Self- and Peer Assessment in Massive Open Online Courses

Wilfried Admiraal[1], Bart Huisman[1] & Maarten van de Ven[1]

[1] Leiden University Graduate School of Teaching, Leiden University, Leiden, the Netherlands

Correspondence: Wilfried Admiraal, Leiden University Graduate School of Teaching, Leiden University, Wassenaarseweg 62A, 2333 AL, Leiden, the Netherlands. Tel: 31-71-5276081. E-mail: w.f.admiraal@iclon.leidenuniv.nl

Abstract

Open online distance learning in higher education has quickly gained popularity, expanded, and evolved, with Massive Open Online Courses (MOOCs) as the most recent development. New web technologies allow for scalable ways to deliver video lecture content, implement social forums and track student progress in MOOCs. However, we remain limited in our ability to assess complex and open-ended student assignments. In this paper, we present a study on the quality of self- and peer assessments in three MOOCs. In general, the quality of self-assessments and peer assessments was low to moderate, suggesting that both self-assessment and peer assessment should be used as assessment *for* learning instead of assessment *of* learning. Based on low correlations with final exam grades and other assessment forms, we conclude that self-assessments might not be a valid way to assess students' performance in MOOCs. Yet the weekly quizzes and peer assessment significantly explained differences in students' final exam scores, with one of the weekly quizzes as the strongest predictor. Future research on MOOCs implies a reconceptualization of education variables, including the role of assessment of students' achievements.

1. Introduction

In recent years, free access has been provided to content which previously had a price: searches, software, music and references, to name but a few. Access to the Internet and broadband has increased rapidly and huge growth in mobile connectivity has brought online content and interaction to a global audience. At the same time, open online distance learning in higher education has quickly gained popularity, expanded, and evolved. Recently, Massive Open Online Courses (MOOCs) appear to be a significant force within higher education.

However, while new web technologies allow for scalable ways to deliver video lecture content, implement social forums and track student progress, we remain limited in our ability to evaluate and give feedback for complex and often open-ended student assignments. Self- and peer assessment might offer promising solutions that can scale the grading of complex assignments in courses with thousands of students. In this paper, we present a study on the general quality of self- and peer assessments in three Leiden University MOOCs in the Coursera platform.

2. Massive Open Online Courses (MOOCs)

A typical MOOC of 2014 might take place over 4 to 10 weeks. Students, on average, dedicate two to six hours a week to the course. Materials are consumed in diminishing volumes throughout the MOOC as many learners' commitment wanes. Course applicants can be numbered in the tens of thousands, while those who complete and obtain certificates are usually numbered in the hundreds. As in regular higher education, the value of a MOOC for student learning highly depends on how learning processes are facilitated, stimulated and assessed.

The most influential categorization of MOOC pedagogy relates to the notion that there are two main kinds of MOOCs, each of which determines a particular pedagogical approach: the connectivist or cMOOC, driven by pedagogical principles of social learning, and the institutionally-focused xMOOC, reliant on video-lecture content and automated assessment. However, there is a move away from the cMOOC/xMOOC dichotomy towards recognition of the multiplicity of MOOC designs, purposes, topics and teaching styles, sometimes using alternative terms such as Distributed Open Collaborative Course (DOCC; Jaschik, 2013), Participatory Open Online Course (POOC; Daniels, 2013), Small Private Online Course (SPOC; Hashmi, 2013) or Big Open Online Course (BOOC; Tattersall, 2013).

