Sentiment Analysis of User Feedback on the HSE Contact Tracing App

Kaavya Rekanar (Kaavya.Rekanar@ul.ie)
University of Limerick

Dr. Jim Buckley
University of Limerick

Sarah Buckley
University of Limerick

Dr. Manzar Abbas
University of Limerick

Dr. Sarah Beechum
University of Limerick

Dr. Muslim Chochlov
University of Limerick

Dr. Brian Fitzgerald
University of Limerick

Dr. Liam Glynn
University of Limerick

Dr. Kevin Johnson
University of Limerick

Dr. John Laffey
National University of Ireland Galway

Dr. Bairbre McNicholas
National University of Ireland Galway

Dr. Bashar Nuseibeh
University of Limerick

Dr. James O'Connell
University of Limerick

Dr. Derek O'Keeffe
National University of Ireland Galway

Dr. Mike O'Callaghan
University of Limerick

Dr. Abdul Razzaq
University of Limerick

Dr. Ita Richardson
University of Limerick

Dr. Andrew Simpkin
Research Article

Keywords: Covid-19, digital contact tracing, usability, transparency, performance, mhealth

DOI: https://doi.org/10.21203/rs.3.rs-96174/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Digital Contact Tracing (DCT) is seen as a key tool in reducing the propagation of viruses such as Covid-19, but it requires uptake and participation in the technology across a large proportion of the population to be effective.

While we observe the pervasive uptake of mobile device usage across our society, the installation and usage of contact tracing apps developed by governments and health services globally have faced difficulties. These difficulties range across the user-populations’ issues with the installation of the apps, usability and comprehension challenges, trust in the efficacy of the technology, performance issues such as the effect on battery life, and concerns about data protection and privacy.

In this work, we present our findings from a comprehensive review of the online user feedback relating to the initial release of the HSE Contact Tracker app, in an effort to inform later iterations and thus help sustain and potentially increase usage of the technology. While this might seem quite tightly scoped to the Irish context only, this app provides the basis for apps in many jurisdictions in the United States and Europe.

Our findings suggest a largely positive sentiment towards the app, and that users thought it handled data protection and transparency aspects well. But feedback also suggested that users would appreciate more targeted feedback, more proactive engagement and also suggested that both the ‘android-battery’ issue and the backward-compatibility issue with iPhones seriously impacted retention/uptake of the app respectively.

1. Introduction

As the coronavirus (Covid-19) continues to spread globally, governments and public health institutions look to contact tracing to help isolate and contain outbreaks. The more traditional manual contact tracing approach initially adopted in Ireland is both time and resource intensive, and may struggle to identify all contacts quickly enough, before they cause further transmission. In contrast, the efficiency and responsiveness of a digital approach using the proximity sensors in smartphone devices has the potential to limit delay and catch a greater number of contacts [1].

Key to the effectiveness of these digital solutions, however, is the percentage take-up of the app across the population, with one study in the UK [2] recommending the epidemic could be suppressed with 80 percent of all smartphone users using the app. This equates to 56 percent of the population overall.

1.1 Motivation

Analysis of the public’s response to the initial release of the HSE Contact Tracker app [3] can help guide the system's evolution towards maintaining, and expanding, the uptake and ongoing use of the app to fight Covid-19 in Ireland. It can do so by informing on user refinements to aspects like its usability, functional effectiveness, and performance. The voluntary nature of participation in the use of the app, combined with the requirement for a critical mass of users across the country to make the app effective, makes such feedback a crucial tool in the campaign to defeat the spread of the virus.
To that end, this research [4] [5] analysed all app reviews of the HSE app on the AppStore [6] and Google Play [7] using seven different aspects of interest: General Characteristics, Usability, Functional Effectiveness, Performance, Data Protection, Autonomy (of users), and Overall (generic comments). This analysis focused on ‘positive’ and ‘negative’ sentiment expressed by the user under each of these aspects, in order to build on areas well received, and to target areas where future releases of the app could be refined.

1.2 HSE App Overview

Ireland’s Health Service Executive released the COVID Tracker app (see Fig 1), developed by Nearform, across the Apple and Google online app stores in early July 2020.

Built on the Google and Apple Exposure Notification API (GAEN), it uses Bluetooth and anonymous IDs to log any other phone with the app it is in close contact with – tracking the distance and the time elapsed. Every 2 hours the app downloads a list of anonymous IDs that have been shared with the HSE by other users that have tested positive for Covid-19.

