



A Framework for Evaluation of Large Online Graduate Level Courses for Professional Learners

Dr. Kerrie A Douglas, Purdue University-Main Campus, West Lafayette (College of Engineering)

Dr. Douglas is an Assistant Professor in the Purdue School of Engineering Education. Her research is focused on improving methods of assessment in large learning environments to foster high-quality learning opportunities. Additionally, she studies techniques to validate findings from machine-generated educational data.

Hillary E. Merzdorf, Purdue University-Main Campus, West Lafayette (College of Engineering)

Hillary E. Merzdorf is a PhD student in Engineering Education at Purdue University. Her research interests are in assessment of digital engineering learning environments, evaluation of educational technology, and the ethical use of student data.

A Framework for Evaluation of Large Online Graduate Level Courses for Engineers

Abstract

Massive open online course (MOOC) platforms have evolved from providing primarily courses that are free or low-cost to working with industries and universities to offer credentials, advanced degrees and professional education. As more engineering schools and corporations develop partnerships with MOOC providers, there is a need for frameworks to guide how to conduct evaluation in the ‘massive’ environment. However, researchers have criticized traditional evaluation metrics are not suitable for MOOC environments. The purpose of this paper is to present an evaluation framework for large online graduate level engineering courses. This framework addresses this need with a comprehensive evaluation plan of practices and outcomes in MOOCs. Modified from Guskey’s (2000) professional development evaluation process, this framework examines learners’ satisfaction and value alongside performance, as well as pedagogies to support learning, application of content, and integration of the course with long-term institutional goals. We present the five levels of criteria, metrics, and data sources and discuss their application to evaluating MOOCs. The five levels of evaluation criteria are: 1) *Learner Satisfaction*, 2) *Learner Outcomes*, 3) *Pedagogical Practices*, 4) *Learner Use*, and 5) *Broader Impacts*.

Introduction

Corporations spend millions of dollars each year on professional development training for their employees [1]. One approach to reducing costs has been to partner with Massive Open Online Course (MOOC) providers, such as edX, Coursera, or FutureLearn. Additionally, engineering schools have begun partnering with MOOC platforms to provide graduate degrees for working professionals. Along with these new academic, industry, and MOOC provider collaborations is the promise of reducing corporate training expenses while increasing skills for on-the-job work [2]. Despite the huge amount of financial investment by all stakeholders, there is still relatively little evidence regarding the quality of these learning opportunities.

The most common approaches to evaluating MOOCs for professional learners yield limited useful information for continuous improvement or a full summative evaluation. One reason is that MOOC platforms enable thousands of learners to enroll in a single course. Some people in the course are learning from MOOCs independently, while others are enrolled for credit or a certificate [3]. Thus, some of the evaluation criteria based on best practices in distance education (e.g., learners introducing themselves to each other in the first class) are simply not feasible. Another major limitation to common evaluation approaches is a focus on course design alone, rather than including what learners are able to achieve as a result of the learning opportunity and what learning materials were useful. Pedagogies previously regarded as “best practice” have been significantly modified to accommodate high enrollment of learners who have little to no interaction with the course instructor [4]. While researchers are prolific at generating findings

regarding innovative pedagogical approaches in MOOCs, there is still not enough evidence to conclusively state which pedagogies can provide a higher quality learning opportunity. Educational technologies that support online teaching will continue to be developed and marketed for use long before any rigorous studies of efficacy can be conducted [5].

Evaluation is, by definition, examining multiple sources of evidence to come to a judgement regarding the quality of what is being evaluated [6],[7]. The quality of any educational resource is determined by characteristics of the learner, the learning materials, and their interaction. The purpose of this paper is to present an evaluation framework for large online graduate-level courses for professional engineers. We propose an evaluation framework for these courses informed by Guskey’s [8] evaluation of professional development, Quality Matters [9], our own research on stakeholder needs, and Davidson’s [6] Genuine Evaluation framework. The five levels of evaluation are:

- 1) Learner Satisfaction
- 2) Learning Outcomes
- 3) Pedagogical Practices
- 4) Learner Use
- 5) Broader Impacts/Return on Investment

We present these five levels of evaluation with criteria, metrics, data sources, and data use necessary to answer the evaluation questions which speak to the quality of the learning opportunity and the outcomes associated with it.

