How do we model learning at scale?

Citation for published version:
Joksimovic, S, Poquet, O, Kovanovic, V, Dowell, N, Mills, C, Gasevic, D, Dawson, S, Graesser, A & Brooks, C 2017, 'How do we model learning at scale? A systematic review of research on MOOCs', *Review of Educational Research*, vol. 88, no. 1, pp. 43-86. https://doi.org/10.3102/0034654317740335

Digital Object Identifier (DOI):
10.3102/0034654317740335

How do we Model Learning at Scale? A Systematic Review of Research on MOOCs

Srećko Joksimović

University of South Australia, Australia

Oleksandra Poquet

University of South Australia, Australia

Vitomir Kovanović

University of South Australia, Australia

Nia Dowell

University of Michigan, United States

Caitlin Mills

University of Notre Dame, United States

Dragan Gašević

University of Edinburgh, United Kingdom

Shane Dawson

University of South Australia, Australia

Arthur C. Graesser

University of Memphis, United States

Christopher Brooks

University of Michigan, United States

Abstract

Despite a surge of empirical work on student participation in online learning environments, the causal links between the learning-related factors and processes with the desired learning outcomes remain unexplored. This study presents a systematic literature review of approaches to model learning in Massive Open Online Courses offering an analysis of learning related constructs used in the prediction and measurement of student engagement and learning outcome. Based on our literature review, we identify current gaps in the research, including a lack of solid frameworks to explain learning in open online setting. Finally, we put forward a novel framework suitable for open online contexts based on a well-established model of student engagement. Our model is intended to guide future work studying the association between contextual factors (i.e., demographic, classroom, and individual needs), student engagement (i.e., academic, behavioral, cognitive, and affective engagement metrics) and learning outcomes (i.e., academic, social, and affective). The proposed model affords further inter-study comparisons as well as comparative studies with more traditional education models.

Keywords: Non-formal education, learning environments, MOOCs, engagement

How do we Model Learning at Scale? A Systematic Review of Research on MOOCs

Massive Open Online Courses (MOOCs), as one of the most prominent ways for facilitating learning at scale, have now been part of the educational landscape for almost a decade. The volume of learners enrolling in MOOCs generated widespread interest among the public, popular press, social and education commentators (Reich, Stewart, Mavon, & Tingley, 2016). Some stakeholders expressed their belief in the groundbreaking effect MOOCs may have on higher education, possibly making traditional brick-and-mortar universities obsolete (Shirky, 2013). Alongside the potential of MOOCs, professionals in educational technology have expressed concerns about widely applied pedagogical models based on the information transmission integrated in many of the MOOCs. Despite a polarized debate (Selwyn, Bulfin, & Pangrazio, 2015), student enrollment numbers and course offerings continued to grow (Jordan, 2015a; Shah, 2015). This has resulted in a wave of interest from researchers and, within a relatively short time frame, we have witnessed a substantial number of research studies and reports on MOOCs (Jordan, 2015b), as well as the formation of two annual MOOC-related scholarly conferences (Haywood, Aleven, Kay, & Roll, 2016; Siemens, Kovanović, & Spann, 2016).

Research has largely focused on students' persistence in MOOCs and the development of models to predict dropout or academic performance. Despite the volume of work to date, commentators have criticized such research as being primarily observational and lacking appropriate rigor. Reich (2015), for example, asserted that MOOC research has failed to provide causal linkages between the observed metrics and student learning, despite the vast amount of data collected on student activity within MOOCs. This limitation is in part due to the lack of theoretically-informed approaches employed in the analysis of MOOCs. Institutional reports on

MOOC provisions as well as special issues on MOOCs have offered some insight into engagement during learning with MOOCs, but have presented little (or no) evidence of the factors contributing to learning *per se* (DeBoer, Ho, Stump, & Breslow, 2014; Reich, 2015).

The limited insight offered by the research thus far can be attributed to a general lack of understanding that non-formal educational settings, such as MOOCs (Walji, Deacon, Small, & Czerniewicz, 2016), differ from those of more traditional forms of education in many aspects. Technology and economies of scale allows for designing courses for unparalleled numbers of students and in ways that were not available in more traditional forms of learning (Reich 2015). Some of the reports indicate that more than 58 million of students enrolled in at least one of almost 7,000 MOOCs, offered by more than 700 universities (Shah, 2015). Students' interactions in such contexts result in a magnitude of data on learning and in various data formats, stored within platforms promote practices that are substantially different from those in traditional face-to-face or online learning (DeBoer et al., 2014; Evans, Baker, & Dee, 2016). The diversity of students represented in MOOCs is also unprecedented. The range in diversity is reflected in students' cultural backgrounds, socioeconomic and employment status, educational level, and importantly, their motivations and goals for registering in a particular course (DeBoer et al., 2014; Glass, Shiokawa-Baklan, & Saltarelli, 2016; Reich et al., 2016). Therefore, DeBoer et al. (2014) and Evans et al. (2016) among others, have argued that MOOCs require a "re-operationalization and reconceptualization" (p.2) of the existing educational variables (e.g., enrollment, participation, achievement) commonly applied to conventional courses.

This study concurs with the argument by DeBoer and colleagues (2014) and posits that a more holistic approach is needed to understand and interpret learning-related constructs (observed during learning) and their association with learning (outcomes). These learning-

related constructs are often observed under the broader concept of *learning* – a term commonly applied across a range of contexts with multiple interpretations and definitions (Illeris, 2007). Conceptually, learning refers to both (1) a complex multilevel process of changing cognitive, social and affective aspects of the self and the group, as well as (2) the outcomes of this process observed through the cognitive, social and/or affective change itself. Distinguishing between the process and the outcomes of learning, along with the contextual elements, is essential when modeling the relationships between them.

The necessity to redefine existing educational variables within new contexts originates from the concept of validity in educational assessment (Moss, Girard, & Haniford, 2006). Validity theories in educational measurement have been primarily concerned with a(1) standardized forms of assessment (e.g., tests); (2) providing a framework for interpretations of assessment scores in a given learning environment; and (3) making decisions and taking actions to support and enhance students' learning (Moss et al., 2006). However, aiming to take a more pragmatic approach to validation, Kane (1992, 2006) posited that performance assessment should not be restricted to "test items or test-like tasks" (Kane, 2006, p.31). Evaluation of students' performance can include a wide variety of tasks, performed in different contexts and situations (Kane, 2006). To make valid interpretations of student performance in MOOCs, it is necessary to have a clear understanding of how evaluation metrics have been defined for a given learning environment and its students (Kane, 2006; Moss et al., 2006).

This study contributes to the development of the "next generation of MOOC research" (Reich, 2015, p. 34) that can aid in explaining the learning process and the factors that influence learning outcomes. The present study critically examines how learning-related constructs are measured in MOOC research, and re-operationalizes commonly used metrics in relation to the

specific educational variables within (1) learning contexts; (2) learning processes (i.e., engagement), and (3) learning outcomes. The study is framed in Reschly and Christenson's (2012) model of the association between context, engagement, and outcome. Reschly and Christenson (2012) defined engagement as both a process and an outcome, therefore aligning the concept of engagement with a broader understanding of learning. In their work, Reschly and Christenson (2012) observed four aspects of student engagement: academic, behavioral, affective and social. The authors conceptualized these as mediators between contextual factors, such as student demographics or intentions, and learning outcomes. Thus, we first examine commonly used learning-related metrics through a systematic review of the literature between 2012 and 2015 inclusive. We then analyze these metrics of observed student activity in light of Reschly and Christenson's (2012) model of associations between context, engagement, and student outcomes. Reschly and Christenson's (2012) model stems from the work on dropout prediction and increasing school completion, observing engagement on a continuum scale (ranging from low to high). By discussing the metrics representing the outcomes and indicators of learning within Reschly and Christenson's model, we demonstrate limitations and strength of current approaches to measuring learning in MOOCs. We then highlight differences that emerge between the Reschly and Christenson model and open online settings, to propose a modified operationalization of how learning in MOOCs can be studied.

We refer to MOOCs are planned learning experiences within *non-formal*, *digital* educational settings, used to facilitate learning at scale. In computer-mediated (networked) settings, as is the context of our research, learning is observed as a dynamic and complex process. Learning, involves student interactions with other students, teachers, and content (Goodyear, 2002; Halatchliyski, Moskaliuk, Kimmerle, & Cress, 2014). By *non-formal,* we

assume any systematic learning activity conducted outside the formal/institutional settings (Eraut, 2000); in MOOCs such activity occurs within the structure prepared by the instructor but is heavily influenced by learner's motivations, actions, and decisions. Finally, *digital (education)*, refers to an emerging approach to learning mediated by various technological methods (Siemens, Gašević, and Dawson, 2015). Digital learning brings online, distance and blended learning under a single concept, and could be structured as formal/informal, self-regulated, structured/unstructured, or lifelong.

Research Questions

The present study identifies student engagement metrics and contextual factors commonly used to model learning and predict learning outcome or course persistence in non-formal, digital educational settings. First, we examine traces of student activity operationalized as indicative of learning processes through a systematic review of the literature. We then use findings from the review to refine a well-established model of student engagement in the context of learning with MOOCs. Finally, we summarize the common methods used to examine the association between the metrics calculated and outcome measured, as means for defining and interpreting eventual association between different elements of the model constructs. To address these aims we posed the following research questions:

RQ1. What are the most common approaches to operationally defining and measuring learning outcomes in the research on MOOCs? Is there misalignment between them with a common model of student engagement?

RQ2. What are the most common approaches to operationally defining and measuring learning context and student engagement in the research on MOOCs? Is there misalignment between them with a common model of student engagement?

RQ3. What are the common approaches to studying the association between the identified metrics and measured outcome?

In contending that the majority of the current MOOC studies focus on the examination of the association between student engagement and course outcomes, Reich (2015) argues that "[d]istinguishing between engagement and learning is particularly crucial in voluntary online learning settings" (p.34, *ibid.*). However, Reich's argument is limited to assessment scores, rather than on the individual and group changes that take place during and over the process of learning. According to Reich, introducing assessment at multiple time points, relying on the assessment methods validated in prior research, and making a better integration of assessment in the course design in general, are important steps in understanding learning in MOOCs (Reich, 2015). In part, we concur with Reich's (2015) premise. However, we also acknowledge that not all MOOCs include (formal) assessment practices, especially those MOOCs designed with connectivist pedagogies (Siemens, 2005). Additionally, the diversity of student intentions for enrolling in voluntary online learning requires additional considerations on how learning might be operationalized in the context of MOOCs in the absence of assessment models. Moreover, Gašević, Dawson, Rogers, and Gašević (2016) stressed the importance of considering contextual factor when trying to predict learning outcome or course persistence. Framing their research around the Winne and Hadwin (1998) model of self-regulated learning, Gašević and colleagues (2016) showed how instructional conditions, as a vital component of external conditions affect the interpretation of learning-related measures. Therefore, we rely on the Reschly and Christenson (2012) model that observes student engagement as a mediator between contextual factors (e.g., intents) and learning outcomes, regardless of their operationalization. The model

offers a broader view on the outcomes of learning, defining engagement as both a process and an outcome (Reschly & Christenson, 2012).

Method

Literature Search and Inclusion Criteria

To derive the extant research literature a computer-based search from 2012 to 2015 (inclusive) was undertaken over three phases (Figure 1). Although the first MOOC was offered in 2008, it was only in 2012 when the major MOOC providers (i.e., Coursera, edX and Udacity) were established, and an inaugural course was launched[1]. Moreover, as noted by Raffaghelli, Cucchiara, and Persico (2015), it was only post 2012 when the MOOC research proliferated, demonstrating a growing maturation of the field.

The first phase involved a search of the following databases: EdiTlib, EBSCOhost (Education Source, ERIC, PsychINFO, PsychArticles, and Academic Search Complete), Scopus, Web of Science, Science Direct, Taylor & Francis, and Willey. The following search criteria were used for defining inclusion in the study:

Title, abstract, and/or keywords must contain at least one of the following terms:

*mooc** **OR** *"massiv* open online"* **AND**

Title, abstract, and/or keywords must contain at least one of the following terms:

predict **OR** *learn** **OR** *associat** **OR** *assess** **AND**

Title, abstract, and/or keywords must contain at least one of the following terms:

*engage** **OR** *outcome** **OR** *retention* **OR** *interact** **OR** *behavi** **OR** *attrition* **OR** *dropout* **OR** *particip** **OR** *complet**.

[1] http://news.mit.edu/2012/edx-faq-050212

The initial search resulted in 1,004 studies. After completing the search, two researchers coded the studies according to the inclusion criteria. The coding process comprised reading the title and abstract for each study and assigning a binary category – relevant/not-relevant. In cases where it was not obvious from the title and abstract whether a given study would be relevant for answering our research questions, the coders examined the article in detail (i.e., reading the methods and results sections). The coding was conducted through several steps. The first step included the joint coding of an initial set of 50 studies, in order to refine the inclusion criteria and to define a set of rules for accepting studies for the review. The changes between the original inclusion and exclusion criteria were minor. Specifically, the initial version of the inclusion criteria did not consider employees (e.g., we were not aware of the significant number of studies focusing on professional medical education), as it was further added to item (6) in the list below. Also, in the initial inclusion criteria, we had not been precise about item (8) from the list below, i.e., exclusion of studies relying on log data and surveys or questionnaires. These were later included as a special sub-set because they contained various learning-related metrics extracted from log-data, often used to describe the datasets of the analyzed studies. In other words, although such studies did not attempt to predict learning outcome of course persistence, they included operationalization of learning-related constructs.

