Towards Low-burden Responses to Open Questions in VR

Dmitry Alexandrovsky
dimi@uni-bremen.de
DMLab, University of Bremen
Bremen, Germany

Susanne Putze
sputze@uni-bremen.de
DMLab, University of Bremen
Bremen, Germany

Alexander Schülke
schuelk1@uni-bremen.de
University of Bremen
Bremen, Germany

Rainer Malaka
malaka@tzi.de
DMLab, University of Bremen
Bremen, Germany

(a) Controller Drumming **(b)** Freehand **(c)** Pinch

Figure 1: The three input methods for open questions investigated in our user study.

ABSTRACT

Subjective self-reports in VR user studies is a burdening and often tedious task for the participants. To minimize the disruption with the ongoing experience VR research has started to administer the surveying directly inside the virtual environments. However, due to the tedious nature of text-entry in VR, most VR surveying tools focus on closed questions with predetermined responses, while open questions with free-text responses remain unexplored. This neglects a crucial part of UX research. To provide guidance on suitable self-reporting methods for open questions in VR user studies, this position paper presents a comparative study with three text-entry methods in VR and outlines future directions towards low-burden qualitative responding.

CCS CONCEPTS

• **Human-centered computing → HCI theory, concepts and models**; **Empirical studies in HCI**.

KEYWORDS

Virtual reality; self-reporting; design space; typing in VR.

1 INTRODUCTION AND RELATED WORK

For Virtual Reality (VR) user studies, most setups rely on mid- or post-experience questionnaires outside VR [24, 25]. Usually, after a tasks participants are required to leave VR and fill out questionnaires on PC or paper. However, this change in realities breaks the study flow and produces a Break in Presence (BIP) [26, 27] which has been associated with disorientation, loss of control and negative emotions [14, 23, 24, 26]. Putze et al. have shown that the switching between realities has a strong physiological effect which holds on for a significant period of time [21]. Therefore, BIPs in VR user studies might lead to uncontrolled biases, in particular right before the questionnaire responses. Moreover, particularly for unsupervised remote (i.e., online) VR studies, the break of the study flow may cause the participants to drop off before finishing the experiment. VR research has started administering questionnaires inside the Virtual Environment (VE); therefore, the participants do not need to leave the VE to fill out questionnaires and can stay closer to the experience which improves the study flow along with the User Experience (UX) of the the study [3, 22, 24]. However, the process of filling out questionnaires is interrupting and burdening, and in many cases it causes participants to terminate the study [17, 29]. Although research has shown that in-VR Questionnaires (inVRQs) minimize the BIPs in VR user studies, they cannot avoid BIPs resulted by the question asking entirely [21].

Schwind et al. [24] contrasted the screen-based questionnaires against VR-embedded questionnaires and found that with embedded assessment the subjective responses in VR are more consistent. In contrast, others have shown that in-VR questionnaires may lead to inconsistencies [12]. To counteract for such inconsistencies, Alexandrovsky et al. [3] presented important usability criteria for in-VR questionnaires. Other tools that allow administering questionnaires in VR are the VR Questionnaire Toolkit [9], VRate [22]. Similarly, MRAT [18] is a toolkit for Augmented Reality (AR) studies. These tools aim for a less-disruptive study flow and target problems of context-dependent forgetting [1, 11] due to environment change [20] which may bias responses. However, to date only closed questions (i.e. mostly Likert-scales) has been implemented and evaluated in VR. Nonetheless, qualitative research methods play a significant role in HCI as they often yield an explanation of quantitative outcomes and thus, provide for a deeper understanding of many phenomena [2, 4, 16]. For instance, in post-experience semi-structured interviews or in surveys with open-ended questions, the participants could explain why the gave a good or a bad usability score when testing a system [16]. HCI shares a variety of qualitative research methods including, but not limited to semi-structured interviews, think-aloud protocols, semi-open questions [16]. All these methods have there specific use cases and differ in the workload the participants exhibit. For example, van den Haak et al. [30] concurrent and retrospective think-aloud protocols. The authors showed that both research methods uncover usability issues similar with similar quality, but the two assessment methods differ in how the usability problems are detected and in their impact on the participants' performance. Therefore, van den Haak et al. argue that concurrent think-aloud methods are not suitable for complex and demanding tasks, and in such scenarios post-experience responses would provide more reliable results.

