End-User Development of Experience Sampling Smartphone Apps - Recommendations and Requirements

Rough, Daniel J.; Quigley, Aaron J.

DOI:
10.1145/3397307

Publication date:
2020

Document Version
Peer reviewed version

Citation for published version (APA):
Rough, D. J., & Quigley, A. J. (2020). End-User Development of Experience Sampling Smartphone Apps - Recommendations and Requirements. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *4*(2), [56]. https://doi.org/10.1145/3397307

End-User Development of Experience Sampling Smartphone Apps - Recommendations and Requirements

DANIEL ROUGH, University of St Andrews, Scotland
AARON QUIGLEY, University of St Andrews, Scotland

Professional programmers are significantly outnumbered by end-users of software, making it problematic to predict the diverse, dynamic needs of these users in advance. An end-user development (EUD) approach, supporting the creation and modification of software independent of professional developers, is one potential solution. EUD activities are applicable to the work practices of psychology researchers and clinicians, who increasingly rely on software for assessment of participants and patients, but must also depend on developers to realise their requirements. In practice, however, the adoption of EUD technology by these two end-user groups is contingent on various contextual factors that are not well understood. In this paper, we therefore establish recommendations for the design of EUD tools allowing non-programmers to develop apps to collect data from participants in their everyday lives, known as "experience sampling" apps. We first present interviews conducted with psychology researchers and practising clinicians on their current working practices and motivation to adopt EUD tools. We then describe our observation of a chronic disease management clinic. Finally, we describe three case studies of psychology researchers using our EUD tool *Jeeves* to undertake experience sampling studies, and synthesise recommendations and requirements for tools allowing the EUD of experience sampling apps.

CCS Concepts: • **Human-centered computing → Smartphones**; **Ubiquitous and mobile computing design and evaluation methods**; *User centered design*; Field studies; • **Software and its engineering → Visual languages**; • **Applied computing** → *Consumer health*; *Psychology*.

Additional Key Words and Phrases: end-user development, experience sampling method, ecological momentary assessment, technology acceptance

ACM Reference Format:
Daniel Rough and Aaron Quigley. 2020. End-User Development of Experience Sampling Smartphone Apps - Recommendations and Requirements. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 0, 0, Article 0 (2020), 19 pages.

1 INTRODUCTION

The Experience Sampling Method (or ESM) is widely employed in ubiquitous and mobile computing to gather user needs, evaluate technologies and understand user intentions. The method originates in social psychology literature, described by Csikszentmihalyi and Larson as *"an attempt to provide a valid instrument to describe variations in self-reports of mental processes"* [6, p. 526]. A similar definition, in the context of behavioural medicine, is *Ecological Momentary Assessment* (EMA), given by Stone and Shiffman in 1994 [24], with the terms used interchangeably to define the repeated, remote assessment of individuals in their everyday lives.

ESM attempts to minimise recall bias and enhance *ecological validity* - the extent to which the findings of a research study can generalise to real-world settings. Participants in ESM studies are asked to provide details on

Authors' addresses: Daniel Rough, University of St Andrews, Fife, Scotland, danieljrough@gmail.com; Aaron Quigley, University of St Andrews, Fife, Scotland, aquigley@st-andrews.ac.uk.

2474-9567/2020/0-ART0 $15.00
https://doi.org/

their current physical or mental states, at various times, as they go about their day. This contrasts with typical laboratory assessment, where participants are asked to recall such details that may have occurred days before, or to perform short, rigorous experiments that fail to capture the natural contexts of emotions or symptoms.

Preceding the advent of smartphones, ESM studies were conducted with paper diaries that participants would fill in throughout the day, often accompanied by a basic signalling device that would randomly prompt them to do so. Since then, smartphones have become ubiquitous across all ages and demographics, overcoming many limitations of paper, by deploying ESM apps directly onto participants' own devices. Using this mobile delivery method, the smartphone acts as both the signalling device and the medium for delivery and completion of surveys.

While there are a number of benefits to doing so - participants are not physically encumbered by carrying around paper diaries or additional devices; researchers can receive real-time feedback as soon as participants complete surveys; smartphone sensors can collect additional objective context (e.g. location) - their implementation remains a significant challenge. Domain experts such as psychologists and clinicians are unlikely to have the requisite programming knowledge to develop ESM apps, and likewise, a professional programmer is unlikely to understand the contextual factors pertaining to a given domain. Further, any necessary changes to an app require additional time and money, reducing its flexibility.

The work in this paper therefore focuses on *end-user development* (EUD), a field of human-computer interaction research focused on the empowerment of end-users to create and modify their software to satisfy their own diverse, dynamic requirements [12]. The underlying motivation is that, by allowing researchers and clinicians to develop and deploy their own ESM apps independent of a professional programmer, this will significantly lower the existing barriers to addressing research questions and healthcare issues that could benefit from such apps. In response, we have developed the *Jeeves* tool to allow end-users with no programming skills to create their own ESM apps, couple them to participants' smartphones, and receive assessment results in real-time. While our previous work has addressed design decisions and lab-based usability studies of Jeeves [20, 22], this paper extends the evaluation conducted in [21] to synthesise necessary functionality and constraints of EUD for ESM apps (hereby referred to as EUD-ESM) addressing the following question:

What factors influence the adoption of technology for researchers and clinicians to develop experience sampling smartphone apps?

Although usability is critical for the adoption of EUD technology, it is highly dependent on the context of use. Moreover, it is a single factor that does not determine whether the technology would actually be useful, nor whether it could be feasibly deployed in practice. In this paper we make the three following contributions:

- Qualitative analysis of interviews with psychology researchers and clinicians, from which salient factors influencing acceptance of EUD are derived
- Further qualitative analysis of three case studies involving psychology researchers and observation at a local clinic, results of which are triangulated with interview feedback
- A summary of design considerations for the development of future EUD-ESM tools, consolidated from the results of our field research

2 RELATED WORK

Previous literature has surveyed the use of smartphone technology for ESM in behavioural research, deriving benefits and barriers. van Berkel et al. present a thorough survey of mobile-based ESM, including the digital evolution of ESM from paper to modern-day smartphone implementations, usage of the method in computer science literature, and considerations for conducting ESM studies on mobile devices, such as privacy and presentation [28]. Similarly, Pejovic et al. review technological developments in smartphone-based ESM, the possibilities of sensor-based augmentation, and practical considerations for implementation of such studies [18].

Both reviews also discuss existing tools for smartphone ESM development, which our work builds upon by deriving considerations and requirements for the developers of such tools in HCI research and practice.

Other work in the HCI domain has sought to improve the compliance of ESM studies through various means, including visualisations [9], personalisation [15], identification of appropriate prompting moments using sensors [19], and incorporation of social media into the sharing of questions and answers [27]. Our work differs in that we target the motivation to employ EUD tools by researchers and clinicians, as opposed to study participants themselves. In doing so, we separate our work from research that aims to understand and improve the motivation of these participants. For example, Musthag et al. determine the effects of "micro-payments" on data quality in field studies where motivation is difficult to maintain [17]. Similarly, van Berkel et al. demonstrated that gamification of ESM studies improved participants' quantity and quality of survey responses [29].

2.1 End-User Development (EUD)

Although app-based ESM studies are increasingly reported in literature, app implementations are bespoke to their respective studies. For universal acceptance of smartphone ESM, the diversity of research questions and end-users necessitates a researcher or clinician being able to adapt these features found in bespoke apps to any ESM study they choose to run. In doing so, they perform EUD activities.

For such "professional end-user developers", often a key challenge is not one of education, but of infrastructure, and how such EUD environments can be incorporated into current working practices [23]. This factor is of key concern in this paper. Developing EUD tools that support users in their current practice, and evaluating the success of these tools in doing so, are major challenges, which require investigation beyond participants interacting with a tool in a lab study. In real-world deployments, individuals' interactions with technology are highly dependent on other individuals who form part of their working practices, and existing technologies already used, in what can be considered as socio-technical systems [8].

