An important tool is peer assessment. The members in the group know better who is working and who is not. Chin and Overton (Chin & Overton, 2005) point out that it can be beneficial to make this assessment anonymous and stress the importance of delivering clear outlines on how students should evaluate one another.

The use of this assessments to attribute marks can be made in a number of ways, such as:

Point distribution: where students are given a total of points to divide among themselves (Waterloo, n.d.-b). This can be harmed by peer pressure and each group can follow different and even emotionally led criteria to determine the division. While this practice can lead to the development of negotiation skills, more assertive people can be rewarded for this characteristic.

Individual weighting: where students are awarded points according to their grading on predefined criteria (e.g. literature searching, analyzing literature, writing report, among others). This grades are joined with the awarded grades in order to compose the average. In this case both the peer pressure and issues with criteria are mitigated, the implementation of this practice without digital assistance is also time consuming (Waterloo, n.d.-b).

Splitting tasks

A more direct approach would be divide task. If this is implemented, then the assessments problems become those of individual assessment. The issue is that the forceful division of work not only goes against team work ethos (Chin & Overton, 2005) but also leads to individualization and students worrying more about their tasks than the group (Waterloo, n.d.-b).

Direct evaluation

By asking questions directly to each student about their work and/or its development is another way to evaluate individual development. This is however time consuming for the instructor and the students (Waterloo, n.d.-b).

If the questions are directed at students during class time, there is a class time consumption, if they are sent as homework questions or work logs, more workload is added in order to answer and evaluate these tasks.

Chin and Overton (Chin & Overton, 2005) propose interviews and observation as means to directly evaluate both groups and individuals, they also state that this is especially relevant in laboratory work.

Use group work rubrics

The Eberly Center from the Carnegie Mellon University (Eberly, 2015), suggests the usage of group work rubrics that include sections that deal with the student participation in the presentation and an individual reflection rubric taking into account contributions and lessons learned.

2.10 GRADING

Assessing practices are used worldwide as means to verify outcomes, attribute academic merit and measure knowledge. This is perhaps the most controversial aspect of evaluating, attributing a number or concept to an abstract knowledge, a competence, a skill, among many others.

What are the objectives of grading? Some books on grading in higher education skip the discussion on what are the assumptions of grading and go directly to the implementations (Marzano, 2006; Walvoord & Anderson, 2010). According to the University of Sydney (Sydney, 2016) “A grade is supposed to provide an accurate indicator of a student's mastery of learning standards” and grades are not supposed to: be used as motivation tool or behavioral contract system; grades are distorted by students’ behavior, character and work habits, grades are not meant to provide feedback. This is one of many examples of one of the key principles of grades, fairness. Race (Race, 2015) proposed the following set of “Values for assessment”: validity, whether the format is valid to assess its purpose; reliability, how much different assessors would agree on grade or marks, this is related to fairness and justice; Authenticity, how related the assessment is to *real-world* practice; Transparency, how well can the students see how the assessment works and marking occurs; Inclusiveness, how well does this practice suits students with special needs, such as dyslexia, dyspraxia and so on. While all these values have some connection to the grading process, reliability and transparency are especially significant to understand grading.

In this line of reasoning, grading should be based on the most efficient methods for *estimating* the outcomes. This leads to some changes in practices and difficulties in achieving consistency.

Considering a test composed of exclusively multiple choice questions and a student that knows little to nothing regarding the intended outcomes. Which score should he

get? Using grading as a ruler to measure outcomes, close to zero. What happens is that that student can guess, and with guessing his expected score changes depending on the number of alternatives of the questions, if there are four alternatives 2.5, if there are five 2,0 and so on. If there it is considered that a difference of two points in ten is significant, these factors should be considered. So, each wrong alternative would decrease the total score by 25% if there are five alternatives and 33% when there are four. What happens is that in this case a source of error is introduced, guessing is statistically now closer to zero gain, but the student is being evaluated for his risk taking and boldness is a relevant factor to determine grade.

A similar discussion can be made for each of the assessment tasks, and the conclusion is that some practices can help. Creating practices composed by different types of tasks can reduce this issue since they require different collateral skills and approximating these to outcome-related practice can also be helpful. The mathematical division of grades should also be carefully considered. Since there can be more than a single learning goal to achieve reliability this would have to be assigned weights.

Another way to use grades is as motivation tools. Many authors either directly or indirectly support this practice (Brookhart, 2013; Marzano, 2006; Mitchell, 2006; Race, 2015; Walvoord & Anderson, 2010).

Considering the section of this work about motivation and the model proposed (Ryan & Deci, 2000) pursuing grades motivate through the factor of external regulation and introjection. This means that they can to some extent motivate students, but in a more limited way. Whether a professor decides or not to use grades consciously as motivation tools, some students will prepare themselves for tests, not for the subject (Marzano, 2006; Race, 2015; Walvoord & Anderson, 2010).

Students in this level of motivation are more driven towards efforts that will affect the grades, and this by itself can be used to promote learning. By approximating the tasks to the outcomes, increasing their authenticity, there is less space for assessment-specific learning. By grading according to clear principle also closer to the desired outcomes, by using rubrics for instance, and clarifying these processes to the students can also motivate towards this development. These practices are however, highly behaviorist, and are much more of a practical concession to situations where the

professor and course (at least in the university level) have limited influence on, than an ideal development.

Balancing motivation and reliability of grading is one of the challenges of assessment practices. It is important to remember that despite the overall importance of grading in the current academic practices, measuring students for comparison purposes and conformity achievement finds little or no support in the more recent learning frameworks.

2.11 FEEDBACK

Feedback is one of the most important activities for formative assessments, since it is a primary driver of change. In this discussion we will consider both the direct comments and observation of the evaluated tasks as well as recommendations for improvement, according to Race (Race, 2015) commonly referred to as “feed-forward”, as feedback. There are many types of gotten feedbacks, they can come from the student herself or himself, from peers, from professors, from the activity itself, observations of other works and many other sources. Despite where they come, feedback can be useful or useless.

Given that there are multiple possible sources of feedback for students, why is that given by a tutor somehow different? Is it? The key differences are that the tutor has a higher level of expertise, disposition to help the student, skill in constructing and devising tasks, evaluative skill regarding student performance and expertise in framing feedback to students according to Sadler (Sadler, 1998). While these are actual differences on the qualities of tutor for providing feedback, what are the impacts in the outcomes that arise from this practice.

One of the main problems with standard tutor feedbacks is that they are passive listening or reading activities, and the result of passive methods for learning is questionable. The same author that preached on the advantages of personal feedback has defended that this approach has little impact, feedforwards being a better alternative, and actively learning how to evaluate a good work by herself or himself even better (Sadler, 2009).

A perceived problem in the quality of assessed tasks is that student often fail to understand what is expected of them (Race, 2015), making expectations-based

feedforwards especially useful. Another positive use of feedbacks and feedforwards is actually praising competent achievement, however difficult it may be not to underplay neither exaggerate this use.

In the same book by Race, a significant contribution is made regarding feedback for competence development. Using the concepts expressed in Figure 6, the author analyses the impact of feedback when the student is in either of the four possibilities regarding having competence or not, and being conscious about it or not. Summarizing the discussion, he proposes that in conscious competence, the feedback appears as praise, the difficulty is in adjusting the dose since understating may lead to discouragement, and over-praising can seem condescending. Conscious incompetence is routine, the suggestion for type of feedback is make the student try to propose the means to improve and if needed build on it. The author states that the most import area for feedback is where the student is unconsciously incompetent, since when exploring a new knowledge area, learners do not even know what they do not know thus the importance of finding out the gaps and deciding which of them should be developed. In this context the competence of a teacher is to help this discovery and provide basis for the choices. The last type of feedback is given to performers that are competent but do not know it and receiving this type of feedback according to the author leads to an increase in self-confidence.

Figure 6 - States in Developing Competences

Source:(Race, 2015)

Another key aspect of feedback practices is that of computer-delivered feedback. Especially concerning structured tasks these can at little time expense help the student understand his or her mistake. It has the advantages of being timely (as the delivery

can be programmed), available when needed and as detailed as wanted (even reviewing passages that explain how to get the task right). The problems are the limited nature of this feedback, since it mainly concerns common and predictable mistakes, if the feedback is not understood there is no further questioning and the effect of forgetting just after one moves to the next screen (Race, 2015).

Each of those feedback practices can contribute to the learning environment in some way, however the development of the student's capacity to evaluate the work for himself is also critical. Peer assessment assisted by clear objectives is one of the ways to improve students' evaluation sense, especially if observed and directed.

2.12 RUBRICS

The definition of an academic purpose rubric provided by Brookheart is *"A rubric is a coherent set of criteria for students' work that includes descriptions of levels of performance quality on the criteria"* (Brookhart, 2013). Rubrics are one of the main and most widely used tools for evaluating and grading tasks taking into account multiples outcomes and/or multiple evaluators. They contain an analyzed outcome attached to a scale with descriptions of achievement levels.

Rubrics can be divided into two different types according to Brookheart (Brookhart, 2013) Analytical and Holistic. In a Holistic rubric all the intended outcomes are put together in one single set of scores/descriptions where the evaluator assesses all the outcomes at once. The analytical separates each outcome in a separated set of scores/descriptions to be later put together in one final score. According to the author, the second type is generally more suitable, despite being more time consuming.

In paper form a complete rubric is composed by the task description and three basic parts that serves as parameters: the scale levels (possibly in the form of grades); the dimensions (the skills, competences, knowledge that are being considered); the descriptions of what constitutes each level of performance (Mitchell, 2006). Figure 7 is an example of rubric grid.

