

Research Ethics of Twitter for MOOCs

Eamon Costello, Enda Donlon, and Mark Brown
Dublin City University

Abstract

This study examined the ethical considerations researchers have made when investigating MOOC learners' and teachers' Twitter activity. In so doing, it sought to address the lack of an evidence-based understanding of the ethical implications of research into Twitter as a site of teaching and learning. Through an analysis of 31 studies, we present a mapping of the ethical practices of researchers in this area. We identified potential ethical issues and concerns that have arisen. Our main contribution is to seek to challenge researchers to engage critically with ethical issues and, hence, develop their own understanding of ethically appropriate approaches. To this end, we also reflected and reported on our own evolving practice.

Keywords: social media, privacy, research ethics, MOOCs, Twitter

Costello, E., Donlon, E., & Brown, M. (2019). Research ethics of twitter for MOOCs. *Online Learning, 23*(3), 252-269. doi:10.24059/olj.v23i3.1564

Research Ethics of Twitter for MOOCs

There is a significant body of research literature related to the implications and applications of Twitter for teaching, learning, and scholarly activity (Gao, Luo, & Zhang, 2012; Williams, Terras, & Warwick, 2013; Dunlap & Lowenthal, 2009; Veletsianos, 2012). The open, online nature of Twitter meant that it was quickly adopted and co-opted by MOOC designers and teachers but also informally by MOOC learners. MOOC teachers have used it as a secondary learning space (Salmon et al., 2015) or as part of the fabric of MOOC itself (Bozkurt et al., 2016). Research has been conducted using Twitter on a wide diversity of topics, such as MOOC learner experience (Kop, 2011), the acquisition of social capital (Joksimović, Dowell, et al., 2015) and comparisons between what happens within a MOOC and what learners say on social media (Joksimović, Zouaq, et al., 2015). Researchers have analyzed hashtag aggregations, creating large Twitter data sets made of many MOOC teachers and learners across multiple courses. Research has been undertaken of learners according to user sentiment (Shen & Kuo, 2015), temporality of learning (Zhang et al., 2015), level and types of learner discussion (Veletsianos, 2017), and the influence of various actors in discussions about MOOCs (Costello et al., 2017).

Twitter opens possibilities for researchers who are not MOOC providers (i.e., not directly involved in the provision of the MOOC itself). MOOC providers have access to discussion forum postings, assessment work, and activity logs. This information, for many reasons that include

strong ethical ones, may be tightly controlled by these stakeholders. In contrast, the tweets that emanate beyond the MOOC walls into Twitter are there for any researcher to potentially gather and analyze.

This treasure trove of data is alluring. It is accessible and free to harvest—a digital data feast that has been described in terms of its latent value as akin to gold or oil (Hirsch, 2013). To pursue the analogy, however, we know that the extraction and exploitation of oil has many downsides. There may be a great many insights we could unlock from mining digital data, but there are equally consequences that this activity could have that may be unforeseen, unintended, or at worst wilfully ignored. This became apparent to members of our research team during a research study of the hashtag #MOOC on Twitter (Costello et al., 2016; Costello et al., 2017). In our study we sought to analyze the discourse of MOOCs on Twitter by conceptualized Twitter users as actors in a form of networked public.

In conducting a literature review of research for a study of Twitter and MOOCs, (Costello et al., 2018), we noted that researchers had varying practices with regard to ethics. This was in truth not an initial major concern of our literature review, and it was only during the project as we began asking questions about how we should best report data and results that we found ourselves contemplating this question. Our initial belief was that Twitter data was in the public domain and did not pose many complications for use in low-risk research. However, practices of some scholars in this area (Veletsianos, 2017; Koutropoulos et al., 2014), and cautions from other research traditions, such as medical health research (Conway, 2014), gave us pause. This, combined with the ethical context of the new European Union General Data Protection Regulation (GDPR), which has increased the legal obligation to protect citizens' data rights, spurred us to conduct a systematic analysis of ethical practices and concerns of the relevant research literature. The overarching aim of the research presented here was to interrogate the underlying ethical assumptions that had informed our initial research design. We aimed to do this by asking what ethical considerations other researchers have made when investigating MOOC learners' and teachers' Twitter activity.

Ethics and Big Data

Ethical consideration can be seen as a cornerstone of any research endeavor. Indeed, it can be argued that ethics are integral to professional academic practice (Bruhn et al., 2002). The need for increased ethical awareness has been highlighted for research into teaching and learning at scale (Slade & Prinsloo, 2013; Ferguson & Buckingham Shum, 2012). This clarion call has been made in an era where the phenomenon of big data has been posited as an interplay of “technology, analysis, and mythology that provokes extensive utopian and dystopian rhetoric” (boyd & Crawford, 2010). Selwyn (2012) cites social media as one example of technology that is socially disruptive and that poses deep “ideological (rather than purely technical) questions” of institutional education. This is in part because big data may be oversimplified as an “educational fix” (Enyon, 2013). In other words, the issues of big data are neither neutral, nor straightforward.

If we follow such warnings of deep sociotechnological entanglement, ethics becomes an imperative cornerstone of research design. It is therefore surprising that there is such a dearth of relevant literature on the ethical considerations of research on MOOCs (the few articles on the subject include Esposito, 2012; Rolfe, 2015; Marshall, 2016). Learning analytics research often deals with MOOCs, and there are some ethical methods and practices discussed in the literature here (Slade & Prinsloo, 2013, 2014; Ferguson & Buckingham Shum, 2012; Siemens, 2013; Manca, Cavaglione, & Raffaghelli, 2016). Moreover, in social media research there is well-developed

research literature concerned with the analysis and discussion of ethical research (Taylor & Pagliari, 2018; Conway, 2014). This is particularly notable in medical health research, where researchers may be working in areas with obvious privacy implications, such as illness, addiction and so on (Conway, 2014). We will now expand on some key themes of this literature.

Privacy and Informed Consent

There are a few key issues that such literature reveals that are relevant for our purposes here—namely, the right to privacy, the complexity of preserving this right, informed consent, and researcher biases. First is the issue of the right to privacy. This can be seen as closely linked to informed consent, when a research participant willingly cedes some privacy through disclosure of information to a research team. The participant is making an informed decision to share their data with the researchers with full knowledge of any potential risks and of the privacy implications. Many ethical review boards or committees see information that is publicly available as essentially unproblematic and not requiring user consent. It has been argued, however, that this conception may be overly simple in complicated modern digital networked environments where what is public and what is private is not always easy for people to discern. It is easy to overshare, to not realize the privacy settings of systems such as Facebook or Twitter:

Given a general lack of deep understanding by most researchers and research subjects of the technical operation of the Internet, private and public spaces on the Internet can really only be understood in term of metaphors. (Anderson & Kanuka, 2009, p. 119)

Best practice from medical social media research holds that data should not be reproduced without being anonymized and de-identified even if it is in the public domain (Conway, 2014).

