

A mixed-methods measurement and evaluation methodology for mobile application usability studies

Paweł Weichbroth

WSB University in Gdansk

Grunwaldzka 238A,82-266 Gdansk, Poland

Email: pweichbroth@wsb.gda.pl

Abstract—Low usability of mobile application is thought to diminish the perceived level of the quality by a user whose experiences substantially determines its market success or failure. However, while a single method of measurement employed to study usability may produce an unreliable or incomplete evaluation outcome, in this paper, contrarily we propose to take advantage of both qualitative and quantitative methods adequate to collect data, that would describe all usability attributes. In particular, in the scope of mobile application usability studies, this paper (i) depicts the main assumptions of elaborated M4MAUME methodology, (ii) describes the self-developed software tool (RVDA) for retrospective video data analysis, (iii) specifies the experimental setup, and (iv) discusses the preliminary results obtained from 20 experiments performed on four different groups of mobile applications. Eventually our findings lead us to the conclusion that mixing different methods have produced reliable and valuable outcomes which may be used to improve and manage usability in current and future projects, as well as to enhance existing software quality assurance (SQA) programs.

I. INTRODUCTION

USABILITY of desktop interfaces has been widely studied. During the past 30 years, the human-computer interaction (HCI) community has developed a plethora of concepts [1,2,3], inspection techniques [4,5,6], frameworks [7,8,9], and heuristics [10,11,12] in order to better understand, measure and evaluate usability, which simply aims at eventually delivering better software products [13].

The long lived marriage of hardware and communication technologies has brought about the inevitable shift from desktop to mobile computing [14,15], leading to new user requirements regarding mobile applications [16,17,18]. As a consequence, usability engineers face the issue of usability measurement and evaluation in the instance of new software settings, including user interface (UI) design [19], connectivity [20], context-awareness [21], as well as hardware capabilities, concerning screen size [22], storage space [23] and overall performance [24].

In light of systematic literature review and analysis, covering the volume of 791 documents, indexed by Scopus database and published between 2001 and 2018, we determined the research methods applied to measure and evaluate usability of mobile applications [25]. Our findings also show that in only a few studies, collecting data concerned more than one technique, since very few utilized the retrospective video data analysis as well. This research

gap inspired us to establish a new laboratory, equipped with both hardware apparatus and software tools, as well as elaborate congruent methodology with all the necessary data collection techniques.

The remainder of the paper is organized as follows. In Section II, a brief research background is described, followed by the M4MAUME methodology (Section III). In Section IV the RVDA tool is specified. In Section V we detail the general settings of the experimental setup, followed by a discussion of the preliminary results obtained from the undertaken experiments. Finally, Section VII concludes the paper.

II. DEFINITIONS, MEASURES, AND METHODS

The generally accepted definition of usability is the one, given by ISO in 92411-11, which states that usability is the “extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [26]. This definition has been adopted in the majority of studies in the subject of mobile application usability.

In our latest study [25] we pointed out the ten most frequent usability attributes, namely: efficiency (70%), satisfaction (66%) and effectiveness (58%), learnability (45%), memorability (23%), cognitive load (19%), errors (17%), simplicity (13%), ease of use (9%) and navigation (8%), where the percentage in brackets show the relative frequency of occurrence of the attribute in reference to all extracted documents.

Efficiency is the ability of a user to complete a task with speed and accuracy [27]. Efficiency is measured in a number of ways, such as the duration to complete a given task, or the duration to finish a set of tasks [27,28]. In general, two methods are put into use: controlled observation, and survey.

Satisfaction is a user's perceived level of comfort and pleasure, or a user's perceived level of fulfillment of their expectations and needs [29]. Satisfaction is measured by using questionnaires and other qualitative techniques, typically used to capture a user's intangible attitude towards an application [29,30].