Figure 7 - Rubric Grid

Task Description

	Scale level 1	Scale level 2	Scale level 3
Dimension 1			
Dimension 2			
Dimension 3			
Dimension 4			

Source: (Mitchell, 2006)

Computer implementations may offer a similar structure but with each dimension being a separate selection similar to a multiple choice question.

Some rubrics are used to measure a single dimension and can be attached to many contexts to evaluate specific outcomes, following there is a rubric on creativity proposed by Wiggins (Wiggins, 2012):

6 The work is unusually creative. The ideas/materials/methods used are novel, striking, and highly effective. Important ideas/feelings are illuminated or highlighted in sophisticated ways. The creation shows great imagination, insight, style, and daring. The work has an elegant power that derives from clarity about aims and control over intended effects. The creator takes risks in form, style, and/or content.

- The problem has been imaginatively re-framed to enable a compelling and powerful solution
- Methods/approaches/techniques are used to great effect, without overkill
- “less is more” here: there is an elegant simplicity of emphasis and coherence
- Rules or conventions may have been broken to create a powerful new statement.
- Common materials/ideas have been combined in revealing and clever ways
- The audience is highly responsive to (perhaps disturbed by) the work
- The work is vivid through careful attention to telling details and deft engaging touches
- There is an exquisite blend of the explicit and implicit

5 The work is highly creative. The ideas/materials/methods used are imaginative and effective. There is attention to detail. A clear and confident voice and style are present.

- Novel approaches/moves/directions/ideas/perspectives were used to good effect
- There is an effective blend of personal style and technical knowledge
- Familiar materials and ideas have been combined in new and imaginative ways
- The work provokes a lively audience response

4 The work is creative. The ideas/materials/methods used are effective. A voice and style are

present.

- Novel approaches/moves/directions/ideas/perspectives were used to good effect
- There are imaginative and personal touches scattered throughout the work
- The work keeps the audience mostly engaged
- There is a discernible and interesting effect/focus/message/style, with lapses in execution
- The work takes some risks in methods/style/content

3 The work is somewhat creative. The ideas/materials/methods used show signs of imagination and personal style.

- Familiar approaches/routines/moves were used, but with a few new twists
- There are places where ideas and techniques are borrowed whole.
- Novel ideas or approaches may be present but they seem stuck on, excessive, out of place and/or not integrated effectively in the work
- Time-tested recipes and clichés are used even where there is a personal voice – the work is pretty “safe”
- The work is a mish-mash of interesting and familiar approaches and effects, but with no coherence OR the work is technically very competent and coherent, without much spark or insight

2 The work is not very creative. The approach is trite and the ideas clichéd, leading to a flat and predictable performance. There is little sense of the creator’s touch, voice, or style here.

- The work offers little in the way of new approaches/methods/ideas
- There is little sign of personal voice, touch, or style
- The work suggests that the creator confuses “creative” and “risk-taking” with “shocking in a juvenile way”
- There is excessive and incoherent use of different materials, techniques, ideas
- The creator may have confused great care and precision with creativity – the work is more polished than imaginative or revealing

1 The work is uncreative.

- The performance re-creates someone else’s performance or relies exclusively on the models/algorithms/moves/recipes/templates/directions/materials provided.
- The work is predictable throughout, relying almost exclusively on hackneyed approaches; there is no apparent personal touch
- The work is timid and lacking in vivid feelings and ideas – so abstract that it has little to say to an audience
- The work is done with care but without direction or insight

In the rubric about creativity it is interesting to notice that the author does not only provide one statement for each of the values, but also some examples of evidences that can be used to support selecting the score.

Some of the most significant advantages of using rubrics is that they allow for more structured peer and self-assessment and according to some authors they help developing both critical thinking skills and build up analyzing and evaluating. In the same line, the students who are required to grade peers' work can also learning both through their colleague answer as well as by the grading process. By using rubrics to assess their own and their peers a student can also understand better what is expected (Brookhart, 2013; Marzano, 2006; Mitchell, 2006; Race, 2015).

As grading tools, they can help achieve reliability and transparency (Brookhart, 2013; Mitchell, 2006; Race, 2015).

2.13 COMPUTATIONALLY ASSISTED ASSESSMENT

Computers, cell phones notebooks and tablets enable a wide variety of assessment practices and provide some advantages especially regarding formative tests. The spread of smartphone and the adaption of many education management systems to incorporate mobile technology facilitated this practice.

Some of the key features of this development are:

Instant feedback – Online platforms and educational software allow professors to design feedback that is presented as soon as the student answer the question or finishes the test (normally for structured tasks). If wanted the professor can design in some platforms a different feedback for each of the possible mistakes. This feedback can also include different learning tools, such as directing the student to learning material, videos, book chapters or even proposing assistance exercises.

Integrated assessment practices – using computers and smartphones, assessment practices can be easily incorporated in classroom both for structure and unstructured tasks. During class students can answer quizzes and in some cases the professor may get immediate feedback, possibly affecting the class outcome since the need to clarify a topic or develop a discussion further may be detected.

Question databank – computer assisted assessments can also use one of the many questions databases available, and in combination with tagging for learning outcomes,

help alleviating the time consumption of assessments and increasing its overall quality (since the professor may focus on the questions that need development instead of trying to develop or adapt all).

Adaptive assessment - Computers allow for tests that rewrite themselves as the students' progresses. Be it through pre-developed questions that change based on previous choices or through the proposition of questions more relevant to the development of that specific student.

Tag based test generation – Given a data bank with tag organized questions, it is possible for the professor to generate an array of tests for each student based on the tags (and how many questions of each type will be placed on a test from each category) or easily find questions to create a test.

Anti-cheating measures – Since computers can reorganize alternatives and questions easily, putting the questions in random order, changing alternatives from place and comparing for plagiarism are some of the ways how computers can help regarding cheating.

Randomly generated questions – Questions with numerical variables can be modeled in order to allow randomly generated questions. When practicing or studying the learner may end up running out of questions before feeling confident that she or he has mastered the concept. With this tool the student can try similar questions many times until satisfaction.

Promoting peer-assessment – It is much faster to let computational platforms handle peer performance than it is to manually distribute and afterwards compile a significant amount of grading. When considering the possibility of using analytical rubrics in classes that are not rather small, this advantage becomes almost a requirement.

Simulations – Students can perform activities and use simulated scenarios in order to deepen their learning experiences. These simulations have the possibility of resembling with more accuracy real-world environments.

2.14 DIRECTIVE VERBS

One of the key tools for requiring the usage and assessing different levels of thinking, both for creativity and critical thinking, is the usage of directive verbs in

inquiries(Cashin, 1987; Criswell & Criswell, 2003; Hoeij et al., 2004; Palmer & Devitt, 2007; Reiner et al., 2002a). These verbs can be used in many different task types, such as multiple choice, multiple response, essay question, reports and oral questions. Table 3 compiles the verbs proposed by different authors while connecting them to a level, and possibly sublevels in blooms taxonomy. The column of the equivalence to the revised Bloom taxonomy was based on Abduljabbar and Omar (2015) on classifying questions, adapted to the revised version (L. W. Anderson & Krathwohl, 2001).

Table 3 - Directive Verbs

Verb	Revised Bloom Taxonomy	Source
Account for	2-Understand	(UEL, 2008)
	2.7- Explaining	
Analyze	4-Analyze	(Abduljabbar & Omar, 2015; Reiner et al., 2002b; UEL, 2008)
Apply	3-Apply	(Reiner et al., 2002b)
Appraise	4-Analyze	(Abduljabbar & Omar, 2015)
	5-Evaluate	
	5.2- Critiquing	
Argue	4-Analyze	(UEL, 2008)
	4.1-Differentiating	
	4.2- Organizing	
Arrange	1-Remember	(Abduljabbar & Omar, 2015)
	1.2-Recalling or	
	2-Understand	
	2.3- Classifying	
Assess	5-Evaluate	(Abduljabbar & Omar, 2015)

Table 3 - Directive Verbs (continued)

Verb	Revised Bloom	
	Taxonomy	Source
Balance	2-Understand 2.3- Classifying 2.6- Comparing	(UELТ, 2008)
Be critical	5-Evaluate 5.2- Critiquing	(UELТ, 2008)
Calculate	3-Application	(Abduljabbar & Omar, 2015)
Categorize	4-Analyze	(Abduljabbar & Omar, 2015)
Clarify	2-Understand 2.1-Clarifying	(UELТ, 2008)
Classify	2-Understand 2.3- Classifying	(Abduljabbar & Omar, 2015; Reiner et al., 2002b)
Compare	2-Understand 2.6-Comparing or 5-Evaluate 5.2-Critiquing	(Abduljabbar & Omar, 2015; Reiner et al., 2002b; UELТ, 2008)
Compose	6-Create 6.1-Generating	(Abduljabbar & Omar, 2015; Reiner et al., 2002b)
Conclude/draw conclusions	2-Understand 2.5-Infering	(UELТ, 2008)
Contrast	2-Understand 2.6-Comparing	(UELТ, 2008)
Criticize	5-Evaluate 5.2-Critiquing	(UELТ, 2008)
Deduce	2-Understand 2.5-Infering	(UELТ, 2008)
Define	1-Remember 1.2-Retrieving	(Abduljabbar & Omar, 2015; UELТ, 2008)

Table 3 - Directive Verbs (continued)