Privacy options of technology may be complicated, confusing, or hidden. Equally, technology can be overly simple and seductive, designed to reel us in, make us feel safe, and hence while away some time in its embrace. It may cause users to overshare. One question that consequently arises is whether it is ethically appropriate not just to collect data without user consent from public spheres but also to reproduce that data. Such reproduction may fail the test of whether such user-generated data is being used in a context intended by its creator.

Preserving Privacy When Publishing Research

Can researchers forgo informed consent if they simply anonymize and de-identify the data? Zheleva and Getoor (2009) showed that making a profile private on a social network may be insufficient to ensure privacy and that metadata, such as group membership, can “leak” information in unforeseen ways. Similarly, data sets that have been “anonymized” in good faith by researchers, such as by the removal of personal identifiers (e.g., names of students), can be reverse engineered to uncover the original identities. This was well recounted by Zimmer (2010) in his analysis of the case of a data set publicly released by researchers of the anonymized Facebook activity of students in an unnamed northwestern university in the United States. Clues quickly led commentators to conclude that the university in question was Harvard, and it soon became apparent that the identity of students could easily be pieced together and either determined or inferred from metadata, including a codebook that described the data set. What this incident revealed was that the sociologists, by their own admission, were not computer or information scientists (Zimmer, 2010). Moreover, it illustrates how linked data can change the nature of information because, once it was known that the university was Harvard, other information, such

as particular class offerings, dorm sizes, and so on, could be used (i.e., linked) to the original data set, which further revealed personal identity likelihoods of individuals.

A number of open data sets, such as those of MOOC learners, have anonymized and de-identified participant data (Manca, Caviglione, & Raffaghelli, 2016) and been published to enable their study by a wide research community. In many ways, Twitter may look like a large open data set and the fact that it is in the public domain may lead us to assume that it does not need to be anonymized or de-identified during research. The development of anonymized and de-identified data is not without particular problems. Firstly, critical meaning may be leached from the data during the process. Daries et al. (2014) showed, using an empirical example, how efforts to obscure the identities of learners in an edX MOOC data set lead to perturbations in the data. They showed how they got different results from an equivalent analysis of raw data set and of a de-identified data set. What this highlights is that ethical designs can have methodological implications or rather that the two are not so easily separable as one might assume.

Twitter is a classic example of data that, because of its open nature, is often assumed to be straightforward to compile. However, in reality the compilation of Twitter data sets requires much expertise, technical computing resources, and access to high-quality data sources, such as a Twitter streaming or firehose API, which can be expensive (Driscoll & Walker, 2014; Morstatter et al., 2013). Moreover, even the very best Twitter data sources cannot yield 100% complete data; hence, researchers are always operating on some sample of the data. The language of streams and fire hoses is apt when considering catching such torrents of digital data, as even the best bucket will incur some spillage. Big data needs to be handled with care. Without proper infrastructure and expertise, there are potentials for data leaks and breaches.

Role of the Researcher

With Twitter and through tweets, accounts of learning can extend beyond the MOOC itself. This has implications for the type of research that is carried out. Researchers who lack access to what is happening within the walled garden of the MOOC may claim to have fewer potential conflicts of interest. Given that they are not teachers, who have particular ethical responsibilities, their role may be one concerned only with research and a fidelity to its process. Much research highlights the pitfall of bias towards positive findings in educational research. One recent study ascribed this bias to factors including the “fuzzy boundaries between learning and teaching research, scholarship and teaching; [and] the positive agendas of ‘learning and teaching’ funding bodies” (Dawson & Dawson, 2016). In other words, the overlapping roles of teacher and researcher can lead to potential conflict. Tensions may arise as to whether ultimate loyalty should lie with the research process, the teaching, the students, or the institution (Dale, 1996; Burman & Kleinsasser, 2004). That is not to say, of course, that teachers cannot be researchers. Positionality is well established in educational research (Merriam et al., 2001). One of its essential aspects is precisely an explicit articulation or conceptualization of the researcher’s role (Denzin, 1989). Indeed, it may do this not to acknowledge that researcher biases *may* exist so much as to assert that they cannot *but* exist (Merriam et al., 2001).

Research Question

Ethics cannot be something bolted onto a research process, some necessary compliance evil. If we consider research as a complicated human activity, and in this case a sociotechnical one, we will see that there are decision points at many stages that have ethical dimensions. Ethical decisions may preserve or betray rights of participants. In the case of the former, do they then

weaken the potential strength of the research? If researchers anonymize, de-identify, and aggregate everything, will the results be less meaningful? Or if they tie themselves up in knots with ethical concerns, will they get anything done?

In an attempt to untangle these questions and provide some handrails, we sought to critically appraise the relevant literature through an analysis of the existing ethical practices of researchers on MOOCs and Twitter as recounted in the literature. The overarching question that guided our research was, hence, formulated as follows: What ethical considerations have researchers reported to have made when investigating MOOC learners' and teachers' Twitter activity in the published literature?

From this, and following from the literature outlined above, we derived the following guiding subquestions:

- Have researchers reported that they sought ethical approval? Have they believed ethical approval to be necessary or exempt by deeming Twitter data to be in the public domain?
- How did the studies report on the collection, processing, and storage of tweets? Did any of the studies make reference to use of linked data?
- How have studies reported conflicts of interest and the role of the researcher or a stance adopted on this role?
- In studies that reproduced tweets, how did they deal with the ethical implications of this, such as by anonymizing and obfuscating the tweets or by seeking the Twitter users' permission to reproduce such tweets?

Methods

The data set for this study was generated by using a systematic literature review approach (Petticrew & Roberts, 2009; Okoli, 2015; Kitchenham, 2004). Our inclusion rules specified that studies must meet five criteria: (1) be concerned with using Twitter (or Sina Weibo, the Chinese version of Twitter) for research into MOOCs, (2) be written in English, (3) be empirical, (4) be published in peer-reviewed conference proceedings or journals, and (5) be published (or available in preprint) during the period between January 2011 and July 2017. We took *empirical* to mean studies that gathered primary data in some form (i.e., not theoretical, conceptual, or opinion pieces). The data could be from Twitter in the form of tweets, or it could comprise tweet and Twitter-user metadata, such as social network structures of Twitter users.

