Assessment approaches in massive open online courses: Possibilities, challenges and future directions

Article *in* International Review of Education · March 2018

DOI: 10.1007/s11159-018-9710-5

CITATIONS
49

READS
1,349

2 authors, including:

Yao Xiong
Pearson Inc
16 PUBLICATIONS **373** CITATIONS

SEE PROFILE

Int Rev Educ
https://doi.org/10.1007/s11159-018-9710-5

ORIGINAL PAPER

Assessment approaches in massive open online courses: Possibilities, challenges and future directions

Yao Xiong[1] · **Hoi K. Suen**[2]

Abstract The development of massive open online courses (MOOCs) has launched an era of large-scale interactive participation in education. While massive open enrolment and the advances of learning technology are creating exciting potentials for lifelong learning in formal and informal ways, the implementation of efficient and effective assessment is still problematic. To ensure that genuine learning occurs, both assessments *for* learning (formative assessments), which evaluate students' current progress, and assessments *of* learning (summative assessments), which record students' cumulative progress, are needed. Providers' more recent shift towards the granting of certificates and digital badges for course accomplishments also indicates the need for proper, secure and accurate assessment results to ensure accountability. This article examines possible assessment approaches that fit open online education from formative and summative assessment perspectives. The authors discuss the importance of, and challenges to, implementing assessments of MOOC learners' progress for both purposes. Various formative and summative assessment approaches are then identified. The authors examine and analyse their respective advantages and disadvantages. They conclude that peer assessment is quite possibly the only universally applicable approach in massive open online education. They discuss the promises, practical and technical challenges, current developments in and recommendations for implementing peer assessment. They also suggest some possible future research directions.

✉ Yao Xiong
yxiong@pitt.edu

Hoi K. Suen
HoiSuen@psu.edu

[1] University of Pittsburgh, 815 Learning Research & Development Center, Pittsburgh, PA 15260, USA

[2] The Pennsylvania State University, University Park, PA, USA

Published online: 06 March 2018

Keywords massive open online course (MOOC) · formative assessment · summative assessment · peer assessment · lifelong learning (LLL)

Résumé Méthodes d'évaluation dans les formations en ligne ouvertes à tous : possibilités, défis et futures orientations – L'essor des formations en ligne ouvertes à tous (FLOT) ouvre la voie à une ère de la participation interactive de masse à l'éducation. Tandis que l'inscription libre et massive ainsi que les avancées des technologies d'apprentissage créent des possibilités prometteuses pour l'apprentissage tant formel qu'informel tout au long de la vie, la réalisation d'une évaluation efficiente et efficace demeure un obstacle. Pour garantir un véritable apprentissage, il est nécessaire d'effectuer à la fois des évaluations *pour* l'apprentissage (évaluations formatives) qui mesurent les progrès actuels des apprenants, et les évaluations *de* l'apprentissage (évaluations sommatives) qui recensent les progrès cumulés des apprenants. La récente tendance des prestataires à attribuer des certificats et insignes numériques sanctionnant la réussite aux cours signale aussi la nécessité de résultats d'évaluation appropriés, sécurisés et précis qui garantissent la responsabilité. L'article examine les approches possibles d'évaluation qui correspondent à la formation en ligne ouverte à tous sous l'angle de l'évaluation formative et sommative. Les auteurs signalent l'importance et les défis d'évaluer les progrès des apprenants des FLOT dans ces deux buts. Ils identifient plusieurs approches d'évaluation formative et sommative en examinant et analysant leurs avantages et inconvénients respectifs. Ils concluent que l'évaluation entre pairs est fort probablement la seule approche universellement applicable dans la formation en ligne ouverte à tous. Ils en présentent les aspects prometteurs, les défis pratiques et techniques, l'évolution actuelle dans la réalisation de ce type d'évaluation ainsi que des recommandations. Ils proposent enfin plusieurs orientations possibles pour de futures études.

摘要 用于慕课的评估方法: 机会, 挑战及未来发展方向 – 慕课开启了大规模互动学习的新时代。 教育科技的进步为终身学习创造了很多机会, 但同时如何实现高效的学习评估也成为一个很大的挑战。 为了帮助学生学习, 形成性评估 (给学生提供阶段性反馈) 和总结性评估 (评估教学的最终效果) 都是必要的手段。 很多慕课开始向课程完成者颁发电子证书, 这一趋势也使得安全有效的评估变得尤为重要。 本文阐述了评估在慕课中的重要性和所面临的挑战, 并介绍了适用于慕课的形成性和总结性评估方法, 并对不同方法的优势和劣势进行了分析。 作者认为在慕课中, 学生互评是一个普遍适用的评估方法。 本文还对学生互评的优势、 挑战、 发展趋势以及实际应用中的问题进行了探讨, 最后提出了慕课评估方法未来的发展方向。

Background

Massive open online courses (MOOCs) emerged with the Open Educational Resource (OER) movement,[1] which ushered in an era of open online education and massive online learning. Distinguished from traditional distance learning and online courses, massive scale and open access are the two main unique features of MOOCs. Several MOOC providers (e.g. *Coursera*, *Udacity* and *edX*) associated with elite universities emerged in 2012, which propelled MOOCs into the spotlight for the higher education community. While many policy makers and educators were excited about the potential of MOOCs for closing the post-secondary attainment gap (Sandeen 2013a) and partly fixing the "cost disease" in higher education (Bowen 2013), some worried that traditional brick-and-mortar higher education would be disrupted negatively by the online education innovation. By 2018, the MOOC excitement seems to have subsided a little. Looking back, we are now able to speak rationally about both the exciting development of MOOCs and the obstacles impeding their further development.

George Siemens (2013), one of the pioneers of the MOOC phenomenon, has summarised three formats of MOOCs, namely *connectivist* or "cMOOCs", *extended* or "xMOOCs" and "quasi-MOOCs". The distinction between the first two, cMOOCs and xMOOCs, is mainly their different pedagogical foundations. In a cMOOC, learners actively collaborate with each other to create knowledge. In an xMOOC, learners watch and listen to video lectures etc. created and uploaded by the course provider. cMOOCs are built on the connectivism theory that knowledge is generated among learners, while xMOOCs utilise a more didactic teaching mode where teachers give instructions as experts and learners consume knowledge. The third category of MOOC, quasi-MOOCs (e.g. *Khan Academy*, *Lynda.com*, *MIT OpenCourseWare*, and *YouTube* how-to videos) are not essentially courses, but are more akin to educational resources.

The pedagogical differences between cMOOCs and xMOOCs result in different methods of assessment. Within cMOOCs, it is difficult to assess learning, knowledge, ability, mastery or achievement of individual students using traditional methods, such as multiple-choice questions, because the core component of a cMOOC is knowledge generation and knowledge sharing (Siemens 2005). The small number of fee-paying students who have enrolled for the purpose of earning credits are usually evaluated by course instructors based on the overall qualities of whatever they have produced or created (assignments, artefacts etc.), while the majority of the same course's participants, having opted for unaccredited participation at a cheaper rate (or even for free), remain unevaluated (Hollands and Tirthali 2014). Students in an xMOOC, on the other hand, are assessed individually mainly by automatic machine grading or by peer assessment. In the case of quasi-MOOCs, no assessment applies, because they are structured online learning contents without any assignment, student interaction or course deadlines.

[1] "The open educational resource (OER) movement has been growing rapidly since 2001, stimulated by funding from benefactors such as the Hewlett Foundation and UNESCO, and providing educational content freely to institutions and learners across the world" (Sclater 2009, p. 485).