Effectiveness is the ability of a user to complete a task in a given context [31]. It is measured by the number of successfully completed tasks, the number of steps required to complete a task, the number of double taps unrelated to the

operation of an application, and the number of times that a back button is used on the mobile device [31,32]. To collect all necessary data to estimate measures, two methods are in common use: controlled observation, and survey.

Learnability is defined twofold: first-time and over-time. The former refers to the degree of ease with which a user can interact with a newly-encountered application without seeking guidance or referring to documentation. It is measured by the number of attempts to solve a task, the number of assists when performing a task, and the number of errors performed by a user [33]. Contrarily, the latter is the capacity of a user to achieve proficiency with an application. Typically, a user's performance during a series of tasks is observed to measure how long it takes these participants to reach a pre-specified level of proficiency [33,34].

Memorability is the degree of ease with which a user can remember how to use an application effectively. It is measured by asking users to perform a series of tasks after having become proficient with the use of the application, and afterwards asking them to perform similar tasks after a period of inactivity. To determine how memorable the application was, a comparison is made between the two sets of results [35,36]. In a few cases, eye tracking techniques were utilized to a greater extent in the usability studies.

Cognitive load refers to the amount of mental activity imposed on a user's working memory during application usage [37]. Cognitive load theory differentiates cognitive load into three types: extraneous, intrinsic and germane [38]. Extraneous refers to instructional and presentation schemas, caused by the mental activities and elements that do not directly support application usage; intrinsic refers to the task complexity, caused by the number of elements in a task and the degree to which these elements are related to each other; germane regards to the amount of mental effort used to form schemas and actively integrate new information with prior knowledge during application usage. In the practice of cognitive load measurement, instruments such as a subjective rating scale, a thinking aloud dual task protocol or eye tracking have been used [32,39].

Errors relates to the amount and type of errors which occur during task performance by a user [40]. On the other hand, it is the ability of an application to recover from occurred errors. Both these definitions also respectively reflect a way the attribute is measured [40,41]. Controlled observation and survey are the only methods used to observe both application and users' performance or to collect a users' perceived level of application usage without error.

Simplicity is the degree of being easy to understand or being uncomplicated in form or design, described by such characteristics as the number of menu levels, the number of performed gestures to reach a destination object, and the duration of searching a button to perform a specific function. On the other hand, simplicity is the level of comfort with which a user is able to complete a task, measured by predefined statements with the Likert-scale rating.

Ease of use is the perceived level of the user's effort related to usage of the application. The survey instrument is used to collect data from users on perceptions concerning their experienced interaction with the application [40,42].

Navigation is the perceived level of the user's effort to access relevant information. Similarly, the survey instrument is applied to collect data from users on perceptions concerning their understanding of the information architecture [43,44,45].

III. THE 3M4MAUME METHODOLOGY

Theoretical background

By definition, methodology is a body of methods, rules and postulates employed by a discipline [46]. In brief, a method is a process of doing something [47], a rule is a prescribed guide for conduct [48] and postulate is a hypothesis advanced as an essential premise of a train of reasoning [49].

Theoretical triangulation is the use of multiple theories or hypotheses when examining a phenomenon [50], while methodical triangulation has been defined as multimethod, methods triangulation, or simply mixed-method [51]. The application of triangulation, in the scope of multiple sources of data, enhances the reliability of results [52]. There is a direct link between data triangulation and data saturation, where the former is a method to establish the latter.

The quantitative paradigm is based on positivism [53], where evidence is characterized by empirical research. This group of methods emphasize objective measurements and the numerical analysis of data collected through polls, surveys and questionnaires. In contrast, the qualitative paradigm is based on interpretivism [54] and constructivism [55]. The three most common methods are: focus groups (group discussions), individual in-depth interviews, and participant observation [56].

Verbal protocol analysis (VPA) is a method for collecting and analyzing verbal data regarding cognitive processing [57]. In other words, VPA is the record of spoken thoughts, provided by subject when thinking aloud during, or immediately after completing a task [58]. It is based on the premise that in order to capture a participants real and authentic experiences, we must allow them to express themselves freely [59]. Alternatively, in a narrow sense, this method (protocol) is also defined as think aloud or thinking aloud, and is relatively free-form [60] and open-ended.