Verb	Revised Bloom	
	Taxonomy	Source
Demonstrate	1-Remember 1.2-Recalling or 2-Understand 2.1-Interpreting 2.2-Exemplifying or 4-Analyze 4.1-Differentiating	(Abduljabbar & Omar, 2015; UELT, 2008)
Describe	2-Understand 2.7-Explaining	(Abduljabbar & Omar, 2015; UELT, 2008)
Determine	3-Apply	(UEL, 2008)
Develop an opinion/view	4-Analyze 4.3-Attributing	(UEL, 2008)
Discuss	4-Analyze 4.2-Attributing or 6-Create 6.1-Generating	(Abduljabbar & Omar, 2015; UELT, 2008)
Elucidate	2-Understand 2.7-Explaining	(UEL, 2008)
Employ	3-Apply	(Abduljabbar & Omar, 2015)
Estimate	3-Apply or 5-Evaluate 5.2-Critiquing	(UEL, 2008)
Evaluate/weigh up	5-Evaluate 5.2-Critiquing	(UEL, 2008)
Examine	4-Analyze	(Abduljabbar & Omar, 2015; UELT, 2008)
Explain	2-Understand 2.7-Explaining	(Abduljabbar & Omar, 2015; UELT, 2008)
Evaluate	5-Evaluate 5.2-Critiquing	(Reiner et al., 2002b)
Express	2-Understand 2.7-Explaining	(Abduljabbar & Omar, 2015)

Table 3 - Directive Verbs (continued)

Verb	Revised Bloom	
	Taxonomy	Source
Give evidence	2-Understand 2.2-Exemplifying 2.7-Explaining	(UEL, 2008)
Identify	1-Remember	(Reiner et al., 2002b; UEL, 2008)
Identify trends	4-Analyze 4.2-Organizing	(UEL, 2008)
Illustrate	2-Understand 2.2-Exemplifying	(Abduljabbar & Omar, 2015; UEL, 2008)
Indicate		
Infer	2-Understand 2.5-Infering	(Reiner et al., 2002b; UEL, 2008)
Interpret	2-Understand 2.7-Explaining or 5-Evaluate 5.2-Critiquing	(Reiner et al., 2002b; UEL, 2008)
Justify	4-Analyze	(Abduljabbar & Omar, 2015; Reiner et al., 2002b; UEL, 2008)
Judge	5-Evaluate 5.2- Critiquing	(Abduljabbar & Omar, 2015)
List	1-Remember	(Abduljabbar & Omar, 2015)
Label	1-Remember	(Abduljabbar & Omar, 2015)
Locate	2-Understand or 4-Analyze	(Abduljabbar & Omar, 2015)
Match	1-Remember 1.1-Recognize or 4-Analyze	(Reiner et al., 2002b)
Outline	2-Understand	(UEL, 2008)
Prove	4-Analyze	(UEL, 2008)
Propose	6-Create	(Reiner et al., 2002b)

Table 3 - Directive Verbs (continued)

Verb	Revised Bloom	
	Taxonomy	Source
Recall	1-Remember	(Abduljabbar & Omar, 2015; Reiner et al., 2002b)
Recognize	1-Remember	(Abduljabbar & Omar, 2015; Reiner et al., 2002b)
Relate	1-Remember	(Abduljabbar & Omar, 2015)
Review	2-Understand 2.4-Summarizing	(UEL, 2008)
Sketch	3-Apply	(Abduljabbar & Omar, 2015)
State	2-Understand 2.4-Summarizing	(UEL, 2008)
Solve	3-Apply	(Abduljabbar & Omar, 2015)
Summarize	2-Understand 2.4-Summarizing	(UEL, 2008)
Synthesize	6-Create 6.1-Generating	(UEL, 2008)
Trace	1-Remember 1.2-Recalling	(UEL, 2008)

Sources: Author's creation.

The intent of this table is not to define a strict relation between verbs used in questions and thinking levels, but to provide a simple tool for supporting professors' decision on how to formulate questions in order to assess the desired levels.

Reiner propose that based on the defining verbs it is possible to decide the most suitable types of task. Figure 8 shows the proposed relation.

Figure 8 - Directive Verbs and Types of Task

Source:(Reiner et al., 2002b)

While this relation should not be taken strictly, this proposition reinforces the importance of choosing an adequate type of question, especially regarding the different intended outcomes.

3 USING INDIVIDUAL ASSESSMENT TO DEVELOP CRITICAL THINKING AND CREATIVITY IN THE ENGINEERING SUBJECTS

3.1 CONNECTING THE PIECES

In this sections some relations from the topic presented in the literature review will be highlighted in order to contextualize their importance for engineering education.

3.1.1 Assessment and Motivation

The two types of tests described in the literature review have different responses when relating to the motivation of a learner.

In summative assessments, the rewards and punishment are clear since a score is given. While the stakes of the test can vary, the grade itself provides an external stimulus, and as such acts on the external regulation, when the main perceived reward is the grade itself, and/or the introjected motivational level, when the student relates grades and achievement as an internal measure of success. The main issue when using grades to motivate students in such a manner is that the expected behavior is to prepare himself or herself to perform well exactly what the test asks, in a way increasing the importance of the perceived demands of the test. An analogy would be a driver that reduces his or her speed because of a speed trap, and as soon as it is over, accelerates again. If there is an approximation of the actual outcomes with the assessment methods, this type of motivation can be better used. Choosing what to reward or not while defining the grading system is more significant for the more external types of motivation than for the internal ones.

Formative assessments do not have such direct external rewards and punishment. Since they are less related to scoring and have little or no classification/labeling purpose, the rewards are more related to learning itself. Without the scoring pressure it is possible to infer that they relate less to the external regulation motivation and more to the more internalized levels of extrinsic motivation. If their importance for the success in the course is understood, the range of possible motivation increase to the full range (external regulation regarding their influence on performance and introjected for their possible influences on self-worth).

It is possible to analyze the learning motivation regarding the subject and the assessment task. One example are crossword puzzles, where people dedicate themselves to the activity many times because of the pleasure of the activity itself, even if this tool also assesses factual knowledge, vocabulary among others. Designing tasks that are more attractive and motivating by themselves is one of the principles of gamification.

3.1.2 Type of tasks, assessment and grading:

The division between structured and unstructured task is useful due to the difference of their characteristics to assessment and grading.

Structured tasks are deemed as much more reliable and transparent (Seçmeliye et al., 2016; Waterloo, n.d.-a), given that they are well designed, than unstructured ones since right and wrong are design decisions. But reliable for what? They are reliable when considering the task itself, when asking whether the chosen response was correct, or if the statement is true. Regarding the real intention of grading the student achievements they range from an acceptable level of reliability to none. True reliability in putting a number in knowledge is not achievable, but reliability enough to serve the purposes such as accreditation and subject approval is more likely. So how should reliability be dealt with? Understanding the assumptions being made is good starting point.

Considering an outcome to be assessed there are assumptions about the importance of the different bloom levels, the body of factual knowledge that is useful, the depth needed to achieve that outcome, the capacity of the assessment tool being used, the format of the assessment, among many others. Each of those can hurt the reliability of the practice.

An assessment practice that consider all those items can lead to more reliable results, using a single type of assessment is hardly the best choice. When looking at the specific case of Engineering Education, there many formative subjects where preoccupation with accreditation and mastery of subjects is higher. The assessment tool used to determine whether the student has sufficient knowledge to be deemed competent in that method is probably not the same that can best be used to develop

the conscience of the impacts of applying the same principle in another context or scale. This understanding lead to significant differences in assessment practices.

3.1.3 Critical Thinking, Creativity Skills and Competence in Engineering

Creativity and critical thinking are skills while in the competence development framework, three items were described as required to develop competences, knowledge, attitudes and skills.

While engineering students may be creative outside the university environment, this does not necessarily mean that they can foster their creativity in engineering context. This difference is the reason why this work deals mainly with the development of competence, because it includes the discussion of the engineering knowledge and the attitudes to make use of this skills.

By proposing activities where critical thinking and creativity are used in engineering context and nurturing attitudes that go in that direction, these skills can be integrated into increasingly more developed competences.

3.2 ASSESSMENT PRACTICES

Each task has different characteristics, they demand different amounts of time to develop, mark and to be answered. While neither is definitely better than the other and a full solution by itself, they do fit different roles differently. In this section an overview of the suitability of the proposed tasks to assess different outcomes in the context of engineering subjects will be discussed.

3.2.1.1 Usage in Engineering Subjects

Both structured and unstructured tasks have their place in engineering context, however some unstructured tasks are especially authentic in engineering environments.

A useful division of the engineering subject for determining which tasks fit each application is: basic sciences, prerequisite subjects and applied subjects.

While the level of importance may change according to the area, projects are one of the most authentic tasks that can be used in engineering. Most of the descriptions of

engineering needs and discussions of the profession include projects, but using them directly in the learning environment is a challenging task. An engineering project is many times a teamwork activity, involving competences from different areas and frequently, if not always, extrapolating divisions of subjects. In classroom, this type of project is especially time consuming for students to develop. Despite these difficulties inserting projects in engineering classes can have a number of benefits, particularly when it is possible to transcend a subject-based division. They require the usage of at least the third level in Bloom's taxonomy, they can foster creativity and critical thinking, they work intrinsically expected professional outcomes, among many others. If the conditions to create a significant project based assessment in the learning environment, this is perhaps the most complete tool for assessing students' competences.

Oral questions are common workplace occurrences not only for engineers, but for nearly every field. In engineering education, they can be used as test for knowledge and application of discipline, but they may be better used as a complementary activity to enhance other tasks. Instead preparing full-blown tests based on this model, professors could use oral examinations to deepen the discussions delivered in presentations, to assess student's understanding of their projects and question deliveries. The main issue of oral questions is that they occupy almost exclusively class time, so they may be better suited, especially in bigger classes for occasional questioning.

Just as with oral questions presentations are an authentic task for many areas, engineering is not an exception. There is little to say about differences in this task for engineers.

