

Tweeting about exams: Investigating the use of social media over the summer 2016 session

Tom Sutch and Nicole Klir Research Division

Introduction

Twitter is an online social networking service, where users can post short messages (known as tweets). In general these tweets are available publicly and can be searched. As well as enabling users to 'follow' other users, thereby receiving all their tweets, Twitter also facilitates discussion among ad-hoc communities (Bruns & Burgess, 2015) by means of a hashtag: users can tag their tweet with a particular word or phrase, preceded by a hash (#) to allow users to search easily for other tweets containing that word or phrase. The list of 'trending topics' (words or phrases that are especially popular at any given time, as determined by an algorithm) presented on Twitter's front page further enhances Twitter's use for real-time discussion among disparate groups of users. Users can opt to 'retweet' any tweets that they encounter, thus passing on a copy of the tweet to their followers.

In 2013, Twitter reported 15 million monthly active users in the UK (Curtis, 2013). Younger people are more likely to use Twitter than older people: in May 2016, 33 per cent of Twitter users in Great Britain were between 15 and 24 years-old (Statista, 2016). Most analysis of social media use concentrates on adults aged 16 and over¹, but according to a survey conducted for the UK Safer Internet Centre, 37 per cent of 11 to 16-year-olds used Twitter weekly (UK Safer Internet Centre, 2015).

In recent years there has been increased tweeting by exam candidates following the end of their exams, discussing the questions on the paper. It seemed that candidates were now conducting 'exam post-mortems' in public, rather than confining them to private conversations on their way out of the exam hall (Lebus, 2016). The ability of candidates to discuss exams they have just taken with others across the country, or indeed the world, has already led awarding bodies to implement additional measures to ensure the security of their assessments. But this phenomenon has also led to coverage in national media. One notorious example from June 2015 concerned a question in an Edexcel General Certificate of Secondary Education (GCSE) Mathematics paper, and became known as 'Hannah's sweets' due to the context of the question. Candidates discussed the question and its difficulty after their exam on Twitter. Their tweets included memes², images and videos and were accompanied by the hashtag #EdexcelMaths, which began to trend on Twitter, prompting national media to run stories (for example, BBC News, 2015).

It is of interest to know the extent of tweeting about exams, and what users are saying for two reasons: first, to give some insight into the views of exam candidates into the assessments they are taking; and secondly, because stories in the national media about exam questions may indirectly shape public perception of exams and standards.

This article presents an analysis of exam-related tweets in the summer 2016 UK examination session, using real-time data from Twitter. We wanted to establish the extent of exam-related tweeting, any patterns over time, the topics being discussed, and the sentiments being expressed.

Methodological considerations

Social media data is generally available free of charge from rich *Application Programming Interfaces* (APIs) under liberal terms of use. This provides a 'free' source of data enabling extensive analysis to be carried out. Before embarking on such analysis, however, it is important to consider two issues in particular: representativeness and ethics.

Representativeness

The conversations taking place on Twitter are unlikely to come from a random subset of candidates taking exams, so there are immediate questions over how representative this data is, and what inferences can be drawn about the wider population of candidates (see Ruths & Pfeffer, 2014 for a detailed discussion of representativeness). In our context, the following are notable sources of unrepresentativeness:

- It is difficult to ascertain whether a user tweeting about GCSEs (for example) is a current candidate, relaying comments of those who are, discussing GCSEs in general, or is exploiting Twitter's currently trending topics to promote their completely unrelated tweet;
- Candidates tweeting about an exam may not be representative of all candidates. This bias can be split into two parts:
 - Twitter users may not be a random sample of candidates taking the exam (there are likely to be biases by socioeconomic background, school type and gender);
 - Conditional on being a Twitter user, candidates may be more likely to tweet about the exam if they have 'something to say': a strong opinion or emotional response, or an observation that they think would interest or amuse other Twitter users;
- Those tweeting may not be expressing their true feelings, for example, using sarcasm or humour in an attempt to connect with people, seek attention (Rui & Whinston, 2012), gain retweets, or develop a personal 'brand'.

Drawing inferences beyond the data we have towards exam candidates in general is not possible without biases being thoroughly investigated, quantified and accounted for. We do not attempt to do this in this article. However, we believe that the 'raw' data from Twitter is still useful as a source, as it shapes media coverage of exams, and thereby has a potential influence on public attitudes to exams.

1. For example, surveys on internet access carried out by the Office for National Statistics.

2. A meme is an image, video or piece of text spread from person to person. They are commonly humorous in nature and may be slightly adapted before being passed on.

Ethics

Ethics are often overlooked when analysing social media data. For example, in a systematic analysis of academic research using Twitter, Zimmer and Proferes (2014) found that only 4 per cent of studies made any mention of ethical issues or considerations in relation to the research design and data collection methods. In this case, ethics are particularly important because the subjects of interest are young people undergoing the stressful situation of taking high-stakes exams.

Henderson, Johnson, and Auld (2013) identify four ethical dilemmas associated with using social media for research purposes, focusing specifically on the educational research sphere and using children as subjects. These are:

1. **Consent in social media.** This requires a consideration of what is public and private, and there is no consensus on this. The two extremes are considering everything that is *actually* accessible as public, or only material that is *perceived* as public by participants.
2. **Traceability** resulting in a loss of confidentiality. Any verbatim text taken from Twitter is traceable (by searching for it), and researchers cannot be sure that information that is currently private is not subsequently made available at a later date.
3. **Research with children and young people.** These groups are generally considered as vulnerable subjects, necessitating greater sensitivity to consent and confidentiality. Young people have different understandings of privacy when using social media sites, and may not be considering the longer-term ramifications of posting content.
4. **Recognising and responding to illicit/reportable activities** which are evident through analysis of social media data. For example, researchers might find evidence of cyberbullying, incitement, or copyright violation.

Given these issues, we carried out our analysis at a high level by picking out numerical trends, keywords and sentiment, then aggregating, rather than looking at individual tweets. We also discarded unnecessary fields (such as real name and location) when processing the data, and did not quote any verbatim tweet content.