If a user has been closer than 2 metres for more than 15 minutes with any of these phones they will get an alert that they are a close contact. The app runs in the background.

Beyond the core contact tracing technology lies additional voluntary self-reporting functionality – users can choose to log daily health status or symptoms via the Check-In option, and also to share their age, group, sex and locality. Also optional is the ability to share a contact phone number so the HSE can contact them.

1.3 Pillar Derivation

The analysis process in this research involved coding user-reviews into 7 aspects, henceforth called pillars: General Characteristics, Usability, Functional Effectiveness, Performance, Data Protection, Autonomy (of users), and Overall (generic comments). These pillars were derived and refined through an iterative 6-phase process:

– A bottom-up approach, where individual contact tracing applications were evaluated for derivation of important app characteristics;

– A parallel academic/grey (media) literature review of app/health-app evaluation papers, to the same end;

– Cluster analysis, to create an amalgamated framework that revealed distinct super-categories (the pillars). This was refined via team review for redundancy and sufficiency;

– A Devil’s Advocacy phase where individual ‘pillar owners’ were challenged by another member of the team to assess the characteristics in that pillar for sufficiency, relocation, and relevance. This was followed by a full team critique of the pillars against the same criteria.

– Application of the resultant framework against the HSE app, to evaluate the framework further, leading to refinement of the pillars and characteristics.
Application of the framework against four other contact tracing apps using different architectures, tracing technologies and data guidelines: SafePaths [8], Corona-Warn [9], Aman [10] and Novid [11].

1.4 Research Questions

In order to inform the evolution of the HSE Covid Tracker app going forward the following two research questions were formulated:

1. How do users perceive the HSE Covid Tracker app version 1.0.0?

2. What are the prevalent issues users have with the HSE Covid Tracker app, V 1.0.0?

An ancillary analysis also probes the commonalities and differences between Apple and Android users to assess if there are any platform-specific issues that arose and to see how common the profiles are across the two sets of users.

1.5 Structure

In the next section, we discuss the method followed for data gathering and analysis, and then we present our results. Finally, the discussion section focuses on our findings, and potential recommendations for improving the efficiency of the app towards limiting the Covid-19 pandemic.

2. Methodology

In order to ensure that the data for the analysis was ecologically valid, it was decided to focus on naturally occurring reviews only. There is no discussion/issue forum on the HSE Contact Tracker’s GitHub page [12] but user reviews are available on the Google Play store and the iTunes App store and these have been used by other researchers in similar analyses [13] [14] [15]. Hence, a python program was developed to scrape reviews of the HSE’s Covid Tracker app from both the Google Play store and the iTunes App Store. This script was executed on the 13th of August 2020 and scraped all reviews up to that date. It should be noted that, at this point in time all the reviews were for version 1.0.0 of the app.

The reviews, thus scraped, were populated into a .csv sheet, which was further converted into .xlsx format for ease of analysis. This spreadsheet is made available for interested readers at [16]. The extracted file was cleaned of duplicates, sorted alphabetically and contained the following fields:

- User name;
- Time stamp;
- Review id (unique identity);
- The number of stars given to the app in the review (the user’s rating);
- The number of ‘thumbs-up’ from other viewers;
- The associated textual review.
2.1 Analysis

Manual sentiment analysis was subsequently performed against the 7-pillar evaluation framework described above. These pillars are referred to in the article by their acronyms: General Characteristics (GC), Overall comment (O), Functional Effectiveness (FE), Usability (U), Data Protection (DP), User Autonomy (A), Transparency (T), and app Performance (P). The sentiment analysis was performed manually because, even though there have been huge improvements to automated sentiment analysis in recent years [17] [18], the precision and recall rates achieved are still not perfect [17] [18] and this would be exacerbated in this instance because here ‘negative sentiment’ aims to capture not just reviews with a negative tone but also, quite positive reviews that request a specific refinement or modification.

Each review was randomly allocated to one of four reviewers, the overall allocations to each reviewer being of equal size. These reviewers were tasked with independently segmenting each review into a set of positive and negative user observations, and classifying each of these observations into their appropriate pillar: essentially a form of content analysis that allows “for the objective, systematic and quantitative description of the manifest content of communication” [19] [20]. Here, as mentioned above, ‘negative user observations’ refer to both comments with negative sentiments and comments suggesting refinements.

A joint discussion session at the start of the analysis ensured that all reviewers had a common understanding of the seven pillars and the sentiments being sought. The subsequent analysis resulted in three new fields being incorporated into the spreadsheet:

- The text segment where a positive, neutral or negative sentiment was detected;
- The sentiment (positive/negative/neutral) associated with that text segment;
- The associated pillar.