Researchers at the University of Illinois Springfield have developed the Assessing MOOC Pedagogies Tool (Swan, Day, Bogle, & Van Prooyen, 2014). They used this tool to characterize the pedagogical approaches taken in 13 MOOCs of Coursera (5), Udacity (7) and EdX (1), which are the three mostly used providers of a MOOC platform. The MOOC pedagogy is described along ten dimensions that are adapted from similar scales developed by Reeves (1996) for describing the pedagogical dimensions of computer-based instruction and by Harris and Hofer (2009) to situate pedagogical decisions on which they suggest technology integration should be grounded. These ten dimensions are

1) Epistemology (objectivist to constructivist);

2) Role of the teacher (teacher centered to student centered);

3) Focus of activities (convergent to divergent)

4) Structure (less structured to more structured)

5) Approach to content (concrete to abstract);

6) Feedback (infrequent and unclear to frequent and constructive);

7) Cooperative learning (unsupported to integral);

8) Accommodation of individual differences (unsupported to multifaceted);

9) Activities and assignment (artificial to authentic), and

10) User role (passive to generative).

Ratings for each set of courses were quite similar, although there were some clear differences between the two platforms Coursera and Udacity. Coursera courses, more than Udacity courses, followed a format that resembles the traditional lecture/text-testing routine of traditional university courses spread over multiple weeks with hard deadlines.

In a review of the literature and debate, Bayne and Ross (2013) extracted three emerging issues for MOOC pedagogy: 1) the role of the teacher, 2) learner participation and 3) assessment. Firstly, the role of the teacher in the MOOC has been under-examined as most research has investigated the learner perspective (Liyanagunawardena, Adams, & Williams, 2013). Two main teacher roles appear from the literature, which are connected to the way the MOOC is designed: the academic celebrity teacher in xMOOCs and the facilitator in cMOOCs. The academic celebrity teacher is the role of a respected authority based in an elite institution. These lecturers are not available to MOOC participants in any interpersonal way but primarily through the recordings of their lectures. The recordings are supplemented with automatically marked quizzes, discussion posts and pass/fail tasks. In cMOOCs, the teachers' role focusses on facilitating self-directed learning. A more sophisticated distinction between teacher roles in MOOCS is necessary in order to get a better understanding of effective pedagogies. Literature on moderator roles in computer conferencing from the 90s (Admiraal, Lockhorst, Wubbels, Korthagen, & Veen, 1998; Paulsen, 1995) might be helpful in this.

Secondly, learner participation is one the most examined aspects in literature and debates about MOOCs. The key dilemmas in MOOCs center on what participation actually means, how it should be measured, and what metrics of success and quality are appropriate. Milligan, Littlejohn and Margaryan (2013) describe a continuum of active, lurking and passive participation, and Hill (2013) distinguishes five archetypes of no-shows, observers, drop-ins, passive participants and active participants. The notion that people might sign up for a course not intending to complete the assessments is common in free courses where the barrier to entry is usually as low as clicking a registration button and entering an email address. This means that new measures of success and quality are required, because participant behaviors and intentions are so diverse.

Assessment is the third emerging issue in literature on MOOCs leading to questions like "What sorts of learning can be assessed at scale?", "How should individuals be authenticated so that the correct person's work is being assessed?", "How can cheating be prevented?", and "Who should decide how much university credit a MOOC is worth?", to name a view (Bayne & Ross, 2013, p.29). It becomes clear that "openness" of a MOOC has a very different future in a system of accreditation than that it does in informal learning settings. Self- and peer assessment - which has been historically used for logistical, pedagogical, metacognitive, and affective benefits - might offer promising solutions that can scale the grading of complex assignments in courses with thousands of students. How to design self- and peer assessments is a challenge in itself as MOOCs have massive, diverse student enrollment. In order to be able to develop effective self- and peer assessments, we first need to gain more insight in the quality of

these grading procedures in MOOCs. More specific, we formulated the following research questions:

1)   What is the reliability of self- and peer assessment implemented in MOOCs?

2)   What is the relationship between self- and peer assessment and quizzes?

3)   To what extent do self- and peer assessment and quizzes explain differences in students' final exams scores?

3. Methods

3.1 Context of the Study

In two MOOCs organized at Leiden University in the Netherlands, intermediate self-assessments and peer assessments were used in addition to final exams. The first MOOC, *The Law of the European Union: An Introduction*, was a 5-8 weeks MOOC, run in June 2013. This course included small video clips, discussion fora, quizzes, a case study and a voluntary exam. The second MOOC, *Terrorism and Counterterrorism: Comparing Theory and Practice*, was a 5-weeks MOOC in Fall 2013 with weekly videos, quizzes and peer assignments as well as a voluntary final exam. This MOOC was rerun February 2014. All three courses required 5 to 8 hours student work per week.