At this point in time, the vast majority of existing MOOCs are of the xMOOC type (Hollands and Tirthali 2014), and the main MOOC development in the past few years has been in xMOOCs (Veletsianos and Shepherdson 2015). Therefore, the rest of this article considers only this kind, and, to facilitate communication, we will henceforth use the term MOOCs to refer to xMOOCs only. Recent developments have included the evolution of a number of business models to finance MOOCs into either certificate-granting programmes (e.g. the "Signature-Track" programme offered by Coursera)[2] or degree programmes.[3] Hybrid MOOCs, which integrate MOOCs into traditional higher education in a "flipped" format,[4] are another common accrediting type of MOOC (Sandeen 2013b).

In addition to the integration of MOOCs in formal education, several unique characteristics of MOOCs make them particularly suitable for lifelong learning (LLL). First, the open enrolment of MOOCs provides free or low-cost learning opportunities, which removes cost barriers for lifelong learners. Second, the sophisticated information and communication technologies (ICTs) embedded in MOOCs can potentially facilitate personalised learning (Dinevski and Kokol 2004; Sandeen 2013a; Smith 2014; Soares 2011). Third, MOOCs provide a good platform to promote interaction and collaboration among learners of similar interests, which is key for LLL, since LLL happens more often than not in a collaborative manner (Sharples 2000). To sum up, the support of massive open access, the availability of low-cost accrediting models (e.g. certificates and digital badges), and the promotion of collaborative learning among peers found in MOOCs foster meaningful and sustainable LLL experiences.

Despite recent advances of MOOCs and their potential to support LLL, many challenges remain. The most crucial challenge is the assessment of student learning. Most traditional brick-and-mortar classroom assessment methods would not work in MOOCs due to the different roles instructors and learners play in a MOOC (Sandeen 2013a; Suen 2014). Instructors in a MOOC only provide learning content and a limited amount of facilitation. General (not-for-credit) MOOC learners are responsible for their own progress, since they set their own learning pace, decide how to use course materials, and how to seek help when needed (Mackness et al. 2010). In a traditional residential course, assessments of student learning are usually carried out by instructors for the purpose of providing feedback to both teaching staff and students. The tools used for this kind of assessment include periodic assignments, quizzes/exams and class projects. However, MOOCs serve a different student population with different intentions. MOOC students in general, and lifelong learners

[2] Signature Track is an "option that … give[s] students in select classes the opportunity to earn a Verified Certificate for completing their Coursera course" (Coursera 2013). It offers "Identity Verification …, Verified Certificates … and Shareable Course Records …, electronic course records [which students can share] with employers, educational institutions, or anyone else through a unique, secure URL" (ibid.).

[3] One example of this is a low-cost Online Master of Science in Computer Science (OMS CS) offered by the Georgia Institute of Technology. See https://www.omscs.gatech.edu/ [accessed 26 January 2018].

[4] The term "flipped" (often used in the expression "flipped classroom") refers to an instructional strategy that inverses the traditional learning environment by delivering instructional content outside of the classroom (usually online) and bringing more activities into the classroom. See https://en.wikipedia.org/wiki/Flipped_classroom [accessed 5 February 2018].

in particular, are not necessarily interested in earning accreditation (Koller et al. 2013). In 2012, many enrolees of a typical MOOC seemed to have joined only to "window shop", with roughly half of them returning for the first lecture and even fewer submitting an assignment (ibid.). The median course completion rate was reported to be 12.6 per cent (Jordan 2015). Daphne Koller et al. (2013) observed that the majority of enrolled students did not intend to pass the course.

Irrespective of whether or not the high dropout rate is a concern, it is, however, widely agreed that it would be useful to understand at what stage and why students drop out (Yuan and Powell 2013). The goal is to provide sufficient support for the diverse audience, including both motivated learners who intend to complete the course, and casual "lurkers"[5] who intend to gain useful information without completing the work, to pass the course. Lack of support has indeed been reported to be an obstacle impeding students' successful participation in a MOOC (Onah et al. 2014). One important component of support is encouraging students by giving them feedback, and that includes assessment. Without assessment, the learning experience in MOOCs is not complete (Suen 2013). Given the diverse and large student population in MOOCs,[6] assessment needs to be self-sustainable and personalised.

Conducting self-sustainable and personalised assessment to fit the MOOC student population is a difficult mission. To provide useful feedback and support, both assessments *for* learning (formative assessments), which evaluate students' current progress, and assessments *of* learning (summative assessments), which record students' cumulative progress, are needed. In this article, we examine possible assessment approaches that fit open online education from both formative and summative assessment perspectives by (a) discussing the importance and the challenges of implementing assessments for both formative and summative purposes in MOOCs; (b) identifying advantages and disadvantages of various formative and summative assessment approaches; and (c) specifically elaborating on the promise as well as practical and technical challenges in implementing peer assessment – a widely applicable approach in MOOCs. With our overview and analyses of the different assessment approaches, we aim to provide a toolbox which might be useful for MOOC instructors and instructional designers in devising appropriate assessments for their courses. We also hope this article will prompt more discussions on assessment in MOOCs – an important and challenging issue in the higher education community.

The importance and challenges of assessment within MOOCs

Importance of assessment

Cathy Sandeen (2013a) has argued that assessment of student learning should be a central feature of a MOOC. Assessment should not be an "after-the-fact add-on",

[5] A "lurker" is someone who watches from the side lines without making his/her presence known.

[6] For recent figures, see https://www.class-central.com/report/mooc-stats-2017/ [accessed 5 February 2018].

but must be integrated at the centre and should permeate every aspect of a MOOC. The purpose of assessment in general can be classified into formative and summative functions. As mentioned earlier, formative assessment refers to "assessment *for* learning" that evaluates student learning at the current stage and supports their learning for the next step, while summative assessment refers to "assessment *of* learning", which is the systematic and continuous recording of student achievement (Harlen and James 1997). In general, formative assessment directly aims at the improvement of learning, while summative assessment aims at providing information for accreditation and evaluation. It is also possible to use one particular assessment for both formative and summative purposes. This general classification of assessment purposes also applies to MOOCs, since the fundamental relation between learning and assessment is the same, irrespective of what format the instruction is delivered in.

The stakes involved in summative assessment are usually higher than those in formative assessment, due to the fact that the former leads to course grade assignment, evaluation of instructor performance, accreditation, and/or other actions that can have a key impact on the lives of students and/or instructors. As such, reliability/accuracy, validity and fairness of outcomes and decisions are critical in summative assessment.

Wynne Harlen and Mary James (1997) have argued that educators may put more effort into formative assessment because it can promote learning with understanding if conducted properly. In their review of literature on the subject, Joyce Wangui Gikandi et al. (2011) found online formative assessment to be an effective pedagogical strategy by (a) providing informative and immediate feedback; (b) fostering engagement with critical learning processes; and (c) promoting equitable education (ibid.). The interactions that take place within formative assessment also have the potential to strengthen students' feeling of connectedness, which is sometimes called "social presence" (Tu and McIsaac 2002). The connectedness further promotes student engagement.

In addition to the aforementioned pedagogical benefits for formative assessment, low-stakes/no-stakes testing in general contributes to both students' cognitive development and their metacognitive development (Roediger and Butler 2011; Roediger et al. 2011). Even a basic retrieval assessment with correct answers given afterwards has been shown to be powerful in enhancing memory and producing large gains in long-term retention (Roediger and Butler 2011).

Challenges of formative assessment in MOOCs

In spite of all its benefits, it is particularly challenging to conduct formative assessment effectively within a MOOC due to the extremely unbalanced instructor-student ratio. A key limitation is that this extreme ratio renders it impossible for the instructor to provide individualised formative feedback to any student beyond a very tiny fraction of the enrolees.