We followed the approaches of existing systematic literature reviews of both MOOC and Twitter research literature (Liyangunawardena, Adams, & Williams, 2013; Veletsianos & Shepherdson, 2016; Gao, Luo, & Zhang, 2012; Williams, Terras, & Warwick, 2013) and derived search terms for MOOCs, Twitter, and their known academic synonyms. This gave us the terms *micro-blogging*, *micro-blog*, *microblogging*, *Microblog*, *Twitter*, and *tweet*, along with *Massive Open Online Course*, *Massively Open Online Course*, and *MOOC*.

Using queries constructed from these terms, we conducted database/indices searches of EBSCO, Web of Knowledge, Google Scholar, IEEE Explore, and Scopus. Google Scholar is permissive in its indexation and is known to include grey literature and non-peer-reviewed work (Haddaway et al., 2015). Hence, we followed Gao et al. (2012) in conducting a full-text search of

Google Scholar articles, sorting them by relevance. After analysing the first 200 results, we determined a lack of further relevant papers.

Results

Overview of the Results

Following the search strategy and application of the inclusion and exclusion criteria described above, we determined that 31 articles could be included in our final corpus. Of these, 20 (65%) were from peer-reviewed journals, and 11 (35%) were from peer-reviewed conference proceedings. The articles described an assortment of methods employed in their studies, which we classified as belong to one of six categories: interviews, surveys/questionnaires, researcher manual qualitative coding of social media content, machine learning analysis of social media content, social network analysis, or other metadata analysis. Figure 1 below shows a high-level mapping of these approaches.

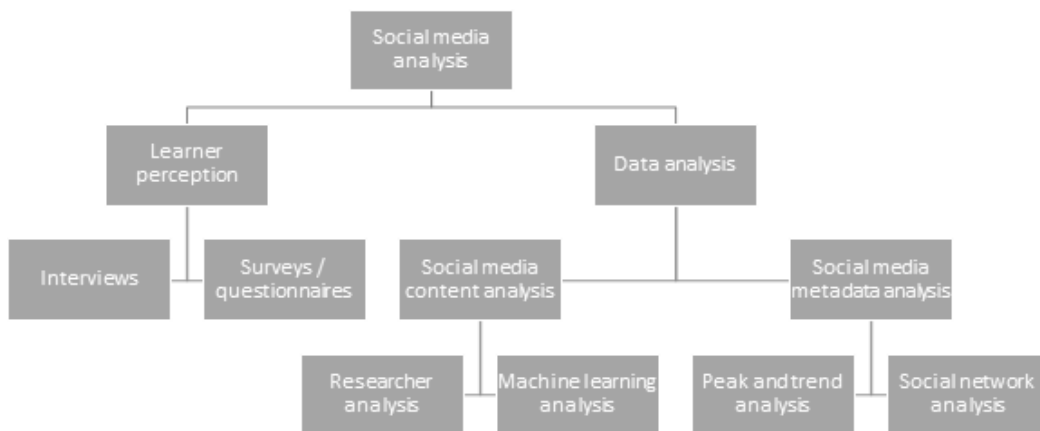


Figure 1. Mapping of the MOOC Twitter literature by research method.

A full detailing of the classes of research method employed by each article is given in Table 1 below. Twenty-one of the studies examined (68%) used more than one research method, while 10 (32%) used only one method. The most common techniques were machine analysis of social media content (employed by 19 studies, or 61%) and metadata analysis (19 studies, or 61%). Other methods were social network analysis (a specialised form of metadata analysis), researcher analysis of social media content, and assessment of learner perceptions via surveys/questionnaires and interviews.

Table 1

Research Methods Employed

	<i>Interviews</i>	<i>Surveys/questionnaires</i>	<i>Researcher analysis of social media content</i>	<i>Machine analysis of social media content</i>	<i>Social network analysis</i>	<i>Other metadata analysis</i>
Abeywardena (2014)			X			
Alario-Hoyos et al. (2014)						X
Alario-Hoyos et al. (2013)	X					X
Bell et al. (2016)	X				X	X
Bozkurt et al. (2016)			X	X		
Chen et al. (2016)						X
Costello et al. (2017)					X	
Costello et al. (2016)			X			
Cruz-Benito et al. (2015)	X	X	X			X
Cruz-Benito et al. (2017)	X	X	X			X
de Keijser and van der Vlist (2014)					X	X
de Waard et al. (2011)	X					
Enriquez-Gibson (2014a)			X			
Enriquez-Gibson (2014b)			X			X
Fournier et al. (2014)	X	X	X	X	X	X
García-Peñalvo et al. (2015)			X			X
Jiang and Kotzias (2016)					X	
Joksimović, Dowell, et al. (2015)			X	X		
Joksimović, Zouaq, et al. (2015)		X	X			X
Knox (2014)	X	X	X			X
Kop (2011)	X	X	X			X
Koutropoulos et al. (2014)			X			
Kravvaris et al. (2016)			X			X
Liu et al. (2016)	X	X				X
Saadatmand and Kumpulainen (2014)	X	X				X
Salmon et al. (2015)	X	X	X			
Shen and Kuo (2015)			X	X	X	X
Skrypyk et al. (2015)					X	
Spilker et al. (2015)	X		X			X
Tu (2014)		X	X	X		
Van Treeck & Ebner (2013)			X			X
Veletsianos (2017)		X	X			X
Yeager et al. (2013)		X		X	X	X
Zhang et al. (2015)						X
Totals	3	14	11	22	12	22

Ethical Approval

Twenty-three (74%) of the articles analyzed did not make any mention of ethics or ethical approval. Only one article (Salmon et al., 2015) reported to have gained approval from an Ethics Review Committee or Institutional Review Board for the research carried out. Six studies (18%) referred to data being in the “public domain” (Chen et al., 2016; Cruz-Benito et al., 2015, 2017; Koutropoulos et al., 2014; Skrypnyk et al., 2015; Bozkurt et al., 2016). Two studies (6%) explicitly stated that due to the public domain nature of the data that their research was exempt from any requirement for institutional approval (Skrypnyk et al., 2015; Bozkurt et al., 2016). Further, Borzut et al. (2015, p. 21) stated that the researchers informed participants that their data would be reproduced in a published study “as a matter of courtesy.” Two studies stated that they were participants of the MOOC they studied (Rhizo14) and that “ethical protocol for use of data was developed in consultation with Rhizo14 participants” (Bell et al., 2016; Bozkurt et al., 2016). Despite not mentioning ethical approval requirements or lack thereof, some studies nonetheless did have some treatment of ethical issues in general (Fournier et al., 2014; Kop, 2011; Bell et al., 2016; de Keijser & van der Vlist, 2014; Skrypnyk et al., 2015). One study stated that that data was collected specifically for the purpose of the study (Skrypnyk et al., 2015). Another study reported a converse situation whereby the researchers “didn’t originally plan to collect Twitter data for this purpose, but since we had this public data we wanted to analyze it” (Koutropoulos et al., 2014, p. 9).

