MOOCs and the Funnel of Participation
Doug Clow
The Open University
Walton Hall, Milton Keynes
MK7 6AA, United Kingdom
+44 1908 654861
Doug.Clow@open.ac.uk

ABSTRACT
Massive Online Open Courses (MOOCs) are growing substantially in numbers, and also in interest from the educational community. MOOCs offer particular challenges for what is becoming accepted as mainstream practice in learning analytics.

Partly for this reason, and partly because of the relative newness of MOOCs as a widespread phenomenon, there is not yet a substantial body of literature on the learning analytics of MOOCs. However, one clear finding is that drop-out/non-completion rates are substantially higher than in more traditional education.

This paper explores these issues, and introduces the metaphor of a ‘funnel of participation’ to recontextualise the steep drop-off in activity, and the pattern of steeply unequal participation, which appear to be characteristic of MOOCs and similar learning environments. Empirical data to support this funnel of participation are presented from three online learning sites: iSpot (observations of nature), Cloudworks (‘a place to share, find and discuss learning and teaching ideas and experiences’), and openED 2.0, a MOOC on business and management that ran between 2010-2012. Implications of the funnel for MOOCs, formal education, and learning analytics practice are discussed.

Categories and Subject Descriptors

General Terms
Algorithms, Management, Measurement, Performance, Design, Economics, Human Factors, Theory, Legal Aspects

Keywords
Learning analytics, participation, MOOCs.

1. INTRODUCTION
“A MOOC is an online course with the option of free and open registration, a publicly shared curriculum, and open-ended outcomes” [38]. MOOCs “have the potential to provoke major shifts in educational practice” [48] and are “officially [...] the higher education buzzword of 2012” [51]. There are two distinct branches: the connectivist MOOCs (cMOOCs) inspired by Downes, Siemens, Cormier, Groom et al, and the more recent, more formal MOOCs (xMOOCs), including Udacity, MITx, EdX, Coursea and Udemy [28]. The pedagogy of these branches are quite distinct: cMOOCs are underpinned by connectivism [52, 33], a sophisticated and innovative reconceptualisation of what it means to know and to learn, whereas xMOOCs “are so far based on a very old and out-dated behaviourist pedagogy, relying primarily on information transmission” [7]. This paper will use MOOC as an umbrella term covering both branches, and will take a broad view of ‘Course’ to include any structured open, online learning opportunity.

2. LEARNING ANALYTICS AND MOOCS
MOOCs – and particularly the cMOOCs closely associated with many learning analytics figures – pose particular challenges for learning analytics practice. Participation in a MOOC is “emergent, fragmented, diffuse, and diverse” [38]; it seems unlikely that the learning analytics process will be any less so.

Much learning analytics work presupposes a formal education context. When learning analytics are most effective, they are an integrated part of a whole system of learner support, which is hard to deliver in a MOOC.

The foundational work on Signals at Purdue [4, 5, 9] is based around a predictive model of likely completion. This is potentially problematic in a MOOC context. Essa and Ayad [22] set out to extend predictive modeling to accommodate “the considerable variability in learning contexts across different courses and different institutions”. Whether such efforts to encompass diversity can include MOOCs is, at present, an open question. A more profound problem with predictive modeling in MOOCs is the lack of human resource to mediate the feedback, and the lack of support available to learners who have come to know that they are at risk. One useful design framework for learning analytics [26] can be readily applied to a MOOC, but the terminology (e.g. ‘teachers’ and ‘students’) may need to be applied loosely.

On the other hand, some learning analytics technologies present fewer issues in a non-formal context, such as recommendation engines and other semantic technologies, content analytics [23], social network analysis and visualisation (for an arresting example applied to a MOOC, see [10]), and social learning analytics [8].

3. LEARNING ANALYTICS OF MOOCS
There has not yet been extensive published research on xMOOCs, partly because they are so new, and partly because of their proprietary nature. On the other hand, cMOOCs have been researched in some depth, including a specific concern with learning analytics.