Data collection

We obtained data from the *Twitter Streaming API*³ which provides continuous access to real-time tweets. The particular API endpoint that we used was `statuses/filter` which returns public tweets matching one or more criteria. Access to this endpoint is free of charge, but the results returned are limited to 1 per cent of global tweet volumes.⁴ Twitter is a proprietary service, and details of the algorithms used to filter and present tweets are not published.

The only criteria we specified were on the `track` parameter, which is used to set the keywords to search for. We used the following keywords in order to capture exam-related discussion:

exam	exams	examination	examinations
resit	re-sit	GCSE	GCSEs
A Level	AS Level	A Levels	AS Levels
OCR	Edexcel	AQA	ocrexams
revise	revision	revised	revising

As well as the names of the exams, these keywords include the names and Twitter handles of the three English awarding bodies (AQA and Edexcel tweet from `@aqa` and `@edexcel` respectively; OCR (Oxford, Cambridge and RSA) tweets from `@ocrexams`).⁵ The filter is applied in such a way that searching for 'AQA' will return tweets containing handles or hashtags consisting of this word alone ('`@aqa`' or '`#aqa`') but not hashtags such as '`#aqamaths`'.

It is possible to apply a filter to return geolocated tweets posted from a specified geographical area, for example the UK. We did not opt to do this because only a minority of tweets has associated geolocation. We anticipated that the number of candidates tweeting about particular exams would be small and did not wish to restrict them further, and by inspection we found that most of our data was related to UK exams.

We used *Apache Storm* to process the stream of tweets in real time, carrying out the following additional filtering:

- Discard tweets that are in a language other than English (according to the tweet language field); retain tweets in English, and those with unspecified language;
- Require a stricter match on keywords. It is not possible to specify exact phrases to search for in the `track` parameter. For example, searching for 'A Level' will retrieve tweets that contain the word 'a' and the word 'level'; there is no requirement that they are together. We address this limitation by specifying that the words must be together, and separated by a space or a hyphen.

We saved the data at the level of individual tweets, retaining only the tweet id, text, date/time of tweet, user id, username, and (for retweets and quoted tweets) details of the original tweet.

We processed data from Saturday 14 May to Thursday 14 July 2016. This included the exam session (from Saturday 14 May to the end of June) as well as some background data afterwards to give context. In total we collected 6.44 million tweets during the exam session (from Saturday 14 May to Thursday 30 June, excluding the Half-term week beginning Monday 30 May when no exams took place).

Results

Twitter activity over time

Figure 1 is a simple plot of the number of tweets per day in our sample.⁶ Weekends are shaded to aid interpretation. A pattern within the week is immediately apparent: there are more tweets during weekdays and fewer at the weekends (especially Saturdays). There are also longer-term trends: high numbers of tweets during the start of the session in mid-May, followed by an overall decrease towards the end of the session in late June. The absolute peak was on Tuesday 17 May, when we captured 337,177 tweets. The volume of tweets was markedly lower during Half-term week (beginning Monday 30 May), during which no exams were scheduled.

3. Details are available at <https://dev.twitter.com/streaming/overview>.

4. This restriction is imposed after the filter criteria are applied.

5. We did not attempt to search for WJEC or CCEA (Council for the Curriculum, Examinations and Assessment), nor any of the associated accounts for each of the awarding bodies (for example subject-specific accounts).

6. On Friday 3 and Saturday 4 June we experienced data extraction problems, so we do not have a full set of tweets for these days.