Portfolios are, regarding authenticity, a niche type of task. While some engineers, especially in areas related to design, have to present works that would fit into a portfolio type of assessments, many others work in areas where the deliverables are more closely related to reports, presentations and projects. Where it is applicable, the usage of portfolios can bring many benefits since it can evidence the student's skill and if designed for this purpose, creativity.

There are many different uses for essay questions in the engineering context. In every engineering field there are subjects of formative and informative nature, some of them laying the basic tools needed in order to allow the learning of the applied disciplines,

others discussing methods and usages and giving information on the engineering practices. Essay questions as authentic tasks can represent defending a project, responding a client questioning, reporting on a taken decision, among many others.

A characteristic feature of engineering tasks is the amount of problem solving essay questions with numerical variables. Using essays to assess this type of exercise, allows the professor to follow the thinking process of the student in trying to solve a specific type of problem or using a specific tool. While this approach is certainly useful, it can easily be overused, accustoming students to well defined and clear problem descriptions, oversimplifying not the physical reality or the scientific basis of the applied tool, but the uncertainties and lack of well-defined parameters present in workplace engineering activities (Duderstadt, 2008; D. H. Jonassen, 1997). The authenticity of essay questions can actually increase by the usage of ill-defined problems.

Discussing the role of structured tasks in the engineering context, is discussing their role in many learning contexts, presented in the literature review, as they lack special connections to the field. They do have their place particularly when taking into account real world limitations such as the time available, the need for accountability, the large amount of content to assess among others. It is important to note that they can assess and be used to promote sometimes more efficiently than unstructured tasks when used within their limitations, and coupled with computer assisted assessment, they can increase the coverage and enable a much more widespread use of formative assessment practices.

3.2.1.1.1 Concerning Critical Thinking

One of the most direct ways proposed to assess critical thinking is the usage of directive verbs to the higher order thinking skills, as already described in Table 2. By asking students to analyze, examine, discuss, evaluate, critique problems or solutions, for example, some aspects of critical thinking are promoted. For informative and more subjective topics, the implementation of critical thinking essay questions is arguably more direct.

An obstacle in the promotion of tasks related to critical thinking especially in engineering courses is the perceived lack of topics that allow this level of discussion. The definition of engineer as a professional includes the usage of science and

mathematics in order to solve real problems illustrates the need of mastering concepts and applying them in real world solutions. When dealing with technical norms, mathematical methods or physical laws, professors and students may believe the content to be so objective that it does not allow for a critical discussion. A possible way out would be the usage of different methods, requiring reflection on their potential impacts, pros and cons, but even those are many times defined by norms, conventions and technical parameters, which contributes to discouraging critical analysis.

While the discussion of the reasons for such technical limits can contribute to the understanding of the subject and even the awareness of the importance of such norms, it may be perceived as less interesting for it does not directly allow a direct change in practice. One way to connect the knowledge of the norms with critical thinking can be justifying its creation. With proper direction, be it a specific part of the norm that is more difficult to defend or from the perspective of an involved group or process, this discussion can lead to reflection, but still not directly to action.

However, these physical laws, mathematical methods and norms are not each, in the engineering context, an end by itself. They are used to solve problems in some manner, and connecting them in some level to these problems and contexts can make them easier to relate to critical thinking. Some possible ways of addressing this difficulty are listed below:

- By creating problems where the student can choose from different methods and justify which he thinks is better. This kind of solution can generate deeper discussion if there are trade-offs involved, a clear “better” alternative would make the use rule of thumbs and memory easier.
- By evaluating or critiquing the norm or method the student can in fact act by proposing changes and discussing shortcomings. The same consideration about alternatives from the previous item fits here.
- By evaluating the importance of the method for different career paths in the engineering profession. “Evaluate the impact that the mastery of this concept has on the industrial engineer professional field. Are they relevant for his or her success?”
- By providing unreliable or more than one data set to apply the method, principle or norm. Exchanging the usage of clear variables and measurements from given

values to conflicting or less accurate data sets, allows to promote critical thinking by forcing the analysis of what to believe.

- In this context it is possible to ask what question(s) should be asked in order to improve the validity of the question – where should the effort be placed to better understand the problem.

Critical thinking questions can contribute to the development of more intrinsic forms of motivation by provoking reflections on the importance of concepts, skills and tools not for the course itself, but for the intended professional practice on that engineering field and even that of the individual. Asking such questions carry a risk however, the student may not see relevant uses for that specific matter, leading to at least two possible outcomes, a clarification of its relevance (through discussion with the professor, colleagues, professionals or other external evidence), or the reinforcement of its irrelevance, consequently a decrease in the students' motivation levels.

Projects, reports, essays, oral questions and multiple choice questions are the tasks that can more directly promote the usage of development in critical thinking. Many of the activities in developing a project involve critical thinking skills, focusing on contexts, forcing reflections, challenging solutions, requiring justification of project decision and promoting peer evaluations are ways to further their impact on this competence. Reports on the other hand can be critical by nature. This competence can be used by proposing tasks such as reports on the environmental or social impacts of real projects. By promoting the development of critical thinking skills it should be expected that they will also be used to reflect on the reality in which the students are inserted, leading, by the very definition of the word “critical”, to discussions, break of paradigms, status quo and consequently, changes.

3.2.1.1.2 Concerning Creativity

To promote creativity competence in the context of engineering considering the discussions presented in the literature review, some tasks are much more suitable than others. While it is possible to argue for structured tasks as means to assess and promote critical thinking, it is more difficult to say the same about creativity.

The last level of the revised Bloom's taxonomy is by the very definition of the model, the most demanding regarding both assessing and developing. Due to this sensibility,

the tasks practices that promote creativity are more limited and the responses that may harm its development more common. From the discussion on developing creativity it is visible that due to its connection and sometimes even dependence on critical thinking, the basic knowledges and lateral thinking makes providing the conditions for the development of this competence a delicate matter.

To some extent, the unstructured tasks in general can be used to promote creativity, but they have to be especially designed to do so. Traditional reports are perhaps the most limited task in using creativity, but it is possible to devise report task that either promote searching for alternatives or improvements in existing realities. Essay questions can promote creativity but normally do not. Having a correct answer, instead of a well thought-off or based one, is an evidence that an answer is not assessing creativity. The main issue with essays that promote creativity is that answering creatively requires time, an amount of time that can vary drastically between students' and is rather difficult to estimate how much. The usage of essay questions for generating divergent and lateral thinking is a way to promote this competence. Presentations can promote creativity regarding both content and format. Students can be asked to use videos, audios, non-computational presentations among others. The issue is that while this can work creativity as skill, its impacts the engineering competence of creativity are more limited. Portfolios and projects are two of the easiest assessment practices regarding creativity. The main difference is the open-ended nature of these activities. While exam questions can be used to discuss solutions to existing projects, or even evaluate results, there is less freedom to roam in comparison to projects since it is challenging to find close-ended projects. In order for both tasks to actually promote creativity, it is necessary focusing on the idea generation practices, thought documentation, and challenge solutions. Changing conditions can be used to promote adaptation and further these developments.

By using only close ended and well defined tasks, students' end up having little space to use their creativity in engineering assessment practices. Divergent questioning, ill-defined questions, and more holistic assessment tasks are some of tools for mitigating this effect.

Another important aspect of creative assessment is its impact on work quality. There are many situations where creative thought leads directly to improvements, but many times it leads towards new and unexplored or underexplored paths, that after due

development reach stability. However, at first many innovative initiatives cannot compete with well-established ones at the same level. In learning it is important to foster both, meaning that perceived quality and current costs or results of more creative works are often less developed. If the intent is to foster creativity it should be rewarded.

4 INTEGRATION IN ENGINEERING CONTEXT

In order to integrate these assessment practices into engineering courses, some steps are proposed.

The first one is clearly defining the objectives and/or learning outcomes to achieve as they will be used as criteria for the rest of the process. Having defined the outcomes, a strategy to achieve them is formulated taking into account constraints such as preparation times, class size, available infrastructure, number of tasks that have to be devised among others. This strategy should help define the nature, number and frequency of assessment, whether there will be ongoing parallel task or not, the feedback methods and the integration in the classroom environment. At the end of the chapter, other practices that can influence positively engineering competence learning are proposed in no particular order.

4.1 DEFINING OUTCOMES

One of the most important practices in order to improve control over the results of a course is the determination of clear learning outcomes. Factors such as expected engineering competences, the purpose of the subject in the context of the program, the students' perceived shortcomings, ethical issues in engineering and the academic context in which the course is emerged are important in this consideration.

In this work, critical thinking and creativity in engineering have been chosen, and being competences, the underlying knowledge to be critical and creative upon is also a key objective. As discussed in the second chapter, other common goals for engineering subjects are problem solving, communication and teamwork skills.

Courses and the development of assessment require preparation; as such, this goals and outcomes can be more easily defined before the classes start. However, understanding the students' expectations and making them actors in deciding what outcomes are worked towards can make them more involved in the entire learning process. While at first glance this may seem overly risky, it is up to professor to guide this process and to determine how deep these changes can get. By using his or her experience, the professor can also anticipate these expectations, keeping the

motivation factor of the active participation while preserving a reasonable amount of prepared material.

4.2 DEVISING STRATEGIES

To decide what type of assessments will be used, it is useful to know what are the constraints being dealt with. Factor such as time available to work on the course, class size, students' timetable, infrastructure, how many tasks are ready for use, assistants available and related subjects can change the decision on assessment strategies.

The time consumption of assessment practices from the professor point view can be of either initial investment of time (setting up systems, writing tasks), maintenance (giving instructions, providing feedforward, tidying up tasks) or marking (grading and evaluating). These time investments are further divided by whether they are flexible or not, since some tasks are preparation for a specific class, some are designed as class activities, others include face-to-face experience and for some the main limitation is dealing with deadlines. Having assistants or not should also be accounted, especially regarding grading and feedback time.

