Formative and Summative Assessment Options

You'll learn about:

- Coursera's assessment options
- Supporting learner self-assessment
  - The importance of elaborative feedback
  - In-video quizzes -- the quickest form of feedback
  - Formative quizzes as a teaching opportunity
  - Peer review for developing expertise
- Measuring learning and expertise
  - Planning summative assessments
  - Practices for summative quizzes
  - Measuring expertise with peer review assessments
  - Programming assignments
- Improving the course using assessment analytics

Overview

New MOOC instructors are often concerned about their options for providing quality practice and feedback for thousands of learners. It is true, given MOOC class sizes, assessments requiring individual grading by instructors or TAs are not feasible.   The Coursera platform addresses this by providing a range of automated assessment opportunities and peer-reviewed assessment opportunities -- supporting both formative and summative assessment. This guide provides advice on matching instructional goals to Coursera platform assessment formats and their usage.

RECOMMENDATION: Invest time in developing quality assessment opportunities with extensive feedback.

From *How Learning Works:*
"We all know that practice and feedback are essential for learning. Unfortunately, the biggest constraint in providing sufficient practice and feedback to students is the time it takes -- both on the part of students and faculty." (Ambrose, Bridges, DiPietro, Lovett, & Norman, 2010, p. 124)

Since learner assessment evaluation on Coursera are either auto or peer-graded -- MOOC instructors can overcome a major concern of offering sufficient assessment opportunities.

Spending time in creating assessment and feedback opportunities for MOOC courses is a good investment -- because these efforts impact thousands of learners.   It is true, these extra assessments will cause learners to spend more time engaging in the course. But, in the world of MOOCs, that is a desired outcome.

Coursera's platform supports a range of assessment formats:
● In-video quizzes (lightweight, ungraded questions presented within a video lecture)
● Quizzes (supporting multiple choice, numerical response, check all that apply)
● Peer review assessments (an approach useful for practice which results in materials not automatically gradable)
● Programming assignments

RECOMMENDATION: Create frequent opportunities for formative and summative assessment.

A key feature of open online courses is the focus on supporting ***mastery learning***.   Benjamin Bloom originally defined mastery learning as when "the students are helped to master each learning unit before proceeding to a more advanced learning task" (Bloom, 1985). As online learners can re-watch lectures and re-engage with learning content as much as they like, they can be supported in mastering smaller modules of content before continuing to the next material.

Effective support for mastery learning entails developing practice and feedback opportunities for learners that:
● Allow them to accumulate a sufficient quantity of practice
● Engage in practice that is goal-directed
● Engage in a range of appropriately challenging tasks that scaffold learners in developing deeper expertise (Ambrose, et. al, 2010, Chapter 5)

On campus, most courses administer a small set of large assessments for logistical and scheduling reasons - e.g., midterms, final exams, final papers, etc. But, research suggests that more frequent assessments are usually far more effective for developing learners' long term memory for lecture content, a phenomenon called "the testing effect" (Roediger & Karpicke, 2006) (Hattie, 2009).

As detailed below, the MOOC course can be designed with smaller, more directly-tied to content formative and summative assessment opportunities. For more information on MOOC course design, including the role of assessment, please see the Course Planning and Design Guide.

Coursera's assessment options

The Coursera platform supports various assessment formats. This section provides a basic pedagogical overview with further recommendations and examples provided in later sections. Platform tool specifics (types of questions supported in quizzes, etc.) can also be found here.

Formats and Learning Goals

RECOMMENDATION: Choose an assessment format based on the learning goal.

A common misconception is that MOOCs cannot provide deep or rich practice and assessment opportunities. Coursera wants to support instructors in developing assessments spanning a range of cognitive levels (e.g., Bloom's taxonomy (Anderson, Krathwohl, Airasian, Cruikshank, Mayer, Pintrich, ... & Wittrock, 2001)). The following table outlines the various formats of assessment options on Coursera and provides examples of learning outcomes which might be appropriate to the format.