To cover a broad range of research methods in VR, contemporary VR-embedded questionnaire toolkits should support low-burdening qualitative responses along with quantitative questions. For qualitative assessments, questionnaires often employ open-ended questions where the participants are asked to elaborate on their rating, to provide additional explanations, or to give further feedback. For VR user studies, open questions with free-text responses remain unexplored. Partially, this is due to the fact that, despite a growing body of work, typing in VR is slower and more tedious than typing on a real physical keyboard [28]. Speicher et al. [28] developed a design space of text input methods in VR. The authors compared six different text input techniques in VR. Namely: head pointing, controller pointing, controller tapping, freehand and discrete & continuous cursor. The study results show clearly that pointing with hand-held controllers exceeds the other text input methods. However, the authors conclude that none of the virtual typing methods is able to compete with physical keyboards. Using physical keyboards in VR has been shown effective. Yet, such mixed reality approaches come at great expense since the setups introduce additional sensors [15, 19] and are prone to failure. This makes such setups unsuitable for remote or field studies. Nevertheless, virtual typing in VR exists in several consumer products such as the Oculus Quest menu, many VR social platforms, and VR games. Therefore, despite its inferiority, virtual typing is frequently used by end-users. Nonetheless, while different typing methods in VR show comparable performances, it remains unclear if study participants are willing to use them and how the typing paradigm affects the participants' compliance to give free-text answers.

To guide the design of low-burdening self-reporting, Yan et al. proposed a design-space of in-situ self-reporting [32]. The design space builds on five requirements that aim do minimize the burden of self-reports: (1) *Minimal Disruptiveness*: the self-reporting should not distract participants from their ongoing experience, (2) *Inclusiveness*: the self-reporting should prevent the participants embarrassment and maintain their privacy, (3) *Low-focus*: self-reporting should require low attention, (4) *Intuitiveness*: the interface should be self-explanatory and easy to use, and (5) *Expressivity* the self-reporting interface should support a variety of question types. While this framework provides valuable insights for desktop and mobile user studies, it misses some opportunities and tradeoffs that emerge with reality-altering interfaces. To provide an initial step towards the design of self-reporting for open questions and to lay a groundwork for a future design space of self-reporting in VR, this work follows the research question *what type of existing virtual keyboards is most suitable for free-text responses in VR user studies*. This paper presents a comparative study of different typing methods in VR and outlines a planned study, which should compare different response modalities that should contribute the standardization of self-reporting methods in VR user studies.

2 EVALUATING SUITABLE VR TEXT INPUT METHODS

Based on literature and existing end-user VR applications we implemented three typing methods. Namely: *Controller Drumming*, *Freehand Typing*, and *Pinch Typing*. Previous research compared these inputs with other existing methods and showed comparable performances and usability which makes them the best candidates for low-burdening text input for free-text responses in VR user studies [6, 8, 28]. To provide a sound comparability, all input methods were operating on the same virtual keyboard. The participants could select from two keyboard layouts: QUERTY and QUERTZ. The keyboard has a total area of $23 \times 60cm$ with a hexagonal shaped keys with diameter of $8cm$. This size determined by recommendations from previous work by Dudley et al. [7] and internal testing during the development. Users can adjust the height and position of the keyboard.

Controller Drumming. Our controller-based method follows Boletsis and Kongsvik's [5] implementation of drum-like text input (c.f., Fig. 1a). This is a symmetric bimanual typing technique with VR controllers. At each controller is a $14cm$ long front-facing drum stick attached. The users perform keystrokes by hitting keys with the sticks.

Freehand. The Freehand typing employs of the Oculus Quest's hand tracking (c.f., Fig. 1b). Inside VR, the users' hands are represented by high-detail meshes. Like controller drumming, this method is bimanually symmetric. At each index finger is a small sphere-shaped ($r=1.5cm$) interaction area attached. The users perform the keystrokes by hitting keys with their index fingers. In line with findings by Dudley et al. [7] who demonstrated that index-only typing provides better performance, we implemented freehand typing with index finger only keystrokes.