3.2 Assessments

In all three MOOCs, four types of assessments were implemented: weekly quizzes, self-assessment, peer assessment and final exam.

3.2.1 Weekly Quiz and Final Exam

The weekly quizzes and final exam were automatically marked multiple-choice quizzes, testing declarative knowledge of the course content. In MOOC 3 (Terrorism 2014), it was possible to follow a certification track, which meant that students who completed all quizzes, self- and peer assessments and the final exam could receive a certificate. Of the total of 18,622 registrants, 410 students signed up for the certification track.

3.2.2 Self- and Peer Assessment

In each of the three MOOCs, students could complete an essay on a topic that was relevant for the particular MOOC. In the first MOOC, this topic was provided; in the other two MOOCs, students could choose from four topics. The essay assignment started with a case description in which an authentic context was pictured, followed by some prompts. Students were encouraged to prepare this assignment with the use of information which was available in the course environment (video, syllabus, background materials). Then the procedures of how to complete the assignment were introduced along with a rubric of how to assess it. Students had to assess their own essay and then the essay of at least two (MOOC 1) or four (MOOC 2 and 3) of their peers. The nature of the rubrics differed slightly between MOOC1, on the one hand, and MOOC 2 and 3, on the other hand. The rubric of MOOC 1 had a pre-structured format with four items with several sub-items on the accuracy of the content of the essay and one item with four sub-items on the structure and presentation of the essay. Each possible score on each sub-item was clearly described. The rubric of MOOC 2 and 3 was structured with four (assignment 1) or five (assignment 2) items. The first three or four items referred to the accuracy and adequacy of the content of the essay; the last item assessed the structure of the essay. Students were instructed about the deadlines and they were reminded that they agreed with the Coursera Honor Code about plagiarism. Students were instructed to assign a score of 0 to plagiarized work.

4. Results

Thousands of participants were registered in each of the three MOOCs, although substantial less data was collected on quizzes, self-assessments, peer-assessment assignments and final exam. In Table 1, we present descriptive indices of each assessment (mean scores, standard deviations in scores, range of scores and number of valid assessments, respectively).

From Table 1 it is clear that in all three MOOCs the number of participants who completed the quizzes decreased over time. The number of participants who completed self-assessments and peer assessments was a small portion of the total student enrollment. Participants who completed the voluntary final exam formed about 10% of the total student enrollment (from 6% in MOOC 1 to 12% in MOOC 3).

Table 1. Descriptive indices of assessment types (N=number of registrants; Mean=mean score; s.d.=standard deviation in scores; Min/Max=range of scores; n=number of valid assessments)

	MOOC1 N= 52559				MOOC2 N= 26890				MOOC3 N= 18622			
	Mean (s.d)	Min	Max	n	Mean (s.d)	Min	Max	n	Mean (s.d)	Min	Max	n
Quizzes												
1	3.75 (1.41)	0	5	7472	8.83 (1.48)	0	10	5399	8.83 (1.69)	0	10	4459
2	3.42 (1.38)	0	5	4322	9.03 (1.37)	0	10	4077	9.01 (1.47)	0	10	3288
3	4.21 (1.18)	0	5	3349	12.07 (2.22)	0	14	3593	8.59 (1.82)	0	10	2810
4	3.80 (1.32)	0	5	3050	13.34 (2.18)	0	15	3230	8.83 (1.72)	0	10	2466
5	--	--	--	--	9.02 (1.56)	0	10	3014	8.98 (1.62)	0	10	2296
Self-assessment												
1	17.94 (4.34)	4	25	397	28.30 (2.99)	10	30	706	18.58 (2.85)	0	20	572
2	--	--	--	--	37.95 (3.71)	5	40	561	37.37 (5.24)	5	40	475
Peer assessment												
1	15.29 (5.42)	0	25	688	25.23 (5.20)	10	30	824	16.38 (4.45)	0	20	635
2	--	--	--	--	32.86 (7.52)	5	40	579	33.71 (6.70)	5	40	491
Final exam												
1	11.44 (5.48)	0	20	3168	17.26 (5.58)	0	25	2988	17.44 (5.81)	0	25	2274