Assessment Format	Learning Goals
In-video quiz	● Connect new content with learner's prior knowledge ● Motivate active video lecture watching ● Allow learners to assess whether they've learned the key video lecture outcomes (recommended Bloom's taxonomy levels: remember and understand)
Quiz	● Engage learners in deepening and applying their knowledge (recommended Bloom's taxonomy levels: understand, apply, analyze, evaluate) ● Increase learning through repeated recall and retrieval of information (Roediger & Karpicke, 2006)
Peer review	● Develop more expert-like evaluative judgment (critique, analyze, provide feedback) ● Demonstrate expertise through creation of non-automatically gradable materials (e.g., video presentations, essays, reports, reflections, designs)
Programming assignment	● Demonstrate ability to create a program to solve a highly specified problem (with replicable behavior for given input/output sets)

Formative v. Summative Assessment

RECOMMENDATION: Use assessment formats for both formative (guiding learning) and summative (measuring learning) tasks.

Instructors and learners most commonly associate assessments with a "final" measurement of learning -- commonly exams. This is still an important function of assessments in MOOCs. However, increased importance should be placed on additional practice and formative assessment opportunities to allow learners to both self-assess their knowledge and to deepen their learning (Ambrose, Bridges, DiPietro, Lovett, & Norman, 2010) (Karpicke & Grimaldi, 2012). Any assessment format (except for in-video quizzes) can be used formatively or summatively -- though the design of the assessment may differ. The table below provides some examples of how a specific assessment format might be used differently for formative or summative purposes.

Assessment Format	Formative	Summative
Quiz	Does not contribute to course grade (or contributes minimally) Can be retaken an unlimited number of times Occurs every 20-30 minutes of content Has 2-10 questions for every 20-30 minutes of content "Teaches" the answer to the question through extensive explanations supplied via "option-level feedback" (a feature of the quiz design system) Contains multiple questions on the same concept to give more practice opportunity Has questions whose	Contributes substantively to course grade May be limited in how frequently learners can re-take it (coming soon in the on-demand platform) Occurs every 1-2 hours of content Has 10-20 questions for every 1-2 hours of content Option-level question feedback refers learners to course materials where this content is taught Includes questions at all cognitive levels Is a subset (with modifications) of the content in formative assessment materials

	cognitive level is appropriate to the learning materials just covered	
Peer Review	Does not contribute to course grade (or contributes minimally) Provides opportunity for preliminary feedback (e.g., a revised product may be submitted in a later peer review assessment) Supports submission of and feedback on preliminary stages in larger project Is framed as a task to help learners develop more expert-like evaluative judgement (e.g. to learn how to critique or analyze) Can be used to promote reflection on learning processes (e.g., might accompany a summative assessment)	Contributes substantively to course grade Allows learner to demonstrate evaluative judgment skills Enables learners to demonstrate knowledge and skill in integrative and more complex tasks. May provide a useful "take-away" for documenting the learning / value of the course (e.g., an item useful in a portfolio, as something to discuss in a job interview, as something to desirable to share with friends or family)
Programming Assignment	Does not contribute to course grade (or contributes minimally) Can encourage concept practice over problem solving (e.g., CodingBat) Can reduce procrastination and/or scaffold larger project development (e.g., low-stakes intermediate deliverables)	Contributes substantively to course grade Allows learner to demonstrate problem solving and programming skills.

Supporting learner self-assessment

The importance of elaborative feedback

A strength of online learning is the ability to provide immediate feedback to learners through automatically evaluated assessments. Research generally suggests that immediate feedback is more helpful for learners than delayed feedback (Shute, 2008, pp 163-165).*   Additionally, the best feedback is formative, not merely evaluative (indicates correct answer only). Formative feedback can be defined as follows:

"Formative feedback is defined… as information communicated to the learner that is intended to modify his or her thinking or behavior for the purpose of improving learning." (Shute, 2008, pp 154)"

RECOMMENDATION: Treat formative assessments as another teaching venue by providing elaborative feedback.

The importance of developing "good" questions or assessment prompts (e.g. peer review and programming assignments) is obvious. Formative assessments need to be aligned with the goal (content and level) for the material just taught. However, it is just as important (and often a second thought) to provide elaborative feedback describing the "what, how, and/or why" of the problem (Shute, 2008, p. 177). Simply indicating that an answer is correct is much less helpful for learning.