Main assumptions

The Mixed-Methods Methodology for Mobile Application Usability Measurement and Evaluation (3M4MAUME) is a body of three integrated methods, namely: (1) survey, (2) participant observation and (3) verbal protocol analysis. In this regard, this multi-faceted approach is designed to take advantage of both quantitative (1) and qualitative (2 & 3) approaches, which in a specified combination are able to

provide a full panorama on an application's quality-in-use characteristics, as well as user's attitudes and perceptions.

In particular, each survey is implemented according to a series of steps, each of which includes an application-specific functionality and defined set of formats and procedures.

Participant observation simultaneously takes place in conjunction with thinking aloud protocol. To preserve privacy, and encourage subject confidence, only an observer is present during the testing session. Moreover, each participant can ask questions to clarify aspects of the task, but all answers given are as brief as possible to minimize the time burden during the application testing session.

The observer has access to an external monitor which displays the participant's interaction with the application in real-time mode, while a third-party application records the session, including both video and voice data.

After the test, the participant is asked to run through the ideas generated and explain the thinking behind them, where not already mentioned.

Task no 1. Data collection

The usability testing procedure aims to collect data for analysis, and is the process described by a fixed sequence of three steps:

- 1) Pre-testing questionnaire collects the demographic data and ten statements regarding a mobile applications usage and observed usability issues, and following the self-assessment of the skills used to perform five tasks which correspond to the main application functionality, as well as to the tasks assumed to be performed afterwards by the participant.
- 2) Application testing session is described by the protocol which assumes participant recording during application usage, coded by the audio/video hardware apparatus in order to collect both voice and video data, without assistance or guidance. Moreover, participants are always encouraged to think aloud about the application properties and behavior, as well as to speak frankly of any other important issues.
- 3) Post-testing questionnaire which aims to reproduce the perceived quality in use, specified by ten usability attributes where each is typified by at least five statements.

The pre and post questionnaires are individually administered to the participants, before and after the test respectively. Therefore, the output of this task concerns two questionnaires, and an audio-video binary file, which gathered from the whole group, are afterwards verified, and eventually serve as the input for the data analysis task.

Task no 2. Data analysis

The data analysis task is the process described by a fixed sequence of the following three steps:

- 1) Inspecting video content that comprises annotation procedures in which the user's actions and application

responses are identified, separated and marked on the timeline.

- 2) Documenting all identified application bugs, defects, errors, and any reported usability issue.
- 3) Extracting numerical values required to calculate usability attribute measures.

It is worth noting that to perform the above task, specific software tools are required to obtain reliable outcomes with acceptable accuracy. Moreover, all necessary calculations are undertaken, regarding quantifying attributes measures and estimating structural model parameters.

Task no. 3. Information visualization

The third stage involves information visualization on the dashboard to empower cognition of the extracted and analyzed data which precedes and facilitates usability evaluation. For that purpose we specified a weighted and labeled graph. Vertices show the sequence of user's actions, and weights are used to represent the duration of application responses. In other words, the graph is a reconstructed image of the interaction which occurred between user and application, enriched by the duration of particular actions and responses.

Task no 4. Usability evaluation

Having measured and estimated all usability measures (see Section II), as well as the visualization of all relevant information, one can further analyze, classify and interpret the obtained outcome. On the other hand, some might go through the audio-video recordings, investigating these cause-and-effect relationships that may lead to a loss of effectiveness and a decrease in satisfaction.

In our opinion, the enriched action-response model provides an effective approach to evaluating particular usability attributes. At the interpretive level of research, the results of the quantitative analysis, provide explanations for evaluators to compile consistent and proper judgments.