4.1 Reliability of Self- and Peer Assessment

4.1.1 MOOC 1 EU Law

The case assignment, which was used for both self-assessment and peer assessment, included five items. The homogeneity of the test in terms of Cronbach's α, was high, both for self-assessment ($α$=.83) and peer assessment ($α$=.90, $α$=.89 $α$=.87 for peer reviewer 1, 2 and 3, respectively). From Table 2, we can see that the correlations between the three peer assessment grades is moderate (between r=.42 to r=.57). In Tables 2 to 8, we include the Pearson's correlation coefficients and the number of valid assessments, and we indicate the significance with *** means p <.001, ** means p= <.01, and * means p=<.05.   The fourth peer assessment was not included as only 6 students had 4 peer grades.

Table 2. Correlations between peer assessments of MOOC1

	Peer2	Peer3
Peer1	.50*** 684	.42*** 80
Peer2		.57*** 80

4.1.2 MOOC 2: Terrorism 2013

The two case assignments, which were used for both self-assessment and peer assessment, included four (assignment 1) or five items (assignment 2). The homogeneity in terms of Cronbach's $α$ was moderate for both self-assessments (for both assignments $α$=.59) and high for all peer assessments (peer assignment 1 between $α$=.72 and $α$=.79 and peer assignment 2 between $α$=.74 and $α$=.80). In all cases, the item that refers to the presentation (structure, layout, and language use) of the completed assignment showed the lowest item-rest correlations (between r=.55 and r=.64). The other items referred to an assessment of the content quality of the completed assignments. The correlations between the assessments of the five peers is moderate (around r = .40 for assignment 1 (see Table 3) and around r = .30 for

assignment 2 (see Table 4)). The assessments of the sixth peer were not included as only 4 students received 6 peer grades. The correlations indicate a low to moderate agreement between peers.

Table 3. Correlations between peer assessments of assignment 1 of MOOC 2

Assignment 1	Peer2	Peer3	Peer4	Peer5
Peer1	.42*** 842	.42*** 842	.40*** 618	.57*** 54
Peer2		.44*** 842	.40*** 618	.38** 54
Peer3			.39*** 618	.28* 54
Peer4				.27 54

Table 4. Correlations between peer assessments of assignment 2 of MOOC2)

Assignment 2	Peer2	Peer3	Peer4	Peer5
Peer1	.30*** 592	.31*** 592	.39*** 573	.38** 64
Peer2		.37*** 592	.33*** 573	.27* 64
Peer3			.32*** 573	.40** 64
Peer4				.36** 64

4.1.3 MOOC 3: Terrorism 2014

The two case assignments, which were used for both self-assessment and peer assessment, included four (assignment 1) or five items (assignment 2). The homogeneity in terms of Cronbach's α was moderate to high for self-assessments ($α$=.60 for assignment 1 and $α$=.75 for assignment 2) and moderate to high for all peer assessments of both assignments (peer assignment 1 between $α$=.59 and $α$=.67 and peer assignment 2 between $α$=.71 and $α$=.79). An exception was the homogeneity of the assessments of the fifth peer of the first assignment: Cronbach's $α$=.18 based on 42 assessments.

Table 5. Correlations between peer assessments of assignment 1 of MOOC 3

Assignment 1	Peer2	Peer3	Peer4
Peer1	.50*** 635	.48*** 635	.38*** 600
Peer2		.50*** 635	.51*** 600
Peer3			.51*** 600

The correlations between the four peer assessments of the first assignment (the fifth peer was left out because of the low reliability) was moderate (around r=.50 ;see Table 5). The correlations between the five peer assessments of the second assignments were generally lower (mostly between r=.30 and r=.40, see Table 6). The assessment of the sixth

peer was not included in both assignments as only 5 (for assignment 1) or 7 (for assignment 2) students received 6 peer assessment grades. The correlations indicate a low to moderate agreement between peers.