Collection, Processing, and Storage of Tweets

Only 22 articles (71%) discussed how tweets were collected, and of those, not all were explicit with regard to the technique used. Where methods were stated, approaches included Crowdfmap (Koutropoulos et al., 2014), Digital Methods Initiative Twitter Capture and Analysis Toolset (de Keijser & van der Vlist, 2014), GNIP API (Costello et al., 2016), gRSShopper (Fournier et al., 2014; Joksimović, Dowell, et al., 2015; Joksimović, Zouaq, et al., 2015; Skrypnyk et al., 2015), NodeXL (Bozkurt et al., 2016; Tu, 2014), TAGsExplorer (Bell et al., 2016), Twitonomy (Enriquez-Gibson, 2014a, 2014b), Twitter API (Cruz-Benito et al., 2015, 2017; Shen & Kuo, 2015) and web crawlers (Chen et al., 2016; García-Peñalvo et al., 2015; Zhang et al., 2015).

Twenty-one studies (68%) outlined their methods for processing Twitter data with various techniques mentioned, including CohMetrix computational linguistic facility (Joksimović, Dowell, et al., 2015), Dedoose (Salmon et al., 2015), Gephi (Costello et al., 2017; de Keijser & van der Vlist, 2014; Tu, 2014; Yeager et al., 2013), Netlytic (Bell et al., 2016), NVivo (Bozkurt et al., 2016; Fournier et al., 2014; Liu et al., 2016), OpinionFinder (Shen & Kuo, 2015), PHP/MySQL scripts (Veletsianos, 2017), R Big Query (Costello et al., 2016, and 2017; Joksimović, Dowell, et al., 2015), Semantria3 (Abeywardena, 2014), spreadsheets (Cruz-Benito et al., 2015, 2017), TagMe (Joksimović, Zouaq, et al., 2015), Twitonomy and Wordle (Enriquez-Gibson, 2014a, 2014b), and WEKA, SimpleKMeans and Weka ClassifierSubsetEval (Kravvaris et al., 2016).

Eleven (35%) studies discussed the extent to which they had obtained full or representative data sets, making mention of the completeness, or otherwise, of their sampling. Of these, some made claims for a complete or at least largely complete data set (e.g., Bozkurt et al., 2016; Kravvaris et al., 2016; Veletsianos, 2017); others acknowledged that they had deliberately obtained only a “snapshot” of the data (e.g., Abeywardena, 2014; Enriquez-Gibson, 2014b), while others lamented the inadequacy of collection tools at their disposal and the extent to which this

compromised the ability of their data to be truly representative (e.g., Fournier et al., 2014; Joksimović, Dowell, et al., 2015; Joksimović, Zouaq, et al., 2015; Koutropoulos et al., 2014).

Seven studies (23%) made mention of how tweets and tweet metadata were stored. Various storage methods included cloud computing (Costello et al., 2016; Costello et al., 2017), Excel spreadsheets (Abeywardena, 2014), raw data being stored in an HTML file (Liu et al., 2016), and JSON format (Joksimović, Zouaq, et al., 2015; Skrypnyk et al., 2015). Of these, only one study (Skrypnyk et al., 2015, p. 212) discussed how security of and access to data were managed: “All analysed data sets are stored in a secure password-protected personal repository.”

It was not fully clear in all studies whether data was linked between Twitter and other sources, and this was typically not prominently discussed. However, two studies made obvious use of linked data that was critical to their research question. One study by Joksimovic, Zouaq, et al. (2015) used data from inside the MOOC (course discussion topics) and data from Twitter to see whether the conversations were being mirrored in both spaces. Another study, of 16 million MOOC learners, linked data from MOOC user profiles with other data generated by the same people from the social networks LinkedIn, Gravatar, GitHub, and StackExchange (Chen et al., 2016).

Role of the Researcher

Of the 31 articles reviewed, 24 (77%) did not disclose whether or not there were any conflicts of interest in their study or give a stated position or conceptualization of the role of the researcher, such as whether they were a teacher of the MOOC. Two articles stated that there were no conflicts of interest. One of these was via a standard unelaborated statement of “no conflict of interest” (Skrypnyk et al., 2015), while the second was more explicit in stating that teachers or those involved in the deployment of the MOOC had been precluded from involvement in the research team (Salmon et al., 2015). Five studies disclosed that at least one of the authors of the research had also played a role in the teaching or facilitation of the MOOC (Knox, 2014; Koutropoulos et al., 2014; Yeager et al., 2013; Saadatmand & Kumpulainen, 2014; Bell et al., 2016). Two of these elaborated on this to articulate a position on the role of the researcher. Saadatmand and Kumpulainen (2014, p. 21) cite “the researcher’s crucial role in ethnography (Creswell, 1998).” Bell et al. (2016) gave more space to this aspect, declaring that the authors were participants in the MOOC that was the subject of their research, seeing themselves “as both insiders and outsiders (Dwyer & Buckle, 2009)” to the research process. Bell et al. (2016) also identified “the dangers of projecting our experiences onto those of others. We did not want to speak ‘for’ others but rather to engage with what was said.” Accordingly, the authors purported to counter this risk by “participant observation and the retrospective study of course archives” (Bell et al., 2016).

Reproduction of Tweets

Six studies (18%) reproduced tweets. Four studies reproduced tweets in their original, unmodified form, and two studies anonymized the tweets in some way. This is shown in Table 2 below, along with the number of tweets. In Veletsianos (2017) tweets were anonymized and de-identified. This was confirmed by searching via Google for the text of the tweets, as shown in the study, which returned no results.