Two coders coded all the studies together and inter-rater agreement (Cohen, 1960) was calculated after coding 250, and 500 studies, as well as at the end of the coding process. All conflicts were resolved at each of the steps. The two coders reached an average inter-rater agreement of 93.6%, with an average Kappa of 0.67. The final set included 96 studies that satisfied the following criteria for inclusion in this review, where the study:

(1) presents an original (primary) research, analyzing MOOC data,

(2) addresses a problem of predicting learning and/or persistence in MOOCs,

(3) analyzed higher or adult education,

(4) was published in 2012 or beyond,

(5) was published in peer-reviewed journal/conference proceedings, available in English,

(6) participants in primary studies were non-disabled undergraduate students, graduate students, and/or employees (e.g., teachers and nurses),

(7) focuses on algorithms that help to identify variables related to learning,

(8) relies on a log data and/or surveys/questionnaires, and the study applies inferential statistics and not primarily descriptive analysis to investigate the data.

Inclusion of both journal and conference papers in our systematic review was necessary. The exclusion of conference papers (and conference proceedings in computer science) would significantly limit the number of studies analyzed. In addition, the analysis targeted studies publicized at the onset of MOOC research, and publishing in conference proceedings would represent the most prominent way for disseminating novel research in a field. Their exclusion would also mean that research published in the main outlet for publication by computer scientist (for whom conference publications are mostly more important than journals), an important constituent group in the field, would be ignored. By integrating the literature from a variety of sources, this review aimed at summarizing the broadest possible set of learning-related metrics used to date. Such a broad overview did not negatively impact on the quality of the analysis. Rather, the extension of the review materials offered a fuller representation of the quantitative measures used to investigate learning at scale.

To ensure a comprehensive and accurate search we manually searched the following journals: Journal of Learning Analytics, Journal of Educational Data Mining, British Journal of

Educational Technology, The Internet and Higher Education, Journal of Computer Assisted Learning, The International Review of Research in Open and Distributed Learning, Journal of Educational Technology & Society, Educational Technology Research and Development, IEEE Transactions on Learning Technologies, Distance Education, International Journal of Computer-Supported Collaborative Learning, ACM Transactions on Computer-Human Interaction, and the International Journal of Artificial Intelligence in Education. A manual search was also conducted for conference proceedings including: International Conference on Learning Analytics and Knowledge, International Conference on Educational Data Mining, International Conference on Computer Supported Collaborative Learning, ACM Annual Conference on Learning at Scale, ACM SIGCHI Conference on Human Factors in Computing Systems, ACM Conference on Computer Supported Collaborative Work, European Conference on Technology Enhanced Learning, and International Conference on Artificial Intelligence in Education Conference. The list of relevant journals and conferences was obtained from Google Scholar metrics list of top publications in the educational technology research category. The manual search resulted in an additional 23 studies, providing a total list of 119 studies selected for further consideration.

In the final phase, we coded the selected 119 studies according to the coding scheme (Table S1 in the supplementary material[2]). The coding scheme was developed with respect to the STROBE Statement[3] recommendations for the observational studies, adapted and extended to account for the specific research questions of this systematic review. Although the STROBE list has been primarily used in medical research, these recommendations for the observational studies are comprehensive, offering a valid basis for coding schemes used in other domains (such as educational research). Nevertheless, given the focus of our study, we removed items such as

[2] http://bit.ly/learning-at-scale-supplement
[3] http://www.strobe-statement.org/index.php?id=strobe-home

"Give reasons for non-participation at each stage", as one of the aspects of describing study participants available in the STROBE recommendations, as well as "Funding" (also available among the STROBE items), as these items were not relevant for the context of the present study. Following the final screening by four independent coders 38 studies were identified that met the above-defined criteria for inclusion (Figure 1).

Analysis

To address research questions, a synthesis of the 38 systematically selected studies was undertaken. The main focus of the systematic review was on the metrics used to assess learning in MOOCs and the outcome variables measured. Thus, each of the studies was coded with respect to these parameters. Moreover, we examined how different studies defined outcome (e.g., learning outcome or dropout), as well as how each of the predictors was extracted. Besides the variables used, we also indicated the statistical methods used to examine the association between predictors and outcome(s), and the noted results (if reported) for each of the analyses applied in the reviewed studies. A definition for each of the coded attributes is provided in Table S1 (please see supplementary material).

Additionally, the studies were coded with respect to (1) the theories they adopted to analyze learning (e.g., online or distance education theories) and (2) study objectives (e.g., predicting final course grade, or predicting drop-out). We also examined whether a study was exploratory of confirmatory, whether authors discussed limitations and generalizability of study findings, and to what extent pedagogical and/or contextual factors were considered. The main study findings across the reviewed literature were summarized to identify common and significant conclusions.

To contextualize the variables, and for further research, we coded the platform where a MOOC was delivered, the educational level suggested for each of the offered courses, course domain, and course completion rates. Due to numerous interpretations of how course completions are calculated (see Section 4.1), here we captured the count of registered, active students, and the number of students who obtained a certificate, if reported. Furthermore, we were interested in the domain of the analyzed courses. That is, whether the courses offered a certificate, and how many xMOOCs or cMOOCs were included in the analyses. The types of MOOCs were labelled based on the categorization commonly found in the literature distinguishing between the connectivist cMOOCs[4] and Coursera-like xMOOCs[5] (Rodriguez, 2012).

We also identified the data sources used for each of the studies included in the review as well as the study focus (e.g., all students, only students who posted to a discussion forum, or students who successfully completed a course).

Limitations

The diversity of terms describing similar concepts and measures presented a significant challenge for this study. Researchers would frequently state that the study examined an association between "learning outcome" and various metrics of student engagement, without a clear description what was considered as an outcome. The lack of specificity in the reviewed studies prompted the need for added interpretations based on a review of the analyzed data. Additional challenges again related to a lack of detail surrounding the metrics used to measure variables associated with any developed predictive model. For example, simply stating that a

[4] Decentralized MOOCs that utilize various platforms to foster interactions between learners, focused on knowledge construction, where teachers' role is primarily focused on the early instructional design and facilitation.

[5] Focused on content delivery to large audiences, utilizing a single platform such as Coursera or edX, where the learning process is teacher-centered.

measure included a "count of discussion activities" is insufficient detail. Simply referring to a broad count of activity does not make it clear if the metric included an aggregation of all possible discussion activities (e.g., posting, viewing, voting) or a specific subset.

The ability to determine measures of time-on-task also presents issues for the review. As Authors (2015c) pointed out, it is important to specify how time-on-task is determined and which (if any) heuristics or approximations were applied. This was not always the case with the studies included in this review. Therefore, the majority of the reviewed studies required detailed investigation of the methods applied and the description of the data analyzed to determine appropriate categorization. The lack of consistency in terminology necessitated further interpretations. Furthermore, we classified variables across the various dimensions of student engagement in light of Reschly and Christenson's model. This classification added a level of subjectivity, which could lead to challenges in ensuring internal validity. Finally, to maintain a quantitative focus, this study excluded often rich observations drawn from qualitative studies which would be more appropriate for a separate literature review.

Quantitative overview of the selected studies

The aim of this section is to present the selected dataset of MOOC research papers. Specifically, here we reviewed 38 studies in relation to their bibliographic information and their overall focus prior to the in-depth analysis of learning-related metrics used in these academic papers.

Table 2 shows the author(s), titles, publication year, publication venue types, the number of courses analyzed, data sources used, and the number of students[6] (registered, active, completed) in the studies included in this review. We observed that, as noted in Figure 2, a

[6] Several studies did not report precise information about the number of participants included or did not report number of students at all, thus we noted "more than" a certain number of participants or noted as "NR".

majority of studies included in the systematic review were published at conferences (Figure 2). Although we reviewed the literature published between 2012 and 2015, only one study published prior 2014 satisfied the inclusion criteria.

Courses delivered on the Coursera platform were most commonly analyzed, followed by the edX platform (Figure 3). We observed that only a few studies examined courses delivered by other MOOC providers. For example, only one study analyzed data delivered via the D2L learning management system (Goldberg et al., 2015), Sakai (Heutte, Kaplan, Fenouillet, Caron, & Rosselle, 2014), UNED-COMA platform (Santos, Klerkx, Duval, Gago, & Rodríguez, 2014), or a course delivered in a distributed environment (i.e., Distributed), using social media (Authors, 2015a). Finally, only Adamopoulos's (2013) study utilized data from MOOCs delivered across various platforms (i.e., Canvas Network, Codeacademy, Coursera, edX, Udacity, and Venture Lab). However, this study was not included in the summary provided in Figure 3, as it was not clear which of the 133 courses analyzed was delivered within the various platforms.

Most of the evidence derived from the modeling of learning behavior in MOOCs was collected from computer science courses (Figure 3). Physical science and engineering, life and social sciences, and arts and humanities courses were also well-represented. In contrast, language learning and personal development courses were rarely examined. This observation is reflective of the sheer volume of MOOC offerings related to the computer sciences compared to other disciplines (Shah, 2015), as well as the technical skills that are required to process MOOC data for analysis.

Only two studies within the dataset analyzed data from connectivist learning environments (Figure 3). Heutte et al. (2014) and Authors (2015a) incorporated data from social media (e.g., Twitter or blogs) in order to understand factors that could explain learning in

cMOOCs. The remaining studies examined MOOCs that were designed in a more structured framework (i.e., xMOOCs).

The systematic review further revealed that typically learning in MOOCs is studied through the analysis of the trace data combined with discussion or survey data, and is generally derived from a single course (Figure 4). Very few studies combined more than two data sources (e.g., survey, trace, and discussion forum data). Moreover, there was only one study that relied on learner-generated data., such as blogs, Twitter, and/or Facebook posts. On the other hand, studies that analyzed two or more courses primarily focused on trace or discussion forum data.

For most the courses analyzed, researchers reported 25,000 to 50,000 registered students (Table 2). This size of cohorts is not surprising given that an enrollment of 25,000 students is commonly referred to as a typical MOOC size (Jordan, 2015b). However, the number of active students or students included in the analyses was generally less than 10,000. As indicated in Table 2, researchers often failed to report the number of registered and active/observed students in their studies.

Results and Discussion

Common Operationalization of Learning Outcomes (RQ1)

As a part of the first research question, our analysis aimed to identify how the reviewed literature defined the results of the learning process, and to discuss their alignment with a common model of student engagement. Specifically, we analyzed how researchers operationalized and measured the outcome variables they were predicting in their various models. Our analysis suggests that learning outcomes have been defined as course completion (e.g., Crossley et al., 2015; Loya, Gopal, Shukla, Jermann, & Tormey, 2015); engagement (Sharma, Jermann, & Dillenbourg, 2015), social interactions (Vu, Pattison, & Robins, 2015);

sociability (Brooks, Stalburg, Dillahunt, & Robert, 2015), and learning gains (Koedinger, Kim, Jia, McLaughlin, & Bier, 2015; X. Wang, Yang, Wen, Koedinger, & Rosé, 2015). The majority of studies use the metrics capturing in-course academic performance and persistence interchangeably with the notions of failure and success within the course (e.g., Adamopoulos, 2013; Santos et al., 2014; Sharma et al., 2015).

Academic performance. Academic achievement in the form of final exam or an accumulated course grade was the predominant variable or proxy for course outcome (Bergner, Kerr, & Pritchard, 2015; Coffrin, Corrin, de Barba, & Kennedy, 2014; Crossley et al., 2015; Gillani & Eynon, 2014; Kennedy, Coffrin, de Barba, & Corrin, 2015; Koedinger et al., 2015; Ramesh, Goldwasser, Huang, Daume, & Getoor, 2014b; Sinha & Cassell, 2015; Tucker, Pursel, & Divinsky, 2014; X. Wang et al., 2015). Alternative to the final grade, a course outcome was defined through basic levels of certification: e.g. 'no certificate', 'normal certificate' and 'certificate with distinction' (e.g., Brooks, Thompson, & Teasley, 2015); potentially complemented with additional categories such as 'completing some exams' and 'completing all exams without passing the course' (Engle, Mankoff, & Carbrey, 2015). In most cases, these levels were derived from the grades, with the exception of Adamopoulos (2013) who asked students to self-report their level of performance from a predefined list.

Cognitive Change. Instead of using grades or categories representing performance to measure the result of learning, several studies employed measures to capture cognitive change of a learner. Champaign et al. (2014) defined course outcome as the improvement of students' ability to succeed on quizzes, i.e., if they were over-performing their prior grades, rather than whether they were receiving high scores. Konstan, Walker, Brooks, Brown, and Ekstrand (2015) took a somewhat similar approach by measuring the change in knowledge through 20-item pre-

and post-class knowledge tests created by the instructor. Finally, Li, Kidziński, Jermann, and Dillenbourg (2015) conducted a study predicting the difficulty of the course content, that in a way reflected that if a learning material required more effort from a learner. Their study established an association between student viewing patterns of the in-course video lectures with student perceived video difficulty.