PLENK2010 has received perhaps the most thorough treatment, with mixed-methods approaches employing a range of qualitative and quantitative sources, including Moodle data-mining, Twitter metrics, content analysis, surveys, and interviews [25, 31]. Other cMOOCs have received similar attention – e.g. CCK08 [24, 35].
and CCK11 [32]. The clear message from these studies is the importance of methods beyond the simply mechanical/quantitative.

Learning analytics is possible in the wider context of open, online learning environments. Pham et al [42] explored two learning ‘blogospheres’, where social network analysis and content analysis yielded interesting results, including a steeply unequal ‘fat tailed’ distribution of posting frequency (which they erroneously label a power law).

There are two significant points of difference in learning analytics in MOOCs compared to formal education: one qualitative, and one quantitative. The qualitative difference is the rationale behind the course and the aspirations of its designers. In a cMOOC, the designers are explicitly not intending to specify end points before the course starts, so a learner who starts but does not complete may well be seen as a success, depending on the reasons. The quantitative difference is, as the old saw has it, one that is sufficiently large to be a qualitative difference: the rate of drop-out is so very much larger in MOOCs. This idea is encapsulated in the funnel of participation.

4. THE FUNNEL OF PARTICIPATION

The funnel of participation is inspired by the ‘marketing funnel’, or ‘purchase funnel’, an idea in widespread use in marketing and sales (see e.g. [29]). This attempts to model a customer’s journey from initial awareness to a sale, typically in four stages: Awareness – they have to know the product exists; Interest – they have to want that sort of thing; Desire – they have to want the specific product; Action – purchase. This model is not without criticism. More recent thinking argues for more sophistication and a focus on what the customer does after purchase [17, 41], but the model remains a widely-used and useful structuring device. The marketing funnel approach is used in higher education in marketing departments, and in alumni fundraising, but is not generally applied to student progress while enrolled.

In the marketing funnel, there is typically significant attrition in numbers through the stages. A vast number of people need to become aware that the product exists; a fraction of those will be interested in that class of product; a fraction of those will form a desire for the specific product; and, finally, a proportion of those will make a purchase.

In formal education, despite concerns about drop-out rates, the total attrition from enrolment/registration to graduation is typically much lower. In a MOOC, the attrition rate is significantly higher – approaching those seen in marketing. This is the basis of the funnel of participation, as shown in Figure 1.

![Figure 1: The funnel of participation](image)

Here, the first step is Awareness: potential learners must know that the MOOC exists. Next is Registration – only a small proportion of those who are aware will want to sign up and succeed in doing so. Then a fraction of those registered will go on to engage in some Activity or other, and some of those will make meaningful learning Progress. The drop off at each stage is large.

The simple funnel as presented here can be extended in granularity. So, for instance, the Activity category can be split into simply making a contribution, and making more extensive or higher-level contributions.

The funnel is intended to be applicable in a range of pedagogical and theoretical contexts, from connectivism to a naïve information transmission model. It is also designed to be congruent with other, broader conceptions of online participation, such as Communities of Practice and legitimate peripheral participation [34, 55]. Dron and Anderson’s collective applications [20], Preece and Shneiderman’s Reader-to-Leader Framework [43], and the Fairy Rings of Participation model [14, 36].

There are also parallels with standard web marketing ideas around ‘conversion’ of visitors through to site-specific goals, such as site registration and online purchases, and click-through and conversion rates for online advertising.

It should be stressed, however, that the funnel of participation does not presuppose a fixed outcome: it requires only a shared form of registration, a shared form of activity, and some notion of what it would mean to progress, however open-ended.

There are two key features of the funnel: steep drop-off from each stage to the next, and steeply unequal patterns of participation. These ‘fat-tailed’ distributions (often mislabelled ‘power laws’) are characteristic of most if not all online social activity. It has been long noted as a feature of open networks [3, 44, 49], and is also seen in formal education where the activity is online [47].