Literature Review

Two of the more common evaluation approaches for professional development and online education are Guskey’s [8] model of professional development and Quality Matters’ evaluation rubric for continuing and professional education [9]. Table 1 summarizes and compares the key evaluation criteria categories included in the two frameworks.

Table 1

Evaluation Frameworks and Criteria Categories for Professional and Online Education

	Learner Satisfaction	Learning Objectives	Knowledge Application	Course Pedagogy	Learning Impact	Accessibility
Evaluating Professional Development (Guskey, 2000)	X		X		X	
Quality Matters: Continuing and Professional Education Rubric (2005)	X	X	X	X		X

In the area of training and professional learning, Guskey's [8] *Evaluating Professional Development* provides the main influence for our evaluation plan. The focus of Guskey's [8] process is on learners' perceptions of a professional development program and outcomes as they relate to participant goals [8]. The first level considers participants' reactions about the content of the program, the process of how it was delivered and who delivered it, and context questions about the training setting or environment. Participant learning is the second level of this plan, indicated by cognitive, behavioral, and affective goals set and assessed by learners reflecting on how the professional development helped them meet their goals. The third level examines organization support and change by measuring wider policies and procedures held by the organization which the professional development aligns with. It evaluates whether the organization endorsed the program through security, openness, interest, administrative support, and recognition. At the fourth level, the framework asks how participants have applied their learning to new areas of practice. Student learning outcomes on a variety of formal and informal assessments are the fifth framework level. Scores should complement the participant learning judgments from the second level. Using pre- and post-tests and multiple measures is important for a targeted and complete understanding of each participant's learning. The goal of evaluation according to Guskey's [8] framework is understanding how the professional development opportunity benefited learners and applying findings towards improvement.

A major strength of Guskey's [8] framework is that the levels of evaluation are clear and focused on measurable outcomes. Thus, the framework works very well if the intended use of the evaluation report is presenting formative or summative outcomes. However, the Guskey [8] framework does not work as well for informing professors or instructional designers about what teaching approaches are effective in engineering courses. Process level evaluation is concerned with how the results were achieved in order to determine specific ways to improve or ensure it continues to happen. While Guskey's [8] framework covers five levels of evaluation, it does not capture information that could be used to inform better ways of teaching the course, apart from learner feedback. Process-level criteria, such as how the course was taught, the pedagogies or curriculum covered, or alignment between the learning objectives and the purported learning outcomes are not part of the evaluation. In addition, the Guskey [8] framework does not explicitly discuss accessibility issues.

A second evaluation program fundamental to online education is the Quality Matters (QM) program [9]. With standards for K-12, higher education, and professional education, QM provides rubrics aligned with eight quantitatively measured standards determining an online course's quality. QM is frequently regarded as the 'gold standard' for evaluation of online courses. There are eight General Standards (GS) for professional education. GS 1 determines how well the overall course design, prerequisites, and expectations are communicated to learners from the start. GS 2 ensures that learning objectives are clear, measurable, and consistent with course objectives. GS 3 evaluates course assessment in terms of learning objective alignment, grading, course fit, and frequency. GS 4 verifies that instructional materials are adequate to support student learning. GS 5 does the same for course activities that provide learning

experiences, as well as ensuring instructor interaction and timely feedback. GS 6 evaluates the use and quality of learning technology. GS 7 ensures that support services are available and useful, and GS 8 judges the overall usability and accessibility of the course. Together, Quality Matters standards evaluate the structure and materials of a course to determine how well students have performed. Poor performance on assessments can be traced to flaws in course design, so evaluators must investigate whether all necessary design components are included to support student learning.

A major strength of the Quality Matters rubrics for professional education [9] is the process-level information collected. The rubrics evaluate course design and structure in comparison to their list of best practices for online education. However, QM is quite expensive, requires institutional membership, and is rather onerous to implement without guidance from QM trained evaluators. Another major limitation is its lesser emphasis on outcomes and student experiences. A QM evaluation would not discover concepts that learners are struggling to learn or whether learners thought some of the assignments were irrelevant. In addition, some of the QM rubrics for professional education do not translate to large online graduate-level courses.