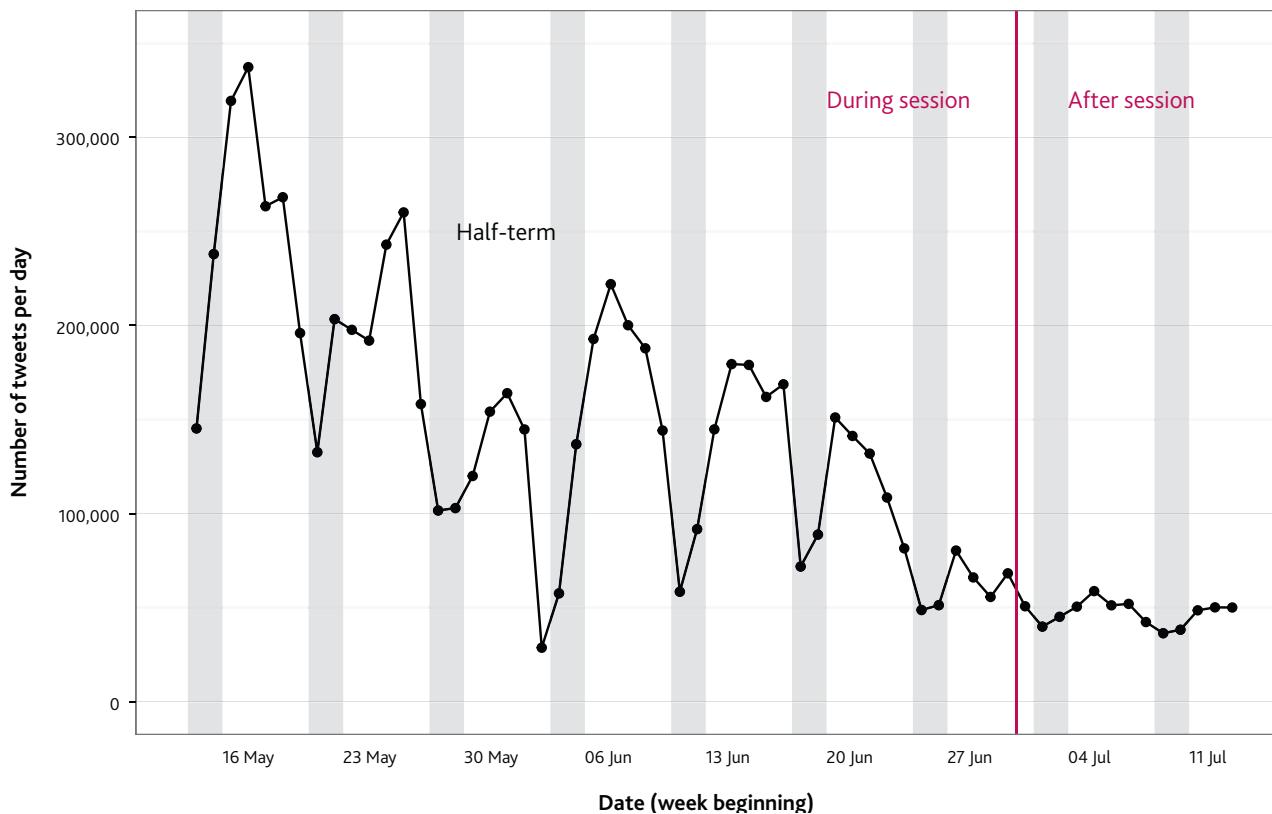


Figure 1: Number of tweets per day, 14 May to 14 July 2016 inclusive

We would expect a certain level of background 'noise'⁷ in the data, because some of our search terms may pick up tweets that are not directly related to UK school exams, and there is general discussion of exams and education on Twitter all year round. Figure 1 shows that after the end of the exam session, the number of tweets we collected was far fewer, so the extent of this background noise is fairly small.

Figure 2 shows the median number of tweets per minute, grouped by weekday.⁸ We have used the median rather than the mean because it gives us a picture of a 'typical' day; the effect of days with extremely high numbers of tweets is minimised.

To aid interpretation, we have highlighted the time periods when exams typically take place. The published starting time of examinations provided by all awarding bodies is either 9:00 a.m. or 1:30 p.m., although exam centres have flexibility to vary these start times by up to 30 minutes. The durations of exams vary; we have highlighted the most common lengths of 1 hour (dark shading) and 1 hour 30 minutes (light shading) after the start time.

The thin grey line shows the typical volume of tweets that we collected during July after the end of the exam session, for comparison. Because the volumes we collected 'in session' are markedly above this baseline, we can have confidence that we were observing genuine trends.

The chart reveals that exam-related Twitter activity follows a distinct pattern. Message volumes gradually increased in the period leading up to an exam, for example, in the evenings and the hours immediately before the scheduled start time. There was much less activity during

exam periods, but message volumes increased after the exam finished. This reinforces the 'public post-mortem' idea (Lebus, 2016).

No examinations were held at the weekends and there was a correspondingly lower number of tweets from Friday evenings onwards. However, message activity did pick up on Sundays as candidates turned their minds to the next day's exams.

Matching search terms

Figure 3 shows the number of exam-related tweets containing selected subject-related search terms on each day during the exam session, represented as a calendar for each subject. Days with more tweets are represented as darker squares. We have scaled the shading in each calendar so that the darkest shade represents the maximum number of tweets for that subject. (The entry for Latin is far lower than for Mathematics, for example, and the number of tweets is similarly lower.)

It is immediately apparent that discussion about exam subjects was concentrated on particular days. The days with the highest number of tweets generally correspond to when exams actually took place, as shown by the coloured frames, and particularly GCSE exams (which have higher entry than Advanced/Advanced Subsidiary levels [A/AS levels]).⁹ In addition, there was often a high number of tweets on the day before the exams. This can be most easily seen for the Sundays preceding Psychology AS level exams on Monday 16 May, and multiple English/English Literature exams on Monday 23 May.

For Biology, there was a large contrast between the number of tweets on Tuesday 17 May and all other days. We recorded 41,806 tweets on this day.¹⁰ This high number is due to students commenting on an AQA GCSE Biology paper which was sat on this day: these comments were picked up by many national media outlets (e.g., Espinoza, 2016).

One notable exception is the high number of tweets about History on Friday 24 June. On inspection, these tweets turned out to be related to

7. That is, tweets that do not directly relate to the exams being sat.

8. We have combined Monday to Thursday because the pattern of tweets is very similar.

9. Subject slots are collectively agreed (Joint Council for Qualifications, [JCQ] 2017, p.4), thus papers are generally timetabled together by all awarding bodies.