After the coders had individually coded the reviews in this fashion, one author was charged with assessing the entire coding for interpreter drift and sentiment inconsistencies. Interpreter drift is where a coder's coding drifts over time [21]. For example, in a coder's initial coding, they may classify a review segment complaining of “the lack of more detailed feedback on the location of cases” as a ‘usability’ issue. But, after fatigue has set in, they may note it as a ‘Performance’ issue. In such cases, the author charged with assessing interpreter drift corrected the drift by re-categorizing the latter comments in line with the categorization of the original comments (in the above example the lack-of-feedback comment would be consistently referred to as ‘usability’).

In terms of sentiment inconsistencies, there were (14) occasions where a reviewer very obviously ticked the incorrect sentiment. In one case, for example, a user complained of battery drain and the coder incorrectly categorized the sentiment as positive. These clear-cut mistakes were also rectified by the author charged with assessing the coding.

In order to assess inter-coder reliability, approximately one seventh of the reviews were coded by more than one reviewer. Figure 2 presents a snippet of the coding spreadsheet, illustrating several user-comments where more than one reviewer coded the reviews.
Inter-coder reliability could then be assessed using these reviews and Cohen’s Kappa [22]. The protocol was as follows: The text segments both coders identified for each review were re-ordered so that, where possible, they were in the same order across coders. Then the pillars and sentiment for each common segment were compared. If they were the same for the same segment, the coders were considered ‘in agreement’. If either the pillar or sentiment were not the same, the coders were not considered in agreement. Likewise if one of the coders noted a segment that the other did not, then this was considered another disagreement between the coders [23].

Figure 3 presents the Kappa analysis, revealing a Kappa of 70.086; a Kappa that suggests ‘good’ agreement between the coders.

Subsequently the individual pillars, sentiment and segments of the entire data-set, as coded by the main coder (not the coders who coded reviews solely for the purposes of the Kappa analysis) were analysed for trends and themes across the reviews: Pillars and sentiment were assessed quantitatively to identify the prevalent types of concerns expressed in the data set. Segments and sentiments were analysed to identify the individual concerns expressed by the app reviewers. The results of this analysis are now presented.

3. Results And Findings

In total 1287 comments were coded, 998 from Android users and 289 from Apple users. Table 1 presents the total number of identified comments per pillar and those totals broken down by positive/negative sentiment. In terms of users’ overall perception of the app, the data suggests that they were largely supportive and impressed. Of the 305 comments classified as ‘overall’ comments, 274 were positive. This perception was cross-platform but slightly higher across Android reviews: Of 223 Android-review comments 203 (91%) were positive and 71 of 82 (86.6%) Apple-review comments were positive. These ‘overall’ comments focused on the actual implementation itself (‘Really well thought out’, ‘Brilliant App’, ‘Well implemented’, ‘it does what it does well’) and, to a lesser degree, on the concept of the app (‘its a no brainer’, ‘great idea’). The ratio of implementation:concept comments was approximately 5:1 in the reviews.

As Table 1 shows, most of the ‘negative’ comments were aimed at performance and usability. The prevalence of performance comments can largely be explained by an Android battery-drain issue that arose on August 8th,

Table 1 Results and Findings
| Apple+Android | Total | Positive | Negative | Negative Corrected |
|---------------|-------|----------|----------|-------------------|
| Overall       | 305   | 274      | 16       | 16                |
| Usability     | 390   | 96       | 260      | 206               |
| Performance   | 448   | 29       | 410      | 0                 |
| Data Protection| 44    | 32       | 11       | 11                |
| Functional    | 70    | 14       | 51       | 51                |
| Effectiveness |        |          |          |                   |
| General       | 7     | 2        | 2        | 2                 |
| Characteristics|      |          |          |                   |
| Transparency  | 23    | 14       | 9        | 9                 |
| Sum           | 1287  |          |          |                   |

where one of the Exposure Notification Service libraries provided by Google was updated (by Google) and that update caused significant battery drainage, overheating and intensive CPU usage, related to the app. Of the 448 performance comments, 444 were from Android users. 361 contained the word 'drain', 53 contained the word 'heat' and five contained the word 'CPU'.