Elaborative feedback can be provided for any of Coursera's assessments. An example of basic evaluative and elaborative feedback is provided below. As per research findings recommending the use of response-specific feedback (Shute, 2008, p. 159) Coursera's various quiz systems provide the opportunity to provide written feedback for each question option (e.g., feedback regarding the specific response).

Example Question: (California Department of Motor Vehicles)
When you are merging onto the freeway, you should be driving:
    A. At or near the same speed as the traffic on the freeway.
    B. 5 to 10 MPH slower than the traffic on the freeway.
    C. The posted speed limit for traffic on the freeway.

Basic Evaluative Feedback (not recommended)	Option-level elaborative feedback (recommended)
A is correct. OR A is correct, when merging onto the freeway you should be driving at or near the same	A is correct. If you are driving at or near the speed of traffic around you, you will be able to merge into the right hand land with minimal disruption to the flow of traffic around you. B is not correct. When merging onto the

| speed as the traffic on the freeway.

Note: This may seem elaborative, but it does not explain, merely restates the answer. | freeway if you are travelling slower than the traffic around you, other drivers will have to brake or change lanes in order to allow you to enter the flow of traffic. This could cause drivers to make sudden changes which may cause accidents.

C is not correct. The posted limit for any roadway is a limit and may not be a safe speed for traffic under current conditions. When merging into traffic, the most important thing is to be travelling at approximately the same speed as those around you so that you can join the flow of traffic with the least disruption. |

*This is especially the case for field (non-laboratory) studies. Laboratory studies of delayed feedback often test a delay of less than a minute -- not the day(s) of feedback delay common in university settings.

In-video quizzes -- the quickest form of feedback

Coursera's platform can engage learners with various formats of quiz questions in two settings:

- **In-video quiz (IVQ)** -- questions that occur during the watching of a video lecture which require the learner to engage before continuing on (e.g., select one or more answers, answer a poll, pause and reflect before continuing on). Data from these questions is not available to be reviewed by the instructor, nor can it be part of the grade in the course. IVQs are purely for formative learning.
- **Quiz** -- a separate course item which contains a number of questions and for which the instructor can set a "passing grade". Data on learner performance on quiz questions can be easily reviewed by the instructor in the analytics dashboard.

This section provides recommendations on uses of in-video quizzes.

RECOMMENDATION: In-video quiz questions should require little learner time (e.g. < 30 seconds).

RECOMMENDATION: In-video quiz questions should encourage learners to keep watching the video.

While it is true that IVQs can encourage learners to be more active in their video watching, if questions are too complex, take too much time to answer, or require higher level application of the material being taught -- learners may use these as a trigger to stop watching the video

lecture. Good IVQs can help form community and help learners leverage their pre-existing knowledge (for more, please see our <u>Video Lecture Practices guide</u>).

However, IVQs can also provide the most basic level by which learners can self-assess.

RECOMMENDATION: Create in-video quiz questions at the end of video lectures to test basic understanding of the learning goals of the video.

RECOMMENDATION: Fully teach the answer to the in-video quiz question by explaining why each answer option is right or wrong using option-level feedback.

The question below provides an example of testing a basic video lecture learning goal and uses option-level feedback to immediately and thoroughly "teach" the content of the question (which was lectured on using almost the exact same verbiage in the lecture).

Lecture learning goal: Learners can describe the relative academic performance of women STEM majors who leave STEM with male STEM majors.

RECOMMENDATION: It is OK for formative assessments to have fewer than 4 option choices.

RECOMMENDATION: Choose incorrect options (i.e., distractors) based on common misunderstandings.

Having a larger number of "distractor" (wrong) multiple-choice question options is important in summative assessment; it reduces the chance the learner answered correctly merely due to chance. Formative assessment through in-video quizzes should focus learners on the goals of exactly what they should have learned and options should be selected to allow explanation of common misunderstandings.