It is conceivable that the tremendous number of dropouts after the first two weeks found in a typical MOOC (Jordan 2015) might also be related to the general lack of formative feedback, which demotivates "leisure learners" from keeping up with the

learning goals. In a residential course, students sitting in the same classroom can naturally feel the connectedness with the course instructor and peers, and the instructor can easily provide instant feedback to student questions. Formative assessment can be conveniently integrated into classroom activities. However, labour-intensive formative assessments used in a traditional course cannot be practically applied to a MOOC with a huge student population. Furthermore, a MOOC's student population differs significantly in terms of learning intentions and goals from that of a residential class. How to provide appropriate feedback to fit the different needs of students remains unsolved. Given all these constraints, sophisticated, self-sustainable and cost-effective formative assessment methods are needed to motivate MOOC students to keep up their engagement.

Among existing formative assessment methods, there are currently two which might meet the needs of a MOOC: (a) assessments that rely on machine grading; and (b) feedback provided by peers. Embedded quizzes and automated essay grading fall within the first category (Balfour 2013). However, machine-graded assessments are not yet sufficiently sophisticated to be able to provide automatic and personalised feedback to each individual MOOC learner. The second category includes peer assessment embedded in open-ended assignments (Luo et al. 2014) and peer interaction using a course discussion platform (Skrypnyk et al. 2015). Some MOOCs have even tried to use "community teaching assistants" who were chosen among top-performing students from the previous iteration of the same course to help provide feedback to students in the current course (e.g. Kulkarni et al. 2015). A systematic administration of this type of heavily human-involved formative assessment throughout a course can be very labour-intensive, because informative and diagnostic feedback needs to be constantly provided to each individual student to guide his or her learning for the next step.

Formative assessment has not yet received the amount of attention it deserves within MOOCs due to all the practical difficulties. Given the immaturity of sophisticated technologies, full implementation of formative assessment throughout a MOOC may put a heavy burden on either instructors or MOOC learners, or on both. It may be costly, or may even demotivate learners with an overwhelming workload, which may worsen student attrition in a course.

Challenges of summative assessment in MOOCs

Unlike assessment for formative purposes, the role of summative assessment within a MOOC is to record a student's performance throughout their course participation. The need for proper, secure and accurate summative assessment scores and decisions has become more urgent in the face of providers' recent shift towards (a) awarding certificates and digital badges for course accomplishments; (b) integrating MOOCs and their course credits into the existing higher education system; and (c) launching MOOC-based degree programmes (Sandeen 2013a, b). Other accreditation models are still evolving. Given the accountability issue attached to summative assessment, stakeholders are calling for reliable and valid assessment results.

Attempts to move traditional summative assessment online are part of an ongoing process. So far, it is more common to only move the teaching and learning online but keep the summative assessment in a traditional standardised setting to ensure security. For example, several MOOC providers, such as *Udacity*, *edX* and *Futurelearn*, partnered with a company specialised in computer-based testing, Pearson VUE,[7] for residential proctored (supervised) testing to grant MOOC credits to those who passed the test (Chung 2015; Lewin 2012).

The main reason for separating assessment from instruction in MOOCs is the lack of online assessment security. Cheating can be expected to be easier in an online setting. There are even paid services that would impersonate students online to do their work and earn grades for them (Newton 2015). Recent developments on proctoring technologies that utilise identity authentication and keystroke recognition together with webcam proctoring aim to promote online testing security (Sandeen 2013a). The proctoring services usually charge a fee ranging from tens of dollars to hundreds of dollars depending on levels of security and testing lengths. The American Council on Education (ACE) initiated a pilot project on credit recommendation for several MOOCs (Sandeen 2013a). Coursera has advertised the ACE credit option to its users who take the credit-recommended MOOCs. In order to obtain the MOOC credit, students need to enrol in the Signature Track programme mentioned earlier and pass the online proctored exam supervised by a company called ProctorU (Hollands and Tirthali 2014). A small number of students have requested the ACE credit option (ibid.). Although online proctored exams have been used for MOOC credits, a lack of research on the effectiveness of the existing proctoring systems threatens the validity of assessment results (Foster and Layman 2013). Deterring online cheating and ensuring secure online assessment is one main challenge faced by MOOC summative assessment.

Another challenge is related to the development of the summative test or exam. Given the higher stakes and accountability in summative assessment, summative tests need to be professionally developed and scientifically evaluated. Course instructors usually are not test development experts and are not equipped to design such high-stakes assessments.

Given the different roles of formative and summative assessments, a balanced combination of the two within a MOOC may be the best approach. Over-reliance on summative assessment may result in negative consequences. These include the washback effect of assessment on instructional practices (e.g. Shohamy et al. 1996), where teachers "teach to the test" and students "study to the test", and the fact of rampant cheating behaviours found in testing (e.g. Garavalia et al. 2007; Newton 2015).

[7] VUE stands for Virtual University Enterprises.

Assessment design options for open online education

There are a large variety of assessment designs that can be potentially used for formative and/or summative assessments in open online courses. These assessment designs differ in terms of test security, and in terms of assessment tasks and response formats. Different combinations of these features may render an assessment more or less suitable for formative or summative assessments in MOOCs.

Assessment security

Based on whether and how the assessment is proctored, three different types of assessment security settings can be found in current MOOCs: (a) unproctored online assessments, (b) remotely proctored assessments, and (c) assessments conducted in brick-and-mortar test centres. These three different security settings are related to assessment purposes, stakes and accountability.

Formative assessment is usually conducted without proctoring, because it mainly serves to support learning and teaching processes within a low-stakes environment. In addition, formative assessments do not need to follow a rigid procedure. They can be simply informal discussions among students, or teaching assistants answering students' questions on a discussion platform. In those situations, there is no need for proctoring. Unexpectedly, however, sometimes summative assessment is also conducted without proctoring. For example, Coursera MOOC end-of-course exams, projects or assignments are often unproctored, even for students pursuing a certificate. However, it is to be noted that the certificate offered by Coursera is a "Certificate of completion" only. It does not certify minimal mastery of subject contents. Unproctored assessment is cost-efficient, but has no test security to speak of. Consequently, it is useful for formative assessments only.

When it comes to awarding course credits or degrees, some form of proctored assessment is necessary to ensure the validity of assessment results. The proctoring can be carried out either online or in person. Brick-and-mortar test centres provide in-person proctoring and are considered to be a secure assessment environment, but not cost-effective. Conversely, remote proctoring is cheaper, but the approach is still immature and in need of further development.

Assessment format

In general, task and response formats which are suitable for summative purposes can be used for formative purposes as well, but the opposite is not always true. For example, summative end-of-course quizzes can be used for formative purposes if scheduled during the middle of the course, aiming to provide diagnostic feedback to students. However, formative assessment formats that aim at providing instant feedback to students, such as discussion forums and question-and-answer (Q&A) sessions, are not suitable for summative purposes.

Formative assessment: discussion forums

One of the most common formative assessment formats is to provide opportunities for online synchronous or asynchronous discussion platforms. Such platforms can provide formative feedback to help students understand the learning content and to support their learning for the next step in the sequence (Breslow et al. 2013; Ramesh et al. 2014). Some top-performing students as well as course teaching assistants can actively provide useful comments to common problems encountered by students (Kulkarni et al. 2015; Skrypnyk et al. 2015). Interestingly, a study on the effect of including professors in the discussion within a MOOC revealed no added value in terms of promoting student learning (Tomkin and Charlevoix 2014). Referring to a cMOOC, Oleksandra Skrypnyk et al. (2015) also reported that the teaching function was not necessarily dominated by the official course instructors, but could be distributed among influential actors who had emerged as such during the course. These results seem to contradict the existing belief regarding the importance of teacher presence and facilitation in an online course (Garrison and Cleveland-Innes 2005; Garrison et al. 2010). One explanation is that the facilitation an instructor can provide is largely diluted in a MOOC with a large number of participants. After all, the impact of facilitation provided by one or two instructors on thousands of students is inherently minimal (Tomkin and Charlevoix 2014). On the other hand, the limited impact of MOOC instructors also suggests that the learning community constituted by MOOC students needs to be self-sustainable with minimal instructor facilitation.