Persistence and Drop-Out. In our review, the studies predicting learning persistence were observed as another approach mainstream to the analysis of learning in MOOCs. Researchers appeared to willingly include course completion or course grade as a point of reference in persistent behavior. Many authors explicitly defined persistence as engagement with both content and assessment and sometimes forum activity as well. For instance, Ye and colleagues (2015) defined a drop-out as a learner who accessed fewer than 10% of the lectures and performed no further assessment activities. Vu and colleagues (2015) integrated participation in more activities than just assessment by operationalizing drop-out events as a stop of engagement in learning events spanning across the course activity including the forums as well as quiz grades. Alternatively, the students not earning a certificate and taking no action between a certain point in time and the time of the issuance of the certificates were defined as 'stop-outs' in the study by Whitehill, Williams, Lopez, Coleman, and Reich (2015). In some of the reviewed articles (e.g., Boyer & Veeramachaneni, 2015), the authors did not explain which learner activity was included as a measure of persistence from one week to the next, i.e., a task and/or a lecture.

In sum, we observed that persistent undertaking of assessment was commonly included as a full or partial indicator of how persistence was measured. Such can be interpreted as an indication of a limited understanding of MOOCs. That is, by defining persistence as a learning

outcome and a predictor of interest, researchers indicate that the mindset guiding such analysis is similar to that applied in a university setting. Specifically, learners undertake courses where their learning is marked by assessments. However, MOOCs nature of open participation does not limit student learning to undertaking assessment, but is varied depending on students' motivation (Eynon, 2014). In a way, using persistence as a proxy for learning ignores the non-formal nature of MOOCs where students are not required to get assessed or follow through the course. For some of the individuals, learning happens outside of continuous in-course assessment if they are sampling content or getting their 'just-in-time' insights relevant to a very specific question they are solving. Currently, these MOOC-specific groups with divergent intentions to learn that reach beyond the formal assessment and prescribed course activities are often grouped within an all-encompassing 'no certificate' category, the one dichotomous to full course completion.

In the analyzed dataset, the study by Sharma et al. (2015) was representative of academic work trying to work around pre-existing formal education assumptions about measuring the outcomes of learning through grades or continuous assessment. The authors expanded course outcomes to include learners who may not be pursuing certification. Measured outcomes were defined by either grades or degrees of interaction with the course material. The authors analyzed the association of clickstream data and performance with two main learner types clearly distinct in their desired course outcomes: active student (submitting graded assignments successfully, or failing) and a viewer (engaging in lectures and/or quizzes without graded assignments).

Social and Affective Aspects of Learning as a Part of Learning Outcome. A focus on social dimensions of learning outcomes was scarce as compared to academic performance or persistence. The majority of studies in this domain focused on the volume of posts or number of connections gained in course forums. Importantly, where social aspects of learning captured

through the numbers of connections or posts were used as measured outcomes, they were included as complementary to grades. The number of forum posts is the most common measure of learning associated with the social interaction. This measure has been typically recorded at the end of the course (Brooks, Stalburg, et al., 2015; Goldberg et al., 2015). Alternatively, Authors (2015a) relied on the concept of social capital to explain the outcome of the learning process. Authors (2015a) used social network analysis to quantify individual positions in networks of learners. Authors (2015a) demonstrated that socially engaged MOOC takers with higher grades and socially engaged participants with higher social capital were not necessarily the same individuals. Such a result supports the premise that MOOCs are used differently by learners, and learning with others is only relevant to some individuals. In relation to students' persistence in participating in MOOC forums, a series of studies focused on student disengagement from posting activity (X. Wang et al., 2015; Yang, Wen, Howley, Kraut, & Rose, 2015). Specifically, Wang and colleagues (2015), as well as Yang and colleagues (2015), found the relationship between the time students joined a MOOC and student difficulty in engaging with others in online discussion forums. This work emphasized the importance of the temporal aspect for modelling aspects of social interaction and collaboration (i.e., learning through the interactions with the others) as an outcome.

Affective aspects of learning outcomes were rarely incorporated into the learning outcomes and were limited to student satisfaction.

Multi-dimensional measures. Some authors used multi-dimensional measures of course outcomes. For instance, Kizilcec and Schneider (2015) predicted learner behavior that was operationalized as a multidimensional construct. The authors approached learning behavior as defined by learner progress in the course, their general performance, and social engagement. The

dimension of learner progress was quantified by the proportion of watched videos and attached assignments (more than 10%, more than 50%, and more than 80%). General performance was operationalized as receiving a certificate of completion. Finally, social engagement was operationalized through a combination of the number of posts (in relation to the most prolific learner) and received votes. Again, although the focus on metrics typical in formal courses is evident, the authors integrated different dimensions that described the learning outcomes.

Overall, in analyzing measured outcomes of learning in the selected studies we observed formal education mindset guiding researchers using measures related to certification, assessment and prediction of drop-out as undesired behavior. Such is not surprising, as the literature stemming from formal educational contexts has validated measures allowing to capture learning as performance, or learning as progress towards completion, or learning as participating in assessment. Hence, operationalizing the learning outcome perceived through an academic (formal education) lens is mostly developed. Few authors maintained focus on measuring cognitive change; whereas the focus on social outcomes of learning is scarce, with the emphasis on the volume of posts or number of connections. Affective aspects of learning outcomes are currently limited to student satisfaction. Few studies employed a more holistic approach using multi-dimensional constructs to measure (and predict) learning outcomes, or by distinguishing that not all learners in MOOCs can be described by a more common university-like profile.

In their model of engagement Reschly and Christenson (2012) described learning outcomes of two broad types. The so-called proximal learning outcomes indicate the product of the learning process that can be proximal and distal. According to the authors, proximal learning outcomes can fall under academic, social and emotional sub-categories (Figure S1 – please refer to the supplementary material). A proximal learning outcome is used to indicate school-related

outcomes, such as grades, relationships with peers, self-awareness of feelings, among others. Distal learning outcomes are observed in post-graduation settings related to adult life. In the model, these are exemplified as for instance related to employment or productive citizenry. Such distinction between what is learnt and applied at school and what is learnt and beyond is fitting in a K12 setting for which the authors developed their model. The MOOC context, however, has some differences. For the majority of their participants, MOOC experiences do not aggregate to ten years of relationships within a community where formal assessment is necessary at different phases. The MOOC participants may be interested in a timely content they need to learn as they engage for a short period of time. Alternatively, they also may undertake the MOOC in its entirety and follow all different learning goals set throughout the entire offering. Therefore, we suggest that proximal learning outcomes are redefined into the immediate and course-level, instead of the school-level, otherwise preserving their academic, social and affective aspects. For the distal learning outcomes, we suggest to redefine them as post-course, instead of referring to them as distal learning outcomes. These suggested modifications are captured in Figure 5 demonstrating the re-operationalized model, whereas the table that summarizes all the studies included in the review along with the learning outcome measured is provided in the supplementary material (Table S3).

Providing means for defining context and engagement types in learning at scale (RQ2)

A challenge for this systematic review involved summarizing a wide variety of variables used to model learning in MOOCs. This was particularly noted in the definition of latent constructs various studies claim to measure. Thus, for example, several studies measured engagement as a latent construct (Ramesh, Goldwasser, Huang, Daume, & Getoor, 2014a; Ramesh et al., 2014b; Santos et al., 2014; Sinha & Cassell, 2015). However, Santos et al.

(2014) focused primarily on metrics extracted from students' interaction within a discussion forum. Ramesh and colleagues (2014a, 2014b), as well as, Sinha and Cassell (2015) also considered students' interaction with other course resources (e.g., quizzes, videos, or lectures). On the other hand, Wang et al. (2015) measured discussion behavior operationalized through the cognitive activities extracted from discussion forum messages. Nevertheless, most studies, although focusing on somewhat similar or same metrics, did not report constructs measured. That is, those researchers focused on the measures of student activity with the course materials or with their peers (e.g., counts of videos watcher, number of messages posted), without necessarily defining such measures as engagement. Although some of the studies used the same operationalization of the measured variable, those metrics were usually labeled in different ways (e.g., discussion behavior, behavior, or engagement). Therefore, in order to provide a more coherent summary of findings, we framed our results around the constructs introduced in Reschly and Christenson's (2012) model of student engagement and adopted in our study (Figure 5).

Contextual variables. A significant number of studies (39.5%) included in the systematic review, observed contextual variables in order to determine to what extent student demographic data (10 studies), course characteristics (5 studies), or student motivation (8 studies) predict learning outcome and/or course persistence. Only one study (i.e., Konstan et al., 2015) observed all three contextual factors. On the other hand, a majority of studies that analyzed demographic data (around 66%) also observed either motivational factors or course-related characteristics.

Demographic variables have been commonly used in understanding factors that influence learning in MOOCs. Age, gender, and level of education were considered in various studies in

terms of predicting course persistence and/or achievement. Some 80% of studies that observed demographic data (i.e., out of 15 studies) included the level of education of course participants. The results somewhat differ across the studies included in the review. Goldberg and colleagues (2015), as well as, Heutte and colelagues (2014) found no significant difference in a likelihood of completing a course across the observed levels of education. The studies observed rather different course settings – health and medicine xMOOC delivered on the Desire2Learn platform Goldberg et al. (2015), and a distributed (cMOOC) version of a humanities course (Heutte et al., 2014). Moreover, Konstan et al. (2015) found no significant association between the level of education and knowledge gain or a final course grade, in a data science xMOOC, delivered using the Coursera platform. However, through the analysis of courses from various disciplines delivered on the Coursera platform, Engle et al., (2015) Greene, Oswald, and Pomerantz (2015), Kizilcec and Halawa (2015), and Koedinger et al. (2015) showed that more educated students are more likely to persist in a course and achieve higher grades.

Existing research does not provide univocal conclusions with respect to the importance of **students' age** for predicting course persistence and achievement. Engle et al. (2015), Koedinger et al. (2015), and Konstan et al. (2015) failed to find an association between students' age and course completion, final course grade, or knowledge gain. Whereas, on the other hand, Greene et al. (2015), Heutte et al. (2014), and Kizilcec and Halawa (2015), showed that older students were more likely to persist with a course. However, Kizilcec and Halawa (2015) also showed that older students achieved lower grades compared to their younger peers.

The prevailing understanding found in the studies included in this systematic review that observed students' **gender** (5 studies) as an important determinant of learning in MOOCs, is that there are no differences between male and female students with respect to the course persistence,

course outcome, and attained knowledge gains (Adamopoulos, 2013; Heutte et al., 2014; Koedinger et al., 2015; Konstan et al., 2015). Only Kizilcec and Halawa (2015) showed that male students were more likely to persist with lectures and assessment, as well as to achieve a grade above 60th percentile, across a wide range of courses (i.e., 21 courses) from various subject domains.

The existing literature on student motivation and engagement in online learning argue that the lack of student affinity to complete a course leads to higher dropout rates, and consequently failure to complete a course (Hartnett, George, & Dron, 2011). Thus, intention to complete a course and number of hours intended to devote to a course work, are commonly considered in predicting course persistence and achievement (i.e., included in 40-50% of studies that observed student motivation). Except for Konstan et al. (2015), who failed to confirm the association between students' intention (i.e., complete a course, and time devoted) and final course grade, findings from other studies (i.e., Engle et al., 2015, Greene et al., 2015, Heutte et al., 2014, and Kizilcec and Halawa, 2015) confirmed general understanding of students' **intrinsic motivation** for persistence and achievement in MOOCs.

Generalizing the findings with respect to the **course** (or **classroom**) characteristics is rather challenging given a diverse set of metrics used in the studies included in this systematic review. For example, Adamopoulos (2013) showed a negative effect of course difficulty, planned workload, and course duration (in weeks) on student retention. It is also interesting that Adamopoulos's (2013) study revealed a negative effect of self-paced courses, compared to more structured course design on successful course completion. On the other hand, Adamopoulos (2013) also showed that peer assessment (compared to automated feedback), and open textbooks, had positive effects on successful course completion. Likewise, Konstan et al. (2015) showed

that being in a specific course track (i.e., programming vs. concepts track[7]) significantly predicts course grade, also being negatively associated with normalized knowledge gains. Finally, Brooks and colleagues (2015) revealed that the fact whether students were paying for a certificate or not, had a minimal predictive power on course grades.

Although original Reschly and Christenson's model (Figure S1 – please refer to the supplementary material) argues for the importance of understanding context through the four factors, namely *family* (e.g., support for learning, goals and expectations), *peers* (e.g., educational expectations, shared common values, aspiration for learning), *school* (e.g., instruction and curriculum, support, management), and *community* (e.g., service learning), contemporary MOOC research suggests somewhat different operationalization of the contextual elements. Therefore, for research of learning at scale we argue that contextual factors should be observed through students' *demographic* data (e.g., age, gender, level of education), *classroom* characteristics (e.g., peers, course characteristics, course platform), and *individual* students' *needs* and motivation (e.g., intent to complete a course, interests in topic), as outlined in Figure 5. It should be noted here that "classroom characteristics" primarily refer to the specific attributes of the given course and not to the notion of the traditional (i.e., face-to-face) classroom.

