Scared, frustrated and quietly proud: Testers' lived experience of tools and automation

Isabel Evans
University of Malta
Department of Computer Information Systems
Msida, Malta
isabel.evans.17@um.edu.mt

Chris Porter
University of Malta
Department of Computer Information Systems
Msida, Malta
chris.porter@um.edu.mt

Mark Micallef
University of Malta
Department of Computer Science
Msida, Malta
mark.micallef@um.edu.mt

ABSTRACT

Software testing is vital, expensive, time-consuming yet a necessary part of software development. Testers perform repeated actions during testing, where automation and tools could reduce costs, timescale and human error. However, challenges to tools adoption have been identified in academic research and industry, which are blockers to success with automation. In attempting to find whether testers were experiencing tool usability shortcomings, we followed an exploratory research path, collecting stories from over 100 test practitioners. We discovered a richer, more complex story than we expected. We realised that usability – while necessary – is not sufficient to enable success, and that other human factors challenge successful automation projects. In answering privately to questions about their experiences of tools and automation, testers expressed themselves in language that was more emotional and linked to their lived experience (LX) than we expected. We uncovered frustrations and fear, as well as pride. In this paper we present our findings so far about *TX: The testers' lived experience of tools and automation*, and we suggest steps for future research.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; *Field studies*; *HCI theory, concepts and models*; **Human computer interaction (HCI)**;

KEYWORDS

user experience, lived experience, usability, human factors, software testing, test automation, test tools

ACM Reference Format:
Isabel Evans, Chris Porter, and Mark Micallef. 2021. Scared, frustrated and quietly proud: Testers' lived experience of tools and automation. In *European Conference on Cognitive Ergonomics 2021 (ECCE 2021), April 26–29, 2021, Siena, Italy.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3452853.3452872

ECCE 2021, April 26–29, 2021, Siena, Italy

ACM ISBN 978-1-4503-8757-6/21/04…$15.00
https://doi.org/10.1145/3452853.3452872

1 INTRODUCTION

The ubiquity of software and the demand for fast delivery of that software present a dual challenge for software teams: greater quality is required, coupled with tight budgets and short time frames to meet the demands of competition [20]. Measuring software quality through software testing is a vital but time-consuming part of the software development process. It is heavily relied on by teams and organisations to provide information for decisions about the readiness of software for its customers. Yet, it is difficult and expensive to carry out [27]. Testing requires many repeated actions, comparisons and data processing. Aspects of testing therefore are candidates for automation. Attempts to automate test execution, and provide tool support for test management monitoring and reporting, are common industry activities. In addition, academic research looks at tools support for a wide range of testing activities.

Some authors argue that the HR resourcing model for traditional testers has reduced, or even been completely removed, while others argue that automation will not replace testers, because people are capable of seeing and understanding issues that automation would not see. The resulting discussion is passionate and wide-reaching, covering the role of testers, the scope of automation, and the craft of testing itself [3]. Meanwhile, some academic research pushes towards 100 per cent automation as a desirable goal [2]. This 100 per cent goal means a tester's experience of tools is not just a matter of skills or usability but a lived experience (LX) of potential redundancy, which could affect their motivation to make the automation project successful.

Automation provides benefits: *"no matter how valuable in-person testing is, effective automation is able to increase the value of overall testing by increasing its (. . .) range"* [15]. However, successfully adopting automation can be problematic; academic research and industry shared experience have identified a number of known problem areas for tools adoption, such as tools acquired but not used (shelfware) and difficulties in sustaining tool and automation use. These include technical issues, for instance maintenance, usability issues, and organisational issues, including management support and staff skill sets [8, 10, 29]. Personal observations during consultancy work and at conferences informed this research; we asked whether focusing on technical and managerial issues was enough to overcome the shelfware phenomenon. Our findings agree that there are challenges, what Wiklund [29] calls *impediments*, to success with test automation and tools projects, including organizational, technical and usability problems. We have already reported in another paper, [6] the implications and possible treatments for the usability issues our participants reported. We have uncovered a richer and more complex story than we expected, including higher emotional content, and a greater range of emotions, affecting the Lived Experience of the testers. In this paper we present data to support an earlier position paper [5]. We have heard stories of pride, fear, frustration, triumph, and anger. We have also found that some participants were happy to share their concerns privately, but not in the open forum of a workshop.