The outcome of this step is the report, which in general presents results and conclusions, as well as a list of recommendations with applicable participant's reviews. In particular, the report categorizes usability issues into three groups: (1) bugs and errors, (2) design and (3) performance. In this line of thinking, the addressed groups may respectively concern testers, designers and developers.

IV. THE RVDA TOOL

To develop the RVDA tool (Retrospective Video Data Analyzer) we used the Electron open source library. Developed by GitHub, and devoted for building cross-platform desktop applications with HTML, CSS, and JavaScript, combines Chromium and Node.js into a single runtime environment [62]. We also used npm to manage packages for JavaScript runtime environment. In general, the above are the major devkit components.

The architecture of the tool has been conceived to allow for an easy and scalable video content analysis, supporting all common video file formats available for playback, where one or more files can be simultaneously opened and analyzed in separate windows.

From a technical viewpoint, there are main modules and rendering modules. The former is responsible for operations on windows, user interface properties, and communication with the operating system. The latter includes tasks on the video content and all associated functions. By design, neither communication nor data exchange occur within modules.

The user interface (UI) is divided into three sections: (1) the video preview area, (2) menu panel, and (3) display manipulators and the timeline. The size of the sections can be easily modified, along with the position of the tool on the computer screen.



Fig. 1 The RVDA Graphical User Interface

There are three media controls available (Stop, Play, Slow) to enact, change or adjust the process of playing video content. These controls support human cognition during processing video data that, in turn, advances the analyzing and extraction of data.

There are four levels available, which are used to break down the tasks to the user's actions and the application's responses (events). Accordingly, the Start and Stop buttons are used to drop pins on the timeline which mark the beginning and the end of each event. Moreover, there are also Save, Reset, and Clear features available.

The Export feature allows timeline data to be saved to the external CSV file which contains the event: identifier, name, start time, end time, and the level number. In this case, data obtained from a series of experiments can be consolidated for further reporting and analysis.

V. EXPERIMENTAL SETUP

The inception of the laboratory was in June 2018, and it is located in the WSB University in Gdansk. Both hardware and software apparatus were chosen to meet specified requirements regarding project scope and budget.

Each application testing session begins with a short introduction, including goals and research agenda. Afterwards the usability testing procedure is described and executed. Our experience shows that a single session lasts approximately 15 minutes.

Hardware apparatus

For real-time image and voice capture and recording we used the Genee Vision 150 document camera. The design and specification combine five major components: optics, camera, lighting system and motherboard with firmware, and software tool.



Fig. 2 The Genee Vision 150 document camera

The built-in camera resolution is 2 megapixel, resulting in image dimensions of 1920 by 1080. While resolution is an important specification, those described above proved to be adequate to extract all required data in the context of the study. Optical and digital zoom is 8x and 10x, and the camera can rotate vertically 180° and 180° horizontally. The device is connected via a USB port to the local computer, equipped with the software apparatus and the latest hardware drivers and software libraries.

Software apparatus

In total we used four software tools. Firstly, the VideoCap captures audio-video data, transmitted by the document camera, to the mp4 file format. Secondly, the input mp4 data is analyzed under the RVDA tool. Thirdly, in order to document and to actively collaborate with all interested parties on the recognized usability issues, on the external virtual server we have installed, configured and deployed self-hosted Git service, namely Gogs. Fourthly, for advanced statistical analysis we take advantage of commerce and open-source tools to make sense of quantitative data.

VI. METHODOLOGY TESTING AND VALIDATION

The elaborated methodology was tested from January to May 2019, in a group of four simultaneous student projects. Each project concerned a different set of mobile applications. Usability testing included five adult participants, both males and females with relevant knowledge and skills, who were using their own smartphones during the session.

Moreover, the methodology underwent one proof of concept (PoC), despite the preliminary stage of its maturity. During the PoC meeting, the interested parties have underlined its practicality, as well as the possible benefits to both designers and developers. Here, it is worth highlighting just one voice of appraisal: *“Despite economic constraints, analysis of video has strong software quality validity, offering rich insight into both application and user behavior. Studying the interactions in a retrospective manner provides valuable lessons and affirmation for quality assurance practices”*.