In order to scale up (i.e. exponentially increase) effective student interaction on discussion forums, Derrick Coetzee et al. (2014) have suggested implementing a reputation system to encourage useful posts and discourage spam posts. Equipped with the ability to vote posts up or down, this kind of system enables the most useful posts to quickly bubble to the top while the not-so-useful ones may be relegated to the bottom (Hollands and Tirthali 2014). The purpose and effect of this is to display the discussion forum in a constructive order and make it efficient for students to search for useful information, seek help and participate in meaningful discussions. Another natural advantage of a MOOC discussion forum is having a group of higher-level students "on board" who may be influential in helping novice learners. This group of students partly plays the role of the instructor, ensuring the scalability of what would otherwise be a MOOC discussion forum consisting only of novice learners. However, implementing such a recommendation system does bear potential risks. For example, some commonly held misconceptions may pop up and rise to the top quickly without correction, thereby further reinforcing these popular misconceptions. In this case, some form of instructor intervention may be necessary.

Formative assessment: question and answer (Q&A) sessions

An alternative to online discussion forums are question and answer (Q&A) sessions offered by instructors from time to time throughout the course. For example, MOOCs offered by edX sometimes provide open office hours to all students who

want to directly raise questions and get instant answers from instructors (Mukta 2015). Both discussion platforms and Q&A sessions have similar strengths and weaknesses. They are open to everybody, but at the same time, given the massive number of students, only the questions from a very tiny fraction of the students can be answered.

Formative assessment: peer assessment

Yet another alternative approach to formative assessment in MOOCs is peer assessment. One possible way of using peer assessment is to integrate it in a long-term assigned project (Cho and Schunn 2007). Students can refine and improve their project throughout, based on peer feedback. The advantages of using peer formative assessment are: immediacy of feedback, scalability, efficiency and self-sustainability. From a pedagogical standpoint, peer assessment is an effective approach to help the peer rater reflect on their own learning through examining others' work, correcting misunderstandings and reading peers' comments (Falchikov and Goldfinch 2000; Kim 2005; Topping 1998; Tsai et al. 2001; van Gennip et al. 2009). One challenge of peer formative assessment is to build trust among peers (van Gennip et al. 2009, 2010). Otherwise, uncomfortable feelings and distrust of assessment results may emerge.

Formative assessment: automated feedback

At this point in time, automated Q&A and feedback for MOOCs is still in its infancy. Although embedded instant questions and feedback surveys are frequently used in online courses (Pappano 2012), they remain useful only for lower-ordered cognitive contents such as definitions of terms, or simple procedural knowledge. Even though sophisticated internal data analytics (e.g. student behavioural data throughout the course) are often performed on MOOC data, they have been performed *post hoc* (retrospectively) to provide insights for future course development only, and not for instant formative feedback to individual learners (e.g. Guo et al. 2014; Khalil and Ebner 2014; Kizilcec et al. 2013; Koller et al. 2013; Xiong et al. 2015). If analytic technology can be optimised and packaged as an intelligent agent for a specific course (Daradoumis et al. 2013), students could benefit most through instant and personalised feedback provided by the sophisticated agent.

Summative assessment formats

For summative assessments in residential classroom settings, an expert-led assessment approach represents the mainstream. From the development of assessments to grading, content experts and/or trained professional graders are involved. However, expert assessment on individual students is not scalable in a MOOC. Machine grading of expert-developed assessment tasks is an alternative, but such an approach is only feasible for highly structured assessment questions, such as multiple-choice questions in general, or specific tasks, such as essay writing

assignments for the evaluation of certain writing skills (see Balfour 2013). These structured assessment tasks are usable only within a very limited scope. Multiple-choice questions are suitable only when knowledge is evaluated based on *recognition* of correct answers, rather than *production* of correct answers. In practice, the use of multiple-choice tests most often results in tapping only low-cognitive-level knowledge. Granted, low-cognitive-level knowledge, which is the foundation for high-cognitive-level skills, is important and needs to be assessed; but an assessment that taps only low-cognitive-level knowledge would be useful only for very specific foundational courses. Similarly, the use of automated essay grading algorithms is currently limited to the assessment of certain writing skills, with algorithms being capable of evaluating mostly mechanical aspects (e.g. grammar, subject-verb agreement, etc.) of writing.[8]

Peer assessment has been used for summative purposes in MOOCs (e.g. Robinson et al. 2015), especially when the test questions are ill-structured[9] and open-ended. Given the immaturity of automated grading technologies and the non-scalability of expert assessment, peer assessment is quite possibly the only universally applicable approach in MOOCs within the foreseeable future (Suen 2014).

Table 1 summarises the different assessment formats, their strengths and limitations.

Peer assessment within MOOCs

Among all assessment methods within MOOCs, peer assessment has received considerable attention (e.g. Admiraal et al. 2014; Goldin 2012; Kulkarni et al. 2015; Luo et al. 2014; Piech et al. 2013; Raman and Joachims 2014; Suen 2014; Xiong et al. 2014). It is probably the only universally applicable assessment method that fits both formative and summative purposes and fits different assignment formats (i.e. well-structured or ill-structured) for MOOCs; in addition to other advantages, such as immediacy, cost-effectiveness, scalability and self-sustainability. However, peers as assessors are generally novice learners. Therefore, the accuracy of peer assessment results is usually a concern (Suen 2013), since novices think differently from experts (e.g. Chi et al. 1981). If peer assessment can be improved to produce credible results, it may well be the most ideal assessment method for MOOCs. As such, peer assessment deserves a more in-depth examination.

A typical peer assessment operation in a MOOC includes three steps. First, all students in a MOOC are instructed to complete and submit an assignment online. Second, after the assignment submission deadline, every student who has submitted the assignment is randomly assigned several completed assignments of other students to evaluate. Third, the evaluation results including the grades as well as the

[8] For more details on what automated essay scoring (AES) can and cannot do, see Shermis et al. (2010).

[9] The term "ill-structured" does not mean that the question has been badly designed, but that it is one which allows for several correct answers and/or ways of arriving at a correct solution. It is this complexity which makes it unsuitable for automated assessment. By contrast, a well-structured question can only lead to one correct answer and is therefore machine-gradable.

Table 1 Summary of formative and summative formats within MOOCs

Purposes	Assessment format	Examples	Strengths	Limitations
Formative	Discussion platform	Synchronous chat room; asynchronous discussion forum	Open to everybody	Usually dominated by a few active students; not all raised questions are followed up; risk of reinforcing errors and misconceptions
	Q&A sessions	Open office hours	Instant feedback available	Can only provide feedback to a very small portion of student population
	Peer assessment	Peers provide feedback to other peers' work during the course	Immediate feedback, scalable, efficient and self-sustainable	Mistrust of peer raters' ability to provide correct feedback
	Embedded instant questions	Multiple choice questions	Instant feedback, automated and efficient	Only applicable to certain skills/levels of knowledge
	Feedback surveys	Questions asking about student experiences	Helpful for further course development	Group information; no direct feedback to individual students
	Internal data analytics	Monitoring of student learning trajectory	Comprehensive analytics and rich information available for further course development	Difficult to provide timely individualised feedback to students
	Intelligent agent	Course agent automatically provides feedback to individual student	Instant feedback, automated and efficient	Expensive to develop an intelligent agent to meet specific course needs
Summative	Embedded end-of-course multiple-choice quiz	Online final exams	Instant feedback, automated and efficient	Only applicable to certain skills/levels of knowledge
	Automated essay grading	Grading of writing assignments	Instant feedback, automated and efficient	Only applicable to evaluation of certain writing skills
	Expert assessment	Instructors provide summative feedback to a small portion of students	Maximised validity	Not scalable
	Peer assessment	Peers provide feedback to other peers' work at the end of the course	Immediate feedback, scalable, efficient and self-sustainable	Lack of credibility of peer raters

written feedback are then returned to the author of the assessed assignment. The whole process is usually double-blind. Each student would only grade a handful of peer assignments; say, four or five.