So essentially all of the negative performance comments were due to this Google update. If this issue was excluded from the analysis, the performance comments would have been entirely positive. However, this seems to have been a very serious issue for Android phone users: Of the 365 reviews associated with these comments 102 mentioned the word ‘uninstall’, suggesting that well over a quarter of those who complained about the issue considered (at least) uninstalling the app.
Interesting also are the data protection and transparency pillars. Users seem to perceive that the HSE has done well on both fronts with 32 of 44 data-protection comments having a positive sentiment and 14 of 23 transparency comments having a positive sentiment. This sentiment trend was consistent across both Apple and Android devices.

3.1 Usability Comments

The prevalence of usability issues across the reviews is to be expected given the forums involved: Users are most typically interested in usability [24]. Their usability feedback is detailed in Figure 4.

The main usability issue, as determined by the number of user suggestions (90), was around the feedback provided by the app regarding the occurrence of Covid across the country. Version one of the app focused on the number of cases in total at national and county level. Users felt that feedback on Covid cases should focus on a more recent time range (36) on active or newly-found cases only (24) and on finer granularity, in terms of geographical location (21). Of this latter category, seven reviewers suggested that hot-spots be identified, but this would be difficult in terms of maintaining the current privacy standards the app embodies.

An interesting idea that arose in two reviews was that the app should also report on the number of close contacts users had per day, where the user could get daily feedback and thus try to minimize that ‘score’ over time. This is analogous to the gamification concept of ‘streaks’ [25] where a user might aim to keep their number of close contacts below a certain daily threshold, over time, and thus continue their ‘streak’.

Another prevalent theme in the usability pillar was a desire of notifications for daily check-ins, where the user self-reports their health status to the app, ideally on a daily basis. 70 comments requested this enhancement or expressed dissatisfaction at it not currently being available in the app.

A surprising finding was that 20 reviewers complained that a town or area was not available for selection when they were profiling themselves during app set-up. Often this was their own town/area, but in five cases reference was made to the exclusion of the six counties in Northern Ireland. Another reviewer noted that it would be interesting to have the possibility of recording two areas, where a user works in one location but lives in another.

Finally, iPhone users complained, in significant numbers, about the apps inability to work on older versions of iOS or older iPhones (the iPhone6 particularly). Overall, 54 of the 289 apple comments were targeted as this issue: by far the most prevalent focus of iPhone users’ concerns. This is unfortunate because, it represents a significant number of potential users who want to install the app but cannot. In addition, as in the case of the battery-drain issue, this is entirely outside of the HSE’s/NearForm’s control: For this to be addressed, Apple would need to incorporate backward compatibility into the associated operating systems. The last column of Table 1, shows the ‘negative’ sentiment corrected for these outside-of-the-HSE’s-NearForm-control issues.

4. Discussion

In general, public sentiment seem positive for the HSE’s Contact Tracking app. The overall comments are largely supportive and, on the critical aspects of data protection, and transparency, public opinion seems favorable, as assessed by the Google Playstore and iTunes App reviews.
Below that positive sentiment though there are some prevalent user requests/concerns that should be addressed. However, the most prevalent platform-specific concerns (and indeed two of the three most prevalent concerns across platforms) are outside of the developers control. The battery drain issue was caused by an update to an underpinning Google library, and was remedied quickly, although not before substantial reputational damage had been done in terms of the public’s perception of the app. Likewise, the backward-compatibility issue for older versions of iOS is a significant issue for potential users and can only be addressed by Apple.

Something that the HSE/Nearform have already worked on, in their new versions of the app, is the information the app gives about the spread of the disease. They have tackled the desire for more timely information with information now provided on new cases within the last day, two weeks, and two months at a national level. They have also expanded the information available at county level, where again users can see the new cases for the last day, two weeks and two months.

But they have not increased the granularity of information in terms of geographical location: That information is still presented in terms of the national picture and in terms of counties. One reviewer suggested that the facility offered by the Central Statistics Office of Ireland, where Covid cases are reported by electoral division(https://census.cso.ie/covid19/), should be incorporated into the app. But the statistics presented on this website are for the 20th of June, and so are out of date now/could not be used.

Another aspect prevalent in the reviews, and one that might facilitate increased user engagement, was the use of daily notification alerts to remind users to check-in with their health status. This, combined with the improved disease-spread feedback provided by the app, and maybe gamification of aspects like recording daily close contacts, may encourage users to retain the app and engage with it for longer time periods [25].