Formative quizzes as a teaching opportunity

The quiz feature can be used to supplement in-video quizzes in helping learners check their learning.

RECOMMENDATION: Create 3-5 formative quiz questions for each 20-30 minutes of video lecture.

RECOMMENDATION: Use option-level feedback to fully teach the answer to each question.

The Coursera "quiz" assessment can be utilized in a formative (rather than summative) way by not including the quiz grade as part of the overall course grade. Additionally, like in-video quiz questions (see prior section), research recommends that option-level elaborative feedback be provided -- fully teaching the material of the question.

RECOMMENDATION: Formative quiz questions can target questions which are more cognitively challenging or take more time to answer (compared to in-video quizzes).

Formative quizzes are a good place to first ask learners to more deeply understand, apply, analyze or evaluate lecture material (Anderson, Krathwohl, Airasian, Cruikshank, Mayer, Pintrich, ... & Wittrock, 2001). Formative quiz questions should build (throughout the course) in terms of cognitive challenge as learners' background and abilities in the course content grows. Do not feel the need to save "hard questions" for summative quizzes. Formative quizzes give learners the chance to practice these abilities and improve them before summative assessment. Well-written option-level elaborative feedback on cognitively challenging questions can illuminate for learners how an expert approaches a challenging issue and the considerations they make in analyzing or evaluating it.

RECOMMENDATION: Formative quiz questions can replicate (perhaps slightly modified) in-video quiz questions.

It is also fine to replicate or ask slight modifications of in-video quiz questions in formative quizzes. A significant research base has shown the value of testing for learning (Roediger & Karpicke, 2006). That is "retrieval is not a neutral assessment of the contents of one's

mind, but the process of retrieval itself contributes to learning." (Karpicke & Grimaldi, 2012, p. 402).

Peer review for developing expertise

Peer review (sometimes called peer assessment) is powerful learning tool which instructors may use for both logistical and pedagogical reasons.

RECOMMENDATION: Use peer review for learner activities which do not produce a result which can be automatically graded (e.g., for essays, reflections, reports, presentations, research papers, portfolios, interviews, etc.).

The peer review system is the only manner of assessing learner work that cannot be automatically assessed through the quiz system or through a programming assignment that is automatically graded.

RECOMMENDATION: Use peer review to help learners develop more expert-like evaluative judgement.

Research on peer review (mostly in traditional university settings) argues that, if implemented well, peer review can fill a vital and rarely addressed goal of higher education -- to develop learner self-regulation through development of evaluative judgement (Greene and Azevedo, 2007).

Although most commonly associated with larger projects (e.g., research papers, etc.), smaller effort peer review assessments can support formative development of the skills necessary for learners to succeed in more extensive projects. For example, formative peer review could be used for:
- Encouraging and providing feedback on preliminary stages of a project (outline, draft, first parts, portions of code).
- Developing supporting skills for other activities (e.g., asking for analysis of buggy code, a worked out solution for a problem (possibly containing an error).
- Eliciting reflection on learning (either generally or in conjunction with other tasks such as problem solving or programming).

RECOMMENDATION: Seek advice on developing formative peer review assessments. The question prompt, rubric design, and contextualization for learners is critical to success.

Research on best practices for question and rubric design, process for peer review and peer review in MOOC settings is very limited (Nicol, 2014) (Kulkarni, Wei, Le, Chia, Papadopoulos, Cheng, Koller, & Klemmer, 2013). This is an area of active interest and research at Coursera and more comprehensive recommendations and examples are forthcoming.

Both general peer review research and Coursera experience recommends the following basics:
- Make the prompt as explicit as possible; especially concerning the format and content of the material to be created
- Create review questions matched to specific sections or parts of the material to be created; include these question with the prompt (e.g, before the learner submits their materials)
- Make sure learners can accurately answer question prompts; don't use words which imply expert judgement (e.g., "does the essay adequately address theories from the course")
  - Operationalize expert-specific terms (e.g., "adequately address" becomes "discusses at least 3 theories from the course" (even better if a list of the theories covered in the course is provided)).