Table 6. Correlations between peer assessments of assignment 2 of MOOC3

Assignment 2	Peer2	Peer3	Peer4	Peer5
Peer1	.35***	.34***	.39***	.32**
	491	491	491	78
Peer2		.27***	.36***	.43***
		491	491	78
Peer3			.40***	.53***
			491	78
Peer4				.58***
				78

4.2 Relationship Between Assessment Types

4.2.1 MOOC 1 EU Law

In Table 7, we present the correlations between student performances in the weekly quizzes, self-assessment, average peer assessment, and the final test. In general, both self-assessment and peer assessment show low to moderate correlations with the weekly quizzes and the final test. The highest correlations are between quiz scores and the final exam (between $r=.50$ and $r=.60$).

Table 7. Correlations between quizzes, self-assessment, peer assessment and final exam

	Quiz2	Quiz3	Quiz4	Self	Peer Total	Final
Quiz1	.54***	.55***	.50***	.17**	.26***	.51***
	4295	3333	3027	394	669	2937
Quiz2		.55***	.53***	.18***	.28***	.52***
		3315	3027	391	658	2871
Quiz3			.60***	.17**	.28***	.53***
			3018	390	648	2842
Quiz4				.26***	.33***	.60***
				384	631	2827
Self					.40***	.30***
					396	383
Peer Total						.42***
						623

4.2.2 MOOC 2: Terrorism 2013

In Table 8, the correlations are presented between all assessments. In general, the correlations of all quizzes are quite high (between $r=.48$ and $r=.68$). The correlations between both self-assessments ($r=.55$) and both peer assessments ($r=.47$) are higher than the correlations between self- and peer assessment of the same assignment ($r=.38$ and $r=.20$ for the first and second assignment, respectively). All assessments are moderately correlated with the final exam (around $r=.40$), except for both self-assessments ($r=.22$ and $r=.16$, respectively).

Table 8. Correlations between quizzes, self-assessments, peer assessments and final exam (MOOC2 Terrorism 2013 above the diagonal and MOOC Terrorism 2014 below the diagonal)

MOOC2 / MOOC3	Quiz1	Quiz2	Quiz3	Quiz4	Quiz5	Self1	Self2	Peer1	Peer2	Final
Quiz1		.63*** 4061	.60*** 3577	.57*** 3214	.48*** 3001	.24*** 684	.13** 559	.27*** 787	.29*** 575	.38*** 2923
Quiz2	.65*** 3275		.64*** 3572	.57*** 3219	.50*** 3003	.29*** 670	.12** 555	.28*** 751	.29*** 572	.43*** 2910
Quiz3	.60*** 2789	.66*** 2789		.65*** 3217	.55*** 3011	.27*** 652	.18*** 552	.26*** 716	.31*** 569	.46*** 2899
Quiz4	.54*** 2458	.54*** 2457	.64*** 2460		.68*** 2999	.24*** 630	.20*** 546	.21*** 690	.26*** 561	.42*** 2885
Quiz 5	.51*** 2288	.52*** 2288	.63** 2289	.77*** 2285		.21*** 611	.21*** 536	.12** 665	.15*** 551	.46*** 2832
Self1	.23*** 560	.11* 548	.10* 530	.14** 510	.11* 490		.55*** 518	.38*** 691	.22*** 525	.22*** 609
Self2	.19*** 470	.17*** 471	.13** 469	.17*** 465	.13** 453	.42*** 445		.17*** 534	.20*** 547	.16*** 535
Peer1	.37*** 617	.31*** 600	.38*** 577	.24*** 554	.28*** 530	.35*** 569	.15** 459		.47*** 548	.40*** 666
Peer2	.36*** 484	.30*** 484	.30*** 481	.27*** 475	.31*** 464	.13** 451	.38*** 471	.43*** 468		.45*** 551
Final	.44** 2191	.43*** 2189	.49*** 2185	.48*** 2168	.52*** 2142	.32*** 492	.26*** 452	.45*** 531	.38*** 462	