10. For comparison, the day with the next highest number of tweets only had 8,359.

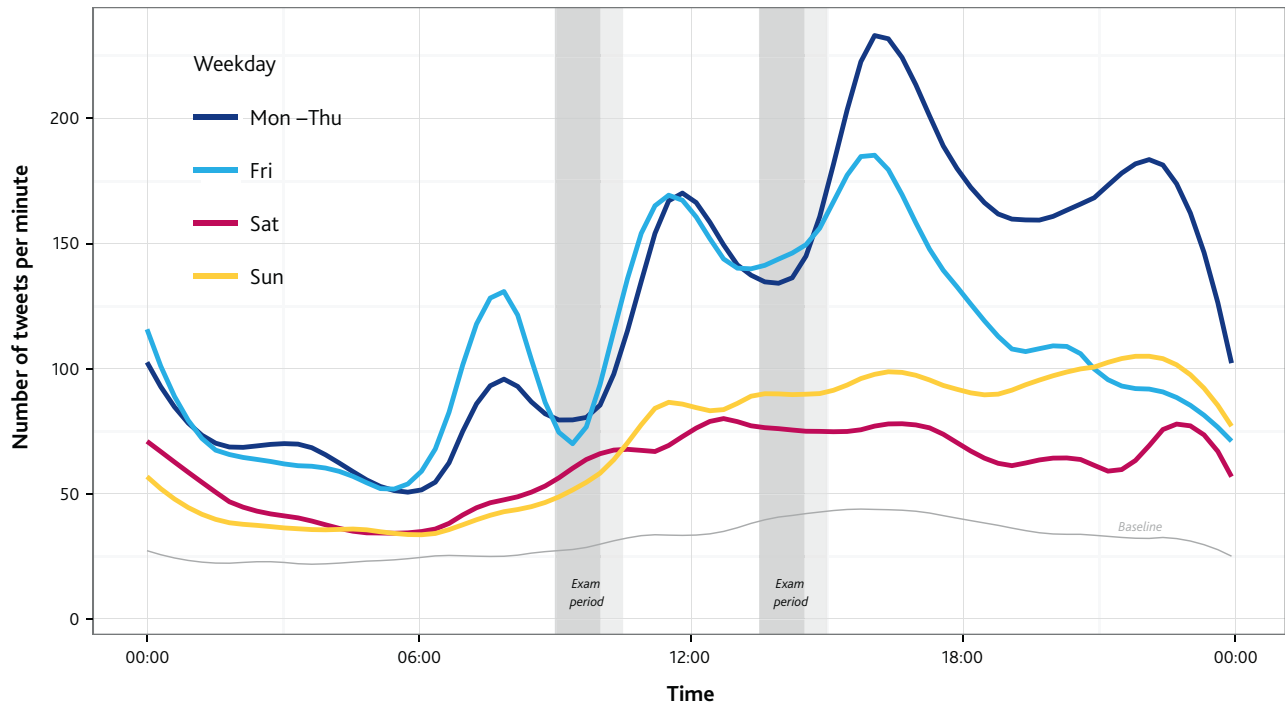


Figure 2: Tweets per time and day of week

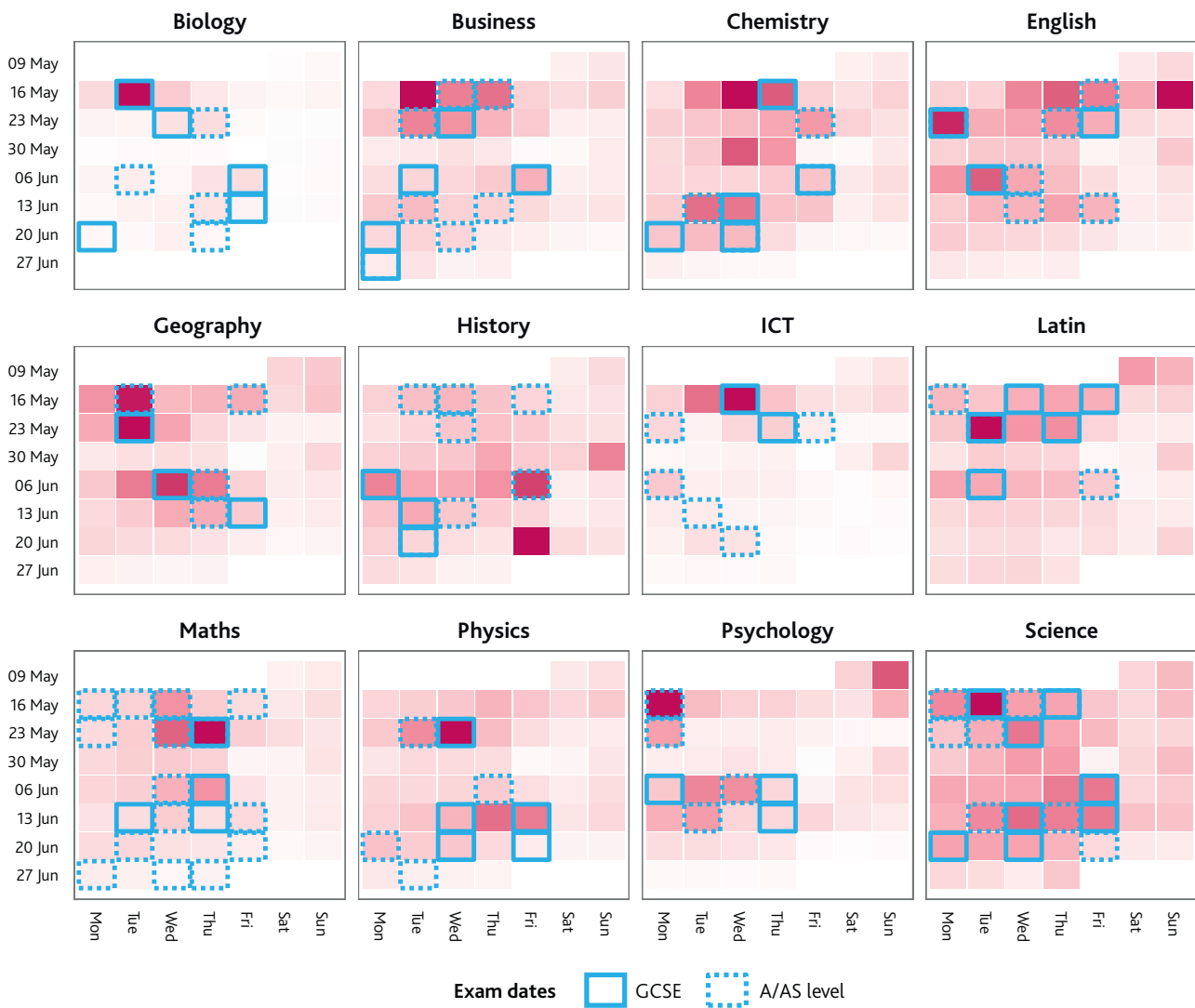


Figure 3: Calendar heatmap showing daily counts of exam-related tweets collected mentioning selected subjects, scaled by the maximum observed for each subject. (Darker squares indicate more tweets)

the result of the EU referendum held on Thursday 23 June, and subsequent political developments: many users were inventing potential GCSE History questions about these events for the future.

Tweet content

To examine the content of the tweets, we counted the occurrence of words over our corpus of tweets. Unsurprisingly the most frequently used words are those we specified as search terms. The top 20 most frequently occurring words are shown below. Those which were not specified as search terms are in bold.

exam	revising	final	school
exams	luck	GCSE	study
revision	day	week	paper
tomorrow	time	students	pass
revise	revised	examination	ur

Other than the search terms we specified, the most frequently occurring words appear to be clustered around forthcoming exams, with the fourth most common word being **tomorrow** and the seventh **luck**. By inspection of a sample of tweets containing these words, we found students were wishing each other luck, or expressing apprehension, before the exam.

