Death by a thousand clicks: Making sense of learner data and sense-making strategies in MOOCs

Mairéad Nic Giolla Mhichíl[1], Conchúr MacLochlainn[1], and Elaine Beirne[1]

[1] Dublin City University

Abstract. MOOCs generate huge amounts of data. The provision of digital learning at scale to a global learner base provides access to a variety of data sets ranging from learner digital activities to learner contributions. This paper sets out the methodological case for adopting particular sense-making strategies and a variety of techniques to aid with data reduction, analysis and presentation with respect to interrogating MOOC data to support specific learning design or redesign objectives. The findings from the initial implementation of these strategies are provided based on a series of MOOCs delivered in the area of Irish Language and Culture by Dublin City University via the FutureLearn platform. Furthermore, it illustrates how the use of a mix of both qualitative and quantitative approaches can provide both rich and measured insights to give MOOC learners an active and vocal voice within the learning design process.

Keywords: Research Design, Sense-making, Mixed Methods

1 Introduction

Of the many ways to die, "death by data asphyxiation" [1, p.281] may be amongst the worst for those engaging in social scientific research. In an increasingly-connected and recorded world, our challenge, once how to define, apply and operationalise key constructs we wish to study, has arguably given way to a form of data saturation. In such a world, where we are confronted at every turn by sources of data unimaginable even 30 years ago, understanding how to build a coherent framework for analysing and making sense of such forms of information may be a key concern. What is the relationship between logs of behaviour /clickstream data and learners' thoughts and emotions? What can we expect to understand from large scale analyses, given that each person behind the screen may be quite different, and interpret their own behaviour differently?

This challenge has been particularly-noted with regards to research into MOOCs: in many cases, the *embarras de riches* that researchers have been presented with, in addition to raising some ethical quandaries [2] also forces us to make decisions upon what is meaningful, and what is actionable. While it is a truism to suggest not all which can be counted counts, understanding the learners behind the screens [3] may be a particular challenge when we have limited access to said persons. As a consequence, in this paper we attempt to move beyond clickstream data, and suggest

how to integrate the wide array of findings we have discovered in our own research to forming a coherent whole. Critically, it is not an argument for the abandonment or limitation of such research: quantitative data is extremely useful, yet it is much more useful when used for specific means. We will use several examples from our own work, and discuss the importance of a broad palette of research methods and strategies as a means of improving MOOC research.

2   Sense-making as a concept and its role in applied MOOC research

A coherent analytical framework is in many ways prerequisite for social research, in that it shapes our inquiry, yet the generally high-level nature of much existing research into MOOCs has not leant itself easily to such consideration. Indeed, the focus has perhaps been on describing at the macro-most level, who MOOC learners are, and what typical patterns of participation they may engage in, rather than asking how this research informs design and re-design, to provide a better learning experience. As argued by Reich [4, p.34] in reviewing studies attempting to analyse learner behaviour using an array of statistical metrics, "It does not require trillions of event logs to demonstrate that effort is correlated with achievement". While understanding that the above correlational relationship exists can be useful, it tells us very little of the means through which this correlation may occur, when it is the latter which we would likely wish to explain. Veletsianos, Collier and Schneider [5, p.571] provide an equally-succinct take on the issue, arguing that though we might understand what is being done by learners, understanding the why and how of this behaviour are considerably more mercurial, if we rely on such data alone.

These concerns relating to lacking a coherent theoretical girding to research are not anything particularly new. In a sense, all social research faces this difficulty from balancing both an approach geared towards description and explaining [6]. Indeed, many qualitative researchers place great emphasis on the analytical role of extremely detailed and local description [7]. A further distinction between variance-based forms of inquiry, where the focus is primarily on understanding differences between persons, and process-based research, understanding how particular events have emerged, has been adopted across a range of disciplines and has strong philosophical underpinnings [8, 9]. The former has been the dominant focus of research into MOOC research, very little research has considered the latter, despite the fruitful means by which this may be done which we will describe below.

Though the above may appear at first glance to map neatly onto a typical quantitative-qualitative divide, with a variance-based model linked with quantitative research, and process-based with qualitative, this need not be the case at all. Both perspectives are of use in analysing different forms of data, which speaks to the importance of being unafraid to mix methods.

