Development of Mobile Data Collection Applications by Domain Experts: Experimental Results from a Usability Study

Johannes Schobel[1(✉)], Rüdiger Pryss[1], Winfried Schlee[2], Thomas Probst[1], Dominic Gebhardt[1], Marc Schickler[1], and Manfred Reichert[1]

[1] Institute of Databases and Information Systems, Ulm University, Ulm, Germany
{johannes.schobel,ruediger.pryss,thomas.probst,dominic.gebhardt,
marc.schickler,manfred.reichert}@uni-ulm.de
[2] Department of Psychiatry and Psychotherapy,
Regensburg University, Regensburg, Germany
winfried.schlee@gmail.com

Abstract. Despite their drawbacks, paper-based questionnaires are still used to collect data in many application domains. In the QuestionSys project, we develop an advanced framework that enables domain experts to transform paper-based instruments to mobile data collection applications, which then run on smart mobile devices. The framework empowers domain experts to develop robust mobile data collection applications on their own without the need to involve programmers. To realize this vision, a configurator component applying a model-driven approach is developed. As this component shall relieve domain experts from technical issues, it has to be proven that domain experts are actually able to use the configurator properly. The experiment presented in this paper investigates the mental efforts for creating such data collection applications by comparing novices and experts. Results reveal that even novices are able to model instruments with an acceptable number of errors. Altogether, the QuestionSys framework empowers domain experts to develop sophisticated mobile data collection applications by orders of magnitude faster compared to current mobile application development practices.

Keywords: Process-driven applications · End-user programming · Experimental results

1 Introduction

Self-report questionnaires are commonly used to collect data in healthcare, psychology, and social sciences [8]. Although existing technologies enable researchers to create questionnaires electronically, the latter are still distributed and filled out in a *paper-and-pencil* fashion. As opposed to paper-based approaches, electronic data collection applications enable full automation of data processing (e.g., transfering data to spreadsheets), saving time and costs, especially in the

context of large-scale studies (e.g., clinical trials). According to [15], approximately 50–60% of the data collection costs can be saved when using electronic instead of paper-based instruments. Besides this, the electronic instruments do not affect psychometric properties [5], while enabling a higher quality of the collected data [14]. In this context, [12] confirms that mobile data collection applications allow for more complete datasets compared to traditional paper-based ones. Additionally, the collected data can be directly stored and processed, whereas paper-based approaches require considerable manual efforts to digitize the data. Note that this bears the risk of errors and decreases data quality. In general, electronic questionnaires are increasingly demanded in the context of studies [11]. However, the development of mobile data collection applications with contemporary approaches requires considerable programming efforts. For example, platform-specific peculiarities (e.g., concerning user interfaces) need to be properly handled. Furthermore, profound insights into mobile data collection scenarios are needed. Especially, if more sophisticated features are required to guide inexperienced users through the process of data collection, hard-coded mobile applications become costly to maintain. Note that adapting already deployed and running mobile applications is challenging, as the consistency of the data collected needs to be ensured.

To relieve IT experts from these challenges and to give control back to domain experts, the *QuestionSys* framework is developed. The latter aims at supporting domain experts in collecting large amounts of data using smart mobile devices. QuestionSys offers a user-friendly configurator for creating flexible data collection instruments. More precisely, it relies on process management technology and end-user programming techniques. Particularly, it allows domain experts without any programming skills to graphically model electronic instruments as well as to deploy them to smart mobile devices. Furthermore, the framework provides a lightweight mobile process engine that executes the individually configured questionnaires on common smart mobile devices.

To demonstrate the feasibility and usability of the QuestionSys framework, this paper presents results from a controlled experiment evaluating the configurator component we implemented. For this purpose, subjects were asked to create data collection instruments. Altogether, the results indicate that domain experts are able to properly realize mobile data collection applications on their own using the configurator. The paper is structured as follows: In Sect. 2, fundamentals of the QuestionSys framework are introduced. Section 3 presents the conducted experiment, while Sect. 4 discusses experimental results. Related work is discussed in Sect. 5; Sect. 6 summarizes the paper.

2   Mobile Data Collection with QuestionSys

This section introduces the fundamental concepts of the QuestionSys framework. In particular, we focus on the configurator component, which will be evaluated in the presented experiment.

2.1   The QuestionSys Framework

The main goal of the QuestionSys framework is to enable domain experts (e.g., physicians, psychologists) that have no programming skills to develop sophisticated data collection instruments as well as to deploy and execute them on smart mobile devices. In particular, development costs shall be reduced, development time be fastened, and the quality of the collected data be increased. Moreover, changes of already running data collection applications shall be possible for domain experts themselves without the need to involve IT experts [21].
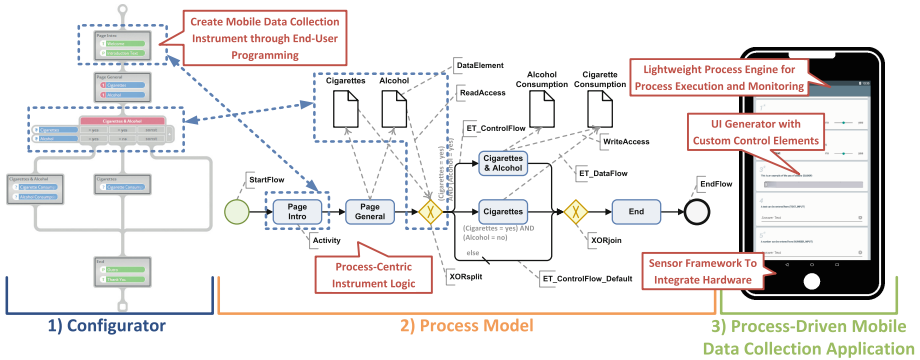

Fig. 1. The QuestionSys approach: (1) modeling a data collection instrument; (2) mapping it to an executable process model; (3) executing it on a smart mobile device.

In order to enable domain experts to develop flexible mobile applications themselves, a model-driven approach is introduced. This approach allows describing the logic of an instrument in terms of an *executable process model* (cf. Fig. 1). The latter can then be interpreted and executed by a lightweight process engine running on smart mobile devices [20]. By applying this approach, process logic and application code are separated [17]. The process model acts as a schema for creating and executing *process instances* (i.e., questionnaire instances). The process model itself consists of process activities as well as the control and data flow between them. Gateways (e.g., `XORsplit`) are used to describe more complex questionnaire logic. Following this model-driven approach, both the content and the logic of a paper-based instrument can be mapped to a process model. *Pages* of an instrument directly correspond to *process activities*; the *flow* between them, in turn, matches the *navigation logic* of the instruments. *Questions* are mapped to *process data elements*, which are connected to activities using `READ` or `WRITE` data edges. These data elements are used to store answers to various questions when executing the instrument on smart mobile devices. Altogether, QuestionSys applies fundamental BPM principles in a broader context, thus enabling novel perspectives for process-related technologies.

To properly support domain experts, the QuestionSys framework considers the entire *Mobile Data Collection Lifecycle* (cf. Fig. 2). The *Design & Modeling*
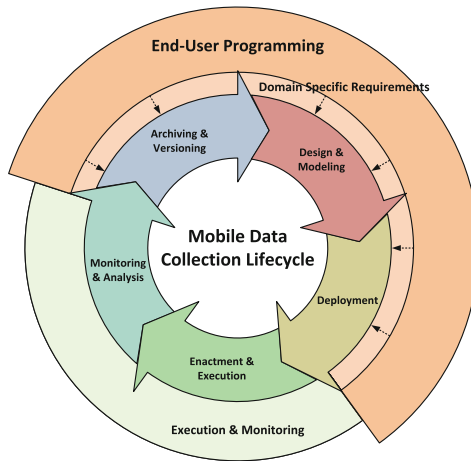
Fig. 2. Mobile data collection lifecycle

phase allows designing sophisticated data collection instruments. During the *Deployment* phase, the modeled instrument is transferred to and installed on registered smart mobile devices. In the *Enactment & Execution* phase, multiple instances of the respective mobile data collection instrument may be executed on a smart mobile device. The *Monitoring & Analysis* phase evaluates the collected data in real-time on the smart mobile device. Finally, different releases of the data collection instrument can be handled in the *Archiving & Versioning* phase. In order to address domain-specific requirements on one hand and to support domain experts on the other, technologies known from end-user programming are applied [21]. The presented study focuses on the configurator component of the presented framework. The latter covers the *Design & Modeling*, *Deployment* and *Archiving & Versioning* phases of the lifecycle.

2.2   Configurator Component

The configurator component we developed (cf. Fig. 3) applies techniques known from end-user programming and process management technology to empower domain experts to create flexible data collection instruments on their own. Due to lack of space, this component is only sketched here [19]:

(a) Element and Page Repository View (cf. Fig. 3a). The element repository allows creating basic elements of a questionnaire (e.g., headlines and questions). The rightmost part shows the editor, where particular attributes of the respective elements may be edited. Note that the configurator allows handling multiple languages. It further keeps track of different element revisions. Finally, created elements may be combined to pages using drag and drop operations.
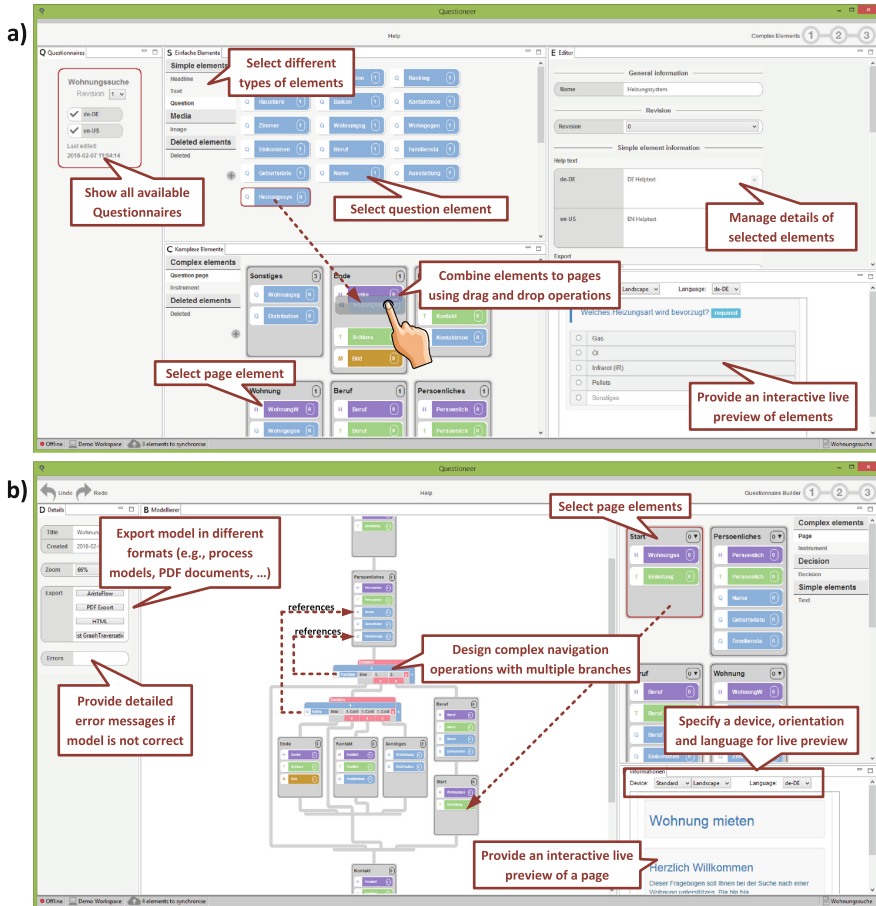
Fig. 3. The QuestionSys configurator: (a) combining elements to pages; (b) modeling a data collection instrument.

(b) Modeling Area View (cf. Fig. 3b). Domain experts may use previously created pages and drag them to the model in the center part. Furthermore, they are able to model sophisticated navigation operations to provide guidance during the data collection process. The graphical editor, in turn, strictly follows a correctness-by-construction approach; i.e., it is ensured that created models are executable by the lightweight process engine that runs on heterogeneous smart mobile devices. When deploying the model to smart mobile devices, it is automatically mapped to an executable process model.

Altogether, the configurator component and its model-driven approach allow domain experts to visually define data collection instruments. Thus, development time can be reduced and data collection applications can be realized more easily.

3 Experimental Setting

In order to ensure that domain experts are able to properly work with the configurator component, the overall concept presented in Sect. 2 needs to be evaluated. This section presents a controlled experiment, whose goal is to evaluate the feasibility and usability of the configurator component. In particular, we provide insights into the subjects and variables selected. Finally, we present the experimental design. Note that the latter constitutes a valuable template for conducting mental effort experiments on mobile data collection modeling approaches in general. Furthermore, when using the presented experimental setting, gathered results may indicate further directions on how to integrate mobile data collection with existing information systems.

3.1 Goal Definition

When developing an application, various software developing models (e.g., waterfall, V-model, SCRUM) may be chosen. Although these models include *testing* or *validation* phases, it cannot be guaranteed that end-users *accept* the final software product. Therefore, additional aspects need to be covered. For example, ISO25010 defines main software product quality characteristics, like *functional suitability*, *performance efficiency*, *usability*, and *security* [16]. The experiment presented in this paper, focuses on the usability of the presented configurator component. In particular, the experiment investigates whether domain experts understand the provided modeling concept and, therefore, are able to work properly with the configurator. For the preparation of the experiment, the *Goal Question Metric* (GQM) [2] is used in order to properly set up the goal (cf. Table 1). Based on this, we defined our research question:

> **Research Question**
>
> Do end-users understand the modeling concept of the questionnaire configurator with respect to the complexity of the provided application?

The subjects recruited for the experiment are students from different domains as well as research associates. [9] discusses that students can act as proper substitutes for domain experts in empirical studies. We do not require specific skills or knowledge from the subjects. The conducted experiment considers two *independent variables* (i.e., factors). First, we consider the *experience level* of the

Table 1. Goal Definition

Analyze	the questionnaire configurator
for the purpose of	evaluating the concept
with respect to the	intuitiveness of the modeling concept
from the point of	developers and researchers
in the context of	students and research associates in a controlled environment.

respective subjects with its two levels *novice* and *expert*. We assign subjects to one of the two groups based on answers regarding prior experience in process modeling given in the demographic questionnaire. In applied settings, novices would be domain experts with little experience in process modeling and experts would be domain experts with more experience in process modeling. Another variable we consider is the *difficulty level* of the task to be handled by the subjects (i.e., *easy* and *advanced* levels). As a criterion for assessing the complexity of a task, we decide to focus on the number of pages and decisions as well as the number of branches of the instrument to be modeled.

Two *dependent variables* are selected to measure an effect when changing the above mentioned factors. The experiment focuses on the *time needed* to solve the respective tasks as well as the *number of errors* in the resulting data collection instrument. We assume that prior experience in process modeling directly influences the subject's time to complete the tasks. In particular, we expect that experts are significantly faster than novices when modeling instruments. In order to automatically measure both dependent variables, a logging