To examine tweets over a period when the sample would be as broad as possible, we focused on a particular day and subject: the morning of Thursday 26 May. All the awarding bodies (including CCEA and WJEC) had timetabled a GCSE Mathematics exam for this date, and the vast majority of 16-year-olds would be sitting an exam. The scheduled start time of the exam was 9:00 a.m., and the duration of the papers ranged

from 1 hour to 2 hours. Figure 3 shows that there was a high number of tweets containing 'maths' on this day (compared to other days in the exam series).

We therefore filtered our dataset further to include tweets containing the text 'maths', sent on Wednesday 25 and Thursday 26 May, and excluding retweets. From this subset, we calculated the number of tweets featuring each word. Words were stemmed using the *SnowballC package* (Bouchet-Valat, 2014) before counting (so, for example, counts for 'revising', 'revise' and 'revised' would all be combined).

Figure 4 presents the number of tweets containing selected words before and after the exam. The time of the exam is shaded in grey.¹¹ We see that tweets referring to revision peak on the day before the exam. The words 'hope' and (especially) 'luck' dominate in the hours before the exam. Immediately after the exam, there was a surge of use of the words 'pass' and 'fail', as candidates were evaluating their performance. Words describing difficulty ('easi'¹², 'hard'), and the details of the paper ('question', 'mark', 'scheme') were also used more frequently.

Figure 5 shows the 300 most frequently occurring words used before and after the exam. Before the exam, words expressing anxiety and wishing luck are most common, whereas after the exam, words related to the exam paper and perceived performance dominate.

Sentiment via emoji

Sentiment analysis (SA) is a family of techniques for computationally determining the emotions in text. Sentiment analysis can be applied at various levels: a whole document, a sentence, or an entity such as a single phrase. In Twitter, SA techniques generally aim to determine the sentiment of the tweet as a whole. Sentiment analysis of tweets is more challenging than in other areas as tweets are limited to 140 characters (prompting users to abbreviate words and phrases), frequently include informal language, and may be on any topic in the users' interest (Giachanou & Crestani, 2016).

11. Due to the variation in duration across papers and awarding bodies, we have presented the maximum two-hour duration.

12. 'easi' is a stemmed version of 'easy'.