Table 2
Tweets Reproduced

	Tweets reproduced	Anonymised & de-identified
Salmon et al. (2015)	1	No
Liu et al. (2016)	3	No
Bozkurt et al. (2016)	6	No
Costello et al. (2016)	17	No
Veletsianos (2017)	20	Yes
Koutropoulos et al. (2014)	60	Yes

Discussion

Almost three quarters of the studies 261ehaviou did not contain any mention of ethics. It may be that ethical considerations were taken by the authors but not mentioned in their studies. In such possible cases we can only say that ethics are considered not important enough to deserve a specific mention. A hint as to this rationale may lie in the six studies that mentioned “public domain data,” considering it essentially unproblematic and to obviate the need for any informed consent. One study, the exception that proved the rule, reported having Ethics Committee/Institutional Review Board approval. Researchers seem to not believe that Twitter could be anything other than a public sphere and not potentially a private space. Moreover, their reasoning can be seen to flow from a harm-based conception of privacy (Bloustein, 1964). However, privacy can also be dignity based:

Such a stance recognizes that one does not need to be a victim of hacking, or have a tangible harm take place, in order for there to be concerns over the privacy of one’s personal information. Rather, merely having one’s personal information stripped from the intended sphere of the social networking profile, and amassed into a database for external review becomes an affront to the subjects’ human dignity and their ability to control the flow of their personal information. (Zimmer, 2010, p. 321)

We found that there was underreporting of data collection techniques and tools. This raises issues around properly interpreting results and the replicability of studies. However, it also has ethical implications, as collecting data from official APIs will not return tweets that users have chosen to delete. For example, the accounts of deceased people can be deleted by Twitter and will not appear via official Twitter API searches but may appear in screen-scraping techniques (Driscoll & Walker, 2014). Official APIs will also be less likely to return tweets which have been removed under European Union laws around the Right to Be Forgotten (European Parliament and Council, 2016).

We found little evidence of protocols around data storage, deletion, or data access. Again, researchers may believe that as the data is in the public domain, then they are simply parsing something that is already freely available. However, analysis itself, specifically using machine learning techniques, generates new data. This is particularly stark in the case of linked data where data from one source is combined with data from one or more other sources. As we found, one

study combined data about users from several social networks. This study by Chen et al. (2016), which conducted a very large-scale analysis of MOOC learners on Twitter (and other social media platforms), is also worth noting for other aspects of its approach. It reported that words in tweets indicating maleness (or by their absence femaleness) were “boxers, shaved, haircut, shave, girlfriend.” Words used in tweets reported to be indicative of younger users were “parents, exams, pregnant, youth, mommy.” If nothing else, these examples serve to illustrate that social media is far from a formal sphere. Rather, it may be one of chat and intimate expression. Certainly, researchers would do well to guard against an overly reductive focus in such contexts (Baruh & Popescu, 2017; Selwyn, Henderson, & Chao, 2018).

The issue arises as to whether use of this data for research into MOOCs constitutes use in a context beyond that intended by the original creator and publisher of that data. In the interest of a research behaviour, a research team may believe they have good reasons to collect personal data; however, linking this data with other data in sophisticated ways raises many issues. Research teams could require ethical practices in the handling and access of such data. More critically perhaps, they would need expertise in protection of that data from hacking, theft, or leaks. Even the most sophisticated and well-resourced of organizations have learned that they cannot control or protect data once in digital form. The studies we 262ehaviou were all conducted before the European Union’s GDPR had come into effect (European Parliament and Council, 2016). There are many aspects of these studies, at least in the reporting of their results, that indicate that they could be in breach of these regulations were they conducted today. It would therefore be interesting to revisit this issue in the future to see whether the GDPR has had any effect on researchers’ practice.

The vast majority of studies that replicated tweets and tweet metadata (such as usernames) in their publication made no attempt to obfuscate this data. Indeed, the two studies that did so stood out in this context (Veletsianos, 2017; Koutropoulos et al., 2014). Though we might consider this best practice for social media research (Conway, 2014), one commentator on a blog post purported to be frustrated that the data was anonymized in one of these studies. He suggested that this made the results less meaningful and lessened the possibility of analysis by other researchers (Downes, 2017). This may illustrate the design decisions that researchers must take and defend in their community and certainly confirms that “ethical issues of digital and open data sets are puzzling and demand careful attention” (Manca, Caviglione, & Raffaghelli, 2016). Given this situation, it is also interesting to note here that so few researchers disclosed their own role in the research process or addressed the issue of researcher positionality.

To fully disclose our own part in this research, we have included our own publications and subjected them to the same levels of analysis as all others, even those that showed what could be argued to be higher levels of ethical best practices. Indeed, this was one of the motivations of the present study—to reflect critically and honestly on our own practice by comparing it with others in the field. For instance, we have republished tweets in full and also tweet metadata in some of our papers. At the time that research was undertaken and published, we believed (as a majority of the studies we 262ehaviou here) that this information was in the public domain and that, as we carefully reviewed the tweets to ensure that they did not contain any sensitive information, this was sufficient. With the benefit of the hindsight afforded by this study, would we have done the same today? Perhaps not, and in later published research (Costello et al., 2018), we modified our practice to anonymize and de-identify republished tweets. However, the issue is not as straightforward as deciding whether to publish information. It may involve weighing a balance of

commitments, priorities, and fidelities to different, potentially conflicting, codes or principles. Our intention has been to critically examine and explore ethical issues with regard to specific research. Hence, we used the research into the effects and implications of the phenomenon of MOOCs as reflected in Twitter as an example. However, this limited the scope of our study. Although ethical issues in Twitter research have been studied (Taylor & Pagliari, 2018; Conway, 2014), much less attention has been paid to the ethical practices of MOOC researchers. This topic requires future study.

A limitation of this study is that absence of evidence is not evidence of absence. Lack of mention of ethical concerns does not mean they have not been considered. Hence, we make a key recommendation that editors and publishers require authors to make explicit details of the following: ethical approvals they have been granted or waived, ethical considerations they have made to protect participant privacy, and information on how data is handled, processed and stored. Moreover, such considerations should be included in the published article or its metadata for readers (Committee on Publication Ethics, 2018).

A key aim of this paper's contribution is to highlight how, when studies are carried out beyond a certain scale, or when data are linked with other data or processed in sophisticated ways, new possibilities but also new responsibilities arise. If big data can be said to have a "social life" (Perrotta & Williamson, 2016), then we argue that researchers would do well to engage as a community in an ongoing response by reflecting critically on our own practice towards its improvement.