To evaluate process quality, outcomes, or effectiveness, Davidson's [6] guidelines provide best practices for evaluation methodology. Evaluation can be for determining productivity, promoting internal growth and progress, or developing quality benchmarks among similar products and services. It should be motivated by the goals and purposes arising from the values of an organization. Truth and certainty of evaluation outcomes are subjective because they are relative to organizational definitions of quality and value. However, they are not arbitrary, unsystematic, or arising from personal bias [6], but integrated with other facts and evidence collected during evaluation. Identifying the fundamental needs of stakeholders leads to the development of evaluative criteria. Importance and merit of criteria are determined within context using multiple measures. Rubrics or benchmarking allow for quantitative comparison of results within and across contexts. Sub-evaluations may assess quality of process content and implementation, value of impacts on participants, cost effectiveness relative to similar options, or exportability of components that could be valuable in other situations [6]. When reporting results, meta-evaluating process itself can attest to the validity and utility of findings, and the conduct, credibility, and cost of evaluation.

Framework Levels

Our framework for the evaluation of large online graduate-level engineering courses is meant to provide both process and outcome level data. Evaluation results are intended to provide instructors concrete feedback on their instructional approaches for continuous improvement and also provide administrators formative and summative evaluation of outcomes. Table 2 contains a description of each level, including the measured criteria, evaluation questions being answered, data sources from which evidence is derived, and the data use that connects evidence with evaluation questions (see Appendix). In this section, we provide details about the framework levels for interpretation and use.

Level 1: Learner Satisfaction (Attitude Level)

The first level investigates learner satisfaction through the attitudes of learners about the course. It measures the construct of learners' perceptions about the experience of participating in the course and what learning they received from it. Research questions at this level are "Was the course a good use of learners' time?", "Was specific content useful to learners?", and "Do learners intend to apply content?" Attitudes about the course are the starting point for understanding whether a MOOC was successful. Evidence of learner perceptions is collected through post course surveys and overall ratings. In questioning whether learners' time was well spent, this level should be supported with information about how course time was spent on learning experiences, as well as whether content was paced effectively and presented at the appropriate degree of difficulty. At this level, we can gain specific feedback about aspects of the course that did or did not work for participants. We are seeking to know the concrete aspects of course design and execution that helped learners meet their self-defined goals.

Level 2: Learning Outcomes (Learning Level)

The second level evaluates learning in terms of course performance. To know whether learners can successfully demonstrate course objectives, it asks the research questions "Did learners gain knowledge and skills from the course?", "Were learners able to demonstrate the knowledge and skills they gained?", and "Do knowledge and skills demonstrate completion of course objectives?" It expects that the course contains well-developed learning objectives, and that learning items are aligned to these objectives. Assignments and quizzes, and corresponding rubrics when applicable, are sources of learning outcome data. Discussion forum contributions are also an indicator of course performance, with learners' perspectives on their learning gathered from post course surveys. This level defines and supports what it means to successfully complete the course by linking outcomes to performance on course items. Assessments, homework, and discussion posts all provide opportunities for students to demonstrate what they know. Curriculum mapping is necessary in order to assert the degree to which students learned what the course creators purported they would learn. As a process for understanding curriculum coverage of a topic, depicting connections among course topics, and linking specific learning items to content [10], curriculum mapping supports evaluation along with learning outcome data.

Level 3: Pedagogical Practices (Support Level)

The third level is concerned with the quality of pedagogical practices to support learning. Its evaluation construct is whether pedagogy has helped learners achieve the established course objectives. Research questions at this level are "Are pedagogical practices aligned with course objectives?", "Do assignments and assessments align to course objectives?", "Are pedagogical practices aligned with best practices (e.g., peer interactions, regular assessment)?", and "Are pedagogical practices compliant with current ADA Standards for Accessible Design?" Compared to Level 2, this level uses rubrics, alignment documents, curriculum, and instructor perspectives of their own practices as data sources. To assess compliance with accessibility standards, course materials and the 2010 ADA Standards for Accessible Design [11] are referenced together. The two goals for this level are (1) identify pedagogies within the course that have proven effective at

helping learners accomplish their course-related goals, and (2) compare course design elements against policy standards to ensure compliance.