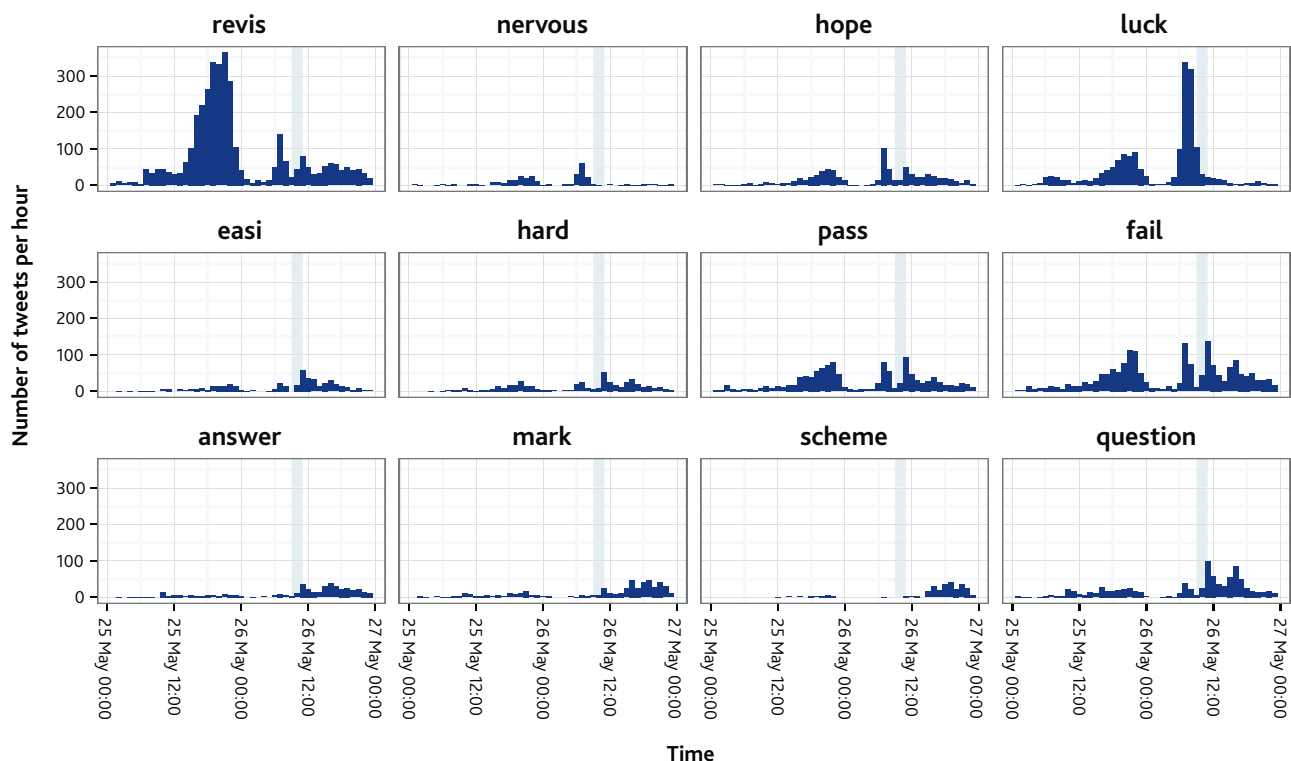


Figure 4: Use of selected words in the period around the GCSE Mathematics exam

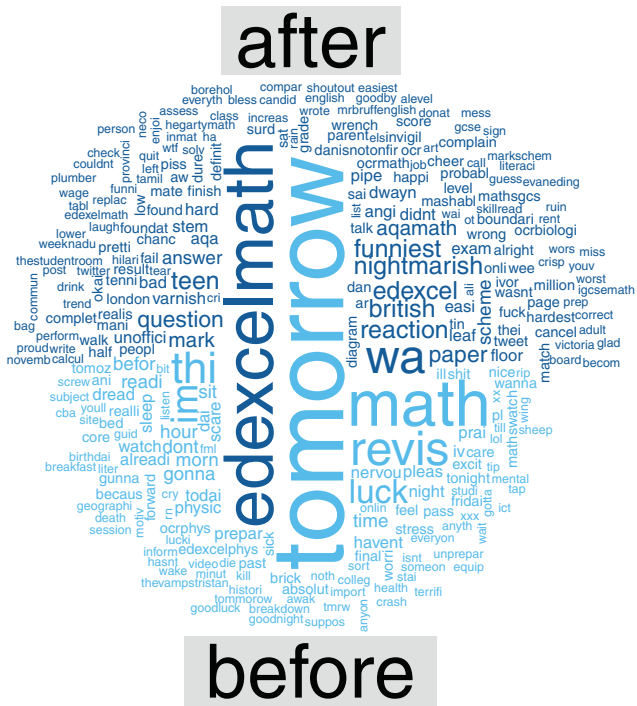


Figure 5: Word cloud of the 300 most frequently used words before (light blue, bottom) and after (dark blue, top) the GCSE Mathematics exam

In this article we focus on a simpler means of determining sentiment: using *emoji*. These are small cartoon images which can be chosen by users to enhance their messages. Many of the most popular emoji are those that express emotion: a smiling face, for example. Hogenboom et al. (2013) found that whenever emoji are used, they can be used as a good proxy for the overall sentiment of a tweet. They are easier to

categorise than standard images accompanying tweets because they are represented as characters in the Unicode character set.

We found that emoji were used in 19 per cent of our tweets, and this proportion was fairly stable over the period we collected tweets. The top emoji in our dataset are shown in Table 2.¹³

Table 2: Top emoji

Image	Description	Number of tweets
😂	face with tears of joy	364,889
😭	loudly crying face	249,542
🙄	upside down face ¹⁴	145,671
😊	smiling face with smiling eyes	140,458
😩	weary face	103,302
🎉	party popper	93,676
🙏	person with folded hands	89,767
🙄	face with rolling eyes	79,922
🙌	person raising both hands in celebration	58,067
🤔	thinking face	57,788

We investigated the trend in the use of emoji over the course of the session, by looking at the percentage of all exam-related tweets (excluding retweets) that included each of the most popular emoji. Figure 6 shows the trend for four selected emoji. There was a gradual decline in the percentage of tweets containing the 'face with tears of joy'