Pinch. This input method is inspired by PinchType [8]. Like freehand, this technique is build upon the hand tracking. However, this method is asymmetric (c.f., Fig. 1c). Like with the freehand method, the index finger of the dominant hand is used for key selection. However, in contrast to the other methods the key selection is not determined by collision of the hand and the key, but rather through proximity; i.e., the closed key to the index finger is highlighted. The actual keystroke is performed with a pinch gesture of the non-dominant hand.

2.1 Study Design

The goal of the study was to determine the usability of the three concurrent text input methods. Since the typing speed differs drastically between individuals we conducted the study within-subjects. The order of the conditions was randomized using Latin square. We conducted the user study remotely. The participants downloaded the study app and participated at home using an Oculus Quest Head-Mounted Display (HMD). The study app guided the participants through the study flow without any communication with an experimenter. For each condition, the task was to write off sentences from the Enron corpus [31] of standardized phrases with a length of 20–28 characters. The target phrases were displayed on a world-anchored User Interface (UI) in VR using the respective input method.

As objective measures of performance we logged each key stroke along with its timestamp, from which we could later calculate the participants' accuracy and the Words per Minute (WPM). We assessed workload on a raw NASA-TLX [13] and usability on UMUX [10] using inVRQs by Alexandrovsky et al. [3]. Afterward, the participants left the VE to answer a conclusive questionnaire and perform a typing test in order to determine their WPM on a regular keyboard. For each typing variant, the participants were briefed. Next, they wrote five sentences with the respective input method and rated workload as well as the usability of the typing method. The total study duration was M=21, SD=4.88 minutes.

2.2 Outcomes

12 participants (self-reported: 8 male, 4 female, age: M=27.08 (SD=1.38)) volunteered for our study. The typing speed on a regular keyboard out-VR was M=63.58, SD=22.13. To determine differences between the typing variants, we conducted RM-ANOVAs with the typing method as within-subjects factor on all measures. The usability on the UMUX (Fig. 2a) differed significantly, between the input methods ($F_{2,22}$=14.03, p<0.01, η_p^2=0.56). Post-hoc comparisons revealed that drumming receives significantly higher scores over freehand (t_{11}=4.29, p<0.01, Cohen's d=1.39) and over pinch (t_{11}=5.29, p<0.01, Cohen's d=1.59). The workload on TLX (Fig. 2b) differed significantly for mental demand ($F_{2,22}$=7.95, p<0.01, η_p^2=0.42) physical demand ($F_{2,22}$=28.90, p<0.01, η_p^2=0.72), effort ($F_{2,22}$=20.07, p<0.01, η_p^2=0.65) and frustration ($F_{2,22}$=9.31, p<0.01, η_p^2=0.46), showing that generally the controller based input method promoted the lowest workload. For the performance analysis, two participants were excluded due to corrupt or missing data. The Net WPM (Fig. 2c), i.e., raw WPM reduced by the uncorrected errors, differed significantly between the conditions ($F_{2,18}$=35.92, p<0.01, η_p^2=0.80) showing that the drumming was significantly faster compared to pinch (t_{49}=13.56, p<0.01, Cohen's d=2.49) as well as freehand (t_{49}=9.35, p<0.01, Cohen's d=1.64) and that freehand was significantly faster than pinch (t_{49}=3.06, p=0.01, Cohen's d=1.10). These corroborated results show clearly that controller based typing is the most suitable text-entry method. Furthermore, these results are in line with related literature confirming that typing in VR by a magnitude slower than out-VR.

3 TOWARDS LOW-BURDENING OPEN QUESTIONS RESPONSES IN VR

The outcomes from the comparison study of different text-entry methods in VR underline that typing in VR is a difficult and burdening task. Most likely the burdening typing methods would disengage the participants from an ongoing user study leading to missing data, incorrect, or incomplete responses to open questions. To address these issues and to provide a clearer picture on the participants' response behavior we plan a user study which compares the compliance of participants of responding to open questions using different input methods. The study is planned as between-subjects design with the conditions (1) VR drumming as the best suitable text-entry method in VR, (2) voice recordings of the responses in VR and as an low-burdening and intuitive interaction, (3) a "traditional" text-entry method using out-VR questionnaires (outVRQs).