The rest of this paper is arranged as follows. We provide the background to our research, describe our research design and method, with reference to the literature review, then discuss our results and their implications, before suggesting further research areas in our conclusion.

2 BACKGROUND

2.1 Defining Automation and Tools

The Cambridge English Dictionary defines a tool as a piece of equipment used by a person to enable them to carry out a task, and automation as the use of tools to complete activities without or with minimal human intervention. Organisations and teams seek to remove or improve activities perceived as difficult, time consuming and expensive, with the hope that automating these will reduce staff levels, decrease time to market, and improve quality. In software testing, the terms 'tool' and 'automation' are sometimes used interchangeably [2, 3, 18]. We define *"test tools and automation"* as any tool used to support any testing activity. In our study, we asked open-ended questions that allowed the practitioner participants to include as a "tool" or "automation" anything they felt was in scope.

2.2 Effect of Automation and Tools

If the *"purpose of the software industry is to provide benefit to its stakeholders by automating activities of all kinds"* [5] then it makes sense to also automate activities within the software development sphere, to provide benefits. However, automation can be a threat as well as a benefit, as it may be mistrusted, misapplied, or may adversely affect people's lives. When we attempt to automate software development tasks like testing, this changes or even removes part of people's roles, which affects their working and personal life, possibly beneficially and possibly negatively. Successful tool support and automation removes repetitive work from people, and supports tasks that have a high cognitive load. When successful, this provides productivity gains, removes tedious or difficult tasks, and potentially is life enhancing. Automation also brings threats to motivation, including fear of redundancy, dissociation, frustration with flawed automation, and problematic job-task mixes where the role for the person becomes more boring and stressful [5, 28].

Organisations which are concerned with the emotional welfare of their staff, as well as with profit and productivity, are both humane and far sighted; concern for staff happiness contributes to health and productivity. Motivation studies show that software testers are more productive when motivated [4, 25, 28].

2.3 Usability, Quality in Use, UX and LX

Figure 1 shows how LX is built up through contributing factors, from the UI design, through usability, to UX and on to LX and how these affect a tester, the user of a test tool. Each of these building blocks contribute to the emotional content of the experience, and hence to the overall LX.

The experience that each person has of the technology they use, is affected by the design of the user interface (UI), however the UI design alone is not enough to support suitable usability characteristics, and not enough to support a good overall user experience (UX). The UI design combines with the interaction design to support usability. There are several definitions of usability in the literature, for example in ISO 9241 and ISO 25010 standards [16, 17]. For this paper we define usability as in ISO 9241, in order to consider user goals, user effectiveness, user efficiency, as well as user satisfaction in a specified context of use. Usability contributes to the ISO 25010 definition of quality in use, which also includes flexibility and freedom from risk, both of which contribute to the overall user experience (UX). Usability and quality in use attributes are built on a foundation of technical software attributes such as maintainability, portability, and reliability. Norman and Nielsen [22] remark, "*[User experience] encompasses all aspects of the end-user's interaction with the company, its services, and its products [meeting the needs of the customer] without fuss and bother.*"

The lived experience (LX) builds on UX to also include "*repercussions extending into the users' daily lives and away from the technology itself*" [23], including the emotional impact of technology. UX attributes such as trust, flow and credibility contribute to the user's LX, and is influenced by past experiences, mental models and personal circumstances. In focusing on LX, we understand that our own lived experience affects our research questions and approach: our *"experiences, choices and options influences one's perception of knowledge"*, the questions we ask and how we interpret the responses to those questions [9]. Emotions are difficult to measure, and the science of emotion offers *"a perplexing multitude of emotion theories"* [1] with disagreement about what emotions are, and how to measure them [14].

2.4 Our Exploratory Journey

Guided by the research question *"What are the testers' experiences of automation and tools?"* we set out on an exploratory journey, asking testers to record their stories of problems and successes in automation and tool use. We intended to look for references to usability and quality in use attributes. However, during that exploratory journey, we were drawn by the data to extend our interest to UX and LX principles, because we found evidence of the emotional impact on testers of using tools. We asked whether tools and automation *"provide meaningful and relevant experiences"* to testers [7].

Looking at the LX of the tool, we consider the affect of the tool on the tester, immediately in their task, but more widely in its affect on their role, their self-perception, and their place in the organisation. Potentially this provides a new way to understand and overcome *"impediments to test automation"* [29] positively both for industry and for the tester.