13. Emoji Keyboard (2016) provided free by <http://emoji.com>

14. This emoji is commonly used to denote sarcasm.

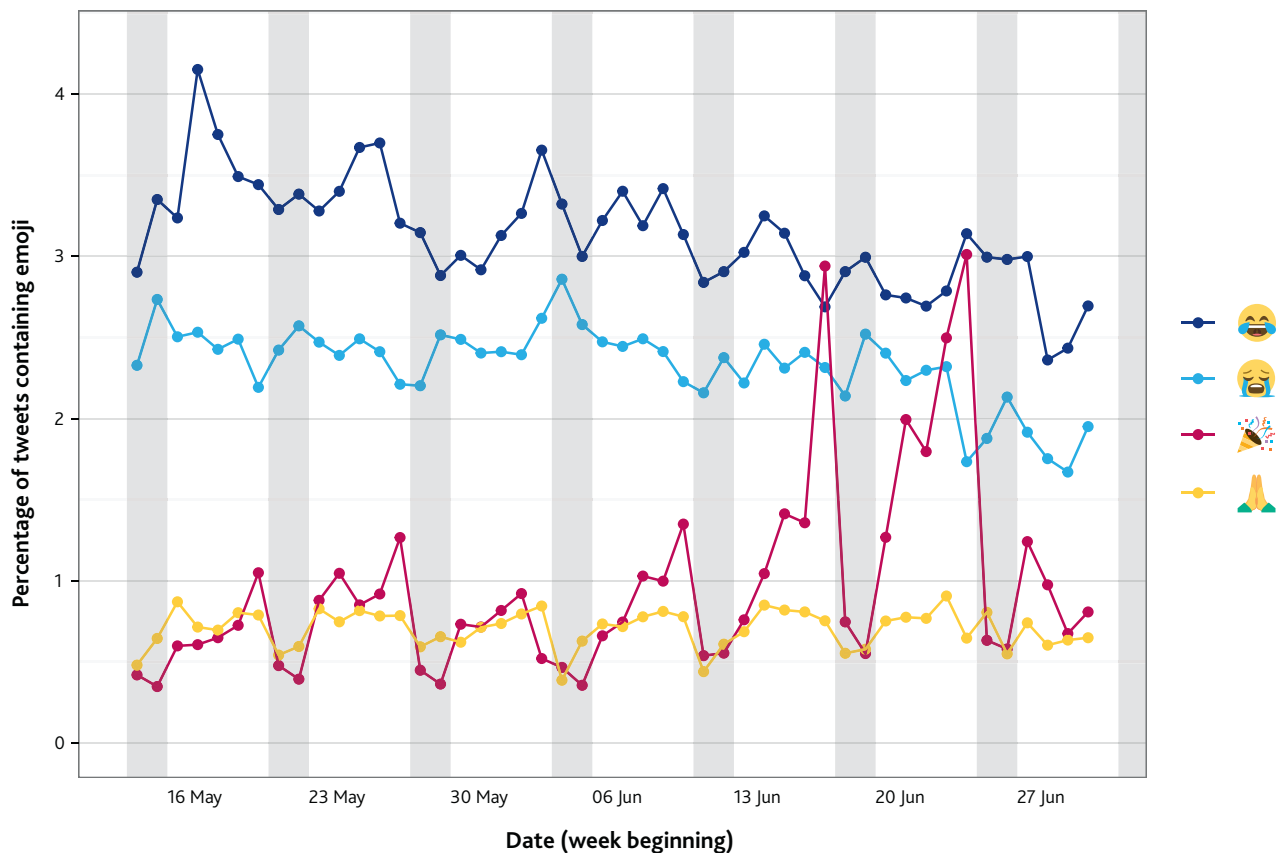


Figure 6: Emoji use over weekday/time

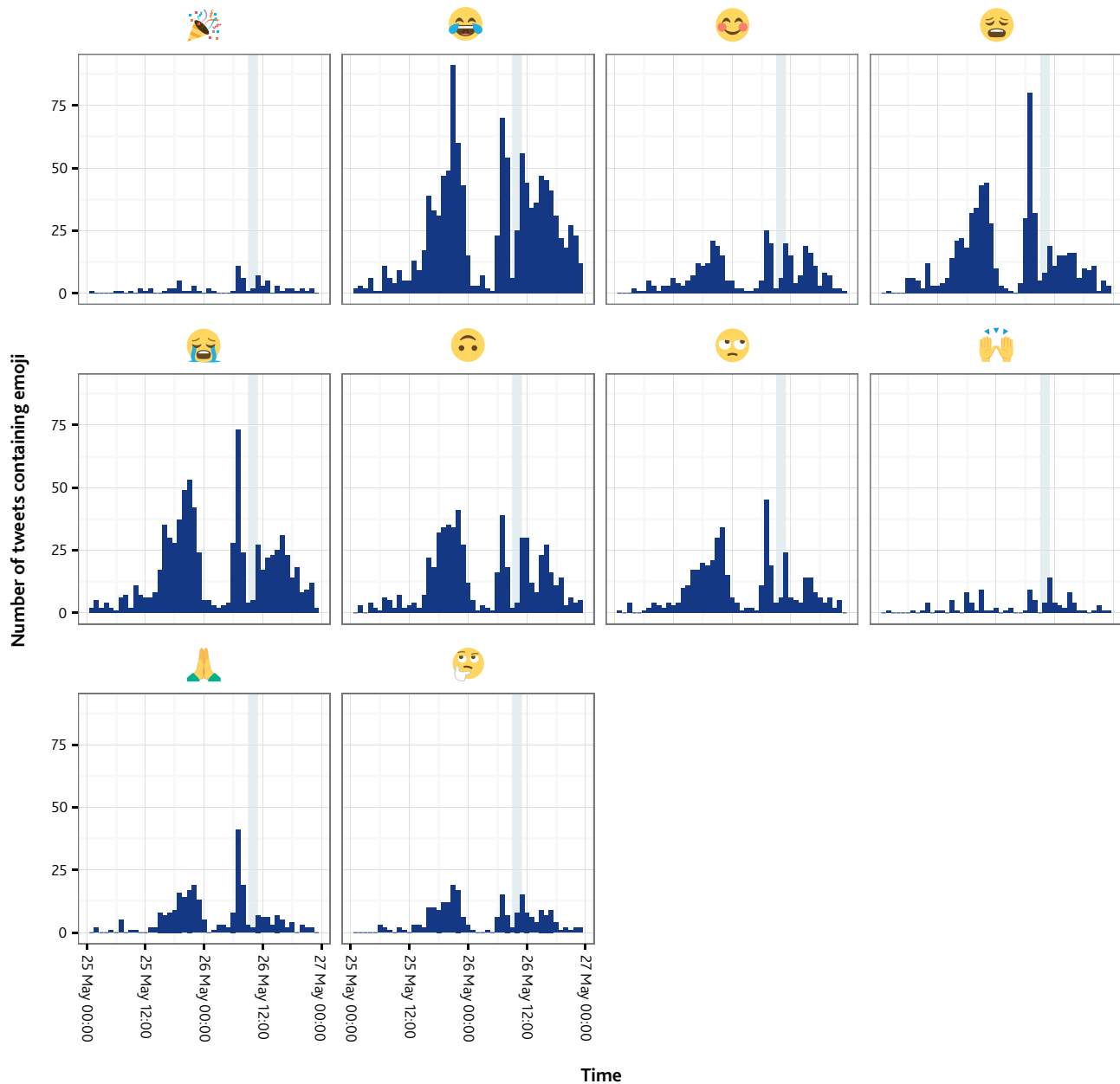


Figure 7: Maths-related tweets containing selected emoji

and 'loudly crying face' emoji over the session. By contrast, the use of the 'party popper' emoji became much more popular towards the exam session, as candidates finished their exams, especially on Fridays. For the 'person with folded hands' emoji (as if in prayer), usage was higher on weekdays when exams were taking place, but there was no evidence of any other trends.

However, mapping overall sentiment in this way does not reveal the full picture, due to the number and frequency of exams throughout the session: any positive sentiment expressed by students after finishing one exam would be counteracted by students expressing nerves and apprehension before a forthcoming exam in the following days.

In addition, we decided to look at sentiment (as measured by emoji) over the GCSE Mathematics example shown earlier, and for the same subset of tweets containing 'maths' on Wednesday 25 and Thursday

26 May. The numbers of tweets containing each of the top 10 emoji (in Table 2) for this period are shown in Figure 7. The most notable pattern here is that several emoji had surges of popularity just before the exam, for example 'person with folded hands' (as if in prayer) and 'weary face'. Immediately after the exam there were spikes in popularity for several emoji, including 'face with tears of joy' and 'loudly crying face'. Overall, there was more use of emoji in the evening before the exam, rather than afterwards.

Discussion

In this article we have described the collection and analysis of 6.44 million tweets from the summer 2016 examination session in the UK.