Student Engagement. Given the purpose of the systematic review and specified search criteria, unsurprisingly, 89.5% of the studies went beyond contextual factors (primarily demographic data) and included engagement-related metrics in predicting retention or achievement in MOOCs. A considerably smaller number of studies (21%), however, attempted

[7] The course design in Konstan, Walker, Brooks, Brown, and Ekstrand (2015) study included two tracks: 1) programming track that included assignments and all the content, and 2) concepts track that was focused on learning programming concepts, without programming assignments and with only few video lectures related to specific programming tasks.

to align extracted metrics with existing educational variables. Such an approach resulted in a wide diversity of variables used to quantify student engagement in non-formal, digital educational settings.

Around 20% of the studies included in the review is the total number of messages students contributed in a discussion forum, during a course. Crossley and colleagues (2015), Engle and colleagues (2015), Goldberg and colleagues (2015), as well as, Vu and colleagues (2015), showed that students who actively participated in the discussion forum (i.e., created a high number of posts) were more likely to complete a course. However, predicting knowledge gain or exam score, yielded somewhat different results. Specifically, Konstan and colleagues (2015) showed that the number of messages posted to a discussion forum was not significantly associated with an increase in knowledge gain. Similar findings were noted by (X. Wang et al., 2015), who showed there was no association between forum participation and knowledge gain. Finally, Vu and colleagues (2015) also showed that the overall activity in discussion forums did not predict the number of quiz submissions nor submission scores. As explained by Vu and colleagues (2015), the relationship between the number of posts and assessment grade seemed to be one-directional. That is, higher grades predicted the number of posts, but the number of posts did not necessarily predict the grade.

A substantial number of studies that measured various forms of student engagement also observed to what extent interaction with course assessment (17.6%) (e.g., the number of total assignment submissions, count of correct quiz attempts) predicted learning outcome or retention. In general, studies showed a significant and positive association between assignment and/or quiz interaction and successful course completion (Brooks, Thompson, et al., 2015; Konstan et al., 2015; Sharma et al., 2015; Ye et al., 2015). Nevertheless, Kennedy and colleagues (2015)

revealed somewhat contradictory results, failing to demonstrate the association between the number of submitted assignments and course performance (i.e., final course grade).

To evaluate the quality of student generated discourse and examine the association between student cognitive behavior and learning, researchers mainly relied on content analysis methods to identify underlying cognitive processes. For example, analyzing cognitively relevant behaviors in discussion forum messages using Chi's ICAP framework (Chi, 2009), Wang and colleagues (2015) showed that active and constructive cognitive processes could predict learning gains. On the other hand, Yang et al. (2015) demonstrated the importance of resolving confusion in the discussion forum in order to reduce student dropout. However, in detecting different confusion states, Yang and colleagues (2015) relied on psychologically meaningful categories of words, extracted from online discussions using the Linguistic Inquiry and Word Count (LIWC) tool (Tausczik & Pennebaker, 2010), as one of the classification features. Whereas, Authors (2015a), as well as Authors (2015b), exemplified how linguistic indices of text narrativity, cohesion and syntax simplicity extracted from online discussion transcripts predict learning outcome and social positioning in various contexts.

Similar to studying cognitive processes, researchers primarily relied on content analysis methods when studying affect in MOOCs, and the association between affect and course persistence or outcome. Thus, Tucker and colleagues (2014) revealed a strong negative correlation between student sentiment expressed in the discussion forum and average assignment grade. Whereas, this correlation was low and positive between student sentiment and quiz grades. Tucker and colleagues (2014) relied on a word-sentiment lexicon (Taboada, Brooke, Tofiloski, Voll, & Stede, 2011), and Adamopoulos (2013) used AlchemyAPI to extract student sentiment from discussion forum messages. Adamopoulos (2013) further showed that student

sentiment towards course instructor, assignments, and course materials have a positive effect on the course retention. Yang and colleagues (2015) on the other hand, highlighted the importance of resolving confusion (expressed in student forum posts) in order to increase retention. However, in order to detect confusion from student contribution to the discussion forum, Yang and colleagues (2015) relied on LIWC features (among others) and word categories that depict student affective processes, including positive and negative emotions.

Through the analysis of the results related to our second research question, we were able to observe a large diversity of metrics used to understand learning and predict student persistence and/or course outcome. Given a large scale and various sources of data, it seems that the first generation of MOOC research (Reich, 2015) primarily focused on understanding "what works" in this new settings, in terms of supporting learning activities and increasing retention. However, another reason for such diversity of metrics used (Table S3 – please refer to the supplementary material) presumably lies in the fact that there is no single commonly accepted analytical method or framework that would allow for studying learning in non-formal, digital educational settings. Failing to provide a common interpretation of observed variables used to understand learning can potentially lead towards limited generalization and low interpretability of results.

Table S3 (please refer to the supplementary material) provides a complete list of metrics, extracted from the studies included in this systematic review, used to model learning in non-formal learning settings. In the following text (Section 5 primarily), we also provided a rationale for conceptualizing learning in MOOCs and definition of the constructs that comprise the adopted model of the association between context, engagement, and proximal learning outcome.

Following the model originally proposed by Reschly and Christenson (2012), we argue that studying learning at scale should observe four engagement types – behavioral, academic,

cognitive, and affective engagement (Figure 5). However, as further discussed in the section "Conceptualizing Learning in MOOCs", we propose different conceptualizations of each type of engagement. Our conceptualizations are tailored towards the specific nature of learning with MOOCs and characteristics of data collected about students' learning. A theoretically-grounded overarching framework of engagement is necessary given the increased interest in MOOCs, and the growing number of metrics used to understand students' behavior and outcomes. A multi-faceted framework of engagement such as the one proposed in this paper provides an infrastructure for researchers to compare and contrast different dimensions of engagement in MOOCs, which can lead to a greater scientific understanding of learning at scale. Indeed, the proposed framework encompasses a vast majority of the commonly used metrics in MOOC research, and can serve as a theoretical basis for future work on the topic.

Association between metrics identified and measured outcome (RQ3)

Systematic literature review of the existing MOOC research also showed differences in statistical approaches to studying the association between engagement metrics and learning outcomes in MOOCs (Table 1), in addition to the inter-study variability in outcomes (Section 4.1) and predictors (Section 4.2). A majority (34.21%) of the included papers used a machine learning approach (e.g., classification using random forest or J48 algorithms). Correlation, chi-square test, regression, ANOVA or MANOVA, social network analysis (SNA), survival analysis, and mixed-effects regression were reported much less often. Five additional papers used statistical methods that occurred less than three times total and thus were classified as "other". These statistical tests included t-test (n = 2), relational event modeling (n = 1), discrete choice model (i.e., random utility model or latent regression model; n = 1), or a structural equation model (SEM; n = 1).

A few insights can be gleaned from Table 1. The most common analysis method adopted was machine learning techniques. Of the papers that used machine learning approaches, only 38% of the 13 also reported another statistical method. The usage of machine learning suggests that a common goal among the papers was to build predictive models (versus explanatory models). Indeed, the goal of predicting students' success in MOOCs is a highly relevant goal for incorporating interventions. It is also important to point out that correlational and regression techniques were also commonly used (36% combined). This may suggest that another important goal among these papers was to not only build predictive models but also explain variance in the dependent variable(s) of interest. Taken together, the statistical methods were quite diverse, perhaps targeting different theoretical or more applied goals.

Conceptualizing Learning in MOOCs

This systematic review of the MOOC research literature involved two related aims. The first aim focused on the development of a summary of the metrics commonly used to measure and model learning in non-formal educational settings. The second aim was to extend these findings and establish a conceptual model that would distinguish between the factors impacting students' learning in a MOOC context. Building on Reschly and Christenson (2012) model of the associations between context, engagement, and student outcomes, we further redefined and re-operationalized these constructs (i.e., context, engagement, and outcome) for research on MOOCs. In so doing, we relied on the insights obtained from the systematic literature review to understand how a diverse set of learning-related constructs is measured in MOOCs, and how these constructs could be linked to an existing model of learning previously validated in educational research. Providing such a model would offer a possibility to compare factors that shape learning in non-formal, digital educational settings with formal (e.g., traditional face-to-

face or online) formats of learning. Specifically, such a model would enable studying whether, and to what extent, factors that contribute to learning differ across various educational settings (e.g., face-to-face; online and MOOCs). Finally, given that a majority of studies in this review, and in MOOC research in general according to Reich (2015), observe (a) certain form(s) of students' engagement in predicting course outcome and/or persistence, it seems reasonable to provide a re-operationalization of this particular concept for the study of MOOCs.

In the context of MOOCs, our systematic review indicated a mainly exploratory nature of the existing research that attempts to investigate the association between various forms of student engagement (or behavior) and learning – defined through learning outcomes or course persistence. In so doing, researchers often failed to account adequately for existing educational frameworks that would allow for more salient interpretations of the results. Even when relying on existing learning theories, researchers generally did not account for a different learning context or a greater diversity of students observed in open non-formal educational context (as compared to online or face-to-face settings).

Following the intention to provide coherence into the diverse analyses of learning-related constructs in MOOCs (Section 4), we framed our inquiry around Reschly and Christenson's (2012) work on dropout prevention and enhancing learning in traditional classroom settings. Similar to Reschly and Christenson (2012), we recognize engagement as a two-fold construct – both a process and an outcome – that mediates the association between a context (e.g., student intent, classroom settings) and a relevant learning outcome (Figure 5). Moreover, we also posit that the student engagement in MOOCs has a mediating role between contextual factors and desired learning outcomes. However, our literature review highlights some of the shortcomings of Reschly and Christenson's original model when applied to MOOCs. For instance, Reschly

and Christenson's model is designed to address systems where children obtain literacies and content while they undergo developmental processes. In that sense, Reschly and Christensen's range of contextual variables is guided towards this particular context, especially in relation to such aspects as learner agency, learner intent and prior knowledge. In a similar manner, Reschly and Christenson's notion of outcomes is not suited to learner-driven learning process where a learner has both power over deciding to which end to engage with the learning activities as well as when to disengage. Learning outcomes in Reschly and Christensen's model address the role of secondary education and how it prepares for future life, whereas in the learning outcomes may have another level of granularity. Finally, although engagement may be defined similarly in both digital and face-to-face settings, the ways of gleaning information about it in digital environments and at scale, require re-operationalization. Therefore, following insights obtained from the systematic literature review, we propose a novel engagement model applicable in the context of learning with MOOCs that considerably differs from Reschly and Christenson's, primarily in the way the model constructs have been operationalized.

The first and foremost difference to Reschly and Christenson's model is the conceptualization of each of the components of the model proposed in this paper (Figure S3 – please refer to the supplementary material). Specifically, whereas the original model observes family, peers, school, and community as main contextual determinant, in the MOOC research, we defined contextual factors as being comprised of: i) demographic information – such as age, gender, or level of education (Goldberg et al., 2015; Heutte, et al., 2014), ii) classroom structure – e.g., course platform, course characteristics (Adamopoulos 2013), and iii) individual needs – e.g., students' intentions (Brooks, Stalburg, et al., 2015; Kizilcec & Halawa, 2015).

Despite an extensive body of research on student engagement in various educational settings, and prevailing understanding of its importance, there is no clear consensus what comprises engagement (Christenson, Reschly, & Wylie, 2012). As noted in the Christenson et al. (2012) review, researchers most commonly refer to two subtypes (i.e., participatory and affective) or include a cognitive engagement as a third subtype. However, there are notable differences in how various subtypes of engagement have been operationalized in a traditional educational context. Thus, the lack of agreement on how engagement has been defined and operationalized in MOOCs (see Section 4.2) perhaps comes as no surprise. Nevertheless, we posit that an attempt to establish a common understanding of how engagement is measured and interpreted in the context of learning in non-formal, digital educational settings is a necessary step towards better understanding learning in this particular context.

Although Reschly and Christenson (2012) observed engagement in traditional learning settings, the theoretical and practical stances considered in conceptualizing the engagement model, seem to align with the general understanding of what important factors of learning in MOOCs are. Specifically, a multidimensional nature of variables observed when assessing learning in non-formal educational settings (Table S3 in the supplementary material) supports the necessity to have multidimensional constructs that include different types of learner activity (e.g., Konstan et al., 2015; Sinha & Cassell, 2015), emotions (e.g., Crossley et al., 2015; Yang et al., 2015), or cognition (Dowell et al., 2015; X. Wang et al., 2015). Finally, similar to Kizilcec and Halawa (2015), Brooks and colleagues (2015), and Reschly and Christenson (2012) argue for the importance of considering a specific learning context (e.g., peers or school) and student agency. In spite of some similarities, operationalizing student agency in Reschly and Christenson's (2012) model is somewhat different from what has been considered in MOOC research included

in this study. Reschly and Christenson (2012) draw on the assumption that "students are able to report accurately on their engagement and environments" (p.9, *ibid.*). Although we agree that "student perspective is essential for change in student learning and behavior" (Reschly & Christenson, 2012, p. 9), we further aim at extracting a majority of evidence of student engagement from the data stored within learning platforms used to deliver courses at scale.