4.2.3 MOOC 3: Terrorism 2014

In Table 8, the correlations are presented between all assessments (below the diagonal). In general, the correlations between all quizzes are quite high (between $r=.51$ and $r=.77$). The correlations between both self-assessments ($r=.42$) and both peer assessments ($r=.43$) are higher than the correlations between self- and peer assessment of the same assignment ($r=.35$ and $r=.38$ for the first and second assignment, respectively). All assessments are moderately correlated with the final exam (between $r=.38$ and $r=.52$), except for both self-assessments ($r=.32$ and $r=.26$, respectively).

4.3 Relationship with Final Exam

In Table 9, the results of the stepwise regression analyses for each MOOC are summarized. As could be expected on the basis of the correlations presented earlier, both self-assessments did not significantly explain differences between students in their final exam grade. The strongest predictor was in all cases one of the quizzes, although peer assessments were also significantly related to the final exam grade.

The correlations between the number of assessment attempts (Quizzes, self-assessment, peer assessment) and the final-exam grade were moderate to low (MOOC 1 $r=.41$ ($p<.001$); MOOC 2 $r=.26$ ($p<.001$); MOOC 3 $r=.30$ ($p<.001$)). This means that there seemed to be a weak relationship between the number of assessments students took and their final exam grade. This finding contradicts other MOOC research that finds a strong positive relationship between the number of student activities and their final course grade (DeBoer, Ho, Stump, & Breslow, 2014).

Table 9. Stepwise regression analyses with final exam as dependent variable (n.a.= not applicable; n.s.= not significant with α=0.05)

	MOOC 1 EU Law 2013		MOOC 2 Terrorism 2013		MOOC 3 Terrorism 2014	
	B (s.e)	R^2_{change}	B (s.e)	R^2_{change}	B (s.e)	R^2_{change}
Weekly quiz 1	1.08 (0.26)	0.04	n.s.		n.s.	
Weekly quiz 2	0.82 (0.20)	0.03	0.91 (0.25)	0.20	1.06 (0.32)	0.02
Weekly quiz 3	n.s.		0.56 (0.15)	0.04	n.s.	
Weekly quiz 4	1.43 (0.22)	0.28	n.s.		n.s.	
Weekly quiz 5	n.a.		0.58 (0.22)	0.01	1.10 (0.26)	0.17
Peer grading 1	0.31 (0.03)	0.08	0.10 (0.04)	0.01	0.25 (0.06)	0.09
Peer grading 2	n.a.		0.17 (0.02)	0.10	0.12 (0.03)	0.03
Self-grading 1	n.s.		n.s.		n.s.	
Self-grading 2	n.a.		n.s.		n.s.	
Adjusted R^2	0.41		0.35		.30	
Degrees of freedom	4, 379		5, 475		4, 414	

In MOOC 3, it was possible to sign up for a certification track, which required completion of all quizzes, self- and peer assessments and final exam in time. Student who registered for a certification track received significantly higher scores on their final exam, compared to the other students ($M_{certification\ track}$ = 19.2 and $M_{other\ students}$ = 17.1; (t(574.5) = 7.41; p<.001)). We repeated the regression analyses for students following a certification track and the other students separately. The results are presented in Table 10.