It's very valuable to ask someone of the approximate background of your anticipated learners to critically read over the peer review prompt and evaluation questions. Ask them:
- Do they know what it is they should produce
- Could they make the judgements required in the evaluation questions

RECOMMENDATION: If possible, provide one or more examples of the product to be produced

Providing an example of what should be produced is especially helpful for MOOC learners of diverse background preparation and who may have English as a second language. Consider providing an example using a different topic, data or input set, or even an outline of the product with placeholders for content (e.g., graphs).

Measuring learning and expertise

Although some of the core advice on formative assessment is also appropriate for summative assessment -- there are some key differences.   This section will repeat core recommendations and highlight critical differences.

Planning summative assessments
Planning summative assessments follows much of the same process as planning formative assessments. This assumes that learning outcomes for the course have been identified first.

RECOMMENDATIONS:
- Use course outcomes to create assessments which measure mastery of those outcomes
- Create assessments to measure **all** learning outcomes
- Provide summative assessments frequently (e.g. after each module or week)

- Choose an assessment format based on the learning outcome

The more distinct separation of formative and summative assessment recommended for MOOC courses places a greater emphasis on the use of summative assessments for learners to gauge their success or mastery. Summative assessments should provide a complete documentation of the depth of learning that has occurred and the expertise level attained.   Summative assessments in MOOCs are less constrained by traditional university assessment logistics including:
- Which content is most important to measure within a 2 or 3-hour setting (e.g. final exam)
- Managing the trade-off in teaching time and assessment time (e.g., having relatively few exams)

Summative assessments in MOOCs can be more comprehensive, in part because they can be administered more frequently.

RECOMMENDATION: Double-check that lecture plans and formative assessments adequately support the learning needed to succeed at these assessments

In MOOC courses, learners have complete access to all instructional materials -- even after taking a summative assessment. When learners struggle with a summative assessment (be it quiz, programming assignment, or peer assessment) they will search through video lectures and any other instructional materials for the answers. If the provided materials don't adequately prepare them, they will ask on discussion forums. Checking that instructional materials support learning what is needed for assessments will reduce instructor engagement time while the course is running.   The next section has specific recommendations for quizzes to address this issue.

Practices for summative quizzes

RECOMMENDATION: Use multiple-choice questions.

Contrary to common beliefs, multiple choice questions have been shown to be as good or better than open-ended questions.
- In a study of college physics exams written by physics instructors, multiple-choice and open-ended exam questions yield almost the same results (Scott, Stetzler, & Gladding, 2006)
- Extensive and detailed analysis of advanced placement exams show multiple-choice questions to be superior for assessing mastery (Lukhele, Thissen, & Wainer, 1994)

RECOMMENDATION: Utilize best practices in multiple-choice question design.

Although instructors rarely receive training in writing quiz questions, an entire field exists which studies assessment practices that produce reliable testing outcomes. A highlight of these practices can be found below.   For a university-focussed set of recommendations with examples see this site from Coursera partner Vanderbilt University.

We organize a summary of best practice either by question structure or question content based on recommendations from (Haladyna & Downing, 1989, pp. 40-41).

Structure:

Question stem:
- Make the stem either a question or a sentence which can be completed (completion only at the end of the sentence is recommended)
- Make sure the directions in the stem are clear
- Keep the stem as short as possible -- avoid excess verbiage
- Don't use negative phrasing (e.g., which of the following is NOT…) The "check all that apply" question format is better for this situation
- Include as much of the necessary phrasing as possible in the stem (rather than repeating it in the options)

Question options:
- Use as many options as feasible (to reduce the chance of random guessing of the correct answer)
- Phrase options positively
- Put options in a logical order; either shortest to longest (text) or smallest to largest (numbers)
- All options should be mutually exclusive (i.e, non-overlapping or independent)
- Options should be as homogeneous in content as possible
- Avoid dramatically different option lengths; specifically try not to have the correct answer be the longest, most specific, using the most technical terms
- Rarely or never use:
    - All of the above
    - None of the above
    - I don't know
    - Never
    - Always

Content:

Question:
- Address a single topic or instructional objective in a given question
- Test for important and significant material - avoid trivial material
- Avoid verbatim phrasing from instructional materials (e.g., video lecture)
- Make sure not to use new vocabulary
- Test higher-level thinking, not just remembering (see more details below)

Question distractors:

- Use plausible distractors; if possible use common student errors or misconceptions
- Use familiar but incorrect phrases as distractors
- Use true statements that are not the correct answer to the question
- Avoid humorous or absurd options

RECOMMENDATION: Design questions which test higher order learning.