Reschly & Christenson's model was designed to analyze formal educational settings. Thus, we further review the consistency of their model's categories in relation to the metrics observed in MOOC studies. First, we find that academic engagement in MOOCs aligns with Appleton, Christenson, Kim, and Reschly (2006) and Reschly and Christenson's (2012) work, and refers to time spent on course activities (e.g., viewing pages, engaging with quizzes and assignments), number of days (weeks, hours) being engaged with a course, assessment (e.g., homework, and quiz), completion rate and accuracy, credit towards course completion, and pre- and/or post-test results (e.g., Boyer & Veeramachaneni, 2015; Li et al., 2015).

Second, our view of behavioral engagement aligns with the original model of engagement (Reschly & Christenson, 2012). A common definition of behavioral engagement "draws on the idea of participation; it includes involvement in academic and social or extracurricular activities and is considered crucial for achieving positive academic outcomes and preventing dropping out" (Fredricks, Blumenfeld, & Paris, 2004, p. 60). For MOOCs, this form of engagement can still be defined through participation in discussion forums, viewing lectures, following course activities, or number of times student accessed course wiki pages (e.g., Li et al., 2015; Santos et al., 2014; Sinha & Cassell, 2015).

Third, cognitive engagement usually refers to students' motivational goals and self-regulated learning skills (Christenson et al., 2012; Fredricks et al., 2004; Reschly & Christenson,

2012). In the context of learning with MOOCs, thus far research has primarily focused on linguistic indicators (e.g., text narrativity or cohesion) of student cognitive engagement, obtained from learner generated artefacts (Authors, 2015a; Authors, 2015b; X. Wang et al., 2015). The rationale behind this subtype of engagement is grounded in the premise that learning and understanding in computer-mediated learning are primarily expressed through the artefacts students generate in the learning process (Goodyear, 2002; Jones, 2008). Thus, studying learning in MOOCs should account for the quality of discourse, as a proxy for students' cognitive engagement.

Fourth, Reschly and Christenson's (2012) model of engagement considers students' affective reactions in the classroom, school identification, valuing learning, and sense of belonging as factors that characterize affective engagement. However, drawing on the premise that language represents a primary means of communication in computer-mediated interactions, as well as the lack of social cues that characterize learning in non-formal, digital educational settings, MOOC research primarily relies on linguistic indices in assessing affective engagement (e.g., positive or negative emotions) in MOOCs (e.g., Adamopoulos, 2013; Tucker et al., 2014). Nevertheless, there has been significant work done recently in assessing student emotions and affect using certain (arguably) more advanced approaches (e.g., Baker, D'Mello, Rodrigo, & Graesser, 2010; D'Mello, Dowell, & Graesser, 2009).

Finally, failing to account for contextual determinants of learning in general (Appleton et al., 2006) or the contextual factors for online and distance education in particular (Gašević et al., 2016; Authors, 2016) could lead towards misinterpretations of the association between engagement and learning, providing an intervention that might not result with an intended outcome. In defining contextual variables, our understanding of factors that frame learning in

MOOCs is defined through demographic data about course participants, classroom settings (e.g., peers and course design), and student individual needs (e.g., intent to complete and interest in topic) (Adamopoulos, 2013; Brooks, Stalburg, et al., 2015; Kizilcec & Halawa, 2015).

Course-level learning outcomes are the most commonly assessed in current MOOC research. They are also further developed as they reach beyond the focus on academic achievement, and include social and affective aspects. Thus, knowledge mastery as the outcome is measured through graded assessment. Alternative metrics are also employed, such as capturing knowledge or skill change. Course-level learning outcomes within the social aspect are limited to engagement with others, rather than the measures of quality of the knowledge construction within the dialogue, or capture of the increased sense of belonging or identity formation. Affective course-level outcomes are limited to course satisfaction only. In contrast, Reschly and Christenson's model defined affective learning outcomes as self-awareness of feelings, emotional regulation, and conflict resolution skills.

Both intermediate and post-course outcomes are not of the main focus in current MOOC research. This is too constraining as such kinds of outcomes seem to be common in non-formal and open settings. For instance, intermediate learning outcomes are of relevance to the vast numbers of just-in-time learners sampling parts of the content. Current approaches to the identification of immediate learning outcomes in MOOC research is limited to academic performance, as the majority of metrics is focused on either predicting module outcomes, or detecting when a student stops engaging with the course. Reschly & Christenson's model, however, argues that engagement can be seen both as the process, as well as the outcome. Thus, it could be hypothesized that engagement metrics could serve as indicators of an intermediate learning outcome for those learners not interested in course completion.

When it comes to post-course outcomes, exemplified as employability and productive citizenry in the original model, they have not been the subject of much MOOC research, with the exception of the focus on employability (E. Y. Wang & Baker, 2015). Again, the lack of focus beyond assessment is limiting, as better measures of post-course outcomes could enrich stakeholders' understanding of the wider impact of MOOCs, and finally evaluate the value of producing MOOCs.

Conclusions

MOOC research has demonstrated significant advances in a relatively short time frame (Raffaghelli et al., 2015; Reich, 2015). Nevertheless, contemporary research in MOOCs almost unequivocally argues for the lack of generalizability of existing results, and for failing to investigate factors that contribute to learning in non-formal, educational settings (DeBoer et al., 2014; Evans et al., 2016). To advance the field of research in non-formal, digital educational settings, there is an imperative to shift the focus from observational studies and introduce more experimental research approaches across different domains and course designs (Reich, 2015). Moreover, we agree with Reich's (2015) assumption that future MOOC research should build on the existing research frameworks, evaluated across educational contexts, in order to provide a basis for comparison between learning in MOOCs and other (more traditional) settings.

Our contribution to the development of the next generation research in non-formal, digital educational settings is twofold. First, we conducted a systematic literature review of the existing body of research in MOOCs that tries to model learning in this particular setting. We were able to identify a wide range of metrics used to predict learning and measure student engagement, across various contexts (e.g., centralized within a single platform, or distributed, using various social media). Nevertheless, usually referred to as a *discussion behavior* (Wang et

al, 2015), *behavior* (Ramesh et al., 2014a, Ramesh et al., 2014b), or *engagement* (Santos et al, 2014, Sinha and Cassell, 2015, Tucker et al., 2014), various researchers tended to observe engagement-related metrics from a single perspective operationalized through students' participation in different activities. Specifically, researchers tend to measure engagement as a form of participation in discussion forums (quantity of contribution) (Vu et al., 2015; X. Wang et al., 2015), watching video lectures (Li et al., 2015), or participating in course assessment activities (Whitehill et al., 2015; Ye et al., 2015). It is also noticeable that the definition of a course outcome is dominated by the formal education mindset for the majority of studies included in this review (Appleton et al., 2006). Regardless of the fact that various researchers have argued for the importance of aligning learning outcomes with students' intentions and interest in completing a course, only a few studies (e.g., Authors, 2015a; Authors, 2015b) made a considerable effort towards the operationalization of social or affective learning outcome (Figure 5).

The second part of our contribution is framed around the redefinition of the existing educational framework in order to account for specific aspects of learning in MOOCs. Specifically, following Reschly and Christenson's (2012) research, we proposed a model for studying the association between context, student engagement and learning outcome (Figure 5). We further suggest that engagement in MOOCs, and learning at scale in general, should be observed as a multi-dimensional construct, comprised of academic, behavioral, cognitive, and affective engagement. Such a definition should bring coherence into MOOC research, providing a common understanding what engagement actually is and how it should be measured in this complex learning context, which seems to lack in the existing studies. We also provided a list of metrics used to operationalize elements of the proposed model (Table S2). However, by no

means, we argue that this is a complete list of metrics used to measure learning (or engagement) in MOOCs.

We contend that for advancing the MOOC research and allowing for comparisons with different (more traditional) forms of education, researchers should align metrics used for assessing learning with the proposed model. Having a generally accepted conceptualization of engagement would allow for obtaining more comprehensive insights into the factors that influence learning with MOOCs as well as how these factors could be generalized across different platforms or compared with diverse context (such as traditional online or face to face learning) (DeBoer et al., 2014). Such a conceptualization would also allow for moving beyond observing student "click data" and exploring how quantity and quality of interactions with the course content, peers, and teaching staff could predict course outcome and persistence, thus providing more salient connection with existing learning theories and practices (Dawson, Mirriahi, & Gasevic, 2015; Gašević et al., 2016; Wise & Shaffer, 2015). Nevertheless, we also acknowledge the lack of metrics in some aspects of the model – i.e., social and affective learning outcomes – that require further conceptualization in the context of learning at scale. Recent advances in the (multimodal) learning analytics research field provide a promising venue for investigation of students' cognition, metacognition, emotion, and motivation using multimodal data, such as eye gaze behaviors, facial expressions of emotions, heart rate and electro-dermal activity, to name a few (Azevedo, 2015; D'Mello, Dieterle, & Duckworth, 2017; Molenaar & Chiu, 2015). Moreover, conducting a systematic literature review of qualitative research conducted in the field would provide complementary insights into the findings introduced here. Being designed to help understanding the process of learning in rich detail, qualitative studies

could potentially provide thick description of the various aspects of social and affective engagement to accompany findings obtained from quantitative research.

Our future research will examine the hypothesized association between context, student engagement and learning outcome. Thus, the proposed model (Figure 5) assumes a mediating effect of student engagement between contextual variables and desired outcome, which is in line with the original model proposed by Reschly and Christenson (2012). Reschly and Christenson (2012) also observed affective and cognitive engagement as mediating factors for the development of behavioral and academic engagement (as indicated with arrows from cognitive and affective to academic and behavioral engagement). However, given the proposed operationalization, this association may not hold in our proposed model. It seems reasonable to expect that direction of the mediating effect would be from behavioral towards cognitive and affective engagement. This assumption is simply due to the fact that in order to reveal traces of cognitive and affective engagement (as currently operationalized) students should first engage with course material and peer learners (i.e., reveal traces of behavioral engagement). Nevertheless, in order to examine those assumptions, we aim to create a statistical model(s) that would allow us to determine the validity of the hypothesized relations.

The original model, as proposed by Reschly and Christenson (2012), also assumes the *Matthew Effect* (Ceci & Papierno, 2005) between the contextual factors and engagement "wherein as students are engaged, contexts provide feedback and support that promote ever greater engagement" (Reschly & Christenson, 2012, p. 9), as indicated with the arrows pointing from context to engagement and vice versa). We posit that in the context of learning at scale, and MOOCs in particular, this association would still hold. Such an implication could be inferred from the existing research on self-regulated learning. Specifically, Winne and Hadwin (1998)

model of self-regulated learning posits that *conditions* (i.e., learning experiences, domain knowledge, motivation, intents), operationalized here through the contextual variables, influence both "standards as well as the actual operations a person performs" (Greene & Azevedo, 2007, p. 336). Through *cognitive evaluation*, students compare *products* and *operations* (here operationalized through the four engagement types) to determine whether a learning goal has been achieved or further adjustments to the cognitive conditions should be applied, completing thus a recursive model of self-regulated learning (Greene & Azevedo, 2007; Winne & Hadwin, 1998).

Acknowledgements

The authors would like to thank the anonymous reviewers for their helpful and constructive comments that greatly contributed to improving the final version of the paper. They would also like to thank the Editor for the generous comments and support during the review process.

References

Authors (2015a)

Authors (2015b)

Authors (2016)

Authors (2015c)

Adamopoulos, P. (2013). What makes a great MOOC? An interdisciplinary analysis of student retention in online courses. In *International Conference on Information Systems (ICIS 2013): Reshaping Society Through Information Systems Design* (Vol. 5, pp. 4720–4740), Retrieved from http://aisel.aisnet.org/icis2013/proceedings/BreakthroughIdeas/13/

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the student engagement instrument. *Journal of School Psychology*, *44*(5), 427–445, doi:10.1016/j.jsp.2006.04.002

Azevedo, R. (2015). Defining and Measuring Engagement and Learning in Science: Conceptual, Theoretical, Methodological, and Analytical Issues. *Educational Psychologist, 50*(1), 84–94. https://doi.org/10.1080/00461520.2015.1004069

Baker, R. S. J. d, D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, *68*(4), 223–241. doi:10.1016/j.ijhcs.2009.12.003

Bergner, Y., Kerr, D., & Pritchard, D. E. (2015). Methodological challenges in the analysis of MOOC Data for exploring the relationship between discussion forum views and learning outcomes. In *Proceedings of the 8th International Conference on Educational Data Mining, Madrid, Spain* (234-241) Retrieved from http://www.educationaldatamining.org/EDM2015/uploads/papers/paper_61.pdf

Boyer, S., & Veeramachaneni, K. (2015). Transfer learning for predictive models in Massive Open Online Courses. In C. Conati, N. Heffernan, A. Mitrovic, & M. F. Verdejo (Eds.), *Artificial Intelligence in Education* (pp. 54–63). Springer International Publishing. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-19773-9_6

Brooks, C., Stalburg, C., Dillahunt, T., & Robert, L. (2015). Learn with friends: The effects of student face-to-face collaborations on massive open online course activities. In

Proceedings of the Second (2015) ACM Conference on Learnin g@ Scale(L@S 2015), 241–244. doi:10.1145/2724660.2728667

Brooks, C., Thompson, C., & Teasley, S. (2015). Who you are or what you do: Comparing the predictive power of demographics vs. activity patterns in Massive Open Online Courses (MOOCs). In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale,* 245–248. doi:10.1145/2724660.2728668

Ceci, S. J., & Papierno, P. B. (2005). The rhetoric and reality of gap closing: when the“ have-nots” gain but the“ haves” gain even more. *American Psychologist, 60*(2), 149.