3 METHOD

3.1 Mixed Method Approach

We used a mixed method approach to explore the interactions of testers with their automation tools, seeking to understand what

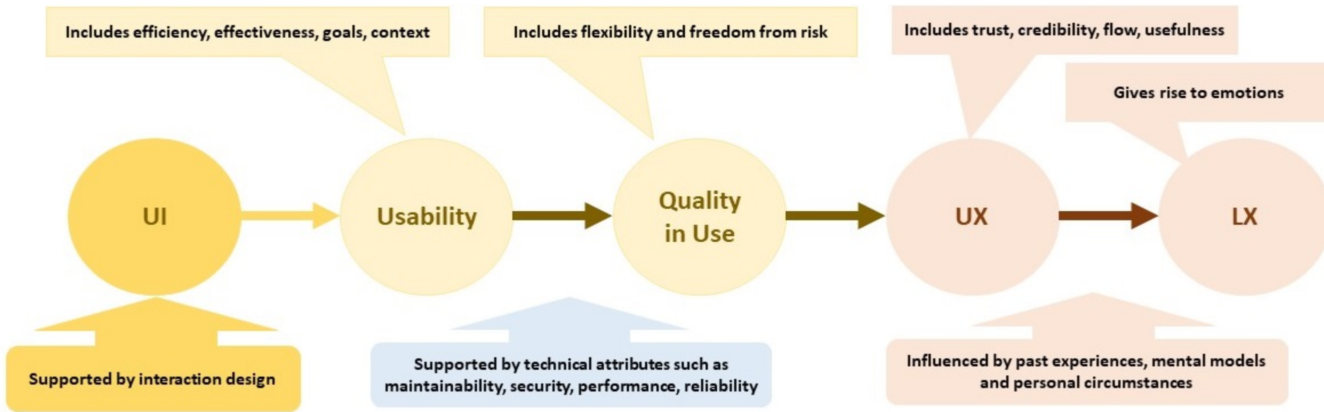

Figure 1: Contributors to LX and Emotions

problems hinder successful tool adoption. Figure 2 shows the outline of the research method. Following preliminary observations during conversations at conferences, we interviewed several testing and automation experts, and carried out a literature review. We then formulated our research question: *"What are the testers' experiences of automation and tools?"* and then held a series of workshops and surveys to collect testers' automation and tools stories.

Our hypothesis at the start of the study was that the challenges would fall into technical, organisational or usability themes [10, 29]. Examples of technical challenges include issues with the IT environment, security, and the tool's performance. Examples of organisational challenges include management support, and testers' skill sets [8]. Examples of usability challenges include poor learnability and poor operability for the tester.

3.2 Review of Literature

We used two main sources for the literature: industry authors and academic research. For industry, we included blogs, tweets, published reports and books. We used Google Scholar to search for academic papers with industry settings.

There are known technical and organisational challenges that Wiklund [29] describes as *"Impediments to Test Automation"*. Other authors identify problems with bias in tool selection, for example [10, 18, 24, 29]. These all appear to be significant factors and provided a starting point for analysis themes. Wiklund [29] analyses problems reported and information requests for one organisation's test automation provided via a user discussion board. Impediments to automation included user errors, usability, and user behaviour. Usability attributes from ISO 25010 were therefore used in our initial themes for coding. Our paper [6] discusses the ways that usability of tools is sometimes misapplied by tools designers. Other literature covers human factors for test automation and tools by examining organisational challenges such as managerial support, tester skill sets, and resourcing, but these do not focus on people's user experience of the tools [8, 10, 18]. Harty [15] comments on fragility and maintainability issues for the tools themselves, and on the skill sets required, and lacking, for testers to engage with tools. This work complements work being done elsewhere [11, 19].

Psychological influences on behaviour when choosing tools is discussed in [24] including the effects of uncertainty when decision making, and the tendency to then choose what is popular rather than optimal for one's own needs. However, the effect of that decision making on the user experience is not covered.