Yet, most EUD research has not primarily addressed this challenge of understanding user needs and evaluating in-situ. Tetteroo and Markopoulos recently conducted a review of research methods in EUD, relating the methods used to the research purpose, such as describing an ideal tool, or evaluating an existing tool [25]. Of particular interest, they discovered that the majority of research into understanding and evaluating EUD tools has been in lab settings, rather than in the field or through qualitative interview and survey feedback, for example.

2.2 EUD for ESM

Bridging these two fields of research, other tools have been developed that aim to facilitate non-programmers' development of ESM apps. To acquire a comprehensive list, the first author systematically searched computer science literature databases including the ACM, IEEE and Scopus digital libraries using the following search term: *"('experience sampling' OR 'ecological momentary') AND ('tool' OR 'platform')"*, resulting in 791 results prior to exclusion criteria. The related work of each of the selected publications were also examined to ensure that similar referenced examples were not omitted. Given that some tools exist in the commercial domain, a standard Google search was used with the search terms listed above, in an attempt to find proprietary examples.

First, articles describing an ESM study or a bespoke application were excluded to limit results to development tools and platforms. Further criteria were applied to limit tools to those relevant to EUD, excluding frameworks intended for use by proficient programmers were excluded, as well as tools that rely on participants to self-initiate assessments. Further details of tools were acquired by reviewing relevant literature, and by testing them when available. For those that could not be tested, authors of publications describing the tool, or the software developers in the case of commercial tools, were contacted to request a trial or further details. If this was not possible and no further information could be obtained without payment, the tool was not included. All search terms and exclusion criteria were established and applied by the first author.

The 11 tools included in Table 1 are those with sufficient information to characterise them with regard to the two key features of smartphone ESM proposed by Intille in [10], namely: *"auto-tailor prompts and questions to context"* (divided into observable context detected through sensors, and implicit context provided by participants) and *"subject can adapt if desired"* (represented in the table as **Participant tailorable**). These two particular features were chosen from Intille's visionary paper because they represent features determined to be useful for sustaining compliance and engagement from related work - namely, context-sensitive prompts and personalisation features. Although the primary contribution of these tools is not necessarily their facilitation of ESM app development, all support this functionality to differing extents.

Table 1. Existing EUD-ESM tools and their implementation of features for future ESM tools proposed by Intille [10]

Tool Name	**URL/citation**	**Auto-tailor to context (sensed)**	**Auto-tailor to context (given)**	**Participant tailorable**	**End-user Evaluation**
iPromptU	www.cogtherapy.com	X	X	X	None
MobileQ	[16]	X	X	X	None
PACO	www.pacoapp.com	✓	■	■	None
ExpSampler	[26]	■	■	■	None
Sensus	[32]	✓	■	■	Usability
PartS	[13]	✓	X	X	Usability
TEMPEST	[1]	✓	✓	X	In-situ
SurveySignal	www.surveysignal.com	X	X	X	None
LifeData	www.lifedatacorp.com	X	✓	X	None
MovisensXS	xs.movisens.com	■	■	■	None
EthicaData	www.ethicadata.com	✓	■	■	None

X = Not implemented ■ = Possible extension/in development ✓ = Implemented

Each tool has a low barrier to entry, allowing simple studies to be set up with very little or no programming involved. However, as shown in Table 1, flexibility is sacrificed, with the inability for researchers to utilise a number of potentially useful features. In particular, the ability for participants to adapt to their own diverse requirements is not present in any of the reviewed systems (although it is potentially feasible as an extension through adapting existing source code in some systems). While usability is clearly a factor in the development of these tools, only two (*Sensus* and *PartS*) report findings of a conducted usability evaluation. In terms of real-world utility, only the *TEMPEST* publication reports on field deployments where the tool has been evaluated in its context of use, including factors pertaining to its adoption.

In summary, research across different domains has explicated potential benefits of smartphone-based ESM for engaging participants and collecting useful, accurate data. Moreover, a number of tools exist to simplify the creation process for non-programmers. However, there is a distinct lack of research into factors for the adoption of these tools by their stakeholders - namely, psychology researchers and clinicians.

2.3 The Jeeves EUD-ESM Tool

In response to the perceived limitations in functionality of existing tools, *Jeeves* was developed as a solution to simplify the use of programmatic concepts, including conditional statements and variables, necessary for development of context-sensitive ESM apps. (App screenshots can be seen in Section 5.) Researchers can create ESM app specifications through a blocks-based programming environment [2]. These consist of *Trigger* blocks that specify times or contexts on which to execute their nested *Action* blocks. These actions include sending surveys and prompts, or updating values of *Attribute* blocks. Attributes hold participant-specific values, through which automatic tailoring to participant-provided context, and allowing participants to tailor their own specifications, are enabled. Finally, as well as triggers based on time or location, researchers can utilise trigger blocks that are activated by in-app functions, such as when a participant completes a survey or presses an on-screen button.

Our previous work described Jeeves in relation to EUD adoption requirements for psychology researchers [21], as well as how usability issues in a lab study of Jeeves were almost entirely separate to real-world usability challenges identified in case studies [22]. In this work, we extend on both of these results by triangulating interview feedback with analysis of our case studies. Further, we incorporate interview feedback from clinicians - another identified stakeholder group for ESM apps - and issues identified from an observation at a local clinic into our synthesised understanding of adoption factors for EUD-ESM in general, not specific to Jeeves itself.

Most closely related to our work with Jeeves is that of Batalas et al., who discuss the engineering-related challenges of EUD-ESM. Their work describes the TEMPEST platform [1], a web-based tool with similar functionality to Jeeves in allowing non-programmers to develop ESM studies. While their work focused on the implementation of TEMPEST, we do not aim to describe our own implementation of the Jeeves framework, and instead seek to build upon their analysis of critical factors for using EUD-ESM through our qualitative research.

3 END-USER INTERVIEWS

To advance our understanding, semi-structured interviews were conducted with eight clinicians and five social psychology researchers. In interviewing these two separate groups, adoption factors specific to the two different domains were obtained, both in terms of their own motivations and constraints, but also those of their participants and organisations. Interviews took place at the interviewees' location of choice, and lasted approximately 45 minutes. Interviews were transcribed, and qualitatively coded for thematic analysis. Coding was structured based on end-users' characteristics, their attitudes towards ESM, and their requirements of an EUD platform.

Although psychology research is the primary domain in which ESM is utilised, a promising application of ESM is in supporting clinical practice, as part of a patient-centred care model for those managing chronic conditions [30, 31]. In this capacity, ESM technology acts to support regular, timely self-monitoring and uncover correlations between different mental/physical states. The interview protocol for both researchers and clinicians was identically structured in three parts:

(1) **Current practices** - Interviewees were asked about their current working practices, with an emphasis on issues they face in gathering ecologically valid information from participants/patients
(2) **Use of technology** - Interviewees discussed their familiarity and experience with different technologies necessary to their work practices
(3) **Preliminary feedback** - Interviewees were given a brief overview of Jeeves, and asked for preliminary feedback on its potential usefulness

3.1 Functionality requirements

Issues regarding software functionality were discussed in far more detail with researchers than with clinicians. While the latter group are constrained to organisation-approved software, researchers use a wider range of technology, and have differing motivations and experience with learning to use new applications.

3.1.1 Collaboration and support. In cooperation with peers, or as supervisors to students, researchers work in teams with varying experience, where community support could improve ease-of-use [8]. P2 explained how research is typically conducted as part of a team with different specialities. Often, the researchers most knowledgeable about study requirements lack the technical skill to implement these requirements themselves.