Level 4: Learner Use (Application Level)

The fourth level evaluates learners' intended and actual application of learning. From this level, we can better understand how the course contributed to learners' current and future careers, interests, and professional development. Research questions at this level are "Do learners intend to continue learning in subject area as a result of this course?", "Do learners intend to apply this material to their career?", "Will this course help learners advance their credentials?", and "Does this course enable new professional opportunities for learners?" Like Level 1, Level 4 regards the individual perspectives of learners as critical for judging whether the course allowed learners to advance in life. From post course surveys or interviews, we can ask detailed questions about what learners valued in the course, what its role was in helping them achieve professional or personal goals, and whether their achievements were recognized outside of the course. Learner use encompasses not only future or intended use of course material, but also the ways that students apply key concepts within the course. Knowledge and skills gained in the course can be applied to later assignments, projects, or discussion board interactions, where learners may demonstrate their understanding of topics by using them in later situations. From this level, we can determine the extent to which course material is relevant to learners' needs, and whether it has immediate and/or long-term value to them.

Level 5: Broader Impacts/Return on Investment (Impact Level)

The fifth level looks comprehensively at the course by evaluating whether its impacts and return on investment outweighed its costs. It looks at the climate of the providing organization towards MOOCs by evaluating how effective the course was in contributing to the open learning initiative or values of the institution. The research questions at this level are "Does the course contribute to the strategic goals of the institution (e.g., certifications, degrees, or research?)", "Did the course enable the institution to reach new audiences?", and "Were instructors motivated to improve their pedagogical practices?" It looks for summative metrics of the course's success in reaching larger educational program goals. Data sources are metrics of completion rates or certificates for the entire program and enrollment in degree paths consisting partially or entirely of MOOCs. If applicable, it evaluates the integration of the course with traditional education programs. It includes end-of-course metrics showing how the course supports long-term goals of individual learners through certifications. It also considers the economic return for providers. In this way, indicators are used to determine the extent to which the course was a good investment for multiple stakeholders.

Discussion

When applying this framework towards student learning and course improvement, it can have both formative and summative purposes. Formative use relies on student data from activity within the course, such as discussion board activity, to conduct ongoing evaluation of how the course is supporting learner needs and goals. Data from surveys, grades, course materials are used to take repeated snapshots of the course over time, capturing variations in learner

perceptions and course delivery. Results and conclusions are used to inform change at all relevant levels, such as instruction methods. On the other hand, summative assessment provides a conclusive look at course success once it has finished. The same data sources and analysis methods are used, but this time from understanding the course as a whole.

We anticipate three intended user groups who can use this framework to evaluate large online graduate-level courses. First, course providers may apply it to a more long-term educational plan. For example, universities who are offering distance instruction as a part of professional engineering education should be interested in evaluating how well individual courses are performing, in the context of the program as a whole. As users, they would be interested in metrics of quality and success that are measurable and context-specific, which allow them to draw summary conclusions. Second, we expect course instructors to apply this framework to evaluate the delivery and quality of their courses. It can provide them with a means of organizing student feedback to make sense of what was or was not successful in the course. Because it addresses curriculum elements, it can help instructors identify materials and practices which are successful or need improvement. They may use it as a tool for ongoing course development, and instructors who are managing multiple courses or a sequence of courses may use the framework for comparing among courses on performance. Third, external researchers and evaluators may use the framework for determining the success of a course, being able to draw conclusions both at the learner level and within a larger administrative context.

Conclusions

MOOC platforms enable thousands of learners to enroll in a professional development or graduate-level engineering course, but there is a dearth of researchers discussing how these courses should be evaluated. When engineering courses are offered purely as a MOOC, free or low-cost short courses with no real investment by learner, they may not need a rigorous evaluation to satisfy stakeholder needs. However, when universities and corporations are offering courses through MOOC platforms for credentialing or credit-bearing purposes, there warrants a higher degree of evaluation. The framework presented in this paper suggests that both process and outcome level evaluation are important for continuous improvement of the courses and to ensure the quality does not suffer. Evaluating large online graduate-level courses for “best fit” with an audience requires looking at not only the course’s content and design, but also its relevance to context. We must evaluate whether the course has met a specific need for learners and helped them accomplish the goals they had upon enrollment. In considering the needs of stakeholders such as instructors and course providers, evaluation should also be able to say how the course outcomes satisfies all stakeholders and supports the values of the learning environment.

We provide a straightforward evaluation framework for the community in engineering education to begin collecting similar types of evaluation data. As more institutions use the framework, the community can begin to benchmark outcomes and develop a list of best practices in large online courses specifically for professional engineers.