Formative peer assessment as an effective pedagogical strategy

Peer assessment in traditional (i.e. brick-and-mortar) classrooms is known to facilitate formative assessment (Topping 1998). It is beneficial to at least three particular aspects of learning: (a) cognitive development (e.g. Tsai et al. 2001), such as understanding of content knowledge (ibid.); (b) metacognitive development, such as self-monitoring and reflection skills (Kim 2005; Sluijsmans et al. 2002a); and (c) social-affective development, such as trust within the peer community (van Gennip et al. 2009, 2010).

In terms of cognitive development, students discover different insights or alternative approaches to problems while reading others' work or observing others' performance (Gielen et al. 2011), which may not happen if they study by themselves. Students may also identify errors or misconceptions of their own by being exposed to others' work. Furthermore, the assessment itself can be a complex cognitive activity, which involves synthesising information, diagnosing misconceived knowledge, considering deviations from the ideal, and providing suggestions for improvement. All of these processes involve higher-ordered thinking. As such, peer assessment promotes thinking.

Besides cognitive gains, students can develop their metacognitive skills by participating in peer assessment. To begin with, peer assessment is reflective (Topping 1998), which can lead to identification of knowledge gaps and to engineering their closure. The process of comparing their own work with that of their peers promotes students' self-assessment and greater metacognitive self-awareness. Students also learn to become better assessors through the practice of peer assessment. Learning how to assess is a crucial part of becoming a lifelong learner, because learners are confronted with the need to assess themselves throughout their lives (Boud 2000). Knowing how to determine appropriate goals and how to assess whether or not one is achieving them is very beneficial to every person.

From a social-affective perspective, actively participating in peer assessment can indeed help build a safe, respectful environment and break the social barriers that may be encountered by novel peer raters (van Gennip et al. 2009). Peer assessment can shorten the social distance and establish a trusting and safe environment among peers if conducted appropriately (Bulu and Yildirim 2008), and further enhance their social relationships (Mak and Coniam 2008).

Despite all the potential benefits of using peer assessment as a formative assessment tool, only a small portion of students in a typical MOOC with a peer assessment assignment have actually participated in it (e.g. Staubitz et al. 2016). Nevertheless, those participants are most likely very motivated learners. A disproportionately low participation rate in MOOCs, compared to high enrolment, has always been a concern (Jordan 2014, 2015). Koller et al. (2013) have explained the different intents of MOOC students. A large number of enrolees just want to be

passive watchers. It is easy to meet the needs of passive MOOC watchers, since they do not need much extra facilitation. The attention should, thus, be put on helping those self-motivated learners who are dedicated to completing the course. Several approaches, such as providing clear guidelines on how to conduct peer assessment (Kulkarni et al. 2015), involving students in the development of assessment rubrics (Falchikov and Goldfinch 2000), and providing peer raters with instant feedback on their performance in peer assessment (Kulkarni et al. 2015), have been demonstrated to be effective approaches to promote student motivation to participate in peer assessment. In order to optimise the learning benefits of formative peer assessment, the first and foremost task is to motivate as many students as possible to participate in the assessment. For MOOCs to realise the expected cognitive benefits of peer formative assessments, ways of increasing student participation in peer assessment are crucial. Perhaps some of the tactics learned from traditional classroom peer assessments can be adapted for use in MOOCs.

Summative peer assessment with accountability

While peer assessment is advocated to be most effective in providing formative feedback, it is not commonly advocated as a summative assessment tool. This is due to concerns for the lack of expertise of peer assessors, potentially leading to erroneous assessment results. Different types of peer assessment errors have been summarised elsewhere (Suen 2014), some of which are due to peer raters' idiosyncratic biases while others can be attributed to random assessment errors. Currently, there are two strands of development aimed at solving the problem of peer rating errors. The first strand of development focuses on training students to be proficient raters, and the second strand targets the correction of peer rating errors *post hoc* as a remedial approach. Both approaches may require monitoring student assessment behaviours, detecting assessment errors, and evaluating student assessment performance in order to correct errors and increase assessment accuracy.

Peer rater training has been minimal in current MOOCs, since most MOOC assessments are not yet sufficiently high-stakes to warrant peer rater training. However, if peer assessment is to serve summative accountability purposes such as awarding digital badges for mastery, rater training might become necessary (a) to ensure that peer raters understand the assessment rubrics, guidelines and processes; and (b) to select competent peer raters who can give credible assessments and to filter out those who cannot. Holding peer raters accountable for their assessment results may be an important approach to motivate students to assess their peers fairly. For example, we may link peer raters' accuracy to their final grades in a MOOC in order to make sure that all peer raters evaluate their peers seriously (Kulkarni et al. 2015).

Approaches focusing on correcting assessment errors *post hoc* as a remedial step have been evolving in recent years (Goldin 2012; Piech et al. 2013; Raman and Joachims 2014; Uto and Ueno 2016; Walvoord et al. 2008; Xiong et al. 2014).

Dependent on what are considered *ground truth* assessment results,[10] these approaches can be classified into two types: one aiming at closing the gap between peers and experts (Walvoord et al. 2008; Xiong et al. 2014) and the other focusing on estimating consensus among peer assessment results after taking into account peer rater errors (Piech et al. 2013; Raman and Joachims 2014; Uto and Ueno 2016). Different statistical modelling approaches which can be used to correct errors in peer assessment have been summarised elsewhere (Suen 2014). The challenges to correct errors *post hoc* are the *noisiness* and *sparsity*[11] of peer assessment data. Complicated data require a complicated modelling approach, which makes the problem difficult to handle. However, if effective, *post hoc* error correction would still represent the least disruptive approach to obtaining accuracy and trustworthy peer assessment results. For example, Chris Piech et al. (2013) have demonstrated that a machine learning algorithm designed to correct peer rater errors can increase rating accuracy by 30 per cent.

Small group peer assessment in MOOCs

A potentially very powerful alternative is to implement some form of small group collaborative peer assessment within a MOOC. Small group collaboration is known to be effective in building knowledge and consensus (e.g. Springer, Stanne and Donovan 1999; Webb et al. 1995; Stahl 2006). Collaboration also coincides with the nature of LLL – as Mike Sharples has pointed out, "lifelong learning is primarily collaborative rather than competitive" (Sharples 2000, p. 178). Designing peer assessment to take place within small groups with interactive feedback may achieve promising learning outcomes.

There can be a number of different ways to form small groups within a MOOC. For example, one possible approach is to group students who have similar interests but different levels of expertise in this interest area.

Students with similar interests would be motivated to work together, and different levels of expertise would foster meaningful learning. Interactions among group members are good sources of formative feedback to guide further progress.

Summative assessment of group work may be conducted in two separate forms. One is to evaluate group member contribution within the group, and the other is to evaluate overall group work quality across groups. The assessment results from the two forms provide insights into both overall group performance and individuals' contribution to group outcomes.