References

- Abeywardena, I. S. (2014). *Public opinion on OER and MOOC: A sentiment analysis of Twitter data*. Paper presented at the International Conference on Open and Flexible Education (ICOFE2014), Hong Kong SAR, China.
- Alario-Hoyos, C., Pérez-Sanagustín, M., Delgado-Kloos, C., Muñoz-Organero, M., Hugo, A., & Parada, G. (2014). Delving into participants' profiles and use of social tools in MOOCs. *IEEE Transactions on Learning Technologies*, 7(3), 260–266. Retrieved from <http://doi.ieeecomputersociety.org/10.1109/TLT.2014.2311807>
- Alario-Hoyos, C., Pérez-Sanagustín, M., Delgado-Kloos, C., Parada, G. H. A., Muñoz-Organero, M. H. A., & Rodríguez-de-las-Heras, A. (2013). Analysing the impact of built-in and external social tools in a MOOC on educational technologies. In D. Hernández-Leo, T. Ley, R. Klamma, & A. Harrer (Eds.), *Proceedings of the 8th European Conference on Scaling up Learning for Sustained Impact – Volume 8095* (pp. 5–18). Berlin: Springer. Retrieved from https://doi.org/10.1007/978-3-642-40814-4_2
- Anderson, T. D., & Kanuka, H. P. (2009). Ethical conflicts in research on networked education contexts. In U. Demiray & R. C. Sharma (Eds.), *Ethical practices and implications in distance learning* (pp. 109–125). Hershey, PA: Information Science Reference.
- Baruh, L., & Popescu, M. (2017). Big data analytics and the limits of privacy self-management. *New Media & Society*, 19(4), 579–596. Retrieved from <https://doi.org/10.1177/1461444815614001>
- Bell, F., Mackness, J., & Funes, M. (2016). Participant association and emergent curriculum in a MOOC: Can the community be the curriculum? *Research in Learning Technology*, 24(1). Retrieved from <http://dx.doi.org/10.3402/rlt.v24.29927>
- Blank, G. (2017). The digital divide among Twitter users and its implications for social research. *Social Science Computer Review*, 35(6), 679–697. Retrieved from <https://doi.org/10.1177/0894439316671698>
- Bloustein, E. J. (1964). Privacy as an aspect of human dignity: An answer to Dean Prosser. *NYUL Rev.* 39, 962.
- boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15(5), 662–679. Retrieved from <https://doi.org/10.1080/1369118X.2012.678878>
- Bozkurt, A., Honeychurch, S., Caines, A., Bali, M., Koutropoulos, A., & Cormier, D. (2016). Community tracking in a cMOOC and nomadic learner 264ehaviour identification on a connectivist rhizomatic learning network. *Turkish Online Journal of Distance Education*, 17(4), 4–30. doi:10.17718/tojde.09231
- Bruhn, J. G., Zajac, G., Al-Kazemi, A. A., & Prescott, L. D., Jr. (2002). Moral positions and academic conduct: Parameters of tolerance for ethics failure. *The Journal of Higher Education*, 73(4), 461–493.
- Burman, M. E., & Kleinsasser, A. M. (2004). Ethical guidelines for the use of student work: Moving from teaching's invisibility to inquiry's visibility in the scholarship of teaching and learning. *Journal of General Education*, 53(1), 59–79. doi:10.1353/jge.2004.0018

- Chen, G., Davis, D., Lin, J., Hauff, C., & Houben, G. (2016). Beyond the MOOC platform: Gaining insights about learners from the social web. In W. Nejdl (Ed.), *Proceedings of the 8th ACM Conference on Web Science, Hannover, Germany, May 22–25, 2016* (pp. 15–24). New York: ACM Press. doi:10.1145/2908131.2908145
- Committee on Publication Ethics. (2018). Guidance for editors: Research, audit and service evaluations. Retrieved from <https://publicationethics.org/>
- Conway, M. (2014). Ethical issues in using Twitter for public health surveillance and research: Developing a taxonomy of ethical concepts from the research literature. *Journal of Medical Internet Research*, 16(12), e290. doi:10.2196/jmir.3617
- Costello, E., Brown, M., Mhichíl, M. N. G., & Zhang, J. (2018). Big course small talk: Twitter and MOOCs—a systematic review of research designs 2011–2017. *International Journal of Educational Technology in Higher Education*, 15(1), 44.
- Costello, E., Brown, M., Nair, B., Mhichíl, M. N. G., Zhang, J., & Lynn, T. (2017). #MOOC friends and followers: An analysis of Twitter hashtag networks. European Conference on Massive Open Online Courses. *EMOOCs 2017: Digital Education: Out to the World and Back to the Campus*, 170–175. Springer. https://doi.org/10.1007/978-3-319-59044-8_19
- Costello, E., Nair, B., Brown, M., Zhang, J., Nic Giolla Mhichíl, M., Donlon, E., & Lynn, T. (2016). *Social media #MOOC mentions: Lessons for MOOC research from analysis of Twitter data*. Paper presented at ASCILITE, 2016. Retrieved from http://2016conference.ascilite.org/wp-content/uploads/ascilite2016_costello_concise.pdf
- Couldry, N. (2014). A necessary disenchantment: Myth, agency and injustice in a digital world. *The Sociological Review*, 62(4), 880–897. doi:10.1111/1467-954X.12158
- Cruz-Benito, J., Borrás-Gené, O., García-Peñalvo, F. J., Blanco, Á F., & Therón, R. (2015). *Detection of non-formal and informal learning in learning communities supported by social networks in the context of a cooperative MOOC*. Paper presented at the 2015 International Symposium on Computers in Education (SIIE).
- Cruz-Benito, J., Borrás-Gené, O., García-Peñalvo, F. J., Blanco, Á F., & Therón, R. (2017). Learning communities in social networks and their relationship with the MOOCs. *IEEE Revista Iberoamericana De Tecnologías Del Aprendizaje*, 12(1), 24–36. doi:10.1109/RITA.2017.2655218
- Dale, H. (1996). Dilemmas of fidelity: Qualitative research in the classroom. In P. Mortensen, & G. Kirsch (Eds.), *Ethics and representation in qualitative studies in literacy* (pp. 77–94). Urbana, IL: NCTE.
- Daries, J., Reich, J., Waldo, J., Young, E., Whittinghill, J., Ho, A., ... Chuang, I. (2014). Privacy, anonymity, and big data in the social sciences. *Communications of the ACM*, 57(9), 56–63. doi:10.1145/2643132
- Dawson, P., & Dawson, S. L. (2016). Sharing successes and hiding failures: ‘Reporting bias’ in learning and teaching research. *Studies in Higher Education*, 43(8), 1405–1416. doi:10.1080/03075079.2016.1258052
- De Keijser, T., & van der Vlist, F. N. (2014). Following UvAMOOC on Twitter – A network analysis of a massive socio technical and cross platform online learning environment. *Online*. Retrieved from <https://www.fernandovandervlist.nl/papers/following-uvamooc-on-twitter.html>