We found that there were more exam-related tweets at the beginning of the session than at the end. This may be because the JCQ seeks to timetable large-entry subjects as early as possible to facilitate marking (JCQ, 2017, p.5). It may also be that students are more excited at the beginning of the session for their first few exams, whereas towards the end they are becoming more accustomed to the rhythms of the exam session. We looked at the sentiments expressed in these tweets using a simple emoji approach, and found a general decrease in the use of the most common emoji ('face with tears of joy' and 'loudly crying face') over the session, perhaps reflecting such a decline in excitement.

By looking for subject-related search terms within the tweets we gathered, we identified that students were more likely to tweet about exams on the day of their exam, or in some cases the day before. We found that during weekdays there was a surge in exam-related tweets immediately before and after exam times, with peak time being the late afternoon when the day's exams were over.

From our investigations of the tweets relating to the GCSE Mathematics exams on Thursday 26 May, we found that it was possible to identify several different phases based on the words used in tweets: on the day before the exam, the dominant topic was revision. Immediately before the exam, the words 'luck' and 'hope' featured prominently; whereas after the exam there was talk of difficulty, and features of the paper (questions and mark schemes) along with an emotional response. When we looked at the sentiments expressed via emoji, we found a similar pattern, although perhaps more negative feeling before the exam than was expressed with words.

Limitations

Twitter data, and social media data in general, is an exciting data source, readily available, which offers the potential to get feedback from a large number of students taking assessments. However, this data should be used with extreme caution, as it is unlikely to be representative of students' views.

The availability and ease of extracting this data may lead us to forget that this is ultimately personal data, collected from young people undergoing stressful high-stakes exams, and ethical issues should be considered before using it.

While social media data offers the opportunity to investigate and understand candidates' views on their experience of exams, and even certain questions, actually using the information gained is fraught with risk. As discussed earlier in this article, discussions on Twitter are unlikely to be representative of students' views. Additionally, paying too much heed to comments on social media could limit awarding bodies' capacity to develop innovative items and assessments that test the full range of candidates' skills (see comments by Professor Rob Coe, cited in Busby, 2016). However, the fact that social media activity does now influence articles in the mainstream national media means that Twitter comment may affect the general public's view of exams and their standards.

References

- BBC News. (2015, June 5). *Tricky GCSE maths exam sees pupils take to Twitter*. Retrieved from <http://www.bbc.co.uk/news/education-33017299>
- Bouchet-Valat, M. (2014). *SnowballC: Snowball stemmers based on the C libstemmer UTF-8 library*. Retrieved from <http://CRAN.R-project.org/package=SnowballC>
- Bruns, A., & Burgess, J. (2015). Twitter Hashtags from Ad Hoc to Calculated Publics. In N. Rambukkana (Ed.), *Hashtag publics: The power and politics of discursive networks* (pp.13–28). New York: Peter Lang. Retrieved from <http://snurb.info/files/2015/Twitter%20Hashtags%20from%20Ad%20Hoc%20to%20Calculated%20Publics.pdf>
- Busby, E. (2016, October 21). Social media fears lead to 'predictable' exam papers. *TES*, 12–13.
- Curtis, S. (2013, September 6). Twitter claims 15m active users in the UK. *The Telegraph*. Retrieved from <http://www.telegraph.co.uk/technology/twitter/10291360/Twitter-claims-15m-active-users-in-the-UK.html>
- Emoji Keyboard. (2016). In *EmojiOne™ Version 3.0*. Retrieved January 19, 2017, from <http://emojione.com/>
- Espinoza, J. (2016, May 17). Students left 'fuming' over GCSE biology exam that contained questions about drunk 15-year-olds. *The Telegraph*. Retrieved from <http://www.telegraph.co.uk/education/2016/05/17/students-left-fuming-over-gcse-biology-exam-that-contained-quest/>
- Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of Twitter sentiment analysis methods. *ACM Computing Surveys*, 49(2), Article 28, 28–41. <http://doi.org/10.1145/2938640>
- Henderson, M., Johnson, N. F., & Auld, G. (2013). Silences of ethical practice: Dilemmas for researchers using social media. *Educational Research and Evaluation*, 19(6), 546–560. <http://doi.org/10.1080/13803611.2013.805656>
- Hogenboom, A., Bal, D., Frasinca, F., Bal, M., de Jong, F., & Kaymak, U. (2013). Exploiting emoticons in sentiment analysis. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing* (pp.703–710). New York, NY, USA: ACM. <http://doi.org/10.1145/2480362.2480498>
- Joint Council for Qualifications. (2017). *Construction of the common timetable: FSMQ, GCE and GCSE qualifications*. Retrieved from <http://www.jcq.org.uk/exams-office/key-dates-and-timetables/construction-of-the-common-timetable%E2%80%9393fsmq-gce-and-gcse-qualifications>
- Lebus, S. (2016). *Summer Series 2016 – Postmortems, petitions and practicalities* [Blog post]. Retrieved from <http://cambridgeassessment.org.uk/blog/summer-series-2016-postmortems-petitions-and-practicalities/>
- Rui, H., & Whinston, A. (2012). Information or attention? An empirical study of user contribution on Twitter. *Information Systems and e-Business Management*, 10(3), 309–324. <http://doi.org/10.1007/s10257-011-0164-6>
- Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science*, 346(6213), 1063–1064. <http://doi.org/10.1126/science.346.6213.1063>
- Statista. (2016). *Age distribution of Twitter users in Great Britain from May 2013 to May 2016*. Retrieved from <https://www.statista.com/statistics/278320/age-distribution-of-twitter-users-in-great-britain/>
- Twitter Streaming API. (n.d.). In *Twitter Development Documentation*. Retrieved January 19, 2017, from <https://dev.twitter.com/streaming/overview>
- UK Safer Internet Centre. (2015). *Friendship in a Digital Age*. Retrieved from <http://www.saferinternet.org.uk/safer-internet-day/sid-2015/friendship-digital-age-new-report-launched>
- Zimmer, M., & Proferes, N. J. (2014). A topology of Twitter research: Disciplines, methods, and ethics. *Aslib Journal of Information Management*, 66(3), 250–261. <http://doi.org/10.1108/AJIM-09-2013-0083>