5. EMPIRICAL UNDERPINNING

Table 1: Summary analytics for three different learning sites, from site opening to 7 Nov 2012. ‘Visits’ and ‘Unique visitors’ from Google Analytics. ‘Registrations’ and ‘Contributors’ (have made at least one contribution) from site databases.

<table>
<thead>
<tr>
<th></th>
<th>iSpot</th>
<th>Cloudworks</th>
<th>openED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visits</td>
<td>1,100,000</td>
<td>275,000</td>
<td>30,000</td>
</tr>
<tr>
<td>Unique visitors</td>
<td>390,000</td>
<td>165,000</td>
<td>15,500</td>
</tr>
<tr>
<td>Registrations</td>
<td>21,000</td>
<td>5,239</td>
<td>1,429</td>
</tr>
<tr>
<td>Contributors</td>
<td>9,000</td>
<td>1,750</td>
<td>198</td>
</tr>
<tr>
<td>Contributor rate</td>
<td>2.3%</td>
<td>1.0%</td>
<td>1.3%</td>
</tr>
</tbody>
</table>

The funnel of participation is underpinned by empirical data. Three examples are presented here, from three entirely separate learning websites. The first is iSpot (www.ispot.org.uk), a social learning community aimed at helping beginners learn to identify nature observations [14, 37, 53]. The second is Cloudworks (www.cloudworks.ac.uk), a professional learning community for educators and educational researchers [15, 16]. The third is openED (www.open-ed.eu), an open online course in business and management aimed at postgraduate/practitioner level [1, 13].

While there is some overlap in the team behind these three sites (including the author), the user communities are entirely distinct, with only a handful of users present on more than one of them.
Table 1 shows summary figures for participation on the three sites. Two features leap out. First is the dramatic fall-off in each step in greater involvement. The second is how similar the rates of attrition are across the three sites.

Closer analysis of these three communities yields further examples of the funnel, with steep drop-off, and highly unequally distributed patterns of activity. For iSpot, this pattern has been explored at length [14]; an updated analysis yielded no significant differences. For Cloudworks, the funnel can be seen in the number of users making different numbers of contributions (Table 2). For openED, the pattern is explored in more detail in Section 6 of this paper.

**Table 2: Number of users who have made given numbers of contributions to Cloudworks.**

<table>
<thead>
<tr>
<th>Contributions</th>
<th>0</th>
<th>1-5</th>
<th>5-9</th>
<th>10-49</th>
<th>50+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users</td>
<td>3,489</td>
<td>1,322</td>
<td>192</td>
<td>192</td>
<td>44</td>
</tr>
</tbody>
</table>

The funnel of participation has been observed on other MOOCs and similar sites. For instance, PLENK2010 had 1,641 registrations, but “about 40-60 individuals on average contributed actively to the course on a regular basis” [25], or 2.4-3.7%—close to the overall contributor rate seen above. On CCK08, there were 2,200 participants [19]; there were 83 respondents to the post-course survey, of whom 15 said they had completed the entire course (and 13 of those were studying formally for credit) [24]. Similarly, on Athabasca Landing, a social learning site for (formal) students at Athabasca University, 78% of the content is created by 21% of the users, and around 18-21% of the users are active [45].

The funnel is also apparent—at least at the very coarsest level—in reported completion rates for xMOOCs. These are variously said to be “less than 10%” of registered students completing the course [28] or “generally between 10 and 20 percent” [30]. The first MITx course, Circuits and Electronics, attracted over 150,000 participants, but “fewer than half look at the first problem set”, and only 7,157 passed, or about 5% [18]. Coursera’s first Software Engineering course enrolled 50,000 students, of whom 3,500 passed, or 7% [40].

These rates are considerably lower than for conventional higher education, where in the UK completion rates are over 90% for highly selective high-status universities, and above 60% for universities with a broader social mission. It is notable, however, that rates for online and distance universities fall somewhere between conventional HE and xMOOCs: University of Phoenix completion rates are 31-36% for undergraduate-level degrees [54], while completion rates for the Universitat Oberta de Catalunya (UOC, Open University of Catalonia) have ranged between 33% and 67% [27].