- Denzin, N. K. (1989). *Interpretive biography* (Vol. 17). Thousand Oaks, CA: Sage.
- Downes, S. (2017). A large-scale study of Twitter use in MOOCs (Reply to Veletsianos). Retrieved from <http://www.veletsianos.com/2017/02/01/a-large-scale-study-of-twitter-use-in-moocs/#comment-42574>
- Driscoll, K., & Walker, S. (2014). Big data, big questions | Working within a black box: Transparency in the collection and production of big Twitter data. *International Journal of Communication*, 8, 1745–1764.
- Dunlap, J. C., & Lowenthal, P. R. (2009). Tweeting the night away: Using Twitter to enhance social presence. *Journal of Information Systems Education*, 20(2), 129.
- Enriquez-Gibson, J. (2014). Following hashtag (#)MOOC: Mobility of online courses on Twitter. In S. Bayne, C. Jones, M. de Laat, T. Ryberg, & C. Sinclair (Eds.) *Proceedings of the 9th International Conference on Networked Learning 2014* (pp. 111–120). Edinburgh: University of Edinburgh.
- Enriquez-Gibson, J. (2014). Tracing digital traces of phatic communion in #Mooc. In T. Arenas Yáñez & O. Saavedra Rodríguez (Eds.), *Proceedings of the 9th International Conference on E-Learning, Technical University Federico Santa Maria Valparaiso, Chile, 26–27 June 2014* (pp. 45–49). Reading: Academic Conferences and Publishing International Limited.
- Enyon, R. (2013). The rise of big data: What does it mean for education, technology, and media research? *Learning, Media and Technology*, 38(3), 237–240. doi:10.1080/17439884.2013.771783
- Esposito, A. (2012). Research ethics in emerging forms of online learning: Issues arising from a hypothetical study on a MOOC. *Electronic Journal of e-Learning*, 10(3), 315–325.
- European Parliament and Council. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. Retrieved from <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32016R0679>
- Ferguson, R., & Buckingham Shum, S. (2012). Social learning analytics: Five approaches. In S. Buckingham Shum, D. Gasevic, & R. Ferguson (Eds.), *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, Vancouver, BC, Canada, April 29–May 02, 2012* (pp. 23–33). New York: ACM Press.
- Fournier, H., Kop, R., & Durand, G. (2014). Challenges to research in MOOCs. *MERLOT Journal of Online Learning and Teaching*, 10(1), 1–15. Retrieved from <http://nparc.cisti-icist.nrc-cnrc.gc.ca/npsi/ctrl?action=shwart&index=an&req=21275387&lang=en>
- Gao, F., Luo, T., & Zhang, K. (2012). Tweeting for learning: A critical analysis of research on microblogging in education published in 2008-2011. *British Journal of Educational Technology*, 43(5), 783–801. doi:10.1111/j.1467-8535.2012.01357.x
- García-Peñalvo, F. J., Cruz-Benito, J., Borrás-Gené, O., & Blanco, Á F. (2015). Evolution of the conversation and knowledge acquisition in social networks related to a MOOC course. In P. Zaphiris & A. Ioannou (Ed.), *Learning and Collaboration Technologies: Second International Conference, LCT 2015. Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015* (pp. 470–481). Cham: Springer.

- Haddaway, N. R., Collins, A. M., Coughlin, D., & Kirk, S. (2015). The role of Google Scholar in evidence reviews and its applicability to grey literature searching. *PloS One*, *10*(9), e0138237. doi:10.1371/journal.pone.0138237
- Hirsch, D. D. (2013). The glass house effect: Big data, the new oil, and the power of analogy. *Maine Law Review*, *66*(2), 373.
- Jiang, S., & Kotzias, D. (2016). Assessing the use of social media in Massive Open Online Courses. *arXiv*:1608.05668. Retrieved from <http://arxiv.org/abs/1608.05668>
- Joksimović, S., Dowell, N., Skrypnik, O., Kovanović, V., Gašević, D., Dawson, S., & Graesser, A. C. (2015). How do you connect?: Analysis of social capital accumulation in connectivist MOOCs. In J. Baron, G. Lynch, J. Jovanovic, S. Lonn, & J. E. Willis III (Eds.), *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge, Poughkeepsie, New York, March 16-20, 2015* (pp. 64–68). New York: ACM Press.
- Joksimović, S., Zouaq, A., Kovanović, V., Gašević, D., Jovanović, J., & Hatala, M. (2015). What do cMOOC participants talk about in social media? In J. Baron, G. Lynch, N. Maziarz, P. Blikstein, A. Merceron, & G. Siemens (Eds.), *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge, Poughkeepsie, New York, March 16-20, 2015* (pp. 156–165). New York: ACM Press. doi:10.1145/2723576.2723609
- Kitchenham, B. (2004). *Procedures for performing systematic reviews*. Staffordshire: Keele University.
- Knox, J. (2014). Digital culture clash: “Massive” education in the E-learning and digital cultures MOOC. *Distance Education*, *35*(2), 164–177. doi:10.1080/01587919.2014.917704
- Kop, R. (2011). The challenges to connectivist learning on open online networks: Learning experiences during a Massive Open Online Course. *The International Review of Research in Open and Distance Learning*, *12*(3), 19–37. doi:10.19173/irrodl.v12i3.882
- Koutropoulos, A., Abajian, S. C., deWaard, I., Hogue, R., Keskin, N. Ö, & Rodriguez, C. O. (2014). What Tweets tell us about MOOC participation. *International Journal of Emerging Technologies in Learning (iJET)*, *9*(1), 8–21. doi:10.3991/ijet.v9i1.3316
- Kravvaris, D., Kermanidis, K., & Ntanis, G. (2016). How MOOCs link with social media. *Journal of the Knowledge Economy*, *7*(2), 461–487. doi:10.1007/s13132-014-0219-2
- Lerman, K., Marin, L., Arora, M., Costa de Lima, L., Ferrara, E., & Garcia, D. (2018). Language, demographics, emotions, and the structure of online social networks. *Journal of Computational Social Science*, *1*(1), 209–225. doi:10.1007/s42001-017-0001-x
- Liu, M., McKelroy, E., Kang, J., Harron, J., & Liu, S. (2016). Examining the use of Facebook and Twitter as an additional social space in a MOOC. *American Journal of Distance Education*, *30*(1), 13–25. doi:10.1080/08923647.2016.1120584
- 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. doi:10.19173/irrodl.v14i3.1455
- Manca, S., Caviglione, L., & Raffaghelli, E. J. (2016). Big data for social media learning analytics: Potentials and challenges. *Journal of E-Learning and Knowledge Society*, *12*(2), 27–39.