3   Types of strategy and design considerations

Perhaps one of the most important considerations in making sense of a complex phenomenon, is to combine many different forms of data to underpin a coherent analysis. Adopting a mixed-methods research strategy provides a means through which different forms of data can be incorporated into an analytical whole (or as contrasting forms of information, equally) [10, 11]. An important corresponding premise is the rejection of viewing different sources of data as fundamentally incompatible [12]. For some, this blending of methods may include relying on qualitative measures to support a statistical argument, while for others some quantitative measures may be used to support a primarily narrative or qualitative design (see [10] for examples of differing taxonomies). In many cases, the specific blending of methods and analysis is the goal, with an array of designs and conceptual categories of how to incorporate varying forms of how this mixing may occur [13]. These can include questions of staggered collection, mixing particular sources of data and iterative design frameworks, none of which have a priori obvious answers.

Regardless of exact design or philosophical stance, there is a general paucity of studies which appear to make use of the kind of explicit mixing which is increasingly-common in other forms of research. Though many MOOC studies make use of terms such as 'mixed-methods' this is rarely explicitly linked to a research paradigm, logic or framework [14, p. 497]. This is despite many articulations of the importance of same to broadening the field and the types of topics that we may consider [15, p. 214].

Our research centres on a suite of Irish language and culture MOOCs. Our research interests span applied linguistics, educational psychology and broader aspects of design and redesign, and have thus been both inter-disciplinary and have drawn upon theoretical frameworks outside the field of MOOC research. We will now discuss the three principles that guided this process.

3.1   Holistic interpretation and data triangulation

In explaining behaviour in the MOOC, we have typically made use of multiple sources of data, such as survey instruments and forum contributions, as complementary, yet distinct, way of sense-making. When measuring the affective reactions of learners to specific tasks, we were (and are) struck by the ways in which the statistical averages we observed (both positive and negative) contrasted with the sentiment expressed in forum contributions. An interesting example on a particular activity was of forum contributions yielding feedback considerably more negative than those expressed in an anonymous survey result on the same activity.

This example presents an actionable research hypothesis, in that seeing negative responses in a forum may encourage others to post similar messages, while an anonymous survey, with no social, external referents, may temper this dynamic considerably. This highlights the potential role of forum participation as a mediator and perhaps even amplifier, in certain contexts, of emotions. It also points to the limitations of findings from this data collection method. Observing the survey results alone, however, would also limit the findings as qualitatively-different types of

feedback expressed in forum posts would be removed and thus limit our understanding of the organic and social nature of participation [16].

Triangulation is also possible, and indeed desirable, in analysis, particularly in "enriching our understanding by allowing new or deeper dimensions to emerge" [17, p.604]. Essential to this is considering how different theoretical conceptions and levels may complement, enrich or perhaps even contradict each other. A vibrant debate exists on the degree to which fully holistic data integration and interpretation is possible, with Bryman [18], in a review of perceived justifications for conducting mixed-methods research, raising 19 different (non-exhaustive) purposes. The range of possibility, as such, is quite great and moving MOOC research forward likely involves considering how this kaleidoscope of options might prove fruitful in analysing data collected on MOOC platforms.

A related point which follows is to under-utilised store of rich qualitative data typically generated over the length of a MOOC, perhaps due to the comparative difficulty of analysing same. For example, the above-mentioned example involving discussion posts provides valuable contextual information in their own right, and can be analysed not purely using a macro-level method such as content analysis, but also in viewing the forum contributions as an ecosystem or socially-constructed form of participation [19] opening MOOC research to different types of questions and methodological frameworks. Some research on these types of dynamics has been conducted (see [20] for a recent ethnographic study of MOOC participation) yet much more could be done. It would also move the field beyond a semi-positivistic view of participation towards a wider base of different types of research.

3.2 Longitudinal research

Understanding change clearly requires a strong awareness for temporal dimensions, and in this regard MOOCs may be perfect for capturing both harder forms of information (such as logs, timestamps, surveys) that can also be paired with the subjective elements of how learners construe their own participation over time, through interviews, open-ended comments or focus groups. Temporality is an important and necessary component of sense making, and is necessary to underpin and conceptualise event sequencing. It further enables us to move from viewing time as the trigger of change to a useful measure through which to explore change [21, p. 98]. Longitudinal research, and considering multiple forms and points of data collection, allows the tracking of how attitudes, beliefs and participation change over an extended period. It follows also that though much research on MOOCs has focused on learner dropout rates, mapping engagement with different types of tasks that make up a MOOC is an equally instructive way to consider how the structure of a MOOC may play a role in participation (and indeed what factors most influence observable dropout). This is particularly the case as several qualitative studies point to time itself as a major factor in whether learners continue [22, p. 143), suggesting that it is not merely satisfaction/dissatisfaction that has an effect but also wider social and personal life circumstances. Our research has involved using staggered survey instruments to measure how different tasks may provoke different types of emotional responses.