To avoid biases, the participants are decepted and tasked to play different variants of a simple shooter to provide qualitative responses about what they liked or disliked about the game. As a primary measure of compliance, we aim to operationalize the response length.

The to be explored modalities cover a wide range of tradeoffs for the study design which require careful consideration. The outcomes of the planned study should give insights about the participants' compliance on Yan et al.'s dimensions of *Disruptiveness*, *Inclusiveness*, and *Intuitiveness*. Furthermore, the study should contribute to the foundation of a design space of low-burden self-reporting in VR which aims to classify the benefits and drawbacks of different design decisions regarding self-reporting in VR user studies and to provide a groundwork for standardized methods for VR research.

REFERENCES

[1] Ethel Mary Abernethy. 1940. The Effect of Changed Environmental Conditions Upon the Results of College Examinations. *The Journal of Psychology* 10, 2 (Oct. 1940), 293–301. https://doi.org/10.1080/00223980.1940.9917005

[2] Anne Adams, Peter Lunt, and Paul Cairns. 2008. A Qualititative Approach to HCI Research. In *Research Methods for Human-Computer Interaction*, Paul Cairns and Anna Cox (Eds.). Cambridge University Press, Cambridge, UK, 138–157.

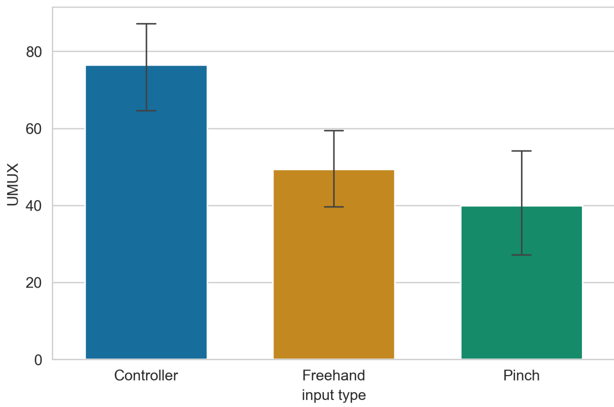
[3] Dmitry Alexandrovsky, Susanne Putze, Michael Bonfert, Sebastian Höffner, Pitt Michelmann, Dirk Wenig, Rainer Malaka, and Jan D. Smeddinck. 2020. Examining Design Choices of Questionnaires in VR User Studies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Honolulu, HI, USA. https://doi.org/10.1145/3313831.3376260

[4] Ann Blandford, Dominic Furniss, and Stephann Makri. 2016. Qualitative HCI Research: Going Behind the Scenes. *Synthesis Lectures on Human-Centered Informatics* 9, 1 (April 2016), 1–115. https://doi.org/10.2200/S00706ED1V01Y201602HCI034

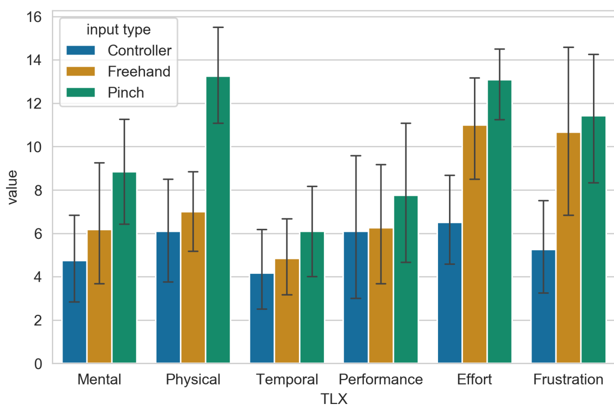
[5] Costas Boletsis and Stian Kongsvik. 2019. Text Input in Virtual Reality: A Preliminary Evaluation of the Drum-Like VR Keyboard. *Technologies* 7, 2 (April 2019), 31. https://doi.org/10.3390/technologies7020031