Table 10. Stepwise regression analyses with final exam as dependent variable for students following the certification track and the other students in MOOC 3 (n.s.= not significant with α=0.05)

	Certification track		Other students	
	B (s.e)	R^2_{change}	B (s.e)	R^2_{change}
Weekly quiz 1	1.34 (0.38)	0.07	n.s.	
Weekly quiz 2	1.42 (0.39)	0.22	n.s.	
Weekly quiz 3	n.s.		n.s.	
Weekly quiz 4	n.s.		n.s.	
Weekly quiz 5	n.s.		1.87 (0.31)	0.18
Peer grading 1	0.20 (0.09)	0.11	0.27 (0.08)	0.07
Peer grading 2	0.10 (0.04)	0.02	0.11 (0.04)	0.02
Self-grading 1	0.26 (0.10)	0.02	n.s.	
Self-grading 2	n.s.		n.s.	
Adjusted R^2	0.42		0.27	
Degrees of freedom	5, 157		3, 256	

From Table 10 it is clear that the assessments explained more differences between students in final exams score if they followed a certification track: the first two quizzes, both peer assessments and the first self-assessment significantly explained differences in the final exam scores. The total amount of variance explained in the final exam score is less for the other students.

5. Discussion and Conclusion

In general, the quality of both the self-assessment and the peer assessment was moderate. These assessments showed a homogenous structure, but the correlations between peer assessments of the same assignments were low to moderate. The latter means that peers did agree on their grades for the assignments only to a limited degree. The correlations between the various peer assessments of the last MOOC were moderate to high. In this MOOC, the procedures and criteria for peer assessment were adapted on the basis of the 2013 run. Moreover, there is only a weak correlation between self-assessment and peer assessment, and the correlations between different self-assessment assignments are higher than the correlations between self-assessment and peer assessment of the same assignments. In addition, self-assessments did not significantly explained variance in students' final exam scores. These results suggest a bias of self-assessments and led us to conclude that self-assessments might not be a valid way to assess students' performance in MOOCs. Yet the weekly quizzes and both peer assessments significantly explained differences in students' final exam scores, with one of the weekly quizzes as the strongest predictor in all three MOOCs. Finally, the number of assessment attempts of students was not significantly correlated with their final exam scores. The latter result does not confirm conclusion from earlier research that found a strong positive relationship between the number of student activities and their course grade (DeBoer et al. 2014).

With this study we provided insight in the quality of the various assessments in MOOCs and how these are related to the final exams. We conclude that self-assessments and peer assessments should be improved if they are used as summative indicators of one's achievements (assessment of learning). In the current MOOCs, they only can be used for self-reflection and peer feedback, emphasizing the formative function of assessment (assessment *for* learning). Future research might go deeper into the quality of assessment assignments of MOOCs including both assessment *of* learning and assessment *for* learning. Due to the massive character of MOOCs summative assessments (so assessment *of* learning) mostly take the form of quizzes or other multiple choice tests, which generate scores automatically. However, this kind of tests does not match with the assessment of more open en more complex assignments. Therefore, other forms of assessment, such as self-assessment, peer assessment or assessment by outside experts, should be develop to make the assessment of the more open assignments possible.

A clear limitation of this study is the limited number of MOOCs examined. Although the data included thousands of participants, only three MOOCs of one host institution (Leiden University) in one platform (Coursera) were studied. Swan et al. (2014) already showed that the pedagogy of MOOCs at different platforms differs, with the Coursera MOOCs emphasizing a more teacher-centered pedagogy. It might be that self- and peer-assessment, which align more with a learner-centered pedagogy, show higher reliability indices in EdX or Udacity MOOCs.

Yet, we agree with DeBoer et al. (2014) that we also should reconceptualize educational variables in research on MOOCs. Differences between traditional classroom data and MOOC data refer to the magnitude of data gathered in terms of numbers of registrants per course, observations per registrant and type of information, the diversity of registrants in reasons for registration as well as in their background, and the registrant use of course tools which is asynchronous and relatively unrestricted in sequence (DeBoer et al. 2014). These authors suggest a reconceptualization of enrollment in MOOCs (e.g., based on registration, course activities, course assignments and assessment, or final exam), participation (the authors show 20 participation metrics which are linked to students' general attendance, their clicks, the hours they spent on course activities, and the assessments), curriculum (curriculum activities showing a variability in sequence), and achievement (which can be based in various indicators of performance and participation). In order to understand the relationship between self- and peer assessment and other gradings and activities in MOOCs we have to think thoroughly what kind of metrics for achievement should be used, how we should define enrollment and participation, in what way the curriculum is implemented, and –therefore- how assessments should be applied.