- Marshall, S. (2016). Exploring the ethical implications of MOOCs. *Distance Education*, 35(2), 250–262. Retrieved from <https://doi.org/10.1080/01587919.2014.917706>
- Merriam, S. B., Johnson-Bailey, J., Lee, M.-Y., Kee, Y., Ntseane, G., & Muhamad, M. (2001). Power and positionality: Negotiating insider/outsider status within and across cultures. *International Journal of Lifelong Education*, 20(5), 405–416. doi:10.1080/02601370120490
- Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013). Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's firehose. In E. Kiciman, N. B. Ellison, B. Hogan, P. Resnick, & I. Soboroff (Eds.), *Proceedings of the Seventh International AAI Conference on Weblogs and Social Media, Cambridge, MA, July 8-11, 2013* (pp. 400–408). Palo Alto, CA: The AAAI Press.
- Okoli, C. (2015). A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems*, 37(1), 879–910. Retrieved from <https://hal.archives-ouvertes.fr/hal-01574600>
- Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438–450. Retrieved from <https://doi.org/10.1111/bjet.12152>
- Perrotta, C., & Williamson, B. (2016). The social life of learning analytics: Cluster analysis and the 'performance' of algorithmic education. *Learning, Media and Technology*, 43(1), 1–14. Retrieved from <https://doi.org/10.1080/17439884.2016.1182927>
- Petticrew, M., & Roberts, H. (2009). *Systematic reviews in the social sciences*. Malden, MA: Blackwell.
- Rolfe, V. (2015). A systematic review of the socio-ethical aspects of Massive Online Open Courses. *European Journal of Open, Distance and E-learning*, 18(1), 52–71. Retrieved from <https://doi.org/10.1515/eurodl-2015-0004>
- Saadatmand, M., & Kumpulainen, K. (2014). Participants' perceptions of learning and networking in connectivist MOOCs. *Journal of Online Learning and Teaching*, 10(1), 16–30. Retrieved from <http://search.proquest.com/docview/1614680811>
- Salmon, G., Ross, B., Pechenkina, E., & Chase, A-M. (2015). The space for social media in structured online learning. *Research in Learning Technology*, 23(1). doi:10.3402/rlt.v23.28507
- Selwyn, N. (2012). Social media in higher education. *The Europa World of Learning 1*, 1–10.
- Selwyn, N., Henderson, M., & Chao, S-H. (2018). 'You need a system': Exploring the role of data in the administration of university students and courses. *Journal of Further and Higher Education*, 42(1), 46–56. doi:10.1080/0309877X.2016.1206852
- Shen, C-W., & Kuo, C-J. (2015). Learning in Massive Open Online Courses: Evidence from social media mining. *Computers in Human Behavior*, 51, 568–577. doi:10.1016/j.chb.2015.02.066
- Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380–1400. <https://doi.org/10.1177/0002764213498851>

- Skrypnyk, O., Joksimović, S., Kovanović, V., Gašević, D., & Dawson, S. (2015). Roles of course facilitators, learners, and technology in the flow of information of a cMOOC. *The International Review of Research in Open and Distributed Learning*, 16(3), 188–217. doi:10.19173/irrodl.v16i3.2170
- Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, 57(10), 1510–1529. Retrieved from <https://doi.org/10.1177/0002764213479366>
- Taylor, J., & Pagliari, C. (2018). Mining social media data: How are research sponsors and researchers addressing the ethical challenges? *Research Ethics*, 14(2), 1–39. <https://doi.org/10.1177/1747016117738559>
- Tu, H. (2014). Exploring the pattern of Coursera network on Twitter: How far are MOOCs reaching out? *Pennsylvania Communication Annual*, 70(1), 113–129.
- Veletsianos, G. (2012). Higher education scholars' participation and practices on Twitter. *Journal of Computer Assisted Learning*, 28(4), 336–349. <https://doi.org/10.1111/j.1365-2729.2011.00449.x>
- Veletsianos, G. (2017). Toward a generalizable understanding of Twitter and social media use across MOOCs: Who participates on MOOC hashtags and in what ways? *Journal of Computing in Higher Education*, 29(1), 65–80. doi:10.1007/s12528-017-9131-7
- Veletsianos, G., & Shepherdson, P. (2016). 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. Retrieved from <http://search.proquest.com/docview/1770070832>
- Williams, S. A., Terras, M. M., & Warwick, C. (2013). What do people study when they study Twitter? Classifying Twitter related academic papers. *Journal of Documentation*, 69(3), 384–410. doi:10.1108/JD-03-2012-0027
- Yeager, C., Hurley-Dasgupta, B., & Bliss, C. A. (2013). cMOOCs and global learning: An authentic alternative. *Journal of Asynchronous Learning Networks*, 17(2), 133–147. doi:10.24059/olj.v17i2.347
- Zhang, J., Perris, K., Zheng, Q., & Chen, L. (2015). Public response to “the MOOC movement” in China: Examining the time series of microblogging. *The International Review of Research in Open and Distributed Learning*, 16(5), 144–159. Retrieved from <http://www.irrodl.org/index.php/irrodl/article/view/2244/3410>
- Zheleva, E., & Getoor, L. (2009). To join or not to join: The illusion of privacy in social networks with mixed public and private user profiles. In J. Quemada (Ed.), *Proceedings of the 18th International Conference on World Wide Web, Madrid, April 20-24, 2009* (pp. 531–540). New York: ACM Press.
- Zimmer, M. (2010). “But the data is already public”: On the ethics of research in Facebook. *Ethics and Information Technology*, 12(4), 313–325. doi:10.1007/s10676-010-9227-5. /