Multiple-choice questions have a bad reputation for being useful only for measuring lower-order thinking skills; recall or remembering - the lowest level on Bloom's revised taxonomy of educational objectives. Though it takes some practice, one can learn to develop multiple-choice questions that assess higher-order skills such as understanding, application, and analysis. Some common types of questions which target higher-order skills might contain:

- Application of knowledge to real world scenarios or case study (e.g., "A nurse is making a visit to a home-bound patient with a history of…" (Brame) ; "You job is to design a bridge for a location with the following characteristics… which type of bridge would you recommend?")
- Analysis (e.g., "Which of the following is a correct conclusion that can be drawn from the graph above?" (Mueller, 2014))
- Solution evaluation (e.g., "Maria produced the following diagram and explanation in response to… how would you judge her solution?" (University of Oregon Teaching and Learning Center, 2014))
- Justification of methods and procedure (e.g., "Why is adequate lighting necessary in an aquarium?" (Brigham Young University Center for Teaching and Learning, p. 2), "Given the scenario… which equation below is what you would set up to start solving the problem?")

RECOMMENDATION: Use quiz question feedback to point learners to video lectures or lessons (in the on-demand platform) where this material can be learned.

Coursera quiz questions allow instructors to provide feedback to learners which they receive after completing the entire quiz. The type of feedback recommended for summative quizzes differs from that recommended for formative quizzes. Formative quizzes should use feedback which teaches the answer to the question. Summative quizzes seek to document learner attainment of mastery and, hence, should direct learners back into the instructional materials where they can learn more about the topic.

Measuring expertise with peer review assessments

RECOMMENDATION: Use peer review for learner activities that generate non-identical (individual) learner-generated items such as the following:

- Essays, research papers, reports and analyses
- Presentations (e.g., video)
- Learner reflection on learning experiences (e.g., on the process or experience of doing something such as writing an essay or program, the learner's personal value of what has been learned)
- Creative materials (e.g., writing, artwork, music, video)

The peer review system is the only manner of assessing learner work that cannot be automatically assessed through the quiz system or through a programming assignment that is automatically graded. But the above activities can be used to engage learners in a range of valuable ways of deepening learning through reflection or the creation of professional or discipline-appropriate documents or items.

A fairly novel format for peer review is to have learners submit videos they have created -- useful when the ability to present is a component of practical expertise in an area.

RECOMMENDATION: Use peer review to develop learners' evaluative judgement

Bensley states "Evaluative judgement, it could be argued, is the cornerstone of critical thinking in all disciplines." (as cited in Nicol, 2014)

Research on peer review (mostly in traditional university settings) argues that, if implemented well, peer review can fill a vital and rarely addressed goal of higher education -- to develop learner self-regulation through development of evaluative judgement (Greene and Azevedo, 2007).

Bensley's elaboration of what constitutes critical thinking in all disciplines provides a useful tool for instructors developing peer review assignments with the goal of developing learner evaluative judgement:

- "distinguishing arguments from assertions,
- finding the central question,
- appraising the form and qualities of evidence,
- making sound predictions from theories,
- generating good hypotheses,
- constructing convincing arguments
- comparing the quality of different things - texts, arguments, objects,
- expressing one's reactions to texts,
- considering multiple perspectives"
  (as cited in Nicol, 2014, formatting added)

RECOMMENDATION: Carefully craft and seek feedback on the peer review prompt and evaluation questions

Research on best practices for question and rubric design, process for peer review and peer review in MOOC settings is very limited (Nicol, 2014) (Kulkarni, Wei, Le, Chia, Papadopoulos, Cheng, Koller, & Klemmer, 2013). This is an area of active interest and research at Coursera and more comprehensive recommendations and examples are forthcoming.