VII. CONCLUSIONS

We are aware that the methodology needs further consultation and discussion with the software industry, with the goal of reaching an optimal cost-benefit ratio. From the economic perspective, the methodology is time-consuming and labor-intensive. While mobile application development projects have relatively low budgets, there may be no allowance to reserve sufficient funds for this area.

However, in conclusion, while the methodology has been acknowledged to be application-agnostic, fully adaptive and replicative, we believe that in the near future some interested parties will decide to implement our methodology, or at least incorporate some of its components.

ACKNOWLEDGMENT

I would like to thank the Rector of the WSB University in Gdansk, Professor Jan Wiśniewski, for his financial support and motivation to continue research, and Professor Małgorzata Rozkwitalska, the Head of the Management Department, for valuable insights and helpful discussions.

REFERENCES

- [1] McGee, M., Rich, A., & Dumas, J. (2004). Understanding the usability construct: User-perceived usability. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 907–911.
- [2] Sung, J., & Yun, Y. (2010). Toward a more robust usability concept with perceived enjoyment in the context of mobile multimedia service. *International Journal of Human Computer Interaction*, 1(2), 12–32.
- [3] Koohang, A. (2004). Expanding the concept of usability. *Informing Science: International Journal of an Emerging Transdiscipline*, 7, 129–141.
- [4] Jeffries, R., & Desurvire, H. (1992). Usability testing vs. heuristic evaluation: was there a contest? *ACM SIGCHI Bulletin*, 24(4), 39–41.
- [5] Zhang, Z., Basili, V., & Shneiderman, B. (1999). Perspective-based usability inspection: An empirical validation of efficacy. *Empirical Software Engineering*, 4(1), 43–69.