An important ethical aspect of human experiences is automony, for example Hamidi et al [13] notes that *"previous research in HCI has long identified the need to support human autonomy as a central ethical value"*. In our studies the importance of testers' (lack of) autonomy and control over their tools also becomes evident as a (de)motivating factor, as part of the emerging themes. We noticed during data collection that allowing participants both the space to express their emotional responses, and the safety of a private, anonymised mechanism, supported their autonomy and control over what they reported. For example, after workshops, several testers privately expressed to the researcher that they felt unable to give their true views with others present, but would be willing to complete an anonymous survey. We believe that by using interviews, conversations, anonymous surveys and workshops, we have added to previous studies work as we enabled participants to express emotional responses and know that those responses are valued by the researcher. This validates the testers' emotions [12], by the researcher being present to the expression of the emotion, reflecting it accurately, and accepting the emotion. Although work exists on emotions in software engineering, for example Wrobel [30], we are not aware of studies that consider the lived experience of software testers with respect to tools and automation. We conclude from this that there is a gap in the literature to specifically cover: that of the testers' LX of tools and automation, and specifically allowing private, autonomous, and safe sharing of experiences.

3.3 Data Collection

Four surveys were made during 2018, via industry conferences. The question sets included open ended questions such as: *tell me a bit about yourself*, and: *tell me a story about an experience with automation / tools.* Two of the surveys were preceded by workshops.

The entire corpus of data arising from these surveys consisted of 180 people's responses. Those responses which did not provide any

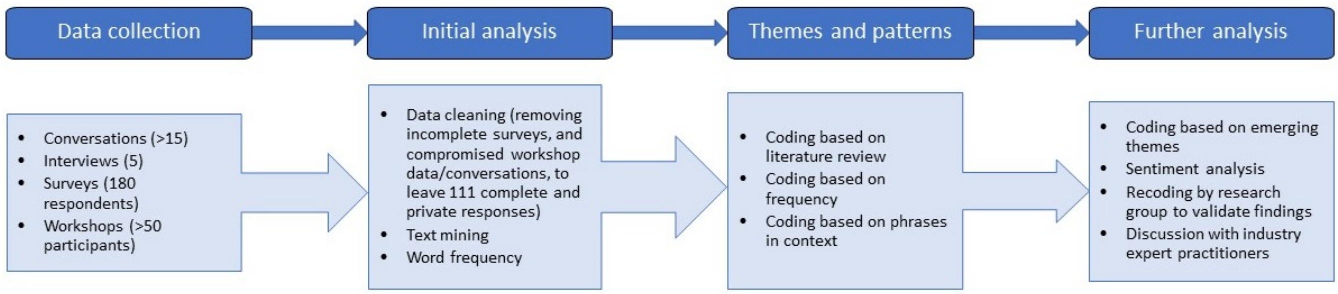

Figure 2: Research Method Flow

data for the *tell me a story* question were removed, leaving a total of 111 survey responses. These per-person records showed a range of responses and were analysed as one group. Additionally, edited transcription [26] was carried out on recorded expert interviews, and key points in conversations were documented. These were not used in the frequency analysis. Relevant quotes from both interviews and conversations are included in this paper. Because some participants were uncomfortable about expressing their views in workshops, we only used data from interviews, conversations and surveys.

3.4 Analysis

Text mining on the survey data was carried out as part of the exploration phase, including a word frequency exercise to determine a high-level weighting of each overarching theme, based initially on categories identified in the literature review ("Managerial/organizational", "Technical", "Usability" See Table 1). We counted positive comments, issues and challenges raised under each code. ISO 25010 software attributes were used to categorise perceived attributes of a good tool, concerns and issues (section 4.2). We performed several iterations of coding, starting with the pre-identifed themes and allowing other themes to emerge during analysis and coding. We identified perceptions, challenges, and commonly held ideas of what is important for successful tool adoption. As well as thematic coding, we used MS Azure Sentiment Analysis to assess the data. As a validity check, we assessed the data to look for patterns and correlations within and between respondents, revisited the text of the answers to check, reclassify, and recode if needed. We also discussed the results with a group of 10 software testing practitioners and experts, none of whom were previously involved.

4 RESULTS

In this section we discuss the results of the data analysis, participant profiles, themes and patterns that emerged. The anonymized dataset with all survey respondent quotes, anonymised participant profiles and code book will be available as an OSF project. Quotes from survey participants are notated [SxPy], which shows the survey and participant number.