Both general peer review research and Coursera experience recommends the following basics:
- Make the prompt as explicit as possible; especially concerning the format and content of the material to be created
- Create review questions matched to specific sections or parts of the material to be created; include these question with the prompt (e.g, before the learner submits their materials)
- Make sure learners can accurately answer question prompts; don't use words which imply expert judgement (e.g., "does the essay adequately address theories from the course")
    - Operationalize expert-specific terms (e.g., "adequately address" becomes "discusses at least 3 theories from the course" (even better if a list of the theories covered in the course is provided)).

It's very valuable to ask someone of the approximate background of your anticipated learners to critically read over the peer review prompt and evaluation questions. Ask them:
- Do they know what it is they should produce
- Could they make the judgements required in the evaluation questions

RECOMMENDATION: If possible, provide one or more examples of the product to be produced

Providing an example of what should be produced is especially helpful for MOOC learners of diverse background preparation and who may have English as a second language. Consider providing an example using a different topic, data or input set, or even an outline of the product with placeholders for content (e.g., graphs).

Programming assignments

Programming assignments are often appropriate for documenting development of programming or software engineering skills. Coursera's "programming assignment" format supports program

solutions which are auto-gradable only. Programs which are not auto-gradable can be assigned through the peer review assessment format.

However, even in the cases where an auto-gradable program is produced, it may be valuable to create a "paired" peer review assessment. This peer review assessment is recommended for evaluating learner ability to explain their program and to engage learners in the process of expert reflection on their design and development processes. Example peer assessments for programming assessment are forthcoming.

Improving the course using assessment analytics

The MOOC environment provides instructors a unique opportunity to improve their courses, not based merely on learner opinion surveys (i.e, student evaluations), but based on the data of how learners perform on various assessments. Much of this data is available to instructors through the analytics dashboards found in the administration bar at the top of any course page.

RECOMMENDATION: Use learner performance on quiz assessments to identify areas for course improvement.

Coursera's analytics dashboard shows per-question information for all quiz questions. For each question, the dashboard shows the following information:
- the percent of learners who got the question correct
- for multiple choice or check-all-that apply question: the percent of learners who selected a particular option
- for short answer questions, the top most commonly entered responses and the percent of learners who entered that response

The instructor can see two data points for each of the metrics listed above. The dashboard shows each metric based on learners first attempt at a question and based on learners last attempt at a question.

As an example, consider a multiple choice question that 85% of learners got correct the first time and 98% got correct the last time they answered it. This question likely is a good question in that, even if not everyone got it correct the first time, learners were able to eventually come up with (hopefully be reviewing course materials) the correct answer.

Examples of question metrics that may indicate concern and possible solutions are shown in the table below.

Quiz Question Metric	Possible Explanations	Possible Actions to

Information		Improve Learner Experience
Relatively few (<70%) of learners get the question correct the first time they attempt it but most (>95%) of learners get it correct the last time they attempt it	The wording of the question is confusion or can be mis-read The question tests a small detail not emphasized in video lecture and/or formative quizzes	Re-word the question Increase clarity or emphasis on question in video lecture Add the question as a learning outcome presented at the beginning of the appropriate video lecture
Relatively few (<70%) of learners get the question correct the first time they attempt it and this number does not increase by much (<10% improvement) by the last attempt	The question content is not sufficiently well taught or explained in the video lectures or instructional material	Create a supplementary video providing further elaboration or explanation of the content (possibly in a video lecture "series" titled "Additional Explanation") Re-record the appropriate video lecture(s) to provide greater clarity or explanation of the content
Very few (<40%) of learners get the question correct the first time they attempt it and this number does not increase by much (<10% improvement) by the last attempt	The question requires skills or understanding above the level of what learners are prepared for by video lectures and instructional materials. The question is a "trick question"	Remove the question

RECOMMENDATION: For peer review, use analytics information on the variance of learner scoring of evaluation questions to identify evaluation questions that are not clear to learners.