"You've usually got somebody who's very up on the evidence, very up on the research, but not necessarily technically...that competent...then you've got somebody else on the team that says right I know how to do [programming]...a minimum of two people" (P2)

In the healthcare context, chronic disease patient care is a cooperative effort between GPs and specialist practice nurses. Patients may also see different doctors, such that it is important to enable **collaboration and understanding** between clinicians, which existing technology does not support: *"if we make a change to patients' medication, the quickest way to communicate that to the GP is to fax them...so, if we email the GP, we have no idea when the GP might pick that up"* (C3)

Further, acceptance of EUD-ESM was contingent on adequate support for learning and applying its features. P1, in transitioning from *SPSS* to *R* statistical software, expressed how such support had enabled her to learn complex functionality: *"By the time I started using it there was that critical mass of people who were developing wikis and Stack Overflow...I'm not very good at using R but I am good at Googling how to do what I want"* (P1)

Thus, as a design consideration, an EUD-ESM tool should be designed for a spectrum of end-users, with **community support** allowing for novices to reuse existing study specifications.

3.1.2 Tailoring and repurposing. To varying extents, all forms of in-situ research intrude on participants' everyday lives. P3 explained how his research involved professionals coping with stressful situations, who were non-compliant due to the time commitment of a long, cross-sectional survey: *"They're very busy people and so getting them to set aside half an hour, to actually do it...is actually really difficult"* (P3)

This requirement was also addressed by clinicians. Patient-centred care relies on the incorporation of patients' own goals, motivations and constraints, which determine whether or not they are likely to engage in treatment. C6 described these factors as the patient's 'agenda': *"Healthcare people, not just doctors, have agendas, but the patients often come in with a totally different agenda"* (C6)

Hence, if EUD-ESM is to be accepted by app end-users, it is necessary to provide a means of **tailoring of protocols and reminders to individuals**, to ensure minimal disruption.

Clinicians were doubtful that certain patient cohorts would embrace self-monitoring. Instead, the ability to implement tailored reminders for appointments and medication into a patient's app was perceived as a significant advantage to an EUD approach: *"...a reminder. 'You have an appointment tomorrow!' Yes, that sort of thing would be fab, because the non-attendance rates are really poor and that costs the NHS an awful lot of money"* (C3)

In addition to the barriers presented by forgetful patients, clinicians identified barriers imposed by those who are expectant of instant feedback. As a resolution, clinicians valued the possibility of **repurposing ESM notifications** to provide reassurance or advice to patients on their readings, while minimising their own workload.

3.2 Organisational requirements

The organisational influences of adoption by both psychology research and clinical practice can be interpreted as non-functional requirements addressing domain-specific constraints:

Affordability was raised by all five psychology researchers, whose motivation to use a technology was contingent on its cost. All researchers referenced use of *Qualtrics* software, for which their university has a site licence, such that this is a prominent reason for its widespread use.

"Affordability is obviously a big thing so...one of the reasons we were speaking about Qualtrics because...we have a university licence for that which is a major, a major issue for using that" (P4)

Indeed, P3 explained that he would be more willing to use *"a good cheap option that requires a little bit of threading to make it work"* as opposed to a more expensive, purpose-built solution.

Technology compatibility was a related concern raised. Outside the functionality and usefulness of the EUD-ESM tool itself, researchers expressed the importance of maximising compatibility with other widely used research software, including *Qualtrics*:

"Inevitably there's gonna be things it can't do...having that capacity to have that inter-operability, plug-in capability, developing sort of thing would be great" (P3)

It is also important for researchers to have participant data in a format compatible with analysis software such as *SPSS* and *R*, simplifying statistical analysis of aggregated results from multiple participants.

Security of sensitive participant information must also be strictly enforced at various levels in both domains, from university ethics to journal policies:

"...reassuring not only our institutional methods committee but publishers now, so for example, PLOS One. I'd need to be able to tell them this is the kind of encryption, this is what's happening" (P1)

Likewise, in healthcare a prerequisite of integration of EUD-ESM requires a guarantee that sensitive patient information will not be accessible outside the organisation. Otherwise, data transfer between mobile devices and clinical computer systems will not be possible, eliminating the benefits of real-time monitoring.

3.3 Summary

From the clinicians' perspective, although the potential benefits of EUD-ESM were recognised, these were contrasted with potential drawbacks that would preclude its adoption with some patient groups. Further, its practical application could be infeasible due to strict barriers imposed by the healthcare service. In contrast, researchers were able to clearly envision the practical application of EUD-ESM. Primary concerns related to affordability and technical support; their areas of research were particularly well-suited to EUD-ESM, and perceived usefulness raised few conflicts. Between both interviewed groups, clear similarities appeared to emerge, with respect to functionality required and organisational constraints (highlighted in bold where they occur in previous subsections). However, to clarify our perceived understanding of the requirements of EUD-ESM in its context of deployment, further in-situ research was required beyond these interviews.

4 CLINICAL OBSERVATION

Clinician interviews revealed that practitioners running chronic disease management clinics are potential stakeholders of EUD-ESM. However, these interviews were removed from the clinical context of interest, so that clinicians did not have access to existing software or relevant materials that represented their practical processes. As another threat to validity, most interviewees were GPs who are seldom involved in the management of chronic conditions. In such cases, Blomberg et al. argue for observation, as opposed to anecdotal evidence, for informing system design [4, p. 130] Thus, to understand potential barriers to EUD-ESM in these settings, a short observation session was arranged at a local clinic. This consisted of one hour assigned to a clinician running a diabetes clinic and another clinician running a hypertension clinic. Within each hour, three patients were seen by each. Brief discussions with clinicians between appointments were used to summarise the preceding appointment, and verify assumptions made in the note-taking process.

All appointments in these clinics were annual reviews of patients. During these reviews, patients were queried about their general health, general management of their conditions, unusual events since their previous review, current medication, and any additional resources required. Informal discussion of lifestyle took place, medication and recent medical results were reviewed, and patients had opportunities to ask questions. Salient factors from interviews are highlighted where they align with observations made during these clinics.

4.1 Information exchange

Where necessary, information provided by clinicians to their patients was educational in content, both of whom explained that a crucial aspect of appointments is to impart knowledge to cope in emergency situations. Although all diabetes and hypertension patients were aware of their respective procedures, both patient groups are burdened by the need to remember different aspects of their treatment regime. This memory burden was observed in these annual review appointments, and the clinician commented that *"it's amazing the things people don't tell you about"* with respect to dangerous situations or symptoms that should be reported. In this capacity, it is possible that EUD-ESM could be **repurposed** to provide reminders to chronic disease patients, as discussed in interviews.

In both clinics, there was a reliance on paper-based assessment forms, self-management plans and information leaflets. In particular, one patient had misplaced her assessment form for blood pressure readings prior to her clinic visit. She instead brought her readings on scraps of paper, which meant a laborious process of transcribing these readings onto a separate piece of paper, then scanning and uploading it into the patient's records. **Technology compatibility** was thus an identified need, as clinicians suggested that the potential for EUD-ESM to integrate with existing patient management software would be useful.

This labour in information exchange was also present between clinicians. For example, changes to medication initiated by GPs were not immediately communicated to other clinicians through their system, such that patients themselves were relied upon to provide up-to-date information. This difficulty in communication was highlighted through interviews, interpreted as a need for **collaboration and support** functionality.

4.2 Self-management

The diabetes clinician explained that, contrary to the stereotype of older patients being averse to technology, they are in fact proactive in self-monitoring their conditions, and often enthusiastic in doing so. However, clinicians explained the need to avoid aggravating the "worried well" - patients whose conditions are stable and well-managed, but who become fixated on quantification of their health. For this reason, patients and clinicians cooperatively devised self-management plans to fit with their lifestyles and habits. If EUD-ESM is to be used for this purpose, it is not adequate to have generalised reminders - a protocol must be **tailorable to the individual.**

If patients have to do their own transfer of readings from glucose or blood pressure monitors into an EUD-ESM tool, there still exists a labour on their part. As such, the **need for minimal intrusion** was emphasised with regards to the existing heavy workload imposed on chronic disease patients.