4.1 Limitations and Threats to Validity

In terms of reliability, the analysis was done by different coders, so there may be a bias or an inconsistency in that coding. To mitigate against this possibility the four coders were all given an introductory session where 15 illustrative examples were coded and discussion undertaken to form a common understanding. The inter-coder reliability assessment undertaken in this study suggests that this introductory session largely achieved its goal, as discussed in section 2, and that the reliability of the analysis was good.

A construct validity issue [26] is that the data obtained may not be totally correct: User opinions may be informed by hearsay and users are not always in an ideal position to report on quality aspects like data protection, or performance [27]. To mitigate against this the analysis focused on sentiments that were more pervasive across the data-set and brought to bear considerable knowledge of the app itself, the researchers having studied it as part of the project goal to derive the pillars beforehand.

Another external validity issue [26] is that this dataset obtained is only a very small sample of the user population. Even considering that a large number of users may have uninstalled the app after the Android battery issue, and the currently quoted download figure of 1.97 million is thus somewhat inflated, it is likely that there are well over 1.3 active users (this is quoted in the app’s User Interface). While a sample of 1287 comments is small by comparison, it can be considered representative [28]. However, a more serious, related
external validity issue is that that sampling protocol in this case is self-selecting, where only vocal reviewers (those that left reviews) are included in the data-set.

5. Conclusion

This report has focused on sentiment analysis of all the reviews that were available on Google Play and Apple collectively before August 14th 2020, towards helping evolve a better contact tracing application to fight the Corona virus. The results suggest that the app is well perceived and seen as sensitive to data-protection concerns. But the reviews also suggest that the feedback it provides regarding occurrence of Covid-19 should be made more granular in terms of location, and that the option of notifications should be included, to remind users to lodge their health status every day. Additionally users seemed upset when their location was not available in the app. Finally, efforts should be directed at prompting Apple to make their Exposure Notification Service available to older iPhones and older versions of iOS.

But some of the suggestions made by users have already been addressed in the next version of the Covid Tracker that was launched by the HSE. Specifically feedback on the status of Covid 19 has been made more current in terms of time-span reported and current cases. This consilience, between our results and HSE updates, not only strengthens the efficacy of our other findings, but also suggests that the health authorities in Ireland are evolving the app in a direction sympathetic to user concerns.

While the analysis proved to be helpful in understanding the sentiment behind the public's opinion regarding the HSE’s Covid Tracker, an automated analysis of users’ sentiments using artificial intelligence could also be developed [29]. This would facilitate understanding wider public opinion about the app over larger data-sets, in a much more time-sensitive manner.

Our future work will be in this direction and already we have trialled an initial approach. The existing reviews scraped and extracted into a .csv file have been cleaned for special characters and unnecessary symbols using Alteryx. Sentiment analysis has been done using R Studio shown in Fig 5.

In this context, sentiment analysis is the computational task of automatically determining what feelings a writer is expressing in text [30]. Sentiment is often framed as a binary distinction (positive vs. negative), but it can also be a more fine-grained, like identifying the specific emotion an author is expressing (like fear, joy or anger) [18].

However, we are conscious that sentiment analysis alone may only help us focus our analysis on dissatisfied users/comments [31]. It will not identify the pillar of interest, identify the users’ specific issues or, indeed, the prevalence of those issues across the data-sets. So additional analysis may be required to facilitate derivation of the users’ issues [32].

Declarations

COMPETING INTERESTS

The authors declare no competing interests.
ACKNOWLEDGMENT

This work was supported, in part, by Science Foundation Ireland grants 16/RC/3918 and 20/COV/0133.

References

1. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser, “Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing,” Science, vol. 368, no. 6491, 2020.

2. Hinch, W. Probert, A. Nurtay, M. Kendall, C. Wymant, M. Hall, and C. Fraser, “Effective configurations of a digital contact tracing app: A report to nhsx,” In:(Apr. 2020). Available here. url: https://github.com/BDI-pathogens/covid-19_instant_tracing/blob/master/Report, 2020.

3. HSE covid tracker-official webpage. HSE. [Online]. Available: https://covidtracker.gov.ie/

4. Covigilant funding proposal to SFI. [Online]. Available: https://lero.ie/sites/default/files/2020-TR-05_Covigilant%20SFI%20Application%20Tech%20Report.pdf

5. E. O’Callaghan, J. Buckley, B. Fitzgerald, K. Johnson, J. Laffey, B. McNicholas, B. Nuseibeh, D. O’Keeffe, K. Rekanar, A. Razzaq et al., “A national survey of attitudes to covid-19 digital contact tracing in the republic of ireland,” 2020.