Regarding time consumption, two more factors are important, the classroom time to dedicated to assessment practices and how much time students can dedicate to the subject. While the first one is often taken into consideration, the second is easily forgotten. The time students have available can sometimes be estimated, but keeping a dialog can help professors dose the assessment activities as well as coordinate them, accompanying moments where time is more limited with more relaxed, so that students are able to dedicate themselves properly.

Class size is a key factor when considering which practices to use since assessment practices have diverse scaling behaviors. Smaller classes facilitate direct relation to students and unstructured tasks, for they reduce workload with assessment. Larger classes favor structured tasks and make the use of peer assessment more attractive for unstructured tasks. Presentations are also a matter to consider, in large classes they are more frequently accommodated in groups, many times larger groups than would be ideal, making it more difficult for professors to individually observe students. The available tasks, activities ready for usage at the beginning of a course, are also a considerable factor since their quality and quantity affect directly assessment results.

Adapting and incorporating tasks is often faster than creating them. Starting a course from scratch requires much more time developing tasks, and such consumption creates some limitations, than building up on previous experiences. Keeping questions and activities organized by tags and modifying them iteratively can also increase the positive impact of assessment. External databases are an increasingly rich option for acquiring activities that can, especially when properly incorporated and adapted, help achieving desired outcomes.

The infrastructure affects assessment in a way that is easier to observe, to implement e-quizz in classes you need internet and the students, either smartphones or computers. Having multimedia projectors allows students to use presentation assisted by software. A room with large tables or individual desks also work differently for group related tasks. However obvious, they do affect which practices are available and is an important consideration.

Whether there are or not subjects to create partnerships with is relevant when considering the possibility of cross discipline projects or tasks. While these partnerships can lead to more holistic projects, approximating assessment from practice, their coordination can be complicated. A broad task can lack connection with the taught concepts but looking at this activity from multiple perspectives can straighten this connection, providing a significant advantage. Different disciplines in the same task can also lead to tradeoff analysis, consequently dealing with thinking skill in the range of critical thinking.

The overall proposition regarding strategy is to use an array of tasks and not excessively a single type. One of the most common formats currently in practice are essay questions tests mostly summative in character, where essays, in engineering frequently numerical or literal problem solving exercises, are used to assess factual learning, understanding, problem solving and whichever other outcomes may be expected. This format of assessment passes out on many opportunities of growth and is not optimal in terms of authenticity, reliability or cost-efficiency. Essay questions are a valuable tool, but is not the best option for every, or even the majority, of assessment conditions.

4.2.1 Choosing the tasks

Given the time requirement of evaluating essay questions, other tools may be better suited for the first levels of Bloom taxonomy. While essays can also do that, the time students take to answer them is significantly greater, limiting their coverage on subjects. All structured questions can be used for greater coverage alleviating the time spent in evaluation, but multiple choices still hold some advantages. First it is possible to find, especially for lower order skills, huge databanks with dozens of questions for many different engineering subjects, the other tasks may have questions banks, but not as widespread.

Exercise lists are other common practice of expositive classes. The proposed exercises are either given and left unchecked or are delivered by the students and assessed. In the first case, if a student does the lists what feedback does he have? Possibly the answer in the back of a book, any other will have to be sought. In the second case, they can be valuable formative tools, but the paper format for these lists, a static list posted online has the same issues, is often not the best implementation.

First, since creating the lists is time demanding, many questions are similar to previous years, fostering the habit of inheriting lists from senior students, especially when the lists are graded. Second, it is difficult to deal with plagiarism. Third, the quantity of exercises is limited to a sequence since it is difficult to implement parallelism on paper through instructions, defining how many exercises of a given type should be made, as students may use tainted criterion for the selection.

How can digitalizing and creating tasks in computer platforms contribute to formative assessment practices?

- For structured tasks, professors can receive the students' results and evaluate their difficulties with much less effort.
- Calculated tasks can generate an infinite amount exercises. The main platforms allow for exercise generation with given numerical sets or randomly generated numbers. To do so the professor enters with the mathematical formula for the result (e.g. $b \cdot h / 2$), the variation parameters for the variable (e.g. b varies in steps of 0.5 from 3 to 7 and h is either 3 or 5), and the required precision for the answer.

- For unstructured tasks with numerical answer, the professor can adapt this question into short answer to check whether the student got the question right or not, and ask the student to upload the solution process. This matches well with calculated tasks.
- The professor can categorize questions via tags and create auto-generated assessments containing a specific number of questions for every tag.
- The platform can be set to provide the student with multiple opportunities to take the test, increasing practice opportunities.
- Activities deadlines can be easily matched with the classroom requirements.
- The alternatives of the multiple choices, the order of the tasks and which task are part of the assessment can be dynamic, making each test different and more difficult for students to cheat.
- It is possible to implement adaptable tasks.

What about critical thinking and creativity? All the advantages described above have little to do with either competence, but they can improve the development of the lower order skills, base for the competences, as well as facilitate the introduction and create space for higher order implementations.

To create and digitalize lists it is important to keep in mind what is the objective of each task. Is it there to assess factual learning? To practice the usage of a tool? To prepare the engineer for using this tool in real practice?

This is a point where critical thinking skills are integrated. Along with these assessments, there should be tasks prepared to get the parts together, to promote reflection on learning, to evaluate possibilities and to attribute importance to what is learned and discussed.

For this purpose, the most widely used tasks are either essays or multiple-choice questions for higher order thinking. The impact of the first option is enhanced through the usage of peer and self-grading, with the added possibility of making the grading to an essay question, which would require students to develop on aspects of their colleagues' answers, working on the analysis and evaluation levels of the revised taxonomy. Another possibility to work critical thinking while peer assessing is the use of rubrics, which is faster to use and clearer, however can lack the depth of a written

review. The second option is also useful due to its practicality and can be effective in assessing, but misses the devising the answer and peer assessment practices.

Regarding creativity, unstructured tasks are much more suitable for their development. There is little to add, besides the incorporation of engineering as a basis, to the discussion of chapter three.

Some steps for transferring lists and creating computer-assisted tasks are listed below:

- Decide the goals of the computer assisted assessment. Is it closing a topic? A provocation for next classes? A milestone used to move to a new concept. A diagnostic assessment to check how well the learning process is going? A contextualization of learned principles?
- If this is not a first-time implementation, build up on the previously used activities. Try to find questions that collaborate with learning outcomes.
- Adapt them, whenever needed, to a task type suitable for the intended outcome (it is important to consider cost-effectiveness). The three most common being multiple choice, short answer and essay.
- Consider making numerical exercises into calculated tasks
- Organize them by tags instead of folder as tag format allows for much more freedom.
- Allocate the tasks in the platform with their instructions and deadlines.

In this section so far the discussion has been limited to isolated tasks brought up together in order to perform content related assessment, whether or not they include higher order skills. While isolated tasks are useful to develop engineering competences, their limited lifespan makes it difficult to develop much or delve deeper in a discussion.

A second type of assessment practice that to be used are more holistic approaches, often in the form of larger problems and projects.

To use projects in parallel to classes can require a significant amount of follow-up but since they are more complex activities, they can incorporate creativity and critical thinking in a variety of ways.

The considerations presented in the review for the development of creativity and critical thinking are applicable rather directly to projects.

Projects are also different considering their frequency. While punctual tasks are designed for coverage and may be implemented in many small tests during the course, the same is not true for projects. Complex projects normally run parallel to the class activities with deliverables design to follow up the process.

4.2.2 The possibility of repetition

Commonly summative assessment tests are a one-time event. The student gets to class, does the task and leave. Some professors ask for a correction of the task that did not meet certain criteria and award partial, or perhaps even full-points for such delivery. Doing so has some benefits and limitations.

On the benefits side, students can use their mistakes and feedback to improve on the subject rather than a statement of non-compliance, perhaps the most significant change. It also promotes practice especially for computer exercises and can also reduce the impact of the behavior of learning only for tests. It can be more coherent with the final grade “measures” achievement mindset since it diminishes the stress, anxiety and other high-stakes test consequences, since what matters in this approach is the result, not the process. This practice also promotes resilience since students are rewarded for trying repeatedly.

On the negative side, if exactly the same task is proposed, the student might just copy or memorize how to complete that task, so similar tasks are better suited. The student may also procrastinate preparation for the exam since there are other opportunities trying again. Students are not rewarded as much for getting it right the first time.

Another consideration is authenticity. Normally tasks in engineering are not done in a single sitting without double checking results or resourcing to any sources. Engineers in working environment have to get the answer right, but this include an entire process that includes developing the solutions, checking, coordinating it with other engineers and departments, prototyping, experimenting, checking commercial viability and so forth. Learning while working, basing solutions on what has been done, and dedicating a reasonable amount of time to the processes is essential for an engineer to work well. Giving the opportunity for students to revisit topics where they had difficulty can foster learning, but require more tasks, time to prepare and considerations regarding how to avoid negative practices such as cheating and over-focused study. Exams with a very

limited for completion may be a way to check how well practiced an engineer is using a tool, but has little relation to real world practice and convey the message that a mistake is a mistake and the grade has already been given.

Summarizing:

The following items offer a descriptive summary of relevant individual assessment practices focusing in engineering education and the development of creativity and critical thinking, based on the literature, conversation with education professors and personal experience.