4.3 Time and burden

Both clinicians explained that time pressure was the greatest burden on their working practice. Patients often have to be scheduled at the last minute, such that when patients fail to attend their appointments, it prevents clinicians from keeping their clinics running to this tight schedule. Although reminders were sometimes made via phone calls, this placed additional responsibility on clinicians who are required to perform this function on top of their standard schedule. If EUD-ESM could be **repurposed** to provide appointment reminders, this alone could alleviate considerable workload.

From the perspective of patients, realistic adoption is also contingent on a non-burdensome process. They often have schedules that are defined by their medication; they may have medication to take at different times of day, as well as less regular processes, such as monitoring their blood glucose or blood pressure. Again, the diversity of patients' schedules and capabilities relates to the requirement of **tailoring to individuals**.

4.4 Summary

While short in duration, this observation session provided further evidence of requirements for incorporation of EUD-ESM into clinical practice that had arisen in interviews. Transfer of information from patients to clinicians, and between clinicians themselves, could be eased if patient readings were captured through an ESM app and synchronised with clinicians' software. Further, an EUD approach would be useful for authoring individually tailored self-management plans that could prompt for readings at appropriate times or even provide medication and appointment reminders. Of course, without a deployment of an EUD-ESM tool into this context, these use cases are hypothetical, and subject to the organisational constraints previously identified, but are indicative of the potential benefits and barriers that would be experienced.

5 CASE STUDIES

As previously discussed, interviews alone were insufficient to determine challenges pertaining to real-world use of EUD-ESM. This section thus describes three applications of Jeeves by psychology researchers, two of which were previously discussed in prior work [22] and one currently in progress. In doing so, we highlight issues that have emerged in practical use that provide concrete suggestions for implementing requirements identified in interviews. For each study, we separate observations into *design challenges*, *testing strategies* and *feature usage*.

5.1 Case Study 1 - ESM During Sport Events

The first case study was conducted with a previously interviewed psychology researcher and a collaborating researcher at a university in Germany. The researchers (hereby referred to as Paul and Oliver) were interested in capturing the dynamics of emotion in members of crowds at sporting events. Their motivation for use of Jeeves was the inherent difficulty in distributing and collecting paper surveys at such events, as well as the static data retrieved by a survey at a single point in time.

5.1.1 Design challenges. In this first real-world application of Jeeves, it was promising that the researchers finished setting up the triggers and actions needed for their study with no mistakes in the blocks-based specification that needed correcting. This was likely due to the simplicity of its structure, which is shown in Figure 1. However, an issue of uncertainty arose wherein a question required two numbers to be input, and had to be split into two separate (but closely related) questions.

Paul: Actually, shall we make this...
Oliver: On one page?
Paul: I'm not sure how that's gonna work with pages, if it's a new page per question, or...
Oliver: That would be good if it's on one page, right?

At this stage, the lack of a previewing feature was noted by both researchers. Paul and Oliver made comparisons with *Qualtrics* - software they were both familiar with in their research that provided this functionality. Without clarification from the observing author, this uncertainty could have caused early abandonment.

Further, during the call, collaborative design between Paul and Oliver was observed to be a challenge in itself, made cumbersome through the screen-sharing function of Skype. Oliver would dictate survey questions from a plan document while Paul created the survey in Jeeves, resulting in a slow and awkward process.

5.1.2 Testing strategies. As described, immediate testing of app functionality was not possible in Jeeves. Prior to the pilot study, Oliver tested the specification on his personal device times to confirm correct operation.

However, while most pilot study participants did not face any technical issues, all voiced concerns about the various app permissions required. Previously, to install the *JeevesAndroid* study app, all permissions, including location and microphone, were requested. It was necessary to explain to participants that this data was not being collected for the study, after which they agreed to continue the installation.

One pilot study participant, running a lower version of Android than was necessary for the study, was unable to install the app and thus had to be excluded. Another participant had certain privacy settings enabled on his device, so that he did not receive surveys at the same time as other participants.

Following the pilot, each survey had to be rewritten in German for the full-scale study. To test these surveys without waiting for them to appear, Oliver added a "button trigger" that would allow the survey to be started on a button press. While this was a useful testing strategy, Oliver forgot to remove the trigger and button he

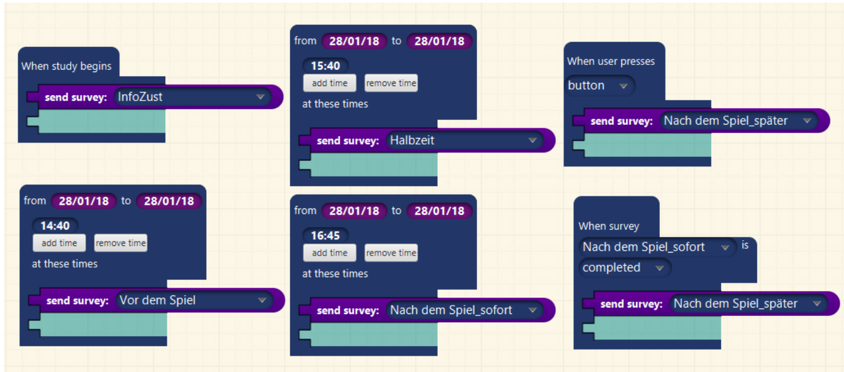

Fig. 1. Paul and Oliver's final specification for the study conducted at the basketball game

created from the version of the app that participants downloaded. Some participants pressed this button through curiosity, initiating and completing the post-match survey too early. (This trigger can be seen in the top-right of Figure 1, showing that when the button is pressed, the post-match survey *"Nach dem Spiel_später"* is sent.)

5.1.3 Feature usage. Although the researchers did not require complex block combinations, issues instead arose when informed consent had to be implemented into the first survey. While the researchers tried to find a workaround by providing a URL link to the informed consent documents, a lack of Internet access was anticipated. Given the sensitive nature of storing and transferring participant data, a means to present informed consent and debriefing documents had to be implemented as a feature to avoid abandonment of the study.

Further, Paul requested a new type of survey question that would present text to participants without requiring an answer. Certain questions had to be given a brief explanation, and it was decided to do so within the survey itself. While these were trivial changes to implement, Paul explained how critical it was that such evolution was possible:

"we were able to shape the design of this at very short notice...I think otherwise we might have logged in the first time and thought 'well it doesn't actually do this', then try and move onto something else".

Although usability-based abandonment could be reduced by incorporating tutorials and FAQs, if primary functionality is not immediately available and ambiguities cannot be clarified, then the researchers would have defaulted to more familiar software or research methods.

5.2 Case Study 2 - ESM in the Menstrual Cycle

This case study describes a collaboration with a second psychology researcher involved in our interviews, and her postgraduate research student. The focus of this case study was an ESM approach required for the student's final year research project. The researcher, who will be referred to as Deborah, identified Jeeves as a useful tool for understanding the contextual factors underlying aggressive behaviour in relationships. Deborah's student, hereby referred to as Lucy, was conducting research to investigate antecedents, consequences, and general variation of female aggression during the menstrual cycle.