Champaign, J., Colvin, K. F., Liu, A., Fredericks, C., Seaton, D., & Pritchard, D. E. (2014). Correlating skill and improvement in 2 MOOCs with a student’s time on tasks. In Proceedings of *the First (2014) ACM Conference on Learning at Scale (L@S 2014),* 11–20. doi:10.1145/2556325.2566250.

Chi, M. T. H. (2009). Active-Constructive-Interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73–105. doi:10.1111/j.1756-8765.2008.01005.x

Christenson, S. L., Reschly, A. L., & Wylie, C. (2012). *Handbook of research on student engagement*. Springer: New York.

Coffrin, C., Corrin, L., de Barba, P., & Kennedy, G. (2014). Visualizing patterns of student engagement and performance in MOOCs. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge (LAK’14),* 83–92. doi:10.1145/2567574.2567586

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37–46.

Crossley, S., Danielle, S., Baker, R., Wang, Y., Paquette, L., Barnes, T., & Bergner, Y. (2015). Language to completion: success in an educational data mining Massive Open Online Class. In *Proceedings of the 8th International Conference on Educational Data Mining (EDM'15)*, 388–391. Retrieved from http://www.educationaldatamining.org/EDM2015/uploads/papers/paper_153.pdf.

Dawson, S., Mirriahi, N., & Gasevic, D. (2015). Importance of Theory in Learning Analytics in Formal and Workplace Settings. Journal of Learning Analytics, 2(2), 1–4.

DeBoer, J., Ho, A. D., Stump, G. S., & Breslow, L. (2014). Changing "Course": Reconceptualizing educational variables for Massive Open Online Courses, *Educational Researcher, 43*(2), 74–84. doi:10.3102/0013189X14523038

D'Mello, S., Dieterle, E., & Duckworth, A. (2017). Advanced, Analytic, Automated (AAA) Measurement of Engagement During Learning. *Educational Psychologist, 52*(2), 104–123. https://doi.org/10.1080/00461520.2017.1281747

D'Mello, S., Dowell, N., & Graesser, A. (2009). Cohesion relationships in tutorial dialogue as predictors of affective states. In Proceedings of *the 2009 Conference on Artificial Intelligence in Education: Building Learning Systems That Care: From Knowledge Representation to Affective Modelling* (pp. 9–16). Retrieved from http://dl.acm.org/citation.cfm?id=1659457.

Engle, D., Mankoff, C., & Carbrey, J. (2015). Coursera's introductory human physiology course: Factors that characterize successful completion of a MOOC. *The International Review of*

Research in Open and Distributed Learning, *16*(2), 46-68. doi:10.19173/irrodl.v16i2.2010

Eraut, M. (2000). Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology*, *70*(1), (pp. 113–136), doi:10.1348/000709900158001

Evans, B. J., Baker, R. B., & Dee, T. S. (2016). Persistence patterns in Massive Open Online Courses (MOOCs). *The Journal of Higher Education*, *87*(2), 206–242. doi: 10.1353/jhe.2016.0006

Eynon, R. (2014). Conceptualising Interaction and Learning in MOOCs. MOOC Research Initiative Final Report [Formato PDF] Http://www. Moocresearch. com/wpcontent/uploads/2014/06/C9146_EYNON_MOOC-ResearchInitiativeEynon9146Final. Pdf (Consultado 30 de Abril de 2014).

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109. doi: 10.3102/00346543074001059

Gašević, D., Dawson, S., Rogers, T., & Gašević, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68–84. doi:10.1016/j.iheduc.2015.10.002

Gillani, N., & Eynon, R. (2014). Communication patterns in massively open online courses. *The Internet and Higher Education*, *23*, 18–26. doi:10.1016/j.iheduc.2014.05.004

Glass, C. R., Shiokawa-Baklan, M. S., & Saltarelli, A. J. (2016). Who Takes MOOCs? *New Directions for Institutional Research*, *2015*(167), 41–55. doi:10.1002/ir.20153

Goldberg, L. R., Bell, E., King, C., O'Mara, C., McInerney, F., Robinson, A., & Vickers, J. (2015). Relationship between participants' level of education and engagement in their completion of the Understanding Dementia Massive Open Online Course. *BMC Medical Education*, *15*(1), 1–7. doi:10.1186/s12909-015-0344-z

Goodyear, P. (2002). Psychological foundations for networked learning. In C. Steeples & C. Jones (Eds.), *Networked Learning: Perspectives and Issues* (pp. 49–75). New York, NY, USA: Springer-Verlag New York, Inc. doi: 10.1007/978-1-4471-0181-9_4

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. https://doi.org/10.3102/003465430303953

Greene, J. A., Oswald, C. A., & Pomerantz, J. (2015). Predictors of retention and achievement in a Massive Open Online Course. *American Educational Research Journal*, *52*(5), 1-31. doi:10.3102/0002831215584621

Halatchliyski, I., Moskaliuk, J., Kimmerle, J., & Cress, U. (2014). Explaining authors' contribution to pivotal artifacts during mass collaboration in the Wikipedia's knowledge base. *International Journal of Computer-Supported Collaborative Learning, 9*(1), 97–115. doi:10.1007/s11412-013-9182-3

Hartnett, M., George, A. S., & Dron, J. (2011). Examining motivation in online distance learning environments: Complex, multifaceted and situation-dependent. *The International Review of Research in Open and Distributed Learning*, *12*(6), 20–38. Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/1030/1954

Haywood, J., Aleven, V., Kay, J., & Roll, I. (2016). L@S ’16: *Proceedings of the Third ACM Conference on Learning @ Scale*. New York, NY, USA: ACM.

Heutte, J., Kaplan, J., Fenouillet, F., Caron, P.-A., & Rosselle, M. (2014). MOOC user persistence. In L. Uden, J. Sinclair, Y.-H. Tao, & D. Liberona (Eds.), *Learning Technology for Education in Cloud. MOOC and Big Data* (pp. 13–24). Springer International Publishing. doi: 10.1007/978-3-319-10671-7_2

Illeris, K. (2007). *How we learn: Learning and non-learning in school and beyond*. Routledge: Oxford, UK.

Jiang, S., Fitzhugh, S. M., & Warschauer, M. (2014). Social positioning and performance in MOOCs. In *Proceedings of the Workshops held at Educational Data Mining 2014, co-located with 7th International Conference on Educational Data Mining (EDM 2014)* (Vol. 1183, p. 14). Retrieved from http://ceur-ws.org/Vol-1183/gedm_paper08.pdf.

Jiang, S., Warschauer, M., Williams, A. E., O’Dowd, D., & Schenke, K. (2014). Predicting MOOC performance with week 1 behavior. In Proceedings of *the 7th International Conference on Educational Data Mining (EDM’14)*, 273–275. Retrieved from http://educationaldatamining.org/EDM2014/uploads/procs2014/short papers/273_EDM-2014-Short.pdf

Jones, C. (2008). Networked learning-a social practice perspective. In *Proceedings of the 6th International Conference on Networked Learning* (pp. 616–623). Retrieved form http://www.lancaster.ac.uk/fss/organisations/netlc/past/nlc2008/abstracts/PDFs/Jones_616-623.pdf

Jordan, K. (2015a). Massive Open Online Course Completion rates revisited: Assessment, length and attrition. *International Review of Research in Open & Distance Learning*, *16*(3), (pp. 341–358). Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/2112

Jordan, K. (2015b, February 13). Synthesising MOOC completion rates | MoocMoocher. [Web log post]. Retrieved from https://moocmoocher.wordpress.com/2013/02/13/synthesising-mooc-completion-rates/

Kane, M. (1992). An argument-based approach to validation. *Psychological Bulletin*, *112*(3), 527–535. doi:10.1037/0033-2909.112.3.527

Kane, M. (2006). Validation. In R. Brennan (Ed.), *Educational measurement, 4th ed.* (pp. 17–64), Westport, CT: American Council on Education and Praeger.

Kennedy, G., Coffrin, C., de Barba, P., & Corrin, L. (2015). Predicting Success: How Learners' Prior Knowledge, Skills and Activities Predict MOOC Performance. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge (LAK'15),* 136–140. doi:10.1145/2723576.2723593

Kizilcec, R. F., & Halawa, S. (2015). Attrition and achievement gaps in online learning. In *Proceedings of the Second ACM Conference on Learning @ Scale (L@S'15)*, 57–66. doi:10.1145/2724660.2724680

Kizilcec, R. F., & Schneider, E. (2015). Motivation as a lens to understand online learners: Toward data-driven design with the OLEI scale. *ACM Trans. Computer-Human Interaction*, *22*(2), 6:1–6:24. doi:10.1145/2699735

Koedinger, K. R., Kim, J., Jia, J. Z., McLaughlin, E. A., & Bier, N. L. (2015). Learning is Not a Spectator Sport: Doing is Better Than Watching for Learning from a MOOC. In

Proceedings of *the Second ACM Conference on Learning @ Scale,* 111–120. doi:10.1145/2724660.2724681

Konstan, J. A., Walker, J. D., Brooks, D. C., Brown, K., & Ekstrand, M. D. (2015). Teaching recommender systems at large scale: Evaluation and lessons learned from a hybrid MOOC. *ACM Trasnactions on Human-Computer Interaction (TOCHI)*, *22*(2), 61-70. doi:10.1145/2728171

Li, N., Kidziński, L., Jermann, P., & Dillenbourg, P. (2015). MOOC video interaction patterns: What do they tell us? *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*, Vol. 9307), 197–210. doi:10.1007/978-3-319-24258-3_15

Loya, A., Gopal, A., Shukla, I., Jermann, P., & Tormey, R. (2015). Conscientious behaviour, flexibility and learning in Massive Open On-Line Courses. *Procedia - Social and Behavioral Sciences*, *191*, 519–525. doi:10.1016/j.sbspro.2015.04.686.

Molenaar, I., & Chiu, M. M. (2015). Effects of Sequences of Socially Regulated Learning on Group Performance. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (pp. 236–240). New York, NY, USA: ACM. https://doi.org/10.1145/2723576.2723586

Moss, P. A., Girard, B. J., & Haniford, L. C. (2006). Chapter 4: Validity in Educational Assessment. *Review of Research in Education*, *30*(1), 109–162. doi:10.3102/0091732X030001109

Raffaghelli, J. E., Cucchiara, S., & Persico, D. (2015). Methodological approaches in MOOC research: Retracing the myth of Proteus. *British Journal of Educational Technology*, *46*(3), 488–509. doi:10.1111/bjet.12279

Ramesh, A., Goldwasser, D., Huang, B., Daume, H. I., & Getoor, L. (2014a). Learning latent engagement patterns of students in online courses. In *Proceedings of the National Conference on Artificial Intelligence* Vol. 2 (pp. 1272–1278). Retrieved from http://www.aaai.org/ocs/index.php/AAAI/AAAI14/paper/view/8571

Ramesh, A., Goldwasser, D., Huang, B., Daume, H., III, & Getoor, L. (2014b). Uncovering hidden engagement patterns for predicting learner performance in MOOCs. In Proceedings of *the First ACM Conference on Learning @ Scale Conference (L@S'14),* 157–158. doi:10.1145/2556325.2567857

Reich, J. (2015). Rebooting MOOC Research. *Science*, *347*(6217), 34–35. doi: 10.1126/science.1261627

Reich, J., Stewart, B., Mavon, K., & Tingley, D. (2016). The Civic Mission of MOOCs: Measuring engagement across political differences in forums. In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale (L@S'16)*, 1–10. doi:10.1145/2876034.2876045

Reschly, A. L., & Christenson, S. L. (2012). Jingle, Jjngle, and conceptual haziness: Evolution and future directions of the engagement construct. In L. S. Christenson, L. A. Reschly, & C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 3–19). Boston, MA: Springer US. doi:10.1007/978-1-4614-2018-7_1

Rodriguez, C. O. (2012). MOOCs and the AI-Stanford Like Courses: Two Successful and Distinct Course Formats for Massive Open Online Courses. *European Journal of Open, Distance and E-Learning*. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ982976&site=ehost-live

Santos, J. L., Klerkx, J., Duval, E., Gago, D., & Rodríguez, L. (2014). Success, activity and drop-outs in MOOCs an exploratory study on the UNED COMA courses. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge (LAK'14),* 98–102. doi:10.1145/2567574.2567627

Selwyn, N., Bulfin, S., & Pangrazio, L. (2015). Massive open online change? Exploring the discursive construction of the "MOOC"in newspapers. *Higher Education Quarterly, 69*(2), 175–192.

Shah, D. (2015b, December 28). MOOCs in 2015: Breaking Down the Numbers [Web log post on EdSurge News]. Retrieved August 19, 2016, from https://www.edsurge.com/news/2015-12-28-moocs-in-2015-breaking-down-the-numbers.