[6] Tafadzwa Joseph Dube and Ahmed Sabbir Arif. 2019. Text Entry in Virtual Reality: A Comprehensive Review of the Literature. In *Human-Computer Interaction. Recognition and Interaction Technologies*, Masaaki Kurosu (Ed.). Vol. 11567. Springer International Publishing, Cham, 419–437. https://doi.org/10.1007/978-3-030-22643-5_33

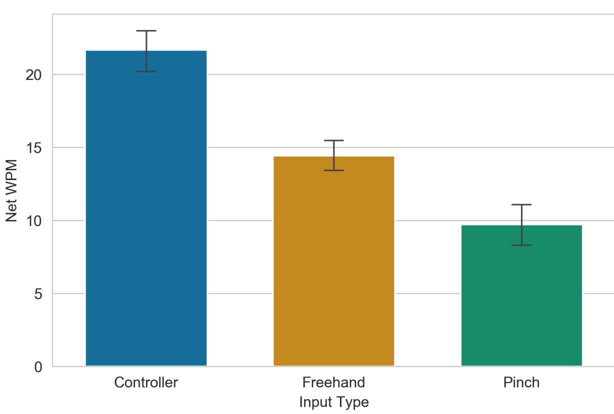
[7] John Dudley, Hrvoje Benko, Daniel Wigdor, and Per Ola Kristensson. 2019. Performance Envelopes of Virtual Keyboard Text Input Strategies in Virtual Reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*.



(a) UMUX



(b) NASA-TLX



(c) Net WPM

Figure 2: User study results.

IEEE, Beijing, China, 289–300. https://doi.org/10.1109/ISMAR.2019.00027

[8] Jacqui Fashimpaur, Kenrick Kin, and Matt Longest. 2020. PinchType: Text Entry for Virtual and Augmented Reality Using Comfortable Thumb to Fingertip Pinches. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–7. https://doi.org/10.1145/3334480.3382888

[9] Martin Feick, Niko Kleer, Anthony Tang, and Krueger Antonio. 2020. The Virtual Reality Questionnaire Toolkit. In *The 33st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings (UIST '20 Adjunct)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3379350.3416188

[10] Kraig Finstad. 2010. The Usability Metric for User Experience. *Interacting with Computers* 22, 5 (Sept. 2010), 323–327. https://doi.org/10/ctr6r4

[11] D. R. Godden and A. D. Baddeley. 1975. Context-Dependent MemoryIn Two Natural Environments: On Land and Underwater. *British Journal of Psychology* 66, 3 (Aug. 1975), 325–331. https://doi.org/10/cjgxhp

[12] Sarah Graf and Valentin Schwind. 2020. Inconsistencies of Presence Questionnaires in Virtual Reality. In *26th ACM Symposium on Virtual Reality Soft- Ware and Technology (VRST '20)*. Virtual Event, Canada. https://doi.org/10.1145/3385956.3422105

[13] Sandra G. Hart. 1986. *NASA Task Load Index (TLX). Volume 1.0; Paper and Pencil Package.* Technical Report.

[14] Jarrod Knibbe, Jonas Schjerlund, Mathias Petraeus, and Kasper Hornb{\ae}k. 2018. The Dream Is Collapsing: The Experience of Exiting VR. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 483:1–483:13. https://doi.org/10.1145/3173574.3174057

[15] Pascal Knierim, Valentin Schwind, Anna Maria Feit, Florian Nieuwenhuizen, and Niels Henze. 2018. Physical Keyboards in Virtual Reality: Analysis of Typing Performance and Effects of Avatar Hands. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 345:1–345:9. https://doi.org/10.1145/3173574.3173919

[16] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research Methods in Human-Computer Interaction* (second ed.). Morgan Kaufmann, Cambridge, MA.