4.1 Participant Profiles

Although to some extent convenience sampling was used, via our own networks, it was important to collect views from a wide range of people, spanning several geographical regions, length and type of experiences, and roles. We wanted to ensure we had not biased the responses we were receiving by only including those practitioners to whom we are closest, in geography or experiences. We therefore launched data collection at events in several countries, and online via a webinar and social media. Answers to survey question 1 indicated that we had responses from a range of participants. Participants came from different countries, including United Kingdom, New Zealand, China, India, Sweden and the USA. Not all participants provided demographic data, however, at workshops that preceded surveys S3 and S4, there was a mix of ages, gender, work experience, and of permanent, contract and freelance participants. A wide range of professional backgrounds was noted, from software engineering degrees to carpentry, and of domains including financial services, avionics and education. While 56 participants described themselves as testers, and 81 as technical/automators, job titles and roles included some quite complex mixes of responsibility and specialisms. These included a *"head of testing (and) tester/automator"*, a *"project manager whilst also tester (manual and strategy)"* and one with responsibilities including *"writing UI tests, security scans and writing applications"*.

Table 1: Themes: Organisational, Technical, Usability

Theme	Participants	Frequency
Organisational Management	101 (90.9%)	377
Technical Attributes	102 (91.8%)	466
Quality in Use / Usability	109 (98.2%)	511

4.2 Text Mining and Frequency Analysis

Table 1 shows the result of the text mining and frequency analysis of the data for software attributes. Tables 2 and 3 show the frequencies of emotional responses by survey question and compared to number of issues raised. By "concern" we mean any mention of that theme as important to the participant, whether a negative or a positive story. The themes of emotion and issues arose from our observations during the frequency analysis, and then was adopted into the coding.

The data from survey respondents shows that about 30 percent used emotional language, even though the questions were not emotional, and they were asked in a survey so the participants would not have been cued by the interviewer's body language or

Table 2: Theme: Emotional Responses

Survey question	Topic	Positive emotions	Negative emotions	Participant distribution
SQ1	About yourself	3	5	7 (6.3%)
SQ2	Story about using a tool	6	22	25 (22.5%)
SQ3	Difficult/easy aspects with tool	1	5	6 (5.4%)
SQ4	Story about avoiding a tool	2	17	12 (10.8%)
SQ5	Attributes of a good tool	1	2	3 (2.7%)
SQ6	Tools used in organisation	4	4	3 (2.7%)
SQ7	Tools used personally	0	0	0 (0%)

Table 3: Themes: Participants, Emotional Responses and Issues

Raised Issue	Showed emotion	No emotion	Comment on emotions
Yes	36 (32.4%)	46 (41.4%)	32 used negative emotions, 8 used positive emotions
No	4 (3.6%)	25 (22.5%)	3 used negative emotions, 1 used positive emotions

other unconscious biasing of their answers (see Table 2). Some responses contained both positive and negative emotions. Even the question *"which tools do you use in your organisation?"* elicited an emotional response from some participants, with one saying *"Many. Too many!"* [S3P4] and another *"Too many . . . a confusing cocktail of tools"* [S5P24]. Only SQ7 (own tools) elicited no emotional responses; many participants simply referred back to their SQ6 answer (organisational tools). The indicators of emotional responses we found in survey response, interviews and conversations included the use of emojis (e.g. frowny face), punctuation to indicate sarcasm (e.g. (!)), use of swear words, body language and facial expressions, obvious emotion words (e.g. frustrated, scary, proud), and giving one response in a workshop and another privately (indicating fear, embarrassment). When we compared the number of issues and challenges raised to the use of emotional language we found that most respondents raised issues, and that the number of issues/challenges did not seem to correlate with the level of emotional language. Our analysis found no correlation between length of story, size of vocabulary used in story, or number of challenges with the emotional language used.

5 DISCUSSION

The testers reported some successes, problems and challenges that were not surprising, and confirmed previous research in the field, and industry expectations. They reported organisational and technical factors, as well as usability as important in the success of test automation and tool adoption, factors that could be impediments or essential building blocks for success. Testers also raised UX challenges and human factors. These were often expressed in language or signalled in ways that were more emotional and more linked to their LX than we expected.

5.1 Corroborating Existing Results

A number of the points raised by respondents confirm enablers and blockers that have been reported elsewhere in academic studies or in industry reports. We confirm the findings of Wiklund [29] that *"neither the research community not practitioners should be surprised"* that managerial and organisational factors, such as management support/expectations, skills and knowledge, support from tools providers, are critical factors for success. These are known to industry trainers and authors [8]. All of the interviewees, and 33 of the survey responses, mentioned management support and expectations as critical to success or failure of a tool/automation project, including problems when a tooling project is not treated as a software acquisition project. This included misplaced expectations and underestimating the time/cost of skills training. Other problems included a lack of support, or changing levels of support from management and from suppliers, as well as short-term cost savings preferred to longer term investment.