Sharma, K., Jermann, P., & Dillenbourg, P. (2015). Identifying styles and path towards success in MOOCs. In *Proceedings of the 8th International Conference on Educational Data Mining (EDM'15),* 408–411. Retrieved from http://www.educationaldatamining.org/EDM2015/uploads/papers/paper_88.pdf

Shirky, C. (2013). MOOCs and Economic Reality. *Chronicle of Higher Education, 59*(42), B2.

Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning, 2*(1), 3–10.

Siemens, G., Gašević, D., & Dawson, S. (2015). *Preparing for the digital university: A review of the history and current state of distance, blended, and online learning* (Research report). *Athabasca, Canada: Athabasca University*. Retrieved from http://linkresearchlab.org/PreparingDigitalUniversity.pdf

Sinha, T., & Cassell, J. (2015). Connecting the dots: Predicting student grade sequences from bursty MOOC interactions over time. In Proceedings of *the Second (2015) ACM Conference on Learning @ Scale (L@S'15),* 249–252. doi:10.1145/2724660.2728669

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, *37*(2), 267–307. doi:10.1162/COLI_a_00049

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(I), 24–54. doi:10.1177/0261927X09351676

Tucker, C., Pursel, B. K., & Divinsky, A. (2014). *Mining student-generated textual data in MOOCS and quantifying their effects on student performance and learning outcomes*. Paper presented at ASEE Annual Conference, Indianapolis, Indiana. Retrieved from https://peer.asee.org/22840

Vu, D., Pattison, P., & Robins, G. (2015). Relational event models for social learning in MOOCs. *Social Networks*, *43*, 121–135. doi:10.1016/j.socnet.2015.05.001

Walji, S., Deacon, A., Small, J., & Czerniewicz, L. (2016). Learning through engagement: MOOCs as an emergent form of provision. *Distance Education, 37*(2), 208–223. https://doi.org/10.1080/01587919.2016.1184400

Wang, E. Y., & Baker, R. (2015). Content or platform: Why do students complete MOOCs?. *Journal of Online Learning & Teaching*, *11*(1), 17–30. Retrieved from http://jolt.merlot.org/vol11no1/Wang_0315.pdf

Wang, X., Yang, D., Wen, M., Koedinger, K., & Rosé, C. P. (2015). Investigating how student's cognitive behavior in MOOC discussion forums affect learning gains. In Proceedings of *the 8th International Conference on Educational Data Mining (EDM'15)*, 226-233. Retrieved from http://www.educationaldatamining.org/EDM2015/uploads/papers/paper_89.pdf.

Wen, M., Yang, D., & Rose, C. (2014a). Sentiment analysis in MOOC discussion forums: What does it tell us? In Proceedings of *of the 7th International Conference on Educational Data Mining (EDM'14),* 130-137. Retrieved from http://educationaldatamining.org/EDM2014/uploads/procs2014/long%20papers/130_EDM-2014-Full.pdf

Wen, M., Yang, D., & Rose, C. P. (2014b). Linguistic reflections of student rngagement in Massive Open Online Courses. In Proceedings of the *Eighth International AAAI Conference on Weblogs and Social Media,* 525–534. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8057

Whitehill, J., Williams, J. J., Lopez, G., Coleman, C. A., & Reich, J. (2015). Beyond prediction: First steps toward automatic intervention in MOOC student stopout. In *Proceedings of the 8th International Conference on Educational Data Mining (EDM'15)*, 171-178. Retrieved from http://www.educationaldatamining.org/EDM2015/uploads/papers/paper_112.pdf

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. *Metacognition in Educational Theory and Practice, 93*, 27–30.

Wise, A. F., & Shaffer, D. W. (2015). Why Theory Matters More than Ever in the Age of Big Data. *Journal of Learning Analytics, 2*(2), 5–13.

Yang, D., Wen, M., Howley, I., Kraut, R., & Rose, C. (2015). Exploring the effect of confusion in discussion forums of Massive Open Online Courses. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale (L@S'15)*, 121–130. doi:10.1145/2724660.2724677

Yang, D., Wen, M., Kumar, A., Xing, E. P., & Rose, C. P. (2014). Towards an integration of text and graph clustering methods as a lens for studying social interaction in MOOCs. *The International Review of Research in Open and Distributed Learning*, *15*(5), 214–234. doi:10.19173/irrodl.v15i5.1853

Ye, C., Fisher, D. H., Kinnebrew, J. S., Narasimham, G., Biswas, G., Brady, K. A., & Evans, B. J. (2015). Behavior prediction in MOOCs using higher granularity temporal information. In In Proceedings of *the Second (2015) ACM Conference on Learning @ Scale (L@S'15*), 335–338. doi:10.1145/2724660.2728687

Tables

Table 1

Overview of statistical approaches reported in reviewed publications

Statistical approach	Number of studies used	Proportion of studies used
Machine learning	13	0.34
Descriptive	9	0.24
Correlational	7	0.18
Regression	7	0.18
Chi-square	7	0.18
MANOVA/ANOVA	6	0.16
Survival analysis	5	0.13
Linear-Mixed models	3	0.08
Other	5	0.13

Figures

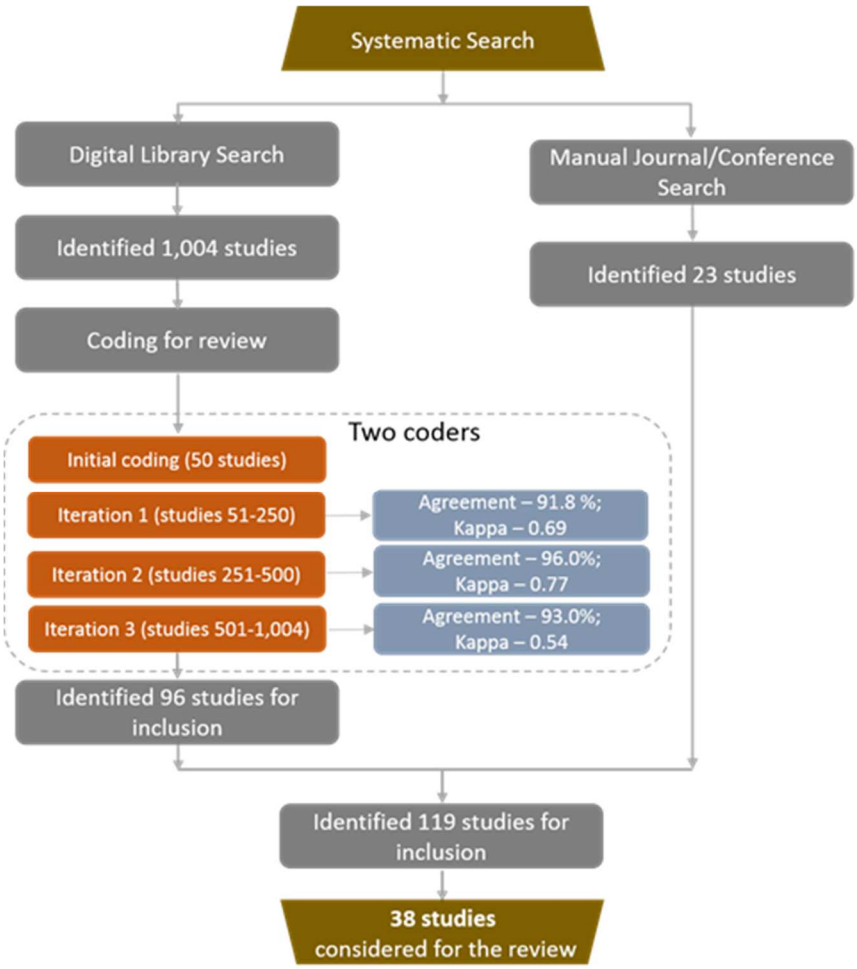

Figure 1. Overview of the systematic search and coding process

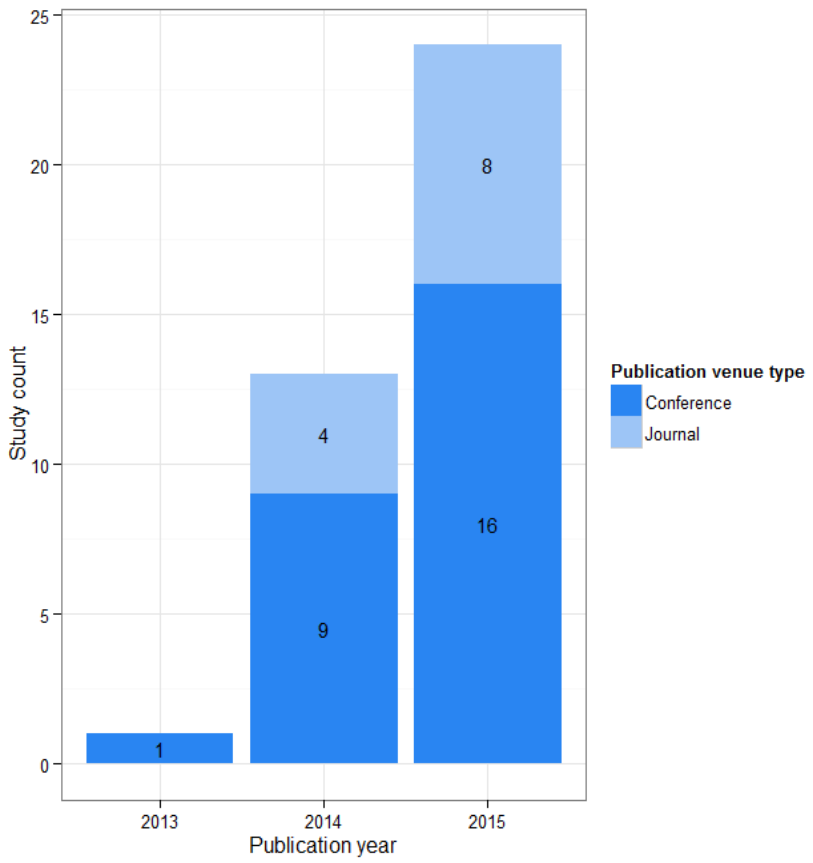

Figure 2. The number of studies per year, with bars showing the respective number of papers published in respective venues (i.e., journal or conference).

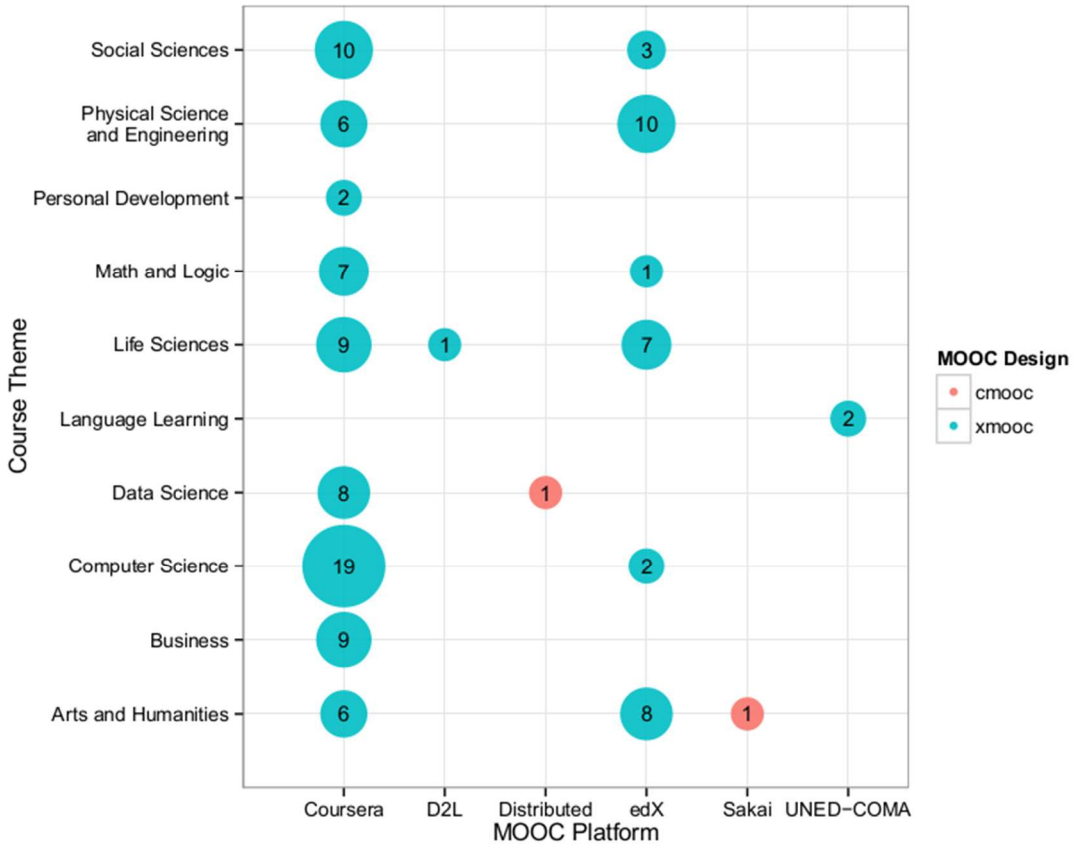

Figure 3. The number of studies within a given topic, delivered on a given MOOC platform, with colors indicating MOOC design (i.e., xMOOC or cMOOC).