**6. THE OPENED 2.0 COURSE**

Having looked broadly at evidence of the funnel of participation, this paper now looks more closely at one specific example: the openED 2.0 course. The openED 2.0 project explored a framework for collaborative, open, multi-institutional development of a course, coupled with an open, online model of delivery. Seven European organisations worked together to create the course, largely based on existing Open Educational Resources (OER). The main learning environment was a customised version of Joomla, with additional learning support provided by email, IRC chats, and Eluminate conferencing. The course was presented three times between 2010 and 2012. The principles and rationale for openED and related courses are articulated extensively elsewhere [39], and an account of the approach to the design of the course has been published [1]. The project’s deliverables included a full report of the evaluation [13]. This section presents a further quantitative analysis of the participation data.

Visit data could be attributed with confidence to 691 individuals; 199 individuals made posts to the course forums. As can be seen from Figures 2 and 3, the distribution of both visits and posts to the forum was steeply unequal, or ‘fat tailed’. Neither, however, follow a power law (see [11]). The visit data could be connected to the forum data for 178 users; there was a clear correlation between the two ($R = 0.86, p < 0.0001$), as would be expected. These patterns of steeply-unequal participation and steep, staged drop out fit the key characteristics of the funnel of participation.

**7. DISCUSSION**

The funnel of participation is a real, significant phenomenon in MOOCs and related courses. Compared to formal learning, there tends to be much higher rates of drop-out, and steeply unequal patterns of participation. This is probably an almost-inevitable consequence of any open, online activity: there is less initial commitment, so the filtering happens at a later stage [38]; and the well-attested tendency for steeply unequal patterns of participation to emerge in online activity is manifest.

The phenomenon shows that MOOCs alone cannot replace degrees or most other formal qualifications. The significant efforts
that institutions put in to supporting their learners to reach a commonality of learning outcome are necessary, and have a real effect. As Siemens [50] argues, the long-term value for universities is likely to lie in precisely those things that cannot be cheaply duplicated through a MOOC.

Does it matter if MOOCs have high drop-out rates? Some argue that it is a positive sign of an exploratory phase [46]; Daniel [18] points out that answers to this question “create a sharp distinction between the xMOOCs providers and other distance learning institutions”, with the xMOOCs observing that early drop-outs do not add significantly to costs [30]. What constitutes drop-out and completion can be a complex problem, particularly for online and distance institutions, such as the UK Open University [6] and UOC [27]: rates are highly sensitive to their precise definition, and vary widely between courses. Is it drop-out, or non-continuation, or climb-out? This is a long-standing issue for distance educators [3], and is a bigger question in MOOCs, because the phenomenon is so much larger. Where we have indications of problems (e.g. the evidence that some learners find eMOOCs confusing [24, 38]), we have a responsibility to do what we can to address them.

Learning analytics offers great potential, but the choice of intervention in a MOOC may be niew problematic. For example, in a formal situation, a prediction of likely failure to complete is instantly meaningful, relevant, and can be mediated by skilled learning professionals, and the learner can be supported by a range of existing resources and specialists.

The value in learning analytics comes from closing the loop effectively to complete the Learning Analytics Cycle [12]. The funnel of participation shows that this is a particular challenge for MOOCs and similar open, online environments: they tend towards steep drop-offs and highly unequal patterns of participation. However, the experience of online and distance teaching institutions, where the rates of drop-out fall somewhere between conventional courses and MOOCs, suggests that it is possible to mitigate the impact of the funnel. There is likely to be significant value in further work to empirically explore and validate how learning analytics can help learners in a MOOC context.

8. ACKNOWLEDGEMENTS

The author wishes to thank James Aczel for his invaluable contribution to the creation and launch of the openED project, and Simon Cross for his work on the project. Thanks are also due to the project partners and the participants in the course. The openED course was funded by the European Commission through the Lifelong Learning Programme as project 505667-LLP-1-2009-1-PT-KA3-KA3MP. This paper remains, however, the sole responsibility of the author.

9. REFERENCES