Tracked over time, this can be used to consider whether patterns may relate to specific steps or tasks, and can then be analysed using supporting sources of data, such as subjective learner accounts or auto-narratives of their own participation.

3.3   Iterative research re-design

Iterative research provides the researcher with the capacity to consider the impact of interventions, changes and refinements on new groups of learners supporting a multi-staged research design. It also follows that small changes in a MOOC over course runs might allow quasi-experimental conditions, which could be taken beyond a simple pre or post-test to incorporate explicitly learner (and lecturer, see [10]) feedback in course redesign. Importantly, this is a process that may reveal tensions, such as the gaps that may exist in learner expectations or desires versus actual technical capabilities on the part of course instructors [23, p. 9]. This co-constructive process may be very valuable in considering multiple perspectives on participation, and equally in listening to the learner voice in considering aspects of redesign.

In our case, the above-mentioned example relating to negative emotional feedback, both articulated by learners and measured using statistical instruments, lead to direct redesign of several steps. We found that providing additional scaffolding, audio support and re-arranging activities, all suggested by a range of learners, useful in subsequent runs of the course, and a clearly more-positive response. This type of research is deeply practical, but can also be embedded strongly in deductive reasoning, experimental principles and seeking gradual improvements as a direct consequence of research.

4   Lessons and conclusion

We would argue that a more philosophically-grounded form of research and sense-making is required in MOOCs than what has typically emerged to this point, and to link to wider disciplines which are readily-implicated in MOOC design, participation and interaction. This may entail in particular forms of staggered, sequential data collection, a mixing of different types of instruments and different paradigm stances. This may even include looking at the tensions that emerge between different stances, and what they may illuminate about the promise (and limits) of various methods and ways of sense making [24]. Conceptually, our lack of access to directly observe learners presents contrasts with that of in-classroom research. Sustained, long-term qualitative research, perhaps encompassing observations of multiple actors within a particular MOOC, would be highly useful and might tell us much about the actual process and dynamics of MOOC learning. Though a category such as 'qualitative research' is necessarily very broad, there is a clear paucity of studies employing the methods and philosophical framework of an interpretive, constructivist paradigm is a clear gap in the research literature on MOOCs. Most MOOC research has veered towards neo-positivistic lines of inquiry and sense-making strategies [15, p. 214]. Qualitative inquiry as a more general subject is of critical importance in "using

methodologies that celebrate richness, depth, nuance, context, multi-dimensionality and complexity rather than being embarrassed or inconvenienced by them" [25, p. 1]. This complexity, locality and embeddedness would augment very well the strong body of existing research into learning analytics, and enable a movement from considering the macro-most aspects of MOOC participation to a sharper, more-limited and contextual focus on why this may be the case in particular contexts. This type of research would also enable us to see how different types of MOOCs may vary radically from each other and the role of participation, membership and wider social life as factors in these important elements of MOOC research. Many of the issues we have raised are central and much-discussed within wider, non-MOOC specific research, such as methodological triangulation, actionable research, and an awareness of the importance of staggered/multiple points of data collection. MOOC research presents much promise and exploiting these opportunities for more varied and grounded research will likely lead to better research, and perhaps as a welcome consequence, better MOOCs.