- *Emotional Responses to Management*: One interviewee remarked *"(the manager) didn't realise software is a bloody difficult thing to build . . . it took three years as opposed to three months"*, while in a conversation at a conference, one automation developer showed strong emotions about management planning, saying *"what the hell were they thinking?"* about underestimated timescales for testers to learn the automation framework. This was also reflected in survey responses, with strong feeling shown about management and vendor support, such as frustration and surprise: *"frustrated (by) lack of internal support [. . .] even after discussions with the 'sponsor'"* [S3P2] and *"Getting the hang of it [. . .] modifying processes [. . .] found it was no longer supported . . . Aargh!"* [S3P15].
- *Emotion and Technical Challenges*: Technical/environmental blockers such as access, test environment set up, and poor maintainability affect testers' ability to function effectively and efficiently, are described by for example [29] and these were also mentioned by 27 survey responses, with 12 people specifically mentioning maintainability of the tests as a concern. Emotional nuances include uncertainty, frustration, for example: *"Trying to identify who had access, and what access they had [. . .] I know how to do it (have learned earlier) but I am still not sure which of the local groups give access to what*

. . . " [S5P12]. Other people's language indicated positive emotions, for example: *"For automation, I have a beautiful portal . . . "* [S5P15].
- *Emotion and Usability Challenges*: Wiklund [29] raise the importance of usability, as did 34 people across our surveys. This included reference to seamful tool support [21], and emotions such as frustration: *"(the best solution I have found) is a combo of (Tool 1) and (Tool 2), but neither does exactly what I want. Very frustrating. [. . .] I don't want to load each page manually and record the results in a spreadsheet so I can show change over time."* [S5P16].

5.2 Emerging Theme: Testers' LX (TX)

In addition to confirming accepted understanding of the enablers and blockers, the data revealed a new theme – *the use of emotion* that led to our paper [5]. The emotional content of LX for testers became clear in the interviews and surveys from the language used and the manner in which the information was conveyed. Aspects such as body language, facial expression, and tone of voice in viva-voce interviews as well as the use of emojis and punctuation in written and on-line answers informed the coding process. This emerging theme is the emotional connection between people and their tool sets, providing a need to emphasise the experiential over the technical. Emotional engagement is shown in Tables 2 and 3, as well as in the quotes. The emotional connection between people and the tools they use is surprising both in its range and depth. This was obvious in the conversations and interviews which all had a high emotional content. However, survey responses also contained emotional language, even when replying to unemotional questions.

During various interviews and conversations, practitioners demonstrated a high level of emotional engagement while discussing their work, reflected in their body language for example, bouncing on their chair, sitting to standing and back, and at times in tears. As with all the emotional responses, we did not set out to observe or record these physical manifestations of emotion; they became apparent as we interacted during conversations. They showed a desire to do more, and the need to feel that their work has significance. For example the point by a trainee automation engineer in one conference conversation: *'(I want to) do something that matters"* commenting on the reason for testing, and in another conversation, an automation expert expressing frustration with their management's favouring speed over quality: *"if we make this crap again, what's the point?"*. This was echoed by survey respondents, for example: *"I think I should leave my job and look for (a) company who actually values testing"* [S3P17]. These quotes shows the personal (lack of) satisfaction in doing a professional engineering job well (or poorly).

Negative emotional language used in responses included: painful; struggling; fear; frustration; infuriating; distrust, uncertainty. Whether impediments are technical and usability, as described by Wiklund [29] or managerial as described by Graham [10] the feelings of powerlessness and personal failure that result enhance the impediments by increasing a desire to delay, or even quit. The emotional impact of being unable to progress when attempting to use automation for testing is also clear: *"it is scary and I always get stuck. I am delaying the inevitable (frowny face)."* [S3P20]. One tester reported finding bugs in a vendor tool with a particularly tester-like positive response to identify problems: *"after being frustrated by it (as it cost me a day and a billing run) I was quietly proud (of breaking it)."* [S4P2]. Even the act of choosing a tool raised negative emotions, which reinforce the findings in [24] that uncertainty and choice led to poor decision making. While speaking about tool vendor marketing claims, one interviewee said these claims *"made my blood boil"*, and a survey participant wrote: *"how to choose test tools in a 'jungle' out there."* [S6P13].