5.2.1 Design challenges. The use of attributes was required in order to tailor the app to each participant's ovulation dates. Attribute creation appeared to be straightforward. Lucy, who had previously reviewed the documentation of Jeeves, was able to clearly grasp the concepts of updating attribute values through survey questions - a key component of the individual tailoring feature of Jeeves. However, attributes and conditional

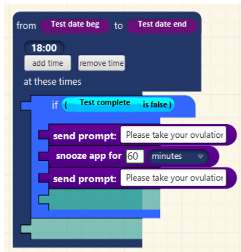

Fig. 2. Trigger to prompt for survey completion

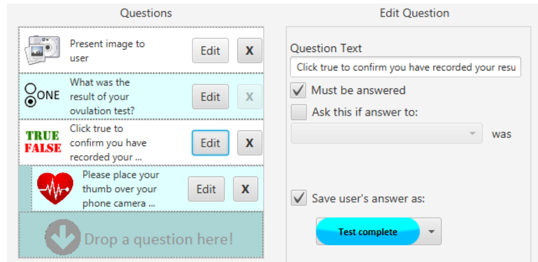

Fig. 3. Researchers' survey showing their 'completion confirmation' question

statements also raised new design challenges. The researchers desired to prompt participants at particular times if they had not completed their survey. Deborah described a proposed workaround strategy to Lucy:

> **Deborah**: I think then there must be some sort of 'this question has been answered' condition or 'this survey is complete' condition and if that's false, then... OK, this is probably not the best way to do it, but IF when we send the user input, when they do their 'please input your ovulation results' survey, we can add a question at the end that is just 'click here to confirm you've taken the test today' then you've got a yes/no answer to save it as an attribute again, which can then be placed into that condition.

In summary, the researchers decided to add a question to their survey (shown in Figure 3) that would update the "Test complete" attribute (shown in Figure 2) to stop reminder prompts being sent. Deborah, who had programming experience, recognised that the Boolean attribute would likely default to be false. Although Lucy understood the concept of attributes, she assumed that not completing the survey would mean that the attribute would not have a value; it was not considered that default attribute values would be implicit programmer knowledge.

Indeed, further programmer-specific knowledge related to the conditional expression, specifically how often the condition would be checked. Deborah and Lucy discussed the trigger shown in Figure 2, deciding whether the second prompt would be executed if the "Test complete" attribute had been updated to true:

> **Deb**: As I understand it, it will send prompt, snooze for 60 minutes, send the prompt again, but will it check that condition again? So if the participant has then put true...
> **Lucy**: I think so because it's underneath this one and you can put another one there...and also will it even have sent the initial prompt and then snoozed if it is true?
> **Deb**: If it's true it won't do anything. Yeah. If it's false it'll prompt, wait, prompt again, but will it check. It doesn't look to me at the moment that it'll recheck the condition before it sends the second prompt

5.2.2 Testing strategy. As with the previous case study, pre-pilot testing and debugging was performed by Lucy on her personal device, simulating a run of the study for a number of days. In this case, the pilot study ran for 21 days, during which participants reported their ovulation result daily for 10 days of the study. While the researchers were satisfied with the app's functionality, participants' experience was unexpectedly negative. Lucy explained that: *"Participants were unsure about the permanent notification that said 'Jeeves running' on their phone...Two participants dropped out of the study saying that the app was 'annoying' them due to this"*

Just as real-world issues of the use of Jeeves could only be ascertained through case studies, some participant issues can only be anticipated through sustained real-world use, regardless of any built-in testing functionality.

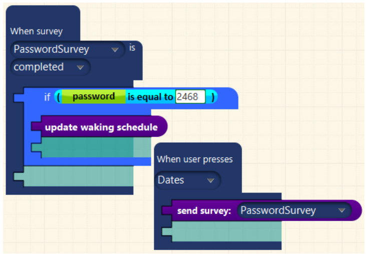

Fig. 4. Triggers for password feature

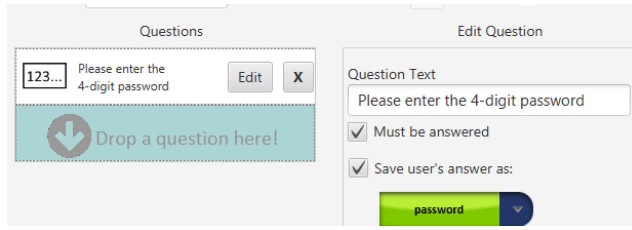

Fig. 5. Survey to assign users' entry to a 'password' attribute

5.2.3 Feature usage. Deborah and Lucy employed the full range of Jeeves functionality that was implemented to provide advantages over alternative EUD-ESM tools. For example, they made use of attributes that allow customisation at an individual participant level, and also used conditional statements in addition to event-based triggers, harnessing more complex trigger functionality. In doing so, this validated interview feedback related to individual tailoring and repurposing ESM functionality to send reminders for survey completion.

Prior to commencing the pilot study, the researchers requested an additional feature for capturing participants' heart rate. While they explained that this was not necessary, it was seen as highly desirable and resolved the burden on participants of manually taking their pulse during assessment sessions.

5.3 Case Study 3: Tracking nurses' dietary habits

The final case study is an ongoing collaboration with a researcher and a student at another nearby university. In their current project, the researchers are investigating the lifestyle habits of nursing staff working long shifts, using ESM to capture in-situ feelings of stress and fatigue in conjunction with recent food and drink intake. The use of Jeeves in this study has necessitated new features and uncovered further considerations.

5.3.1 Design challenges. The nature of this study has required repurposing the functionality of Jeeves towards reducing the risk of data loss. The researchers confirmed that it must be necessary to allow nurses to update their schedules at short notice, but that such a feature might be accessed accidentally. It was suggested that password protection of scheduling functionality would be useful, the implementation of which is given by the two triggers in Figure 4. When a participant presses the 'Dates' button, a survey called *PasswordSurvey* is sent, shown in Figure 5, containing a single numeric question that saves the answer in an attribute called *password*. Completion of this survey activates the conditional statement that checks whether the *password* attribute is equal to 2468 (the researchers' chosen password). If so, then the *"update waking schedule"* action is executed. This workaround had to be described and implemented for the researchers, however, suggesting the potential for worked examples that could be built upon in such cases. Further design challenges were related to the schedule variability of nurses' shift patterns, discussed below.

5.3.2 Testing strategy. The testing process of this study has been more thorough than the previous two due to additional time and budget. For example, fifteen smartphones have been purchased for use in the study to ensure compatibility. The researchers have deployed their specification on these devices and performed multiple test runs over a number of days. This rigour has been necessary to isolate issues in the new scheduling functionality, which requires the app to function differently for each individual on each day of the study, intended to last for approximately three weeks.

However, understanding whether the scheduling is functioning as intended has been aggravated by difficulty in inspecting trigger schedules. As such, our involvement has again been necessary for testing and debugging researchers' uncertainties regarding correct functionality.

5.3.3 Feature usage. As with the previous two case studies, new functionality has been implemented in response to bespoke requirements. The researchers explained how nurses' shift patterns are constantly changing and subject to alteration at short notice, including regular switches from day-shift to night-shift. As such, while the individual tailoring function of Jeeves was seen as useful, it was inadequate to handle the diverse waking/sleeping times of participants, who would have to manually adjust these times for every day of the study.

To rectify this, a special type of "schedule survey" has been enabled, allowing participants to enter start and end dates of their study, then specify their waking and sleeping times throughout the scheduling period. Further, while the previously mentioned 'password' implementation aims to protect inadvertent scheduling mistakes, the researchers requested a means to view and edit participants' schedules remotely through Jeeves. This represents a desire for collaboration support not just between researchers, but also involving their participants.

5.4 Summary

The case studies of Jeeves exposed the diverse means by which ESM research studies are conducted, and how EUD could be applied in this domain. Previously, very little research has been undertaken into introducing EUD of ESM into practice like this, with exceptions of the *TEMPEST* platform by Batalas et al. [1], and *PartS* by Ludwig et al [13]. Here, we summarise observations into themes of complexity, time-saving, and participant burden.