One challenge in implementing small group collaboration is to assign or recommend appropriate group members to students. A promising approach is to use a recommender system (Dascalu et al. 2016; Ricci et al. 2011), which has been investigated within the machine learning community. For a particular assigned project, a student can initiate a request to the recommender system to form a collaborative group. The recommender system can then analyse the information

[10] Ground truth assessment results are the results that are considered to be most reliable and accurate.

[11] In the context of data, noisiness refers to the characteristic of the data that include different types of errors, while sparsity refers to the fact that only limited information is available.

from student profiles, which may include which previous projects and/or courses a student has participated in, what specific interests and/or preferences a student has indicated in a pre-course survey, and which students are close to his/her social interactive network generated from his/her previous interactions within a MOOC. The recommender system would then match a student profile with other students' profiles and recommend group members to the student. Conversely, a student may be looking for an existing collaborative group to join. The recommender system can match the student's profile with those of existing open groups. This can proceed iteratively until groups reach some optimal size.

While assigning students to certain groups is already a considerable task, finding effective ways of managing group processes and ensuring effective group collaboration is even trickier. The challenge of creating effective collaborative learning and peer assessment groups has been documented in traditional online education (distance learning) when students are enrolled to pursue degrees (Brindley et al. 2009). With a massive student population uncommitted to any accreditation programme, it is even more difficult to manage the groups effectively. One possible solution may be to monitor group activities by using learning analytic techniques so that instructors or teaching assistants can view a summary of group activities and can intervene in groups that need help (Sharples et al. 2016). In addition, *social loafing*[12] might be a problem when some group members are not contributing (Karau and Williams 1993). Incorporating peer assessment in the collaboration and allowing peers to evaluate group members' contributions may be a good way to alleviate social loafing. For formative assessment purposes, it might be necessary to assign group leaders to lead group collaboration and to reward them for their effective leadership.

Summary and future directions

Assessment is one crucial component in education. We have discussed different assessment approaches that fit massive open online education for both formative and summative assessment purposes. In a massive open online education world, we are facing new challenges and opportunities. The problem of how to conduct assessment to meet different needs within a MOOC environment is relatively new, and finding a viable solution still requires considerable research. The need for a high-quality, trustworthy and effective assessment approach remains a major gap in the burgeoning open online learning world of MOOCs.

Personalised assessment is one possible future direction. Given that enrolees within a MOOC typically have very diverse intentions and backgrounds, one-size-fits-all standardised formative assessments are probably not the best way towards the achievement optimal learning goals. We may need to rely on the development of artificial intelligence (AI) to help us build intelligent agents who can provide personalised feedback to meet each individual student's need. Research results from

[12] Social loafing is a term used in social psychology; it refers to the phenomenon of a person contributing minimal effort to achieve a goal when he/she works in a group setting.

computerised-adaptive[13] assessment (Gershon 2005) may be adjusted to serve summative assessment purpose in MOOCs.

Authentic assessment[14] is advocated to be superior to a multiple-choice type of assessment for many content areas (Gikandi et al. 2011). By providing real-life scenarios, it more readily activates higher-ordered skills and problem-solving skills for assessment than is possible in multiple-choice testing. But meanwhile, modern technologies are providing opportunities to simulate authentic scenarios. Therefore, the development and integration of authentic assessment in a MOOC is likely to be an interesting future research topic.

Among all the options and issues, peer assessment remains key, and is the most versatile approach in the toolbox for assessment in open online education due to its applicability to most contexts and for different purposes. It is scalable and valuable for both formative and summative assessments. The benefits of peer assessment can be optimised by the addition of a number of features. One such feature is to employ a small-group peer assessment design as discussed earlier. Another feature is to include a rater training protocol within the course. Besides increasing peer assessment accuracy, training can lead to development of student reflection skills and a better understanding of the assessment criteria (Sluijsmans et al. 2002b). Implementing some form of *post hoc* correction of peer rating scores will moreover improve the accuracy of results, which can also be used to further calibrate the knowledge and skills of peer raters. Opinions of knowledgeable raters thus identified can then be given extra weight when they occur in formative discussion chat rooms or forums. Peer assessments may also be implemented in a repeated, iterative manner within a course. Under this design, students can keep revising and developing their work iteratively after acting upon the feedback from peers (e.g. Cho and Schunn 2007; Hwang et al. 2014). Finally, combining two or more features within a peer assessment design may lead to an optimal design in which the advantages of each feature are cumulative, while the limitations of each feature are compensated by other features.

References

Admiraal, W., Huisman, B., & Van de Ven, M. (2014). Self- and peer assessment in massive open online courses. *International Journal of Higher Education, 3*(3), 119–128.

Balfour, S. P. (2013). Assessing writing in MOOCs: Automated essay scoring and Calibrated Peer Review™. *Journal of Research and Practice in Assessment, 8*(1), 40–48.

Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education, 22*(2), 151–167.

Bowen, W. G. (2013). *Higher education in the digital age*. Princeton, NJ: Princeton University Press.

[13] Computerised-adaptive assessment (often also referred to as computerised-adaptive testing; CAT) is a computer-based test form that adapts to examinee's ability levels. In CAT, an examinee next question is dependent on his/her responses to the previous questions.

[14] Authentic assessment is assessment that mirrors real-life problems, usually focusing on contextualised tasks.

Breslow, L., Pritchard, D. E., Deboer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research and Practice in Assessment, 8*(1), 13–25.

Brindley, J., Blaschke, L. M., & Walti, C. (2009). Creating effective collaborative learning groups in an online environment. *The International Review of Research in Open and Distributed Learning, 10*(3). Retrieved 18 January 2018 from http://www.irrodl.org/index.php/irrodl/article/view/675/1271.

Bulu, S. T., & Yildirim, Z. (2008). Communication behaviors and trust in collaborative online teams. *Educational Technology and Society, 11*(1), 132–147.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121–152.

Cho, K., & Schunn, C. D. (2007). Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system. *Computers and Education, 48*(3), 409–426.

Chung, C. (2015). Futurelearn partners with Pearson VUE for proctored testing. *Class Central* (web post, 8 May). Retrieved 18 January 2018 from https://www.class-central.com/report/futurelearn-pearson-vue-proctored-testing/.

Coetzee, D., Fox, A., Hearst, M. A., & Hartmann, B. (2014). Should your MOOC forum use a reputation system? In *Proceedings of the 17th ACM conference on computer-supported cooperative work and social computing* (pp. 1176–1187). New York: Association for Computing Machinery (ACM) Press.

Coursera. (2013). Introducing signature track. *Coursera* (blog post 9 January). Retrieved 26 January 2018 from https://blog.coursera.org/signaturetrack/.

Daradoumis, T., Bassi, R., Xhafa, F., & Caballé, S. (2013). A review on massive e-learning (MOOC) design, delivery and assessment. In *Proceedings of the eighth international conference on P2P, parallel, grid, cloud and internet computing* (pp. 208–213). Compiègne: IEEE.

Dascalu, M.-I., Bodea, C.-N., Mihailescu, M. N., Tanase, E. A., & de Pablos, P. O. (2016). Educational recommender systems and their application in lifelong learning. *Behaviour and Information Technology, 35*(4), 290–297.

Dinevski, D., & Kokol, P. (2004). ICT and lifelong learning. *European Journal of Open, Distance and E-Learning, 7*(2), Article 136. Retrieved 18 January 2018 from http://www.eurodl.org/?p=archives&year=2004&halfyear=2&article=136.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*(3), 287–322.