In one case the evident lack of human autonomy [13] indicated a lived experience as demotivating and depressing as any factory production line: *"I don't really count what I did there as testing though, I ran scripts I didn't write, generated reports I didn't understand and quite honestly I don't think I could really tell you what the product I was testing did in any great detail. My only role was to make sure that the QC test run numbers were met and a spreadsheet was green."* [S5P8].

5.3 Validity

An external validity threat exists from convenience sampling via the industry community, mitigated by using an online survey launched internationally. The data used for this paper were selected from records that included rich and meaningful stories about tools, removing records that had incomplete or missing stories. We sought to overcome validity threats by repeating thematic coding as a group, and by reviewing results with industry experts.

6 CONCLUSIONS AND FUTURE WORK

Humans are emotional, sentient beings, and testers are human. When they use and experience objects, they have a range of perceptions and emotions. Depending on the object and any problems encountered using it, the perceptions may cluster around particular emotions, and these depend also on the individual tester and their particular lived experiences.

In this paper, we have explored problems testers encounter when using test automation and test tools. We set out to explore the usability of these tools. During our exploration, our data led us to consider the Lived Experience of the tools and automation. We contribute evidence for a greater emotional connection between people and their tool sets than expected, and for the importance of the testers' lived experience. We show the evidence we have collected for the lived experience of testers (TX) [5], discuss testers' perceptions of automation, and their frustration with their tools. TX - the Testers' Experience - is a concept built from a complex mix of technical, organisational, usability, UX and human factors, which informs our understanding of the outcomes of attempts to automate testing.

Wiklund's call [29] for *"investigations into how to systematically prevent commonly occurring impediments"* with the results *"packaged in a way that is attractive, available, and useful to practitioners"* now becomes not just a call for automation projects to be technically successful. With our findings on the emotional content of the testers' world, we recommend that organisations and tools builders take into account as a potential enabler the importance of human autonomy, and the totality of people's experience. We also suggest research to investigate further the ways that test automation of a role can affect motivation, of testers and other roles. This includes motivation in relation to redundancy and dissociation, job-task mix, and dealing with flaws in automation. We bring these ideas together under one concept, a new problem area for research, which we are calling Tester Experience (TX). This provides a rich seam of HCI research including multidisciplinary strands across ethics, sociology, psychology and organisational change management.

ACKNOWLEDGMENTS

The authors thank the paper reviewers for their useful and stimulating comments, Julian Harty for early concept discussions, the participants in workshops, surveys, interviews, industry conference discussions and meetups: ANZTB, UKSTAR, EuroSTAR, STARWEST, STAREAST, STARCanada, Women Who Test and the UK Software Testing Retreat.