[17] Andreas Möller, Matthias Kranz, Barbara Schmid, Luis Roalter, and Stefan Diewald. 2013. Investigating Self-Reporting Behavior in Long-Term Studies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Paris France, 2931–2940. https://doi.org/10.1145/2470654.2481406

[18] Michael Nebeling, Maximilian Speicher, Xizi Wang, Shwetha Rajaram, Brian D. Hall, Zijian Xie, Alexander R. E. Raistrick, Michelle Aebersold, Edward G. Happ, Jiayin Wang, Yanan Sun, Lotus Zhang, Leah E. Ramsier, and Rhea Kulkarni. 2020. MRAT: The Mixed Reality Analytics Toolkit. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–12. https://doi.org/10/ghbdtp

[19] Duc-Minh Pham and Wolfgang Stuerzlinger. 2019. HawKEY: Efficient and Versatile Text Entry for Virtual Reality. In *25th ACM Symposium on Virtual Reality Software and Technology*. ACM, Parramatta NSW Australia, 1–11. https://doi.org/10.1145/3359996.3364265

[20] Rüdiger Pohl and Rüdiger F. Pohl. 2004. *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory.* Psychology Press.

[21] Susanne Putze, Dmitry Alexandrovsky, Felix Putze, Sebastian Höffner, Jan D. Smeddinck, and Rainer Malaka. 2020. Breaking The Experience: Effects of Questionnaires in VR User Studies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Honolulu, HI, USA. https://doi.org/10.1145/3313831.3376144

[22] Georg Regal, Jan-Niklas Voigt-Antons, Steven Schmidt, Johann Schrammel, Tanja Kojić, Manfred Tscheligi, and Sebastian Möller. 2019. Questionnaires Embedded in Virtual Environments: Reliability and Positioning of Rating Scales in Virtual Environments. *Quality and User Experience* 4, 1 (Dec. 2019), 5. https://doi.org/10.1007/s41233-019-0029-1

[23] Klaus R. Scherer and Heiner Ellgring. 2007. Are Facial Expressions of Emotion Produced by Categorical Affect Programs or Dynamically Driven by Appraisal? *Emotion* 7, 1 (Feb. 2007), 113–130. https://doi.org/10/d6c9vn

[24] Valentin Schwind, Pascal Knierim, Nico Haas, and Niels Henze. 2019. Using Presence Questionnaires in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 360:1–360:12. https://doi.org/10/gfz8sr

[25] Richard Skarbez, Frederick P. Brooks, Jr., and Mary C. Whitton. 2017. A Survey of Presence and Related Concepts. *ACM Comput. Surv.* 50, 6 (Nov. 2017), 96:1–96:39. https://doi.org/10/gfz8jf

[26] Mel Slater. 2003. A Note on Presence Terminology. *Presence connect* 3, 3 (2003), 1–5.

[27] Mel Slater and Anthony Steed. 2000. A Virtual Presence Counter. *Presence: Teleoperators and Virtual Environments* 9, 5 (2000), 413–434. https://doi.

org/10/cd5kxg

[28] Marco Speicher, Anna Maria Feit, Pascal Ziegler, and Antonio Krüger. 2018. Selection-Based Text Entry in Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 647:1–647:13. https://doi.org/10.1145/3173574.3174221

[29] Niels van Berkel, Jorge Goncalves, Katarzyna Wac, Simo Hosio, and Anna L. Cox. 2020. Human Accuracy in Mobile Data Collection. *International Journal of Human-Computer Studies* 137 (May 2020), 102396. https://doi.org/10.1016/j.ijhcs.2020.102396

[30] Maaike van den Haak, Menno De Jong, and Peter Jan Schellens. 2003. Retrospective vs. Concurrent Think-Aloud Protocols: Testing the Usability of an Online Library Catalogue. *Behaviour & Information Technology* 22, 5 (Sept. 2003), 339–351. https://doi.org/10.1080/0044929031000

[31] Keith Vertanen and Per Ola Kristensson. 2011. A Versatile Dataset for Text Entry Evaluations Based on Genuine Mobile Emails. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '11*. ACM Press, Stockholm, Sweden, 295. https://doi.org/10.1145/2037373.2037418

[32] Xinghui Yan, Katy Madier, Sun Young Park, and Mark Newman. 2019. Towards Low-Burden In-Situ Self-Reporting: A Design Space Exploration. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion - DIS '19 Companion*. ACM Press, San Diego, CA, USA, 337–346. https://doi.org/10/ghfcj8