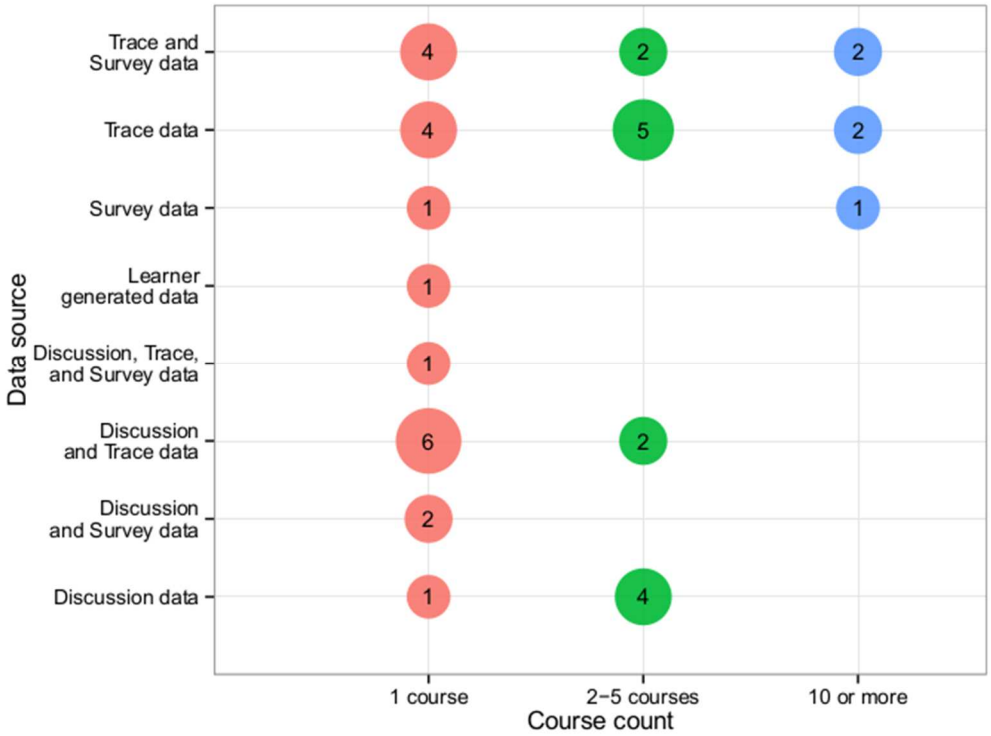

Figure 4. The number of courses using different data sources with the number of courses included in the analyses.

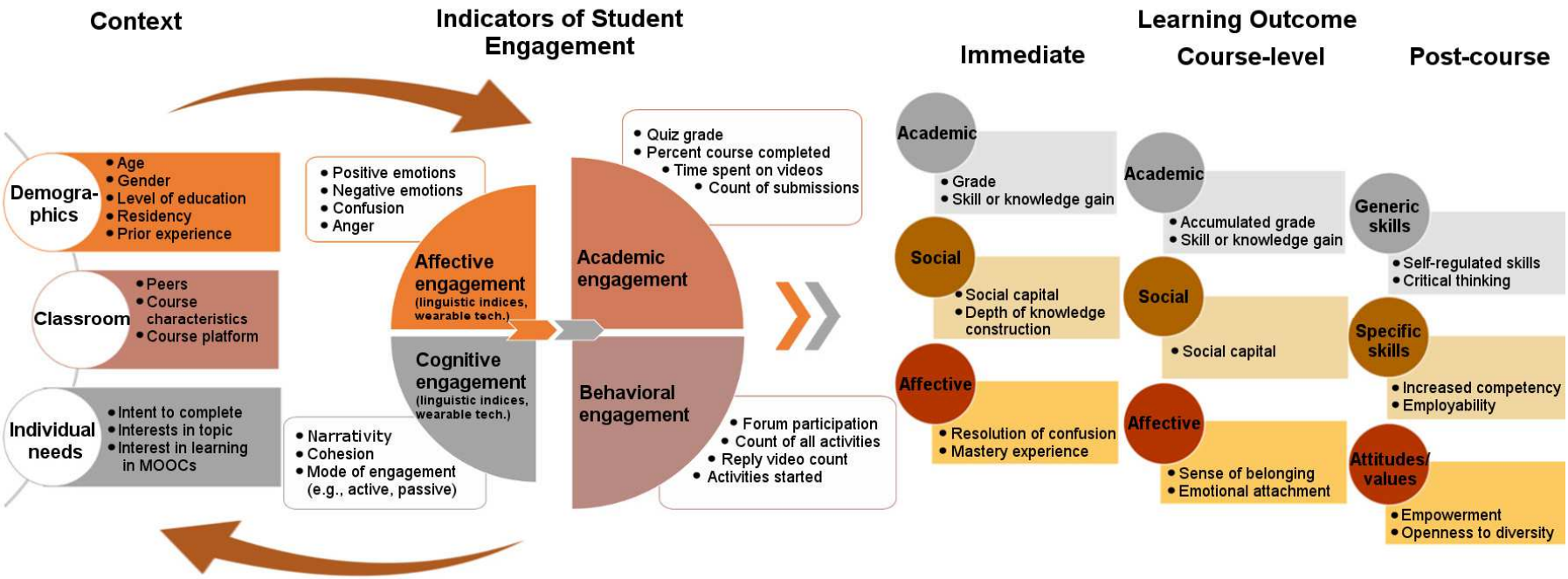

Figure 5. The adopted model of the association between context, engagement, and proximal learning outcome, originally developed by Reschly and Christenson (2012), with indicators specific for learning in non-formal, digital educational settings. Figure S1 depicts the original model, as proposed by Reschly and Christenson.

Appendix A. Overview of the studies included in the systematic review

Table 2

Studies included in the analysis, with the information about the author(s), title, publication venue type, number of courses analyzed, data sources used, number of participants (registered, active, completed), and publication year

#	Study	Title	Publication Venue Type	Num. Courses	Data Sources	Num. of Students Registered	Active/Observed	Completed
1.	Adamopoulos (2013)	What makes a great MOOC? An Interdisciplinary Analysis of Student Retention in Online Courses	Conference	133	S	NR	842	NR
2.	Bergner et al. (2015)	Methodological Challenges in the Analysis of MOOC Data for Exploring the Relationship between Discussion Forum Views and Learning Outcomes	Conference	1	T-D	154,753	~50,000	7,157
3.	Boyer and Veeramachaneni (2015)	Transfer Learning for Predictive Models in Massive Open Online Courses	Conference	3	T	235,197	NR	11,243
4.	Brooks, Stalburg, et al. (2015)	Learn with Friends: The Effects of Student Face-to-Face Collaborations on Massive Open Online Course Activities	Conference	1	T-S	NR	NR	NR
5.	Brooks, Thompson, et al. (2015)	Who You Are or What You Do: Comparing the Predictive Power of Demographics vs. Activity Patterns in Massive Open Online Courses (MOOCs)	Conference	1	T-S	61,820	23,818 (4,130)	NR
6.	Champaign et al. (2014)	Correlating Skill and Improvement in 2 MOOCs with a Student's Time on Tasks	Conference	2	T-S	NR	6,960	8,187
7.	Coffrin et al. (2014)	Visualizing Patterns of Student Engagement and Performance in MOOCs	Conference	2	T	91,994	55,329	2,207
8.	Crossley et al. (2015)	Language to Completion: Success in an Educational Data Mining Massive Open Online Class	Conference	1	T-D	> 48,000	13,314	638
9.	Authors (2015b)	REMOVED FOR THE REVIEW	Conference	1	D	16,091	1,754	517
10.	Engle et al. (2015)	Coursera's Introductory Human Physiology Course: Factors that Characterize Successful Completion of a MOOC	Journal	1	T-D-S	33,378	15,000	NR
11.	Gillani and Eynon (2014)	Communication Patterns in Massively Open Online Courses	Journal	1	D-S	8,700	4,337	NR
12.	Goldberg et al. (2015)	Relationship between Participants' Level of Education and Engagement in their Completion of the Understanding Dementia Massive Open Online Course	Journal	1	D-S	13,950	NR	6,520

#	Study	Title	Publication Venue Type	Num. Courses	Data Sources	Num. of Students Registered	Active/Observed	Completed
13.	Greene et al. (2015)	Predictors of Retention and Achievement in a Massive Open Online Course	Journal	1	T-S	33,938	3,875	1,097
14.	Heutte et al. (2014)	MOOC User Persistence Lessons from French Educational Policy Adoption and Deployment of a Pilot Course	Journal	1	T-S	1,189	917	*NA
15.	Jiang, Warschauer, Williams, O'Dowd, and Schenke (2014)	Predicting MOOC Performance with Week 1 Behavior	Conference	1	T-D	37,933	NR	2,522
16.	Jiang, Fitzhugh, and Warschauer (2014)	Social Positioning and Performance in MOOCs	Conference	2	D	163,100	4,706	NR
17.	Authors (2015a))	REMOVED FOR THE REVIEW	Conference	1	L	NR	1,426	*NA
18.	Kennedy et al. (2015)	Predicting Success: How Learners' Prior Knowledge, Skills and Activities Predict MOOC Performance	Conference	1	T	37,777	22,731	774
19.	Kizilcec and Halawa (2015)	Attrition and Achievement Gaps in Online Learning	Conference	21	T-S	513,098	120,854	NR
20.	Kizilcec and Schneider (2015)	Motivation as a Lens to Understand Online Learners: Toward Data-Driven Design with the OLEI Scale	Journal	14	T-S	295,355	71,475	NR
21.	Koedinger et al. (2015)	Learning is Not a Spectator Sport: Doing is Better than Watching for Learning from a MOOC	Conference	1	T	27,720	14,264	1,154
22.	Konstan et al. (2015)	Teaching Recommender Systems at Large Scale: Evaluation and Lessons Learned from a Hybrid MOOC	Journal	1	S	28,389	21,357	5,643
23.	Li et al. (2015)	MOOC Video Interaction Patterns: What Do They Tell Us?	Journal	2	T-S	NR	31,880	5,539
24.	Loya et al. (2015)	Conscientious Behaviour, Flexibility and Learning in Massive Open On-Line Courses	Journal	1	T	50,335	29,950	10,398
25.	Ramesh et al. (2014a)	Learning Latent Engagement Patterns of Students in Online Courses	Conference	1	T-D	NR	1,665	826
26.	Ramesh et al. (2014b)	Uncovering Hidden Engagement Patterns for Predicting Learner Performance in MOOCs	Conference	3	T-D	> 65,000	27,500	7,000
27.	Santos et al. (2014)	Success, Activity and Drop-outs in MOOCs an Exploratory Study on the UNED COMA Courses	Conference	2	T	56,876	6,252	2,722
28.	Sharma et al. (2015)	Identifying Styles and Paths toward Success in MOOCs	Conference	4	T	NR	NR	NR
29.	Sinha and Cassell (2015)	Connecting the Dots: Predicting Student Grade Sequences from Bursty MOOC Interactions over Time	Conference	13	T	NR	10,000	NR
30.	Tucker et al. (2014)	Mining Student-Generated Textual Data in MOOCS and Quantifying Their Effects on Student Performance and Learning Outcomes	Journal	1	T	NR	NR	NR

#	Study	Title	Publication Venue Type	Num. Courses	Data Sources	Num. of Students Registered	Active/Observed	Completed
31.	Vu et al. (2015)	Relational Event Models for Social Learning in MOOCs	Journal	1	T-D			
32.	X. Wang et al. (2015)	Investigating how Student's Cognitive Behavior in MOOC Discussion Forums Affect Learning Gains	Conference	1	T-D	66,286	33,527	NR
33.	Wen, Yang, and Rose (2014b)	Linguistic Reflections of Student Engagement in Massive Open Online Courses	Conference	3	D	27,750	491	NR
34.	Wen, Yang, and Rose (2014a)	Sentiment Analysis in MOOC Discussion Forums: What does it Tell us?	Conference	3	D	NR	5,512	NR
35.	Whitehill et al. (2015)	Beyond Prediction: First Steps Toward Automatic Intervention in MOOC Student Stopout	Conference	10	T	245,034	NR	20,056
36.	Yang, Wen, Kumar, Xing, and Rose (2014)	Towards an Integration of Text and Graph Clustering Methods as a Lens for Studying Social Interaction in MOOCs	Conference	2	T-D	NR	NR	NR
37.	Yang et al. (2015)	Exploring the Effect of Confusion in Discussion Forums of Massive Open Online Courses	Journal	3	D	NR	NR	NR
38.	Ye et al. (2015)	Behavior Prediction in MOOCs using Higher Granularity Temporal Information	Conference	2	T	NR	NR	NR

Note: Publication Venue Type – Conference (C), Journal (J). NR – Note Reported.

* Given that those studies analyzed courses based on the connectivist pedagogy, we reported a number of students who completed a course as NA (Not applicable).

** Data Sources: T – trace data, D – discussion data, S – survey data, L – learner generated data (e.g., blogs, tweets, Facebook posts).