- [6] Goldberg, J. H., & Wichansky, A. M. (2003). Eye tracking in usability evaluation: A practitioner's guide. *The Mind's Eye*, 493–516.
- [7] Norman, D. A., & Draper, S. W. (1986). *User centered system design: New perspectives on human-computer interaction*. CRC Press.
- [8] Desmet, P. M., & Hekkert, P. (2007). Framework of product experience. *International journal of design*, 1(1), 57–66.
- [9] Plattner, H., Meinel, C., & Leifer, L. (Eds.). (2010). *Design thinking: understand–improve–apply*. Springer Science & Business Media.
- [10] Instone, K. (1997). Usability heuristics for the Web. *Web Review*, October.
- [11] Nielsen, J. (2005). Ten usability heuristics.
- [12] Paz, F., Paz, F. A., Pow-Sang, J. A., & Collantes, L. (2014). Usability heuristics for transactional web sites. *11th International Conference on Information Technology: New Generations*, 627–628. IEEE.
- [13] Weichbroth, P. (2018). Delivering Usability in IT Products: Empirical Lessons from the Field. *International Journal of Software Engineering and Knowledge Engineering*, 28(07), 1027–1045.
- [14] Manchester, J. (2013). Software Revolution, Part IV: Computing in a Mobile World. <https://www.forbes.com/sites/truebridge/2013/10/22/software-revolution-part-iv-computing-in-a-mobile-world/#325625c71cd5> [2019-05-14].
- [15] Truong, A. (2016). China saw a massive shift from desktop to mobile computing in 2015. <https://qz.com/622543/china-saw-a-massive-shift-from-desktop-to-mobile-computing-in-2015/> [2019-05-14].
- [16] Varshney, U. (2003). Issues, requirements and support for location-intensive mobile commerce applications. *International Journal of Mobile Communications*, 1(3), 247–263.
- [17] Wasserman, T. (2010). Software engineering issues for mobile application development. *FoSER 2010*.
- [18] Parsons, D., Ryu, H., & Cranshaw, M. (2006, July). A study of design requirements for mobile learning environments. In *Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06)*, 96–100. IEEE.
- [19] Nilsson, E. G. (2009). Design patterns for user interface for mobile applications. *Advances in engineering software*, 40(12), 1318–1328.
- [20] Hoebeke, J., Moerman, I., Dhoedt, B., & Demeester, P. (2004). An overview of mobile ad hoc networks: applications and challenges. *Journal-Communications Network*, 3(3), 60–66.
- [21] Anagnostopoulos, C. B., Tsounis, A., & Hadjiefthymiades, S. (2007). Context awareness in mobile computing environments. *Wireless Personal Communications*, 42(3), 445–464.
- [22] Raptis, D., Tselios, N., Kjeldskov, J., & Skov, M. B. (2013, August). Does size matter?: investigating the impact of mobile phone screen size on users' perceived usability, effectiveness and efficiency. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*, 127–136. ACM.
- [23] Nkosi, M., & Mekuria, F. (2011). Improving the capacity, reliability & life of mobile devices with Cloud Computing. In *2011 IST-Africa Conference Proceedings*, 1–9. IEEE.
- [24] Wajahat, N. (2019). Mobile App Performance Testing: CheckList, Tools (Android & iOS). <https://www.guru99.com/mobile-app-performance-testing-strategy-tools.html> [2019-05-17].
- [25] Weichbroth, P. (2019). Usability of mobile applications: a systematic literature study (in print).
- [26] ISO 9241-11:2018(en). Ergonomics of human-system interaction. Part 11: Usability: Definitions and concepts.
- [27] Harrison, R., Flood, D., & Duce, D. (2013). Usability of mobile applications: literature review and rationale for a new usability model. *Journal of Interaction Science*, 1(1).
- [28] Ghazizadeh, F. Z., & Vafadar, S. (2017). A quantitative evaluation of usability in mobile applications: An empirical study. *International Symposium on Computer Science and Software Engineering Conference (CSSE)*, 1–6. IEEE.
- [29] Kabir, M. A., Salem, O. A., & Rehman, M. U. (2017). Discovering knowledge from mobile application users for usability improvement: A fuzzy association rule mining approach. In *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, 126–129. IEEE.
- [30] Krzewińska, J., Indyka-Piasecka, A., Kopel, M., Kukla, E., Telec, Z., & Trawiński, B. (2018). Usability Testing of a Responsive Web System for a School for Disabled Children. In *Asian Conference on Intelligent Information and Database Systems*, 705–716. Springer.