Foster, D., & Layman, H. (2013). Online proctoring systems compared. Retrieved 5 February 2018 from https://ivetriedthat.com/wp-content/uploads/2014/07/Caveon-Test-Security.pdf.

Garavalia, L., Olson, E., Russel, E., & Christensen, L. (2007). How do student cheat? In E. M. Anderman & T. B. Murdock (Eds.), *Psychology of academic cheating* (pp. 33–55). Burling, MA: Elsevier Academic.

Garrison, D. R., & Cleveland-Innes, M. (2005). Facilitating cognitive presence in online learning: Interaction is not enough. *American Journal of Distance Education, 19*(3), 133–148.

Garrison, D. R., Cleveland-Innes, M., & Fung, T. S. (2010). Exploring causal relationships among teaching, cognitive and social presence: Student perceptions of the community of inquiry framework. *The Internet and Higher Education, 13*(1), 31–36.

Gershon, R. C. (2005). Computer adaptive testing. *Journal of Applied Measurement, 6*(1), 109–127.

Gielen, S., Dochy, F., Onghena, P., Struyven, K., & Smeets, S. (2011). Goals of peer assessment and their associated quality concepts. *Studies in Higher Education, 36*(6), 719–735.

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers and Education, 57*(4), 2333–2351.

Goldin, I. M. (2012). Accounting for peer reviewer bias with Bayesian models. In J. Kim & R. Kumar (Eds), *Proceedings of the full-day workshop on intelligent support for learning groups at the 11th International conference on intelligent tutoring systems (ITS 2012)* (pp. 27–34). Chania: (n. p.). Retrieved 31 January 2018 from https://sites.google.com/site/islg2012/.

Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos. In *Proceedings of the first ACM conference on learning at scale* (pp. 41–50). New York: Association for Computing Machinery (ACM) Press.

Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education, 4*(3), 365–379.

Hollands, F. M., & Tirthali, D. (2014). *MOOCs: Expectations and reality*. Full report. New York: Center for Benefit-Cost Studies of Education (CBCSE), Teachers College, Columbia University. Retrieved

18 January 2018 from http://cbcse.org/wordpress/wp-content/uploads/2014/05/MOOCs_Expectations_and_Reality.pdf.

Hwang, G. J., Hung, C. M., & Chen, N. S. (2014). Improving learning achievements, motivations and problem-solving skills through a peer assessment-based game development approach. *Educational Technology Research and Development, 62*(2), 129–145.

Jordan, K. (2014). Initial trends in enrolment and completion of massive open online courses. *The International Review of Research in Open and Distributed Learning, 15*(1), 133–160. Retrieved 18 January 2018 from http://www.irrodl.org/index.php/irrodl/article/view/1651.

Jordan, K. (2015). Massive open online course completion rates revisited: Assessment, length and attrition. *The International Review of Research in Open and Distributed Learning, 16*(3), 341–358. Retrieved 18 January 2018 from http://www.irrodl.org/index.php/irrodl/article/view/2112/3340.

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology, 65*(4), 681–706.

Khalil, H., & Ebner, M. (2014). MOOCs completion rates and possible methods to improve retention: A literature review. In *Proceedings of EdMedia. EdMedia: World conference on educational media and technology*, 23 June Tampere, Finland (pp. 1305–1313). Waynesville, NC: Association for the Advancement of Computing in Education (AACE).

Kim, M. (2005). The effects of the assessor and assessee's roles on preservice teachers' metacognitive awareness, performance, and attitude in a technology-related design task. DPhil Dissertation. Electronic Theses, Treatises and Dissertations, Florida State University. Retrieved 18 January 2018 from http://purl.flvc.org/fsu/fd/FSU_migr_etd-3051.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses categories and subject descriptors. In *Proceedings of the third international conference on learning analytics and knowledge* (pp. 170–179). New York: Association for Computing Machinery (ACM) Press.

Koller, D., Ng, A., Do, C., & Chen, Z. (2013). Retention and intention in massive open online courses. *Educause Review, 48*(3), 62–63. Retrieved 18 January 2018 from http://www.educause.edu/ero/article/retention-and-intention-massive-open-online-courses.

Kulkarni, C., Wei, K. P., Le, H., Chia, D., Papadopoulos, K., Cheng, J., et al. (2015). Peer and self-assessment in massive online classes. In H. Plattner, C. Meinel, & L. Leifer (Eds.), *Design thinking research: Building innovators* (pp. 131–168). Cham: Springer.

Lewin, T. (2012). Colorado State to offer credits for online class. *The New York Times*, 6 September. Retrieved 18 January 2018 from http://www.nytimes.com/2012/09/07/education/colorado-state-to-offer-credits-for-online-class.html?_r=0.

Luo, H., Robinson, A. C., & Park, J.-Y. (2014). Peer grading in a MOOC: Reliability, validity, and perceived effects. *Journal of Asynchronous Learning Networks, 18*(2), n2. Retrieved 18 January 2018 from http://onlinelearningconsortium.org/sites/default/files/429-2286-1-LE.pdf.

Mackness, J., Mak, S., & Williams, R. (2010). The ideals and reality of participating in a MOOC. In L. Dirckinck-Holmfeld, V. Hodgson, C. Jones, M. de Laat, D. McConnell & T. Ryberg (Eds), *Proceedings of the seventh international conference on networked learning* (pp. 266–275). Lancaster: University of Lancaster.

Mak, B., & Coniam, D. (2008). Using wikis to enhance and develop writing skills among secondary school students in Hong Kong. *System, 36*(3), 437–455.

Mukta, G. (2015). Experimenting with open online office hours. *edX Blog* (blog post 21 July). Retrieved 18 January 2018 from http://blog.edx.org/experimenting-with-open-online-office-hours.

Newton, D. (2015). Cheating in online classes is now big business. *The Atlantic*, 4 November. Retrieved 18 January 2018 from http://www.theatlantic.com/education/archive/2015/11/cheating-through-online-courses/413770/.

Onah, D. F. O., Sinclair, J., & Boyatt, R. (2014). Dropout rates of massive open online courses: Behavioral patterns. In L. Gómez Chova, A. López Martínez & I. Candel Torres (Eds), *Proceedings of the 6th international conference on education and new learning technologies (EDULEARN 14)* (pp. 5825–5834). Valencia: International Academy of Technology, Education and Development (IATED). Retrieved 31 January 2018 from http://wrap.warwick.ac.uk/65543/1/WRAP_9770711-cs-070115-edulearn2014.pdf.

Pappano, L. (2012). The year of the MOOC. *The New York Times*, 2 November. Retrieved 18 January 2018 from http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html?pagewanted=all.

Piech, C., Huang, J., Chen, Z., Do, C., Ng, A., & Koller, D. (2013). Tuned models of peer assessment in MOOCs. In S. K. D'Mello, R. A. Calvo & A. Olney (Eds), *Proceedings of the 6th international conference on educational data mining (EDM 2013)* (pp. 153–160). Worcester, MA: International Educational Data Mining Society. Retrieved 31 January 2018 from http://www.educationaldatamining.org/EDM2013/proceedings/EDM2013Proceedings.pdf.

Raman, K., & Joachims, T. (2014). Methods for ordinal peer grading. In *Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1037–1046). New York. Association for Computing Machinery (ACM) Press.

Ramesh, A., Goldwasser, D., Huang, B., Daumé Iii, H., & Getoor, L. (2014). Understanding MOOC discussion forums using seeded LDA. In *Proceedings of the ninth workshop on innovative use of NLP for building educational applications* (pp. 28–33). Baltimore, MD: Association for Computational Linguistics.

Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook. In F. Ricci, L. Rokach, B. Shapira, & P. B. Kantor (Eds.), *Recommender systems handbook* (pp. 1–35). Boston: Springer.