- Considering what the assessment is for. Detecting improvement opportunities, practice tools, providing feedback or feedforward, checking the achievement of outcomes, among others
- Making students as active as possible in the assessment process. From defining outcomes to selecting activities and performing assessments, there is little discussion, active is frequently better.
- Considering the strengths of different type of tasks. While an essay can be used to measure retention, perhaps the ten multiple-choice questions that can be accommodated in the same time assessment can achieve that better.
- Varying types of task. Each task has its own limitations and bias, by varying them it is possible to serve a larger diversity in students' styles.
- Consider the different time demands when selecting assessment practices, they can be preparation, classroom, student, maintenance or assessing/grading time. More holistic activities do consume more time, but they also contribute in a more complete way.
- Projects can invaluable since in their development and delivery much can be assessed, but they also provide background for reports, critical essays, presentations, inquiry and other practices.
- Integrating specific tasks into more holistic ones. Especially projects can integrate many assessment practices, allowing for a more complete understanding of students' development.

- Using formative assessment to promote improvement. The information gotten from assessment can and should be used for discussions in class, changes in propositions, deeper analysis and many times more inquiries. Personal feedback and feedforward are also excellent examples. If the assessment led to little or no contribution to the student, professor or class it is probably not a good formative assessment practice.
- Use computers especially for lower order skills. The amount of questions that can be put at the student disposal and the possibility of generating exercises on demand is especially useful in an engineering environment, where many tools and methods have to be not only learned, but practiced. Computers can also facilitate managing courses, deliveries, analyzing tests results and grading.
- Use ill-defined problems. This type of problem has imbedded the need to make decision using judgement and analysis, thus being beneficial for the development of critical thinking. By requiring adaptation and solving in innovative ways, they can foster creativity. They are also more authentic than well-defined questions.
- Challenge the ways of thinking and the preconceptions as often as possible. The need to prepare engineers as actors in a changing society, responsible for their influence on society requires the ability to think in depth.
- Require the student to give up their solutions for new ones. Developing alternatives and changing perspectives are key in creativity development.
- Make students contextualize their own practice.
- Allow and reward holistic thought.
- Be prepared to change the course, assessment practices and even content. Trying to promote critical thinking and creativity in the learning environment while not ready to accept their impact on the classroom itself, is a controversial attitude. Professor should make their decision based on reason and be able to discuss them, otherwise the proposed critical thinking being developed is limited. This is not to say that constraints should be ignored or that student's opinion matter more, but that all elements participating in the learning environment are partners in developing a learning experience. Improving this

experience should be the common goal and the base for debates and discussions on the process.

- Treat lower order thinking as pre-requisites, not as final outcomes. Engineers generate value for society through the application, implementation, analysis, evaluation and creation of solution. Remembering and understanding are just that, prerequisites.

While there are many more possible tasks to improve assessment practices, these are some of the most relevant found when considering their importance and relation to the discussed topics.

5 CONCLUSION

The topic of assessment is a rather controversial one, on one hand there is the perceived need to certify oneself, institutions and the society in general that something has been achieved, that the graduated engineer is ready to perform her or his tasks. This thought contributed to engineering courses highly focused on the technical side of engineering, with many times overly demanding summative test being considered as the effective method to say whether someone is or is not ready to perform as an engineer, however little connection these tests have with real work environment.

On the other hand, when considering the most advanced theories and discussions on the education field, any type of measurement for classificatory external labeling is considered, for good reason, punitive and at best inaccurate. In those theories each person should be prepared to know whether he or she is capable of assuming a responsibility or not, if the individual, through the use of his faculties and based on experience, feels that he or she is prepared to do something, to proceed with his or her learning or to work, that should be enough. Teachers in this paradigm should stand in and through interaction with students, and mediating the interaction between students and peers, promote learning. This approach to education requires conditions that are not yet found anywhere, at least virtually anywhere, regarding higher education.

This is no excuse to accept archaic methods as enough and halt attempts to move to a more humane and responsible form of education, whether in primary schools, higher or engineering education.

This work has reviewed literature from different areas in order to propose practices that walk in the direction of developing skills and competences for the sake of graduating more complete engineers that can promote change. The propositions made deal with both the rationalization of resources in engineering education assessment as well as the usage of wider variety of task and methods to promote the development of two essential skills in promoting change, critical thinking and creativity.

To change the current state of engineering education in Brazil and in the world, there is perhaps a bigger need for dissemination of practices than of developing research in depth, and while some professors may be willing to go “all-in” to improve their courses, many more are willing to adapt their practice in a more progressive way.

Of all the topics presented in this work, some draw especial attention.

The understanding of the responsibility in designing assessments and the decisions made regarding outcomes is one of those topics. By choosing which skills and competences to focus the course development on, educators make decision that affect many students, it is important to be conscious of what these decisions are based on.

On the topic of motivation, students have a great deal of influence on what they bring to the classroom but despite the wish for more motivated students, professors have to work with real world constraints. There are many ways to influence students' motivation, but despite whatever practices are implemented, the individual own mindset is important. However important, there is little point in focusing effort in a matter that belongs to the circle of concern, not of influence.

Another important topic is that of the available tasks. More and more digital tools are incorporated in learning but this impact can be either positive or negative. These tools should be considered just that, tools to incorporate practices. They provide new opportunities for implementing perceived improvements, but the development of students is the focus of assessment and educational practices, not the usage of certain types of tools.

The combination of the knowledge of end activities and their influences, without losing sight of their purpose is the key to developing significant assessment practices.

By attempting to integrate concepts from different areas in a more readily available format for practicing engineering professors, I hope that this work contributes to the approximation of the engineering education fields to the practice.

5.1 LIMITATIONS

Given the huge variety of involved fields and the complexity of the learning related subjects, in order to be able to proceed with the discussion, this work had to select theories and use their propositions, solutions and frameworks without discussing in depth their limitations, critiques and alternative theories.

The sheer amount of practices available makes a selection necessary, but this selection may result in overlooking improvement opportunities.

Much research is still needed to better understand the real effects of the proposed practices, since the theories on the development of both studied concepts still have a long way to go.

5.2 FUTURE RESEARCH

There are many opportunities for implementing new assessment practices and much interest in understanding how they relate to preparing engineers for their future life. Mixed practices in PBL, active learning practices and accreditation practices are commonly published topics that relate with this work.

A special opportunity for future research regards the relation of motivation and assessment, a question which answer would contributed to this work.

BIBLIOGRAPHY

- Abduljabbar, D. A., & Omar, N. (2015). Exam Question Classification Based on Bloom's Taxonomy Cognitive Level Using Classifiers Combination, *78*(3), 447–455.
- Aljarallah, B. M. (2011). Evaluation of Modified Essay Questions (MEQ) and Multiple Choice Questions (MCQ) as a tool for Assessing the Cognitive Skills of Undergraduate Medical Students Moeen-uz-Zafar-Khan and Badr Muhammad Aljarallah, *5*(1), 45–51.
- Ally, M. (2004). Foundations of educational theory for online learning. In *Theory and practice of online learning* (Vol. 2, pp. 15–44). http://doi.org/10.1111/j.1467-8535.2005.00445_1.x
- American Society for Engineering Education. (1996). Assessment White Paper, (June), 1–7. Retrieved from <https://www.asee.org/papers-and-publications/publications/Assessment-White-Paper.pdf>
- Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. (L. W. Anderson & D. R. Krathwohl, Eds.). New York.
- Baartman, L. K. J., & De Bruijn, E. (2011). Integrating knowledge, skills and attitudes: Conceptualising learning processes towards vocational competence. *Educational Research Review*, *6*(2), 125–134. <http://doi.org/10.1016/j.edurev.2011.03.001>
- Bailin, S., & Battersby, M. (2015). Teaching Critical Thinking as Inquiry.
- Baleni, Z. G. (2013). Online formative assessment in higher education. *The Electronic Journal of E-Learning*, *13*(4), 43–48. Retrieved from <http://dl.acm.org/citation.cfm?id=2305228>
- Barnett, J. E., & Francis, A. L. (2011). Using higher order thinking questions to foster critical thinking: a classroom study a classroom study, *3410*(May). <http://doi.org/10.1080/01443410.2011.638619>
- Becker, W. E., & Johnston, C. (1999). The Relationship between Multiple Choice and Essay Response Questions in Assessing Economics Understanding, *75*(231), 348–357.
- Beghetto, R. A. (2005). Does assessment kill student creativity?, *69*(3), 254–263.
- Ben, C. (2001). Is this a trick questions? *Reading 2.1, MTD VA*, 1–69.
- Bennett, M., & Brady, J. (2014). A Radical Critique of the Learning Outcomes Assessment Movement. *Radical Teacher*, *100*(100), 146–152. <http://doi.org/http://dx.doi.org/10.5195/rt.2014.171>
- Berry, F. C., DiPiazza, P. S., & Sauer, S. L. (2003). The Future of Electrical and Computer Engineering Education. *IEEE Transactions on Education*, *46*(4), 467–476. <http://doi.org/10.1109/TE.2003.818757>
- Black, P., & Wiliam, D. (1998). Inside the Black Box : Raising Standards Through Classroom Assessment By Paul Black and Dylan Wiliam. *Phi Delta Kappan*, *80*(2),

139–144, 146–148.