REFERENCES

[1] Lisa Feldman Barrett. 2016. Navigating the science of emotion. In *Emotion measurement*. Elsevier, 31–63.
[2] Antonia Bertolino. 2007. Software testing research: Achievements, challenges, dreams. In *Future of Software Engineering (FOSE'07)*. IEEE, 85–103.
[3] Michael Bolton. 2016. *A context driven approach to automation in testing*. Retrieved June 15, 2020 from https://www.developsense.com/blog/2016/01/a-context-driven-approach-to-automation-in-testing/
[4] Anca Deak, Tor Stålhane, and Guttorm Sindre. 2016. Challenges and strategies for motivating software testing personnel. *Information and software Technology* 73 (2016), 1–15.
[5] Isabel Evans, Chris Porter, Mark Micallef, and Julian Harty. 2020. Stuck in Limbo with Magical Solutions: The Testers' Lived Experiences of Tools and Automation. In *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. SCITEPRESS-Science and Technology Publications, 195–202.
[6] Isabel Evans, Chris Porter, Mark Micallef, and Julian Harty. 2020. Test Tools: an illusion of usability?. In *2020 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW)*. IEEE, 392–397.
[7] Interaction Design Foundation. 2019. *What is User Experience (UX) Design*. Retrieved June 15, 2020 from https://www.interaction-design.org/literature/topics/ux-design
[8] Seretta Gamba and Dorothy Graham. 2018. *Test Automation Patterns*. Retrieved June 15, 2020 from http://testautomationpatterns.org
[9] Lisa M Given. 2008. *The Sage Encyclopedia of Qualitative Research Methods*. Sage publications.
[10] Dorothy Graham and Mark Fewster. 2012. *Experiences of Test Automation: case studies of software test automation*. Addison-Wesley Professional.
[11] Thomas R. G. Green and Marian Petre. 1996. Usability analysis of visual programming environments: a 'cognitive dimensions' framework. *Journal of Visual Languages & Computing* 7, 2 (1996), 131–174.
[12] Karen Hall. 2012. *Understanding Validation: A Way to Communicate Acceptance*. Retrieved June 15, 2020 from https://www.psychologytoday.com/gb/blog/pieces-mind/201204/understanding-validation-way-communicate-acceptance
[13] Foad Hamidi, Morgan Klaus Scheuerman, and Stacy M Branham. 2018. Gender recognition or gender reductionism? The social implications of embedded gender recognition systems. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.
[14] Cindy Harmon-Jones, Brock Bastian, and Eddie Harmon-Jones. 2016. The discrete emotions questionnaire: A new tool for measuring state self-reported emotions. *PloS one* 11, 8 (2016).
[15] Julian Harty. 2011. Finding Usability Bugs with Automated Tests. *Commun. ACM* 54, 2 (2011), 44–49.
[16] ISO. 2011. *ISO/IEC 25010:2011 Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models*. Standard. International Organization for Standardization, Geneva, CH. Retrieved June 15, 2020 from https://iso25000.com/index.php/en/iso-25000-standards/iso-25010
[17] ISO. 2018. *ISO 9241-11:2018 Ergonomics of human-system interaction - Part 11: Usability: Definitions and concepts*. Standard. International Organization for Standardization, Geneva, CH. Retrieved June 15, 2020 from https://www.iso.org/obp/ui/iso:std:iso:9241:-11:ed-2:v1:en
[18] ISTQB. 2018. *ISTQB Certified Tester Foundation Syllabus*. https://www.istqb.org/downloads
[19] Brittany Johnson, Yoonki Song, Emerson Murphy-Hill, and Robert Bowdidge. 2013. Why don't software developers use static analysis tools to find bugs?. In *Proceedings of the 2013 International Conference on Software Engineering*. IEEE Press, 672–681.
[20] Harlan D Mills and Richard C Linger. 2002. Cleanroom software engineering: Developing software under statistical quality control. *Encyclopedia of Software Engineering* (2002).
[21] Andrew M'manga, Shamal Faily, John McAlaney, Chris Williams, Youki Kadobayashi, Daisuke Miyamoto, UK Poole, and UK Porton Down. 2018. Qualitative adaptation: informing design for risk-based decision making. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference*. BCS Learning & Development Ltd., 216.
[22] Don Norman and Jakob Nielsen. 2020. *The Definition of User Experience (UX)*. Retrieved June 15, 2020 from https://www.nngroup.com/articles/definition-user-experience/
[23] Chris Porter. 2015. *Designing for Experience - a requirements framework for enrolment based and public facing e-government services*. Ph.D. Dissertation. University College London.
[24] Päivi Raulamo-Jurvanen, Mika Mäntylä, and Vahid Garousi. 2017. Choosing the right test automation tool: a grey literature review of practitioner sources. In *Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering*. ACM, 21–30.
[25] Stuart Reid. 2015. *Motivated or Motivating? What sort of tester are you?* Retrieved June 15, 2020 from http://www.stureid.info/home-page/white-papers/motivation-testing-roles
[26] Speechpad Team. 2019. Types of transcription. Retrieved June 15, 2020 from https://www.speechpad.com/blog/types-of-transcription
[27] Gregory Tassey. 2002. The economic impacts of inadequate infrastructure for software testing. *National Institute of Standards and Technology, RTI Project* 7007, 011 (2002). https://www.nist.gov/system/files/documents/director/planning/report02-3.pdf
[28] RH Warden and IA Nicholson. 1996. Motivational Survey of IT staff. *Software Futures* (1996).
[29] Kristian Wiklund. 2015. *Impediments for Automated Software Test Execution*. Ph.D. Dissertation. Mälardalen University.
[30] Michal R Wrobel. 2013. Emotions in the software development process. In *2013 6th International Conference on Human System Interactions (HSI)*. IEEE, 518–523.