A quantitative measure of valuable improvements to make to peer review assessments is the amount of variance in learner response to specific evaluation questions. Each learner evaluates the work of their peers through a process of answering a set of evaluation questions. The analytics dashboard for peer review assessments shows two numbers for each evaluation question:

- the average score for that question as defined by the average of all reviews of all submissions (e.g., for each submission, for example, 3 different scores might have been recorded by 3 different reviewers)
- the grader variation for that question as defined by the average amount that the reviewers of a specific submission differed in their grading of this question - as averaged over all submissions

Instructors seeking to improve their peer review metric can start with those questions with the greatest amount of variance. Instructors seeking to improve learner performance on peer reviews can start with the questions where learners scored the lowest and consider whether the peer review prompt could be clarified or whether additional materials should be created to support learners in preparing for the peer review assessment.

References

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., Norman, M. K., (2010). *How Learning Works: 7 Research-Based Principles for Smart Teaching*. San Francisco: Jossey-Bass.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., ... & Wittrock, M. C. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives, abridged edition. White Plains, NY: Longman.

Bloom, B. S. (ed). (1985). *Developing Talent in Young People*. New York: Ballantine Books.

Brame. *Writing Good Multiple Choice Test Questions*. Retrieved from: http://cft.vanderbilt.edu/guides-sub-pages/writing-good-multiple-choice-test-questions/

Brigham Young University Center for Teaching and Learning. *14 Rules for Writing Multiple-Choice Questions.* Retrieved from: https://www.schreyerinstitute.psu.edu/pdf/14RulesforWritingMultiple-ChoiceQuestions2009.pdf

California Department of Motor Vehicles. *Sample 1 of C Class Driver Written Test.* Retrieved from: http://www.dmv.ca.gov/portal/dmv/detail/pubs/interactive/tdrive/exam

Greene, J. A., & Azevedo, R. (2007). A theoretical review of Winne and Hadwin's model of self-regulated learning: New perspectives and directions. *Review of Educational Research, 77(3)*, 334-372.

Haladyna, T. M. & Downing, S. M. (1989) Validity of a taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education, 2(1)*, 51-78.

Hattie, John. (2009) *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* Routledge.

Karpicke, J. D., & Grimaldi, P. J. (2012). Retrieval-based learning: A perspective for enhancing meaningful learning. *Educational Psychology Review*, *24(3)*, 401-418.

Kulkarni, C., Wei, K. P., Le, H., Chia, D., Papadopoulos, K., Cheng, J., Koller, D. & Klemmer, S. R. (2013). Peer and self assessment in massive online classes. *ACM Transactions on Computer-Human Interaction (TOCHI), 20(6)*, 33.

Lukhele, R., Thissen, D., & Wainer, H. (1994). On the Relative Value of Multiple-Choice, Constructed Response, and Examinee-Selected Items on Two Achievement Tests. *Journal of Educational Measurement, 31(3)*, pp. 234-250.

Mueller, J. (2014). *Assessing More Than Facts.* Retrieved from:
http://jfmueller.faculty.noctrl.edu/toolbox/tests/morethanfacts.htm

Nicol, D.J. (2014). Guiding Principles for Peer Review: Unlocking Learners' Evaluative Skills. In C. Kreber, C. Anderson, N. Entwistle, and J. McArthur (Eds.), *Advances and Innovations in University Assessment and Feedback* (pp. X - Y). Edinburgh: Edinburgh University Press.

Roediger, H. & Karpicke, J. (2006). Test-Enhanced Learning; Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, *17(3)*, 249-255.

Scott, M., Stelzer, T., & Gladding, G. (2006). Evaluating multiple-choice exams in large introductory physics courses. *Physical Review Special Topics - Physics Education Research V. 2,* 020102.

Shute, V. J. (2008). Focus on formative feedback. *Review of educational research, 78(1),* 153-189.

University of Oregon Teaching and Learning Center (2014). *Techniques for Writing Multiple-Choice Items that Demand Critical Thinking.* Retrieved from:
http://tep.uoregon.edu/resources/assessment/multiplechoicequestions/sometechniques.html#premise