Robinson, A. C., Kerski, J., Long, E. C., Luo, H., DiBiase, D., & Lee, A. (2015). Maps and the geospatial revolution: Teaching a massive open online course (MOOC) in geography. *Journal of Geography in Higher Education, 39*(1), 65–82.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20–27.

Roediger, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. Mestre & B. Ross (Eds.), *Psychology of learning and motivation* (pp. 1–36). Oxford: Elsevier.

Sandeen, C. (2013a). Assessment's place in the new MOOC world. *Journal of Research and Practice in Assessment, 8*(1), 5–12.

Sandeen, C. (2013b). Integrating MOOCS into traditional higher education: The emerging "MOOC 3.0" era. *Change: The Magazine of Higher Learning, 45*(6), 34–39.

Sclater, N. (2009). The organizational impact of open educational resources. In U. D. Ehlers & D. Schneckenberg (Eds.), *Changing cultures in higher education* (pp. 485–497). Berlin: Springer.

Sharples, M. (2000). The design of personal mobile technologies for lifelong learning. *Computers and Education, 34*(3–4), 177–193.

Sharples, M., Kirsop, L., & Kholmatova, A. (2016). Designing small group discussion for a MOOC platform. In *Proceedings of the third conference on learning with MOOCS (LWMOOCs'16): Being and learning in a digital age* (pp. 11–12). Philadelphia: University of Pennsylvania. Presentation slides retrieved 31 January 2018 from https://www.slideshare.net/sharplem/small-group-learning-for-a-mooc-pplatform.

Shermis, M.D., Burstein, J., Higgins, D., & Zechner, K. (2010). Automated essay scoring: Writing assessment and instruction. In E. Baker, B. McGaw & N. S. Petersen (Eds), *International encyclopedia of education* (3rd ed., pp. 75–80). Oxford: Elsevier. Retrieved 30 January 2018 from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.652.7014&rep=rep1&type=pdf.

Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing, 13*(3), 298–317.

Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning, 2*(1), Article no. 1. Retrieved 18 January 2018 from http://er.dut.ac.za/handle/123456789/69.

Siemens, G. (2013). Massive open online courses: Innovation in education? In R. McGreal, W. Kinuthia & S. Marshall (Eds), *Open educational resources: Innovation, research and practice* (pp. 5–15). Vancouver: Commonwealth of Learning and Athabasca University.

Skrypnyk, O., Joksimović, S., Kovanović, V., Gašević, D., & Dawson, S. (2015). Roles of course facilitators, learners, and technology in the flow of information of a cMOOC. *The International Review of Research in Open and Distributed Learning, 16*(3), 188–217. Retrieved 18 January 2018 from http://www.irrodl.org/index.php/irrodl/article/view/2170/3347.

Sluijsmans, D. M. A., Brand-Gruwel, S., & van Merriënboer, J. J. G. (2002a). Peer assessment training in teacher education: Effects on performance and perceptions. *Assessment and Evaluation in Higher Education, 27*(5), 443–454.

Sluijsmans, D. M. A., Brand-Gruwel, S., van Merriënboer, J. J. G., & Bastiaens, T. J. (2002b). The training of peer assessment skills to promote the development of reflection skills in teacher education. *Studies in Educational Evaluation, 29*(1), 23–42.

Smith, P. (2014). The coming era of personalized learning paths. *Educause Review, 49*(6). Retrieved 18 January 2018 from http://er.educause.edu/articles/2014/11/the-coming-era-of-personalized-learning-paths.

Soares, L. (2011). The "personalization" of higher education: Using technology to enhance the college experience. Center for American Progress (web post, 4 October). Retrieved 18 January 2018 from https://www.americanprogress.org/issues/labor/report/2011/10/04/10484/the-personalization-of-higher-education/.

Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research, 69*(1), 21–51.

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press.

Staubitz, T., Petrick, D., Bauer, M., Renz, J., & Meinel, C. (2016). Improving the peer assessment experience on MOOC platforms. In *Proceedings of the third ACM conference on learning at scale* (pp. 389–398). Edinburgh: Association for Computing Machinery (ACM) Press.

Suen, H. K. (2013). Role and current methods of peer assessment in massive open online courses (MOOCs). In *Presented at the first international workshop on advanced learning sciences (IWALS)*, 21–22 October, University Park, PA.

Suen, H. K. (2014). Peer assessment for massive open online courses (MOOCs). *The International Review of Research in Open and Distributed Learning, 15*(3), 312–327. Retrieved 18 January 2018 from http://www.irrodl.org/index.php/irrodl/article/view/1680/2904.

Tomkin, J. H., & Charlevoix, D. (2014). Do professors matter? Using an a/b test to evaluate the impact of instructor involvement on MOOC student outcomes. In *Proceedings of the first ACM conference on learning at scale* (pp. 71–78). New York: Association for Computing Machinery (ACM) Press.

Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249–276.

Tsai, C.-C., Liu, E. Z.-F., Lin, S. S. J., & Yuan, S.-M. (2001). A networked peer assessment system based on a Vee heuristic. *Innovations in Education and Teaching International, 38*(3), 220–230.

Tu, C.-H., & McIsaac, M. (2002). The relationship of social presence and interaction in online classes. *American Journal of Distance Education, 16*(3), 131–150.

Uto, M., & Ueno, M. (2016). Item response theory for peer assessment. *IEEE Transactions on Learning Technologies, 9*(2), 157–170. Retrieved 30 January 2018 from https://www.computer.org/csdl/trans/lt/2016/02/07243342.pdf.

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review, 4*(1), 41–54.

van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction, 20*(4), 280–290.

Veletsianos, G., & Shepherdson, P. (2015). Who studies MOOCs? Interdisciplinarity in MOOC research and its changes over time. *The International Review of Research in Open and Distributed Learning, 16*(3). Retrieved 18 January 2018 from http://www.irrodl.org/index.php/irrodl/article/view/2202/3348.

Walvoord, M. E., Hoefnagels, M. H., Gaffin, D. D., Chumchal, M. M., & Long, D. A. (2008). An analysis of Calibrated Peer Review (CPR) in a science lecture classroom. *Journal of College Science Teaching, 37*(4), 66–73.

Webb, N. M., Troper, J. D., & Fall, R. (1995). Constructive activity and learning in collaborative small groups. *Journal of Educational Psychology, 87*(3), 406–423.

Xiong, Y., Goins, D., Suen, H. K., Pun, W. H., & Zang, X. (2014). A proposed credibility index (CI) in peer assessment. In *Presented at the 76th annual meeting of the National Council on Measurement in Education (NCME)*, 2–6 April, Philadelphia, PA.

Xiong, Y., Li, H., Kornhaber, M. L., Suen, H. K., Pursel, B., & Goins, D. D. (2015). Examining the relations among student motivation, engagement, and retention in a MOOC: A structural equation modeling approach. *Global Education Review, 2*(3), 23–33.

Yuan, L., & Powell, S. (2013). *MOOCs and open education: Implications for higher education.* Bolton: Centre for Educational Technology and Interoperability Standards (CETIS). Retrieved 18 January 2018 from http://publications.cetis.org.uk/wp-content/uploads/2013/03/MOOCs-and-Open-Education.pdf.

The authors

Yao Xiong, PhD, is a Research Associate at the Learning Research and Development Center of the University of Pittsburgh. Her research interests include web-based peer assessment, applying Bayesian methods in educational measurement, and educational data mining.

Hoi K. Suen, EdD, is a Distinguished Professor Emeritus of Educational Psychology at the Pennsylvania State University. His research interests include psychometrics, educational assessment, and programme evaluation.