5.4.1 Complexity. Paul and Oliver created a simple specification that did not utilise much of the feature set considered important for Jeeves. In contrast, Deborah and Lucy capitalised on these possibilities, creating multiple attributes and using them to configure conditional functionality in their app specification. Finally, our third case study has exemplified both - beginning as a simple specification that periodically sent a survey during normal working hours, and evolving to require new features and novel use of existing features.

Further, a resulting implication of this is that more complex functionality requires more thorough testing. Without adequate support for testing and debugging, these researchers were still reliant on the authors to assist with this process. This arose as a salient factor in all three studies, implying a need for EUD-ESM tools to include accessible, explainable testing techniques from the sub-field of end-user software engineering [11].

5.4.2 Time saving. With regards to saving time, Paul appreciated that the data for all participants was integrated into one spreadsheet. However, because of the row-column format in which Jeeves stored this data, he had to spend time in first creating the necessary SPSS data file, explaining that *"It took me an hour or so to build this manually, so if as much of this work could be done by the app as possible would be great"*. Such compatibility features, available in research software like *Qualtrics*, were supported in both interviews and clinic observations.

Time-saving functionality for participants themselves has been a necessary requirement for the third study, while simultaneously saving the researchers additional effort in study design. Further, in Deborah and Lucy's study, the heart-rate capturing feature was seen as a time-saver, albeit sacrificing the accuracy of commercial heart rate apps. For ESM apps that require accurate physiological readings, EUD-ESM tools should allow integration with existing monitoring technology.

5.4.3 Participant burden. Finally, the first two pilot studies was insightful in determining the quality and reliability of the developed Jeeves app. Although all participants found its use simple and straightforward, problems were caused in Paul and Oliver's study by the large number of sensitive permissions asked for at the beginning of installation. Participants questioned this, and without a researcher's presence to resolve their concerns, it is likely that they would have not complied with installation.

In Deborah and Lucy's study, participants were compelled to uninstall the app due to the constant presence of the Jeeves icon. Interestingly, participants did not report that the notifications themselves were annoying, and the researchers expressed that *"notifications are effective and non-intrusive"*.

Primarily, our design and development efforts were focused on the Jeeves visual programming tool itself, rather than the apps developed with it. While app functionality is fully tailorable by the researcher, its presentation both in terms of credibility and professional appearance were unforeseen issues leading to potential abandonment.

6 DISCUSSION

The goal of our work was to understand how psychology researchers and clinicians could be supported to create their own ESM apps for research participants or patients. Addressing our primary research question, ***What factors influence the adoption of technology for researchers and clinicians to develop experience sampling smartphone apps?***, below we summarise recommendations for future EUD-ESM tools, derived from participants' suggestions and existing EUD tools in other domains. The identified factors of influence, with our corresponding design recommendations, are summarised in Table 2, and discussed below under the case study themes of time-saving, complexity and participant burden. These factors primarily stemmed from interview analysis, with our observation sessions and case studies providing further evidence of their influence. The recommendations for each factor were then derived from design features of existing widely-used software, design features of Jeeves that were applied in practice, or direct suggestions from interview and case study participants. Finally, revisiting existing tools identified in Table 1, we quantify our findings through determining the extent to which these tools implement our design recommendations, and thus propose avenues for future work.

Table 2. Functional and non-functional EUD-ESM factors, with design recommendations

EUD-ESM Factor	Design Recommendation
Collaboration support (F1)	Shared editing space with cooperation facilities
Peer-oriented assistance (F2)	Global repository for example projects
Individually tailorable (F3)	Protocols that can be adapted to participant preferences
Repurposed functionality (F4)	Consent/debriefing/feedback/reminders in addition to surveys
Comprehensive testing (F5)	Ability to test both appearance and contextual behaviour
Safety-criticality (N1)	Impossible to create unstable specifications
Software compatibility (N2)	APIs/output file formats supported by existing software
Hardware compatibility (N3)	ESM app available on both Android and iOS devices

6.1 Time-saving

Time appears to be the most critical barrier faced by researchers and clinicians, thus the time EUD-ESM would ultimately save is an overarching factor for adoption. First, the time EUD-ESM would save is contingent on the specific goals of researchers and clinicians. For example, clinicians desired reminders to be used for appointments, and researchers desired a means to obtain informed consent from participants. When researchers used Jeeves in practice, such time-saving features emerged through direct use, and were not previously considered. We therefore suggest that *EUD-ESM tools should provide flexibility for survey and scheduling functionality to be repurposed* **(F4)**.

More specifically, integration with existing technology (such as statistical software packages or patient management software) would streamline domain experts' work practices. We suggest *EUD-ESM tools should utilise APIs, and compatible output formats, of existing research or clinical software* **(N2)**.

Further, shared repositories would enable clinicians and research groups to build upon, or replicate, each other's work. We observed this in our third case study where the password-protection workaround had to be implemented for the researchers, but could be a useful feature for future studies by other researchers. As such, *EUD-ESM tools should provide a global repository for stakeholders to build on each other's studies* **(F2)**.

6.2 Complexity

Foremost, the quality of an EUD tool, in terms of the functionality it can enable, is a determinant adoption factor. Researchers developing study specifications do so as teams of different expertise, and may also be physically separated. Indeed, it was occasionally an issue that the researchers in our case studies would modify the study simultaneously, causing the later save to overwrite earlier changes. Clinicians further explained how patient management is conducted by multiple stakeholders with different knowledge. As such, *EUD-ESM tools should be engineered as collaborative spaces that allow team members to cooperate and share ideas remotely* **(F1)**.

While needs between end-users vary, the ability to tailor at both the functionality level by domain experts and the personalisation level by individual participants/patients was considered highly important, such that we suggest that *EUD-ESM tools should allow study specifications to be tailored to individual participants* **(F3)**.

6.3 Participant burden

Quality in terms of reliability is another primary requirement. Case study researchers expressed a desire to test their developed apps for correctness. Without this feature, a level of frustration was caused by uncertainty as to whether their specifications satisfied the intended requirements. Direct collaboration with the case study researchers meant that personal assistance was given with app testing, but end-users need to be assured of app resilience prior to deployment if they are to consider adopting EUD-ESM independent of this firsthand support. Thus, as an emergent requirement from these case studies, *EUD-ESM tools should incorporate facilities for testing the context-specific behaviour of apps* **(F5)**.

Reliability is also a critical factor for organisations. Particularly in the healthcare context, all apps must undergo rigorous evaluation to ensure that they will do no harm to patients. This is a difficult implication to address for an EUD tool, given that adoption is contingent on not just the reliability of one particular app, but on the reliability of all apps that could be developed by clinicians. Nevertheless, *EUD-ESM tools should **not** allow researchers to develop apps that will function incorrectly on deployment* **(N1)**.

Finally, the need for participants to have an Android device for use of Jeeves was considered a significant drawback. Without providing potentially expensive study devices, a significant percentage of the population who own iPhones are unable to participate. While this may be a superficial suggestion, it is nonetheless critical for researchers that *EUD-ESM tools should be compatible with all dominant smartphone operating systems* **(N3)**.

6.4 Revisiting existing tools

Having derived various design recommendations, here we aim to tabulate the extent to which existing tools apply these recommendations in Table 3, and elaborate on the implications of blank columns.

6.4.1 Lack of shared workspace and projects. As shown in Table 3, no tool provides a collaborative editing environment, with the majority of tools built on the assumption of a single researcher responsible for editing and maintenance. Also, with the exception of the five tools that include some representation of a community forum, there are no shared spaces for users' projects to be openly accessed or built upon. *EthicaData* provides simple examples, but with a call for replicability in psychology studies, future work should consider a shared repository of ESM specifications that researchers can access for applying identical protocols to their participants.