References

1. Pettigrew, A. Longitudinal field research on change: Theory and practice. Paper presented at the National Science Foundation Conference on Longitudinal Research Methods in Organizations, Austin (1988).
2. Harriman, S., and Patel, J. The ethics and editorial challenges of interest-based research. BMC Medicine 12, (2014).
3. Macleod, H., Haywood, J., Woodgate, A., and Alkhatnai, M. Emerging patterns in MOOCs: Learners, course designs and directions. Techtrends 59(1), 56-63 (2015).
4. Reich, J. Rebooting MOOC research. Science Magazine 347 (62197), 34-35 (2015).
5. Veletsianos, G., Collier, A. and Schneider, E. Digging deeper into learners' experiences in MOOCs: Participation in social networks outside of MOOCs, notetaking and contexts surrounding content consumption. British Journal of Educational Technology, 570-587 (2015).
6. Punch, K. Introduction to Social Research: Quantitative and Qualitative Approaches. Thousand Oaks: SAGE (2014).
7. Geertz, Clifford. The Interpretation of Cultures. New York: Basic Books (1973).
8. Mohr, L. Explaining Organisational Behaviour. San Francisco: Jossey-Bass (1982).
9. Maxwell, J. and Mittapalli, K. Realism as a stance for mixed methods research. In Tashakkori, Abbas and Teddlie, Charles (Eds.) 'SAGE Handbook of Mixed Methods Research'. Thousand Oaks: SAGE, 145-168 (2010).
10. Creswell, J. and Plano-Clark, V. Designing and Conducting Mixed Methods Research. Thousand Oaks: SAGE. 3rd Edition (2018).

11. Mason, J. Mixing methods in a qualitatively driven way. Qualitative Research 6 (9), 9-25 (2006).

12. Howe, K. Against the quantitative-qualitative incompatibility thesis, or Dogmas die hard. Educational Researcher 7(8), 10-16 (1998).

13. Greene, J., Caracelli, V. and Graham, W. Towards a conceptual framework for mixed-methods evaluation designs. Educational Evaluation and Policy Analysis 11, 255-274 (1989).

14. Raffaghelli, J, Cucchiara, S. and Persico, D. Methodological approaches in MOOC research: Retracing the myth of Proteus. The British Journal of Educational Technology 46(3), 488-509 (2015).

15. Veletsianos, G. and Sheperdson, P. A Systematic Analysis and Synthesis of the Empirical MOOC Literature Published in 2013–2015. International Review of Research in Open and Distributed Learning 17(2), 198-221 (2016).

16. Chua, Shi Min, Tagg, Caroline, Sharples, Mike and Rienties, Bart. Discussion Analytics: Identifying Conversations and Social Learners in FutureLearn MOOCs. Working Paper, 7th International Learning Analytics and Knowledge Conference, 36-62 (2017).

17. Jick, Todd. Mixing Quantitative and Qualitative Research: Triangulation in Action. Administrative Sciences Quarterly 24, 602-611 (1979).

18. Bryman, A. Integrating quantitative and qualitative research: how is it done? Qualitative Research 6(1), 97-113 (2006).

19. Saadatmand, M. A new ecology for learning: An online Ethnographic Study of Learners' Participation and Experience in Connectivist MOOCs. Doctoral Thesis. Helsinki Studies in Education 9 (2017).

20. Loizzo, N., Ertnmer, P., Watson, W. and Watson, S. Adult MOOC learners as self-directed: Perceptions of motivation, success and completion. Online Learning 21(2), (2017).

21. Ployart, R., and Vandenberg, R. Longitudinal Research: The Theory, Design, and Analysis of Change. Journal of Management 36(1), 94-120 (2010).

22. Eriksson, T., Adawi, T., and Stohr, C. ''Time is the bottleneck'': a qualitative study exploring why learners drop out of MOOCs. Journal of Computing in Higher Education 29, 133-146 (2017).

23. Michos, K., Herández-Leo, D. and Jiménez, M. How educators value data analytics about their MOOCs. Proceedings of EMOOCs 2017: Work in Progress Papers of the Experience and Research Tracks and Position Papers of the Policy Track. 77-82 (2017).

24. Ortega-Arranz, A., Sanz-Martínez, L., Alvarez- Alvarez, S, Munoz-Cristóbal, J., Bote-Lorenzo, M., Martínez-Monés, A and Dmitriadis, Y. From Low-Scale to Collaborative, Gamified and Massive-Scale Courses: Redesigning a MOOC. EMOOCs 2017: Digital Education: Out to the World and Back to the Campus, 77-87 (2017).

25. Greene, J. and Hall, J. Dialectics and Pragmatism: Being of consequence. In Tashakkori, Abbas and Teddlie, Charles (Eds.) 'SAGE Handbook of Mixed

methods in Social & Behavioural work research'. Thousand Oaks: SAGE, 119-144 (2010).