- [31] Widyanti, A., & Ainizzamani, S. A. Q. (2017). Usability evaluation of online transportation' user interface. *International Conference on Information Technology Systems and Innovation*, 82–86. IEEE.
- [32] Saleh, A., Ismail, R., & Fabil, N. (2017). Evaluating usability for mobile application: a MAUEM approach. *Proceedings of the 2017 International Conference on Software and e-Business*, 71–77. ACM.
- [33] McDonald, K., Abell, W., Smith, C., & Gibbs, S. (2016). Lessons learned from evaluating a mobile app out in the field. *4th International Conference on User Science and Engineering (i-USEr)*, 5–10. IEEE.
- [34] do Nascimento Mendes, I., & Dias-Neto, A. C. (2016). A Process-Based Approach to Test Usability of Multi-platform Mobile Applications. *International Conference of Design, User Experience, and Usability*, 456–468. Springer.
- [35] Aljaber, T., & Gordon, N. (2017). A guidance and evaluation approach for mHealth education applications. *International Conf. On Learning and Collaboration Technologies*, 330–340. Springer.
- [36] Parsazadeh, N., Ali, R., & Rezaei, M. (2018). A framework for cooperative and interactive mobile learning to improve online information evaluation skills. *Computers & Education*, 120, 75–89.
- [37] Lyons, K. M. (2014). U.S. Patent Application No. 13/995,514.
- [38] Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition and instruction*, 8(4), 351–362.
- [39] Chynał, P., Szymański, J. M., & Sobecki, J. (2012). Using eyetracking in a mobile applications usability testing. *Asian Conf. On Intelligent Information and Database Systems*, 178–186. Springer.
- [40] Ryan, C., & Gonsalves, A. (2005). The effect of context and application type on mobile usability: an empirical study. *Proceedings of the Twenty-eighth Australasian Conf.* 38, 115–124.
- [41] Wang, Y., & Brennan, R. (2017). Building High Usability Consumer-oriented Linked Data Mobile Apps. *AICS*, 295–307.
- [42] Korhan, O., & Ersoy, M. (2016). Usability and functionality factors of the social network site application users from the perspective of uses and gratification theory. *Quality & Quantity*, 50(4), 1799–1816.
- [43] Biel, B., Grill, T., & Gruhn, V. (2010). Exploring the benefits of the combination of a software architecture analysis and a usability evaluation of a mobile application. *Journal of Systems and Software*, 83(11), 2031–2044.
- [44] de Paula, D. F., Menezes, B. H., & Araújo, C. C. (2014). Building a quality mobile application: A user-centered study focusing on design thinking, user experience and usability. In *International Conference of Design, User Experience, and Usability*, 313–322. Springer.
- [45] Tomaschko, M., & Hohenwarter, M. (2018). Usability Evaluation of a Mobile Graphing Calculator Application Using Eye Tracking. In *International Conference on Learning and Collaboration Technologies*, 180–190. Springer.
- [46] Merriam-Webster English Dictionary, <https://www.merriam-webster.com/dictionary/methodology> [2019-04-06]
- [47] <https://www.merriam-webster.com/dictionary/method>
- [48] <https://www.merriam-webster.com/dictionary/rule>
- [49] <https://www.merriam-webster.com/dictionary/postulate>
- [50] Denzin, N. K. (2017). *The research act: A theoretical introduction to sociological methods*. Routledge.
- [51] Thurmond, V. A. (2001). The point of triangulation. *Journal of nursing scholarship*, 33(3), 253–258.
- [52] Brown, P. A. (2008). A review of the literature on case study research. *Canadian Journal for New Scholars in Education*.
- [53] Sale, J. E., Lohfeld, L. H., & Brazil, K. (2002). Revisiting the quantitative-qualitative debate: Implications for mixed-methods research. *Quality and quantity*, 36(1), 43–53.
- [54] Altheide, D. L. & Johnson, J. M. (1994). Criteria for assessing interpretive validity in qualitative research. In: Denzin, N. K. & Lincoln, Y. S. (Eds), *Handbook of Qualitative Research*. Thousand Oaks, CA: Sage Publications, 485–499.
- [55] Holtz, P., & Odağ, Ö. (2018). Popper was not a Positivist: Why Critical Rationalism Could be an Epistemology for Qualitative as well as Quantitative Social Scientific Research. *Qualitative Research in Psychology*, 1–24.
- [56] Mack, N. (2005). *Qualitative research methods: A data collector's field guide*.
- [57] Miller, R. L., & Brewer, J. D. (Eds.). (2003). *The A–Z of social research: a dictionary of key social science research concepts*. Sage.
- [58] Kasper, G. (1998). Analysing verbal protocols. *Tesol Quarterly*, 32(2), 358–362.
- [59] Guterman, O., & Neuman, A. (2017). Different reasons for one significant choice: Factors influencing homeschooling choice in Israel. *International Review of Education*, 63(3), 303–318.
- [60] Lockton, D., Harrison, D., Holley, T., & Stanton, N. A. (2009). Influencing interaction: development of the design with intent method. *4th International Conf. on Persuasive Technology*, ACM.
- [61] Van Den Haak, M., De Jong, M., & Jan Schellens, P. (2003). Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. *Behaviour & information technology*, 22(5), 339–351.
- [62] Electron Documentation, <https://electronjs.org/docs/tutorial/about>.