Table 3. Existing EUD-ESM tools and their implementation of derived design recommendations

Tool Name	**F1**	**F2**	**F3**	**F4**	**F5**	**N1**	**N2**	**N3**	**Interface**	**Availability**
iPromptU	-	-	-	-	-	-	-	✓	Native form	Free
MobileQ	-	■	-	-	-	-	✓	-	Web form	Free
PACO	-	■	-	✓	-	-	✓	✓	Web form	Open
ExpSampler	-	■	✓	✓	-	-	✓	✓	Code	Open
Sensus	-	■	-	-	-	-	-	✓	Native form	Open
PartS	-	-	-	✓	-	-	-	-	Visual	Unavailable
TEMPEST	-	-	✓	✓	-	-	✓	✓	Web form	Unavailable
Jeeves	-	-	✓	✓	-	-	✓	-	Visual	Open
SurveySignal	-	-	-	-	-	-	✓	✓	Web form	Paid
LifeData	-	-	✓	✓	-	-	✓	✓	Web form	Paid
MovisensXS	-	-	✓	✓	-	-	✓	-	Visual	Free Tier
EthicaData	-	■	✓	✓	-	-	✓	✓	Web form	Free Tier

6.4.2 Difficulty in comprehensive testing. Both the functional recommendation of a testing environment for simulating different times and contexts, and the non-functional recommendation for ensuring that faulty studies cannot be implemented, are not clearly satisfied by any tool. While most tools allow surveys to be previewed, there is no way to ensure these will be deployed at appropriate times or contexts without rigorous field testing. Future tools would benefit from an integrated testing framework, or as a step forward, recommendations built into tools that advise against faulty or burdensome specifications.

6.4.3 Availability issues. Ultimately, researchers will select a tool to use based on convenience and budget, for which compromises must currently be made. Scaling up studies beyond 10 participants incurs costs for professional support. Of the non-commercial tools, six are publicly available although have their own limitations as outlined in Table 3, as well as issues in code stability and support. Jeeves itself has a number of limitations and we advocate for collaboration of tool developers - for example, applying the visual language of Jeeves to simplify creation of cross-platform apps with *ExperienceSampler* [26] or *TEMPEST* [1].

6.5 Future work

Considering the gaps in Table 3 that highlight design features which are absent in state-of-the-art tools, including entire columns that illustrate how no tool implements a particular feature, there is clear scope for future work on EUD for ESM applications. Here, we briefly discuss our envisioned future work to extend Jeeves as an EUD tool for this purpose, as well as other potential context-sensitive data collection applications.

Foremost, we see features for end-user testing as a complex but ultimately necessary extension in a future version of Jeeves, implementing both **F5** and **N1** recommendations. As an example, recent work by Manca et al. proposes a means of supporting end-user developers to test and debug their trigger-action rule-based applications [14]. Given the issue of debugging behaviours that are not 'immediate' such as those relying on time-based triggers, their solution includes a context-of-use simulation environment. While there may be a number of external reasons why an ESM app fails to deliver a notification (such as the Android operating system killing a necessary service) we anticipate that a similar contextual simulation environment would reduce unexpected app behaviour issues.

Throughout the deployment of Jeeves in the three small-scale case studies described, we continued to add new requested features, provided debugging support for unexpected behaviours, and suggested workarounds for encountered issues. An important lesson learned from this is that we as meta-designers must eventually devolve these responsibilities to power users if Jeeves is to scale up to support a wider variety of studies by an expanding number of end-users. Implementing recommendations **F1** and **F2** is part of this ongoing work in allowing end-users to develop roles as power users, including *domain developers* and *maieuta-designers* as defined by Cabitza et al. [5]. Such users will ensure the long-term sustainability of Jeeves through developing domain-specific functionality and supporting the social conditions required to motivate others to contribute.

Finally, related to the need to allow end-users to extend the functionality of Jeeves independently, there remains untapped potential to integrate with other open-source data capture tools. Jeeves and other openly available platforms have unique features that are currently siloed within their respective code bases. As future work, we envision expanding the capabilities of Jeeves with additional data collection functionality for enhanced context-aware experience sampling. Integration with the AWARE open-source context-awareness framework [7] is one means through which we intend to make Jeeves a useful data collection tool for a wider ubiquitous computing audience and thereby increase participation of domain experts in the evolution of Jeeves.

6.6 Limitations

Although this work uncovered a number of recommendations for EUD-ESM, some limitations exist in the research process undertaken. Primarily, by not involving researchers and clinicians at earlier stages, the development process was not user-centred from an early stage. Nevertheless, the functional prototype shown in interviews acted as a means of both *creativity* and *communication*, as defined by Beaudouin-Lafon and Mackay [3]. Jeeves acted as a common ground for communication, where researchers and clinicians could conceptualise potential applications in terms of available blocks.

Additionally, half our interviewed clinicians were GPs - an unintended consequence of convenience sampling in our university's School of Medicine. It later emerged that GPs are minimally involved in the chronic disease management of patients, with which EUD-ESM could facilitate. Ideally, further interviews should have been conducted with clinicians from our observation sessions. Finally, our review of existing tools was limited by having only one author conduct the review process. While this may have resulted in us missing related work, its purpose was to highlight (to the best of our knowledge) a dearth in evaluation research.

6.7 Conclusion

Researchers' goals and clinical practice methods are continuously evolving, as are the technologies on which these goals and methods rely. Our contribution in this paper is the derivation of recommendations for EUD-ESM in two different settings through synthesising qualitative feedback from interviews, case studies and observation. Our work, while applied to the specific domain of ESM apps, is generalised in its research process that extracts requirements for EUD that cannot be obtained in traditional lab-based usability studies.

ACKNOWLEDGMENTS

The authors would like to thank the clinicians and psychology researchers at our university who voluntarily participated in interviews, observations and case studies. We also thank the researchers currently using Jeeves for their ongoing feedback and suggestions for improvements.

Finally, the authors would also like to thank the anonymous reviewers for their positive feedback and recommendations, as well as the reviewers of our previously submitted version of this article for their detailed critique, which helped shape and focus this revised version.

REFERENCES

[1] Nikolaos Batalas, Marije aan het Rot, Vassilis Javed Khan, and Panos Markopoulos. 2018. Using TEMPEST: End-user programming of web-based ecological momentary assessment protocols. *Proceedings of the ACM on Human-Computer Interaction* 2, EICS, Article 3 (June 2018), 24 pages. https://doi.org/10.1145/3179428

[2] David Bau, Jeff Gray, Caitlin Kelleher, Josh Sheldon, and Franklyn Turbak. 2017. Learnable programming: Blocks and beyond. *Commun. ACM* 60, 6 (May 2017), 72–80. https://doi.org/10.1145/3015455

[3] Michel Beaudouin-Lafon and Wendy Mackay. 2003. Prototyping tools and techniques. In *The Human-computer Interaction Handbook*. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 1006–1031. https://doi.org/10.1201/9781420088892

[4] Jeanette Blomberg, Jean Giacomi, Andrea Mosher, and Pat Swenton-Wall. 1993. Ethnographic field methods and their relation to design. In *Participatory design: Principles and practices*. Lawrence Erlbaum, Hillsdale, NJ, USA, 123–155. https://doi.org/10.1201/9780203744338-7

[5] Federico Cabitza, Daniela Fogli, and Antonio Piccinno. 2014. "Each to his own": Distinguishing activities, roles and artifacts in EUD practices. In *Smart Organizations and Smart Artifacts*. Springer International Publishing, Cham, 193–205. https://doi.org/10.1007/978-3-319-07040-7_19

[6] M Csikszentmihalyi and R Larson. 1987. Validity and reliability of the experience-sampling method. *The Journal of nervous and mental disease* 175, 9 (Sep 1987), 526–36. https://doi.org/10.1097/00005053-198709000-00004