6. HSE Covid Tracker-App Store. HSE. [Online]. Available: http://apps.apple.com/ie/app/covid-tracker-ireland/id1505596721

7. HSE Covid Tracker-Google Play. HSE. [Online]. Available: google.com/store/apps/details?id=com.covidtracker.hse

8. Safepaths [Online]. Available: https://safepaths.mit.edu/

9. Corona-warn. [Online]. Available: https://www.coronawarn.app/en/

10. Aman. [Online]. Available: https://amanapp.jo/en

11. Novid. [Online]. Available: https://www.novid.org/

12. HSE github page for the covid tracker app. [Online]. Available: https://github.com/HSEIreland/covid-tracker-app/pulse

13. Guzman and W. Maalej, “How do users like this feature? a fine grained sentiment analysis of app reviews,” in 2014 IEEE 22nd international requirements engineering conference (RE). IEEE, 2014, pp. 153–162.

14. McIlroy, N. Ali, H. Khalid, and A. E. Hassan, “Analyzing and automatically labelling the types of user issues that are raised in mobile app reviews,” Empirical Software Engineering, vol. 21, no. 3, pp. 1067–1106, 2016.
15. Maalej and H. Nabil, “Bug report, feature request, or simply praise? on automatically classifying app reviews,” in 2015 IEEE 23rd international requirements engineering conference (RE). IEEE, 2015, pp. 116–125.

16. Spreadsheet containing the data for analysis. Lero. [Online]. Available: https://www.lero.ie/sites/default/files/journal%20copy%20of%20scraped%20and%20cleaned%20data.xlsx

17. P. Kirilenko, S. O. Stepchenkova, H. Kim, and X. Li, “Automated sentiment analysis in tourism: Comparison of approaches,” Journal of Travel Research, vol. 57, no. 8, pp. 1012–1025, 2018.

18. Andreevskaia, S. Bergler, and M. Urseanu, “All blogs are not made equal: Exploring genre differences in sentiment tagging of blogs.” in ICWSM, 2007.

19. Berelson, “Content analysis in communication research.” 1952.

20. Y. Sharif, M. English, N. Ali, C. Exton, J. Collins, and J. Buckley, “An empirically-based characterization and quantification of information seeking through mailing lists during open source developers’ software evolution,” Information and Software Technology, vol. 57, pp. 77–94, 2015.

21. Herold, M. Blom, and J. Buckley, “Evidence in architecture degradation and consistency checking research: preliminary results from a literature review,” in Proceedings of the 10th European Conference on Software Architecture Workshops, 2016, pp. 1–7.

22. Cohen, “Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit.” Psychological bulletin, vol. 70, no. 4, p. 213, 1968.

23. P. Hartmann, “Considerations in the choice of interobserver reliability estimates,” Journal of applied behavior analysis, vol. 10, no. 1, pp. 103–116, 1977.

24. Ha and D. Wagner, “Do android users write about electric sheep,” Examining consumer reviews in google, 2013.

25. Buckley, T. DeWille, C. Exton, G. Exton, and L. Murray, “A gamification–motivation design framework for educational software developers,” Journal of Educational Technology Systems, vol. 47, no. 1, pp. 101–127, 2018.

26. K. Wright, M. Kim, and D. E. Perry, “Validity concerns in software engineering research,” in Proceedings of the FSE/SDP workshop on Future of software engineering research, 2010, pp. 411–414.

27. Welsh, K. Rekanar, J. Buckley, and et. al, “Towards a taxonomy for evaluating societal concerns of contact tracing apps,” in 7th International Conference on Behavioral, Economic, and Socio-Cultural Computing (BESC), Bournemouth, UK, 2020.

28. J. Oates, Researching information systems and computing. Sage, 2005.

29. Yang, A. Willis, A. De Roeck, and B. Nuseibeh, “A hybrid model for automatic emotion recognition in suicide notes,” Biomedical informatics insights, vol. 5, pp. BII–S8948, 2012.
30. K. Bakshi, N. Kaur, R. Kaur, and G. Kaur, “Opinion mining and sentiment analysis,” in 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), 2016, pp. 452–455. Please give a shorter version with: \authorrunning and \titlerunning prior to \maketitle

31. Hatzivassiloglou and J. Wiebe, “Effects of adjective orientation and gradability on sentence subjectivity,” in COLING 2000 Volume 1: The 18th International Conference on Computational Linguistics, 2000.

32. Romesburg, Cluster analysis for researchers. Lulu. com, 2004.