- Bloom, B. S., Englehard, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals: Handbook I Cognitive Domain*. New York, 16, 207. http://doi.org/10.1300/J104v03n01_03
- Boud, D. (2007). Rethinking Assessment in Higher Education, 220.
- Boud, D., & Falchikov, N. (2007). Rethinking Assessment in Higher Education, 224.
- Brookhart, S. M. (2013). *How To Create and Use Rubrics*.
- Burton, S., Sudweeks, R., Merrill, P., & Wood, B. (1991). How to prepare better multiple-choice test items: Guidelines for university faculty. *Brigham Young University Testing Services and The Department of Instructional Science*, 36. Retrieved from <http://testing.byu.edu/info/handbooks/betteritems.pdf> <http://www2.swccd.edu/~mseteachingresources/assets/multiplechoice/multiplechoiceguidelines.pdf>
- Busato, V. V., Prins, F. J., Elshout, J. J., & Hamaker, C. (1998). The relation between learning styles, the Big Five personality traits and achievement motivation in higher education. *Personality and Individual Differences*, 26(1), 129–140. [http://doi.org/10.1016/S0191-8869\(98\)00112-3](http://doi.org/10.1016/S0191-8869(98)00112-3)
- Busato, V. V., Prins, F. J., Elshout, J. J., & Hamaker, C. (2000). Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education. *Personality and Individual Differences*, 29(6), 1057–1068. [http://doi.org/10.1016/S0191-8869\(99\)00253-6](http://doi.org/10.1016/S0191-8869(99)00253-6)
- Cashin, W. E. (1987). Improving Essay Tests.
- Chandrasekaran, S., Stojcevski, a, Littlefair, G., & Joordens, M. (2012). Learning through Projects in Engineering Education. *SEFI 40Th Annual Conference*, (March), 8. [http://doi.org/10.1675/1524-4695\(2008\)31](http://doi.org/10.1675/1524-4695(2008)31)
- Chin, P., & Overton, T. (2005). Assessing group Work, (March).
- Commission, E. A. (2016). Criteria for Accrediting Engineering.
- Corporation, T. L. M. (2008). Writing Effective Questions, 1–12.
- Cottrell, S. (2005). Developing Effective Analysis and Argument.
- Craft, A. (2006). Creativity in Schools.
- Crawley, E. F., & Lucas, W. a. (2011). The CDIO Syllabus v2.0 An Updated Statement of Goals for Engineering Education. *Engineering Education*, 24, 1–4. Retrieved from http://files.conferencemanager.dk/medialibrary/59856d54-6d1c-4deb-ac21-0dd98ccd0470/images/CrawleyEtAlCDIOSyllabus2.0Paper_17June2011.pdf
- Criswell, J. R., & Criswell, S. R. (2003). Asking Essay question answering to contemporary needs.
- Davis, M. (1996). Defining “Engineer”. How To Do It and Why It Matters. *Journal of Engineering Education*.
- Dewey, J. (1909). How We Think. In *How We Think* (p. 6).

- Drejer, A. (2001). How can we define and understand competencies and their development? *Technovation*, 21(3), 135–146. [http://doi.org/10.1016/S0166-4972\(00\)00031-6](http://doi.org/10.1016/S0166-4972(00)00031-6)
- Drejer, A., & Riis, J. O. (1999). Competence development and technology. *Technovation*, 19(10), 631–644. [http://doi.org/10.1016/S0166-4972\(99\)00064-4](http://doi.org/10.1016/S0166-4972(99)00064-4)
- Duderstadt, J. J. (2008). Engineering for a changing world, A roadmap to the future of engineering practice, research, and education. *The Millennium Project the University of Michigan*, 131. Retrieved from <http://deepblue.lib.umich.edu/handle/2027.42/88645>
- Eberly, C. (2015). How can I assess group work. Retrieved June 8, 2016, from <https://www.cmu.edu/teaching/design/teach/instructionalstrategies/groupprojects/assess.html>
- Edwards, M., McGoldrick, C., & Oliver, M. (2006). Creativity and curricula in higher education.
- Ennis, R. H. (1993). Critical Thinking Assessment. *Theory into Practice*, 32.
- Eraut, M. (1994). *Developing professional knowledge and competence. Knowledge Creation Diffusion Utilization*. <http://doi.org/9780203486016>
- Felder, R. M., & Brent, R. (2004). the Abc ' S of Engineering Education : Abet , Bloom ' S Taxonomy , Cooperative Learning , and So on.
- Felder, R., & Silverman, L. (1988). Learning and teaching styles in engineering education. *Engineering Education*, 78(June), 674–681. <http://doi.org/10.1109/FIE.2008.4720326>
- Feletti, G. I. (1980). Reliability and validity studies on modified essay.
- Feletti, G. I., & Smith, E. K. M. (1986). Modified Essay Questions : are they worth the effort ?, (Feletti 1980).
- Fernandes, D. (2006). Para uma teoria da avaliação formativa, 19(2), 21–50.
- Fransson, A. (1977). On qualitative differences in learning: IV. Effects of intrinsic motivation and extrinsic test anxiety on process and outcome. *British Journal of Educational Psychology*, 47(3), 244–257. <http://doi.org/10.1111/j.2044-8279.1977.tb02353.x>
- Freire, P. (1968). *Pedagogia do oprimido* (50th ed.). Rio de Janeiro: Paz e Terra.
- Freitas, S. L., Da Costa, M. G. N., & De Miranda, F. A. (2014). Avaliação educacional: Formas de uso na prática pedagógica. *Meta: Avaliacao*, 6(16), 85–98.
- Fryer, M. (2006). Facilitating creativity in higher education.
- Fureti, F. (2012). The unhappiness principle.
- Gilbuena, D. M., Sherrett, B. U., Gummer, E. S., Audrey, B., & Koretsky, M. D. (2015). Feedback on Professional Skills as Enculturation into Communities of Practice, 104(1), 7–34. <http://doi.org/10.1002/jee.20061>
- Goldberg, D. E., & Somerville, M. (2015). The Making of A Whole New Engineer : Four Unexpected Lessons for Engineering Educators and Education

- Researchers. *Journal of Engineering Education*, 104(1), 2–6. <http://doi.org/10.1002/jee.20064>
- Grunert, K. G., & Hildebrandt, L. (2004). Success factors, competitive advantage and competence development. *Journal of Business Research*, 57(5), 459–461. [http://doi.org/10.1016/S0148-2963\(02\)00312-0](http://doi.org/10.1016/S0148-2963(02)00312-0)
- Hager, P., Gonczi, A., & Athanasou, J. (1994). General Issues about Assessment of Competence. *Assessment & Evaluation in Higher Education*. <http://doi.org/10.1080/0260293940190101>
- Harlen, W., & Deakin Crick, R. (2002). A systematic review of the impact of summative assessment and tests on students ' motivation for learning (EPPI-Centre Review, version 1.1*). *In: Research Evidence in Education Library*, (1).
- Harlen, W., & James, M. (1997). Assessment and Learning: differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*. <http://doi.org/10.1080/0969594970040304>
- Helena, E., & Santiago, B. (2015). AVALIAÇÃO : Teoria & Prática - Um Repensar da Ação Docente, 1–17.
- Hinkin, T. R., & Tracey, J. B. (2010). What Makes It So Great?: An Analysis of Human Resources Practices among Fortune's Best Companies to Work For. *Cornell Hospitality Quarterly*. <http://doi.org/10.1177/1938965510362487>
- Hoeij, M. J. W. van, Haarhuis, J. C. M., Wierstra, R. F. A., & Beukelen, P. van. (2004). Developing a classification tool based on Bloom ' s taxonomy to assess the cognitive level of short essay questions European Veterinary Education: Structuring Future Development Developing a Classification Tool Based on Bloom ' s Taxonomy to Assess the C, (February). <http://doi.org/10.3138/jvme.31.3.261>
- Illinois, U. of. (2016). True-False Test Items.
- ISCO. DEFINITIONS OF MAJOR GROUPS, SUB-MAJOR GROUPS, MINOR GROUPS AND UNIT GROUPS (2008).
- Jackson, N., & Shaw, M. (2006). Developing subject perspectives on creativity in higher education.
- Jonassen, D. H. (1997). Instructional Design Models for Well-Structured and Ill-Structured Problem-Solving Learning Outcomes. *Educational Technology Research and Development*, 45(1), 65–90. <http://doi.org/10.1007/BF02299613>
- Jonassen, D., Strobel, J., & Lee, C. (2006). Everyday problem solving in engineering: Lessons for engineering educators. *Journal of Engineering Education*, 9(2), 139 – 151. <http://doi.org/10.1002/j.2168-9830.2006.tb00885.x>
- Khan, S. (2011). Khan Academy. <http://doi.org/10.5860/CHOICE.48-2433>
- Kohn, A. (2006). The Trouble with Rubrics.
- Kong, S. C. (2014). Developing Information Literacy and Critical Thinking Skills through Domain Knowledge Learning in Digital Classrooms : An Experience of Practicing Flipped Classroom Strategy . *Computers & Education*, 78(May), 160–178.

<http://doi.org/10.1016/j.compedu.2014.05.009>

- Krishnamurthy, K., & Subbarao, K. S. V. K. (2015). Assessment , Evaluation and Examinations ; Points to Ponder Introduction :, (2), 60–68.
- Krosnick, J. A., Judd, C. M., & Wittenbrink, B. (2005). The Measurement of Attitudes. In *Handbook of attitudes and attitude change* (pp. 21–76). Retrieved from <http://psycnet.apa.org/books/11352/055>
- Lizzio, A., & Wilson, K. (2004). Action Learning in Higher Education: an investigation of its potential to develop professional capability. *Studies in Higher Education*. <http://doi.org/10.1080/0307507042000236371>
- Llano, S. M. (2015). Debate's Relationship to Critical Thinking.
- Maher, A. (2004). Learning outcomes in higher education: Implications for curriculum design and student learning. *Leisure, Sport and Tourism Education*, 3(2), 1–9. <http://doi.org/10.3794/johlste.32.78>
- Mann, C. R. (1918). A Study of Engineering Education. *Science (New York, N.Y.)*, 48(1243), 420–1. <http://doi.org/10.1126/science.48.1243.420-a>
- Marzano, R. J. (2006). *Classroom Assessment and Grading*. *Journal of Chemical Information and Modeling* (Vol. 53). Alexandria, Virginia, United States of America. <http://doi.org/10.1017/CBO9781107415324.004>
- McAlister, D. (2005). Assessment Handbook, (August), March 2006. Retrieved from <http://www.canterbury.ac.uk/support/learning-teaching-enhancement-unit/assessment/assessment-handbook/>
- Mcmillan, J. H., & Turner, A. B. (2014). Understanding Student Voices About Assessment :
- Mezirow, J. (1990). How critical reflection triggers transformative learning. In *Fostering critical reflection in adulthood: A guide to transformative and emancipatory learning* (pp. 1–18). Retrieved from <http://www.graham-russell-pead.co.uk/articles-pdf/critical-reflection.pdf>
- Min-sik, Y. (2015). Multiple choice testing can smother creativity. Retrieved June 20, 2016, from <http://www.koreaherald.com/view.php?ud=20151221001056>
- Mitchell, A. A. (2006). *Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback and Promote Student Learning (review)*. *Journal of College Student Development* (Vol. 47). <http://doi.org/10.1111/j.1541-4329.2006.00011.x>
- Mizukami, M. da G. N. (1986). *Ensino, as Abordagens do Processo* (11th ed.). São Paulo: LTC.
- Mujeeb, A. M., M. L. Pardeshi, & Ghongane, B. B. (2010). Comparative Assessment of Multiple Choice Questions Versus Short Essay Questions In Pharmacology Examinations, 64(3), 118–124. <http://doi.org/10.4103/0019-5359.95934>
- Nguyen, D. Q. (1998). The Essential Skills and Attributes of an Engineer: A Comparative Study of Academics, Industry Personnel and Engineering Students. *Business*, 2(1), 65–76. Retrieved from

- <http://www.eng.monash.edu.au/uicee/gjee/vol1no3/paper4.htm> \n <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.124.1502&rep=rep1&type=pdf>
- Niewoehner, R. J. (2006). A Critical Thinking Model for Engineering. *CDIO Linkoping*, (June), 1–12.
- Oliver, M., Shah, B., McGoldrick, C., & Edwards, M. (2006). Students' experiences of creativity.
- Outcomes, M. L., Education, H., & Role, T. (2013). Measuring Learning Outcomes in Higher Education : The Role of Motivation, (9).
- Palmer, E. J., & Devitt, P. G. (2007). Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? Research paper. *BMC Medical Education*, 7, 49. <http://doi.org/10.1186/1472-6920-7-49>
- Paul, R., & Elder, L. (2008). Critical Thinking. *The Miniature Guide to Critical Thinking Concepts and Tools*.
- Paulsson, K., Ivergård, T., & Hunt, B. (2005). Learning at work: Competence development or competence-stress. *Applied Ergonomics*, 36(2), 135–144. <http://doi.org/10.1016/j.apergo.2004.09.008>
- Petress, K. (2004). Critical Thinking: An Extended Definition. *Education*, 124(3).
- Race, P. (2015). *The Lecturer's Toolkit - A Practical Guide to Assessment, Learning and Teaching*.
- Reiner, C. M., Bothell, T. W., & Sudweeks, R. R. (2002a). How to Prepare Effective Essay Questions: Guidelines for University Faculty.
- Reiner, C. M., Bothell, T. W., & Sudweeks, R. R. (2002b). Preparing Effective Essay Questions.
- Robinson, K. (2009). *Ken Robinson says schools kill creativity*. *TED Talks*. Retrieved from http://www.ted.com/talks/ken_robinson_says_schools_kill_creativity.html
- Rodríguez-Díez, M. C., Alegre, M., Díez, N., Arbea, L., & Ferrer, M. (2016). Technical flaws in multiple-choice questions in the access exam to medical specialties ("examen MIR") in Spain (2009–2013). *BMC Medical Education*, 16(1), 47. <http://doi.org/10.1186/s12909-016-0559-7>
- Rugarcia, A., Rugarcia, A., Felder, R. M., Felder, R. M., Woods, D. R., Woods, D. R., ... Stice, J. E. (2000). the Future of Engineering Education I. a Vision for a New Century. *Chemical Engineering Education*, 34(1), 16–25.
- Runco, M. a., & Jaeger, G. J. (2012). The Standard Definition of Creativity. *Creativity Research Journal*, 24(1), 92–96. <http://doi.org/10.1080/10400419.2012.650092>
- Ryan, R., & Deci, E. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, 25(1), 54–67. <http://doi.org/10.1006/ceps.1999.1020>
- Sadler, D. R. (1998). *Formative assessment: revisiting the territory*, *Assessment in Education: Principles, Policy and Practice*.
- Sadler, D. R. (2009). Beyond feedback : Developing student capability in complex

appraisal, (April).

- Scott, I. (2011). The Learning Outcome in Higher Education: Time to Think Again? *Worcester Journal of Learning and Teaching*, (5). Retrieved from <http://www.worc.ac.uk/adpu/1124.htm>
- Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, 453–272. <http://doi.org/10.1037/0278-7393.30.3.656>
- Scouller, K. M., & Prosser, M. (1994). Students experiences in studying for multiple-choice questions examinations. *Studies in Higher Education*, 19(3), 267–279. Retrieved from <Go to ISI>://A1994PL14900003
- Seçmeliye, Ç., Öğrencilerinin, L., Bir, G., Gupta, C., Jain, A., & Souza, A. S. D. (2016). Essay Versus Multiple-Choice : A Perspective from the Undergraduate Student Point of View with its Implications for Examination, 2015–2017.
- Seel, R. (2005). Creativity in Organisations: An Emergent Perspective. Retrieved June 4, 2016, from www.new-paradigm.co.uk/creativity-emergent.htm
- Shigley, J. E., Mischke, C. R., & Budynas, R. G. (2002). *Shigley's Mechanical Engineering Design - 9th Ed. Mechanical Engineering* (Vol. New York,). <http://doi.org/10.1007/s13398-014-0173-7.2>
- Simpson, E. J. (1966). The Classification of Educational Objectives in the Psychomotor Domain. *Education*, 3(3). Retrieved from <http://eric.ed.gov/ERICWebPortal/recordDetail?accno=ED010368>
- Stalnaker, J. M. (1951). The Essay Type of Examination. *Educational Measurement*, 495–530.
- Stevens Kevin; Garrison, Lari; Jocuns, Andrew; Amo, Daniel M., R. O. (2008). Becoming an Engineer: Toward a Three Dimensional View of Engineering Learning. *Journal of Engineering Education*, 97(3), 355–368. <http://doi.org/10.1002/j.2168-9830.2008.tb00984.x>
- Sydney, U. of. (2016). MARKING AND GRADING. Retrieved from https://sydney.edu.au/education-portfolio/ei/assessmentresources/marking_grading.htm
- Toolkit, A. (2008). Assessing by Multiple Choice Questions.
- Tosey, P. (2006). Interfering with the interference.
- Trevelin, A. T. C., & Neiva, J. S. F. da S. (2011). Estilos de Aprendizagem e Avaliação no Ensino Superior. *VI Workshop de Pós-Graduação E Pesquisa Do Centro Paula Souza*, 1.
- Trewin, S. A. (2007). Robert Yerkes' multiple-choice apparatus, 1913-1939. *American Journal of Psychology*.
- UCLT. (2008). Instruction verbs in essay questions. *Learnhigher*, 1–2.
- Vesselinov, R., & Grego, J. (2012). Duolingo Effectiveness Study. *City University of New York, USA*, (December 2012). Retrieved from http://static.duolingo.com/s3/DuolingoReport_Final.pdf

- Villanueva, I., & Taxonomy, A. B. (2015). An exploration of Bloom ' s knowledge , skills , and affective-based goals in promoting development of freshmen engineering students ' professional identities, 0–4.
- Walvoord, B. E., & Anderson, V. J. (2010). *Effective Grading* (2nd ed.). San Francisco, CA, United States of America.
- Waterloo, U. of. (n.d.-a). Exam questions: types, characteristics, and suggestions. Retrieved from <https://uwaterloo.ca/centre-for-teaching-excellence/teaching-resources/teaching-tips/developing-assignments/exams/questions-types-characteristics-suggestions>
- Waterloo, U. of. (n.d.-b). Methods for assessing group work. Retrieved June 8, 2016, from <https://uwaterloo.ca/centre-for-teaching-excellence/teaching-resources/teaching-tips/developing-assignments/group-work/methods-assessing-group-work>
- Waterloo, U. of. (2014). Designing multiple-choice questions. Retrieved May 18, 2015, from <https://uwaterloo.ca/centre-for-teaching-excellence/teaching-resources/teaching-tips/developing-assignments/assignment-design/designing-multiple-choice-questions>
- Wendland, M. W., Robinson, C., & Williams, P. A. (2015). Thick Critical Thinking: Toward a New Classroom Pedagogy.
- Whiteley, T. R. (2006). Using the Socratic Method and Bloom's Taxonomy of the Cognitive Domain to Enhance Online Discussion, Critical Thinking, and Student Learning, 33, 65–70.
- Wiggins, G. (2012). Rubric-Creative. Retrieved May 15, 2015, from <https://grantwiggins.files.wordpress.com/2012/02/creative.pdf>
- Williams, K. C., & Williams, C. C. (2011). Five key ingredients for improving student motivation. *Research in Higher Education Journal*, 12, 1–23. Retrieved from <http://search.ebscohost.com.ezproxy.liv.ac.uk/login.aspx?direct=true&db=ehh&AN=70547700&site=eds-live&scope=site>
- Wisconsin, U. of. (2011). Assessment Toolkit Writing True / False Test.
- Zimmaro, D. M. (2004). Writing good multiple-choice exams. *University of Texas*, (512).