[7] Denzil Ferreira, Vassilis Kostakos, and Anind K Dey. 2015. AWARE: Mobile context instrumentation framework. *Frontiers in ICT* 2 (2015), 6. https://doi.org/10.3389/fict.2015.00006

[8] Gerhard Fischer and Thomas Herrmann. 2015. *Meta-design: Transforming and enriching the design and use of socio-technical systems*. Springer London, London, 79–109. https://doi.org/10.1007/978-1-4471-6720-4_6

[9] Gary Hsieh, Ian Li, Anind Dey, Jodi Forlizzi, and Scott E. Hudson. 2008. *Using visualizations to increase compliance in experience sampling*. Association for Computing Machinery, New York, NY, USA, 164–167. https://doi.org/10.1145/1409635.1409657

[10] Stephen S Intille. 2007. Technological innovations enabling automatic, context-sensitive ecological momentary assessment. In *The science of real-time data capture: Self-reports in health research*. Oxford University Press, Oxford, UK, 308–337. http://hdl.handle.net/1721.1/87152

[11] Andrew J Ko, Brad Myers, Mary Beth Rosson, Gregg Rothermel, Mary Shaw, Susan Wiedenbeck, Robin Abraham, Laura Beckwith, Alan Blackwell, Margaret Burnett, Martin Erwig, Chris Scaffidi, Joseph Lawrance, and Henry Lieberman. 2011. The state of the art in end-user software engineering. *Comput. Surveys* 43, 3 (2011), 1–44. https://doi.org/10.1145/1922649.1922658

[12] Henry Lieberman, Fabio Paternò, Markus Klann, and Volker Wulf. 2006. *End-user development: An emerging paradigm*. Springer Netherlands, Dordrecht, 1–8. https://doi.org/10.1007/1-4020-5386-X_1

[13] Thomas Ludwig, Julian Dax, Volkmar Pipek, and Dave Randall. 2016. Work or leisure? Designing a user-centered approach for researching activity "in the Wild". *Personal Ubiquitous Comput.* 20, 4 (Aug. 2016), 487–515. https://doi.org/10.1007/s00779-016-0935-7

[14] Marco Manca, Fabio Paternò, Carmen Santoro, and Luca Corcella. 2019. Supporting end-user debugging of trigger-action rules for IoT applications. *International Journal of Human Computer Studies* 123 (Mar 2019), 56–69. https://doi.org/10.1016/j.ijhcs.2018.11.005

[15] Panos Markopoulos, Nikolaos Batalas, and Annick Timmermans. 2015. On the use of personalization to enhance compliance in experience sampling. In *Proceedings of the European Conference on Cognitive Ergonomics 2015* (Warsaw, Poland) *(ECCE '15)*. Association for Computing Machinery, New York, NY, USA, Article 15, 4 pages. https://doi.org/10.1145/2788412.2788427

[16] Kristof Meers, Egon Dejonckheere, Elise K Kalokerinos, Koen Rummens, and Peter Kuppens. 2019. mobileQ: A free user-friendly application for collecting experience sampling data. https://doi.org/10.31234/osf.io/ynj7e

[17] Mohamed Musthag, Andrew Raij, Deepak Ganesan, Santosh Kumar, and Saul Shiffman. 2011. Exploring micro-incentive strategies for participant compensation in high-burden studies. In *Proceedings of the 13th International Conference on Ubiquitous Computing* (Beijing, China) *(UbiComp '11)*. ACM, New York, NY, USA, 435–444. https://doi.org/10.1145/2030112.2030170

[18] Veljko Pejovic, Neal Lathia, Cecilia Mascolo, and Mirco Musolesi. 2016. Mobile-based experience sampling for behaviour research. In *Emotions and Personality in Personalized Services: Models, Evaluation and Applications*. Springer International Publishing, Cham, 141–161.

[19] Veljko Pejovic and Mirco Musolesi. 2014. InterruptMe: Designing intelligent prompting mechanisms for pervasive applications. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Seattle, Washington) *(UbiComp '14)*. Association for Computing Machinery, New York, NY, USA, 897–908. https://doi.org/10.1145/2632048.2632062

[20] Daniel Rough and Aaron Quigley. 2015. Jeeves - A visual programming environment for mobile experience sampling. In *2015 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE Computer Society, Washington DC, USA, 121–129. https://doi.org/10.1109/VLHCC.2015.7357206

[21] Daniel Rough and Aaron Quigley. 2018. End-user development in social psychology research: Factors for adoption. In *2018 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*. IEEE, IEEE Computer Society, Washington DC, USA, 75–83. https://doi.org/10.1109/VLHCC.2018.8506573

[22] Daniel Rough and Aaron Quigley. 2019. Challenges of traditional usability evaluation in end-user development. In *End-User Development*. Springer International Publishing, Cham, 1–17. https://doi.org/10.1007/978-3-030-24781-2_1

[23] Judith Segal. 2007. Some problems of professional end user developers. In *Visual Languages and Human-Centric Computing, 2007. VL/HCC 2007. IEEE Symposium on*. IEEE Computer Society, Washington DC, USA, 111–118. https://doi.org/10.1109/VLHCC.2007.17

[24] Arthur Stone and Saul Shiffman. 1994. Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine* 16, 3 (Jan 1994), 199–202. https://doi.org/10.1093/abm/16.3.199

[25] Daniel Tetteroo and Panos Markopoulos. 2015. A review of research methods in end user development. In *End-User Development*. Springer International Publishing, Cham, 58–75. https://doi.org/10.1007/978-3-319-18425-8_5

[26] Sabrina Thai and Elizabeth Page-Gould. 2018. ExperienceSampler: An open-source scaffold for building smartphone apps for experience sampling. *Psychological Methods* 23, 4 (2018), 729–739. https://doi.org/10.1037/met0000151

[27] Konrad Tollmar and Chengcheng Huang. 2015. Boosting mobile experience sampling with social media. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '15*. ACM Press, New York, New York, USA, 525–530. https://doi.org/10.1145/2785830.2785894

[28] Niels van Berkel, Denzil Ferreira, and Vassilis Kostakos. 2017. The experience sampling method on mobile devices. *ACM Comput. Surv.* 50, 6, Article 93 (Dec. 2017), 40 pages. https://doi.org/10.1145/3123988

[29] Niels van Berkel, Jorge Goncalves, Simo Hosio, and Vassilis Kostakos. 2017. Gamification of mobile experience sampling improves data quality and quantity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 107 (Sept. 2017), 21 pages. https://doi.org/10.1145/3130972

[30] Jim van Os, Simone Verhagen, Anne Marsman, Frenk Peeters, Maarten Bak, Machteld Marcelis, Marjan Drukker, Ulrich Reininghaus, Nele Jacobs, Tineke Lataster, Claudia Simons, Richel Lousberg, Sinan Gülöksüz, Carsten Leue, Peter C. Groot, Wolfgang Viechtbauer, and Philippe Delespaul. 2017. The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. *Depression and Anxiety* 34, 6 (Jun 2017), 481–493. https://doi.org/10.1002/da.22647

[31] M. Wichers, C. J. P. Simons, I. M. A. Kramer, J. A. Hartmann, C. Lothmann, I. Myin-Germeys, A. L. van Bemmel, F. Peeters, Ph. Delespaul, and J. van Os. 2011. Momentary assessment technology as a tool to help patients with depression help themselves. *Acta Psychiatrica Scandinavica* 124, 4 (Oct 2011), 262–272. https://doi.org/10.1111/j.1600-0447.2011.01749.x

[32] Haoyi Xiong, Yu Huang, Laura E. Barnes, and Matthew S. Gerber. 2016. Sensus: A cross-platform, general-purpose system for mobile crowdsensing in human-subject studies. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Heidelberg, Germany) *(UbiComp '16)*. Association for Computing Machinery, New York, NY, USA, 415–426. https://doi.org/10.1145/2971648.2971711