References

1. J. Bersin, "A Bold New World of Talent, Learning, Leadership, and HR Technology Ahead," 2016. Available: <http://marketing.bersin.com>.
2. D. Shah, "7 New MOOC-Based Fully Online Master's Degrees Announced by edX," Available: <https://www.classcentral.com/report/7-new-mooc-based-fully-online-masters-degrees-announced-edx/>.
3. Author et al. 2018.
4. Author et al. 2019.
5. J. Lai and M. Bower, "How is the use of technology in education evaluated? A systematic review," *Computers & Education*, vol. 133, pp. 27-42, 2019.
6. E. J. Davidson, "Evaluation methodology basics," Sage Publications, Inc., Thousand Oaks, CA, 2005.
7. M. Scriven., "The logic of evaluation," Edgepress, 1981.
8. T. R. Guskey, "Evaluating professional development," Corwin Press, Thousand Oaks, CA: Corwin, 2000.
9. Quality Matters, "Non-annotated Standards from the QM Continuing and Professional Education Rubric (2nd ed.)", 2015. Available: <https://www.qualitymatters.org/sites/default/files/PDFs/StandardsfromtheQMContinuingandProfessionalEducationRubric.pdf>
10. H. H. Jacobs, *Mapping the big picture: Integrating curriculum & assessment, K-12*. Alexandria, VA: Association for Supervision and Curriculum Development, 1997.
11. Department of Justice, "2010 ADA Standards for Accessible Design," 2010. Available: https://www.ada.gov/regs2010/2010ADAStandards/2010ADAStandards_prt.pdf

LEVEL	CONSTRUCT	QUESTIONS	DATA SOURCES	DATA USE
1. Learner Satisfaction (Attitude Level)	How satisfied learners are with the experience and takeaways of the course	Was the course a good use of learners' time?	<ul style="list-style-type: none"> • Post course surveys • Course ratings 	Investigating learner perceptions of the course in terms of their goals and experiences
		Was specific content useful to learners?		
		Do learners intend to apply content?		
2. Learning Outcomes (Learning Level)	Whether learners successfully demonstrate course objectives	Did learners gain knowledge and skills from the course?	<ul style="list-style-type: none"> • Course grades • Discussion forums • Assignments and quizzes • Rubric items • Post course surveys 	Assessing student achievement of course learning objectives
		Were learners able to demonstrate the knowledge and skills they gained?		
		Do knowledge and skills demonstrate completion of course objectives?		
3. Pedagogical Practices (Support Level)	Whether pedagogy helped learners achieve course objectives	Are pedagogical practices aligned with course objectives?	<ul style="list-style-type: none"> • Course rubrics, assessments, or other alignment documents • Instructor evaluation of 	Identifying pedagogies that are effective at helping students accomplish their course goals
		Do assignments and assessments align to course objectives?		

		Are pedagogical practices aligned with best practices (e.g., peer interactions, regular assessment)?	<ul style="list-style-type: none"> their own practices Instructor surveys or interviews 	
		Are pedagogical practices compliant with current ADA Standards for Accessible Design?	<ul style="list-style-type: none"> Course materials 2010 ADA Standards for Accessible Design 	Comparing course design with policy standards for alignment
4. Learner Use (Application Level)	How the course contributed to learners' current and future careers, interests, and professional improvement	Do learners intend to continue learning in subject area as a result of this course?	<ul style="list-style-type: none"> Post course surveys 	Determining the extent to which course material is relevant to learners' needs, and has both immediate and long-term value to learners
		Do learners intend to apply this material to their career?		
		Will this course help learners advance their credentials?		
		Does this course enable new professional opportunities for learners?		
5. Broader Impacts/Return on Investment (Impact Level)	Effectiveness of course as a whole in contributing to the open-learning strategic plan of the institution	Does the course contribute to the strategic goals of the institution (e.g., certifications, degrees, research)?	<ul style="list-style-type: none"> Completion rates for programs or certifications Enrollment in degree paths that include MOOCs 	Determining the extent to which the course was a good investment for stakeholders
		Did the course enable the institution to reach new audiences?		

		Were instructors motivated to improve their pedagogical practices?	<ul style="list-style-type: none">• Integration of course with traditional education• Instructor interviews	
--	--	--	--	--

© 2019 SEED Research Lab, Purdue University, Director Kerrie Douglas