References

Admiraal, W. F., Lockhorst, D., Wubbels, T., Korthagen, F. A. J., & Veen, W. (1998). Computer-mediated communication environments in teacher education: computer conferencing and the supervision of student teachers. *Learning Environments Research, 1*, 59-74. http://dx.doi.org/10.1023/A:1009936715640

Bayne, S., & Ross, J. (2013). *The pedagogy of the Massive Open Online Course: the UK view.* York, UK: The Higher Education Academy.

Daniels, J. (2013). MOOC to POOC: Moving from massive to participatory. *JustPublics@365*. [Online] Avaliable: http://justpublics365.commons.gc.cuny.edu/2013/02/05/mooc-to-pooc-moving-from-massive-to-participatory/ [Accessed 29 May 2014].

DeBoer, J., Ho, A. D., Stump, G. S., & Breslow, L. (2014). Changing 'course': reconceptualizing educational variables for Massive Open Online Courses. *Educational Researcher, 42*, 74-84. http://dx.doi.org/10.3102/0013189X14523038

Harris, J., & Hofer, M. (2009). Grounded tech integration. *Learning & Leading with Technology, September/October*, 23-25. [Oline] Avaliable: http://www.learningandleading-digital.com/learning_leading/20090910/#pg24 [Accessed 30 May 2014].

Hashmi, A. H. (2013). HarvardX set to launch second SPOC. *Harvard Crimson.* [Online] Avaliable: http://harvardx.harvard.edu/links/harvardx-set-launch-second-spocharvard-crimson-amna-h-hashmi-september-16-2013 [Accessed 29 May 2014].

Hill, P. (2013). Emerging student patterns in MOOCs: A (revised) graphical view. *e-Literate*. [Online] Avaliable: http://mfeldstein.com/emerging-student-patterns-in-moocs-arevised-graphical-view/ [Accessed 12 September 2013].

Jaschik, S. (2013). Feminists challenge Moocs with Docc. *Times Higher Education.* [Online] Avaliable: http://www.timeshighereducation.co.uk/news/feminists-challenge-moocswith-docc/2006596.article [Accessed 29 May 2014].

Liyanagunawardena, T. R., Adams, A. A., & Williams, S. A. (2013). MOOCs: A systematic study of the published literature 2008-2012. *The International Review of Research in Open and Distance Learning, 14*(3), 202–227.

Milligan, C., Littlejohn, A., & Margaryan, A. (2013). Patterns of engagement in connectivist MOOCs. *Journal of Online Learning and Teaching, 9*(2), 149-159.

Paulsen, M. F. (1995). Moderating educational computer conferences. In Z.L. Berge & M.P. Collins (Eds.), *Computer-mediated communications and the online classroom* (Vol. III; pp. 81–103). Cresskill, NJ: Hampton Press.

Reeves, T. (1996). Evaluating what really matters in computer-based education. [Online] Avaliable: http://www.eduworks.com/Documents/Workshops/EdMedia1998/docs/reeves.html#ref10 [Accessed 30 May 2014]

Swan, K., Day, S, Bogle, L., & Van Prooyen, T. (2014). AMP: A tool for characterizing the pedagogical approaches of MOOCs. *E-Mentor, 52*(2). [Online] Avaliable: http://www.e-mentor.edu.pl/artykul/index/numer/54/id/1098 [Accessed 31 May 2014] http://dx.doi.org/10.15219/em54.1098

Tattersall, A. (2013). Gold rush or just fool's gold - A quick look at the literature. *ScHARR MOOC Diaries.* [Online] Avaliable: http://scharrmoocdiaries.blogspot.co.uk/2013/07/scharr-mooc-diaries-part-xvii-gold-rush.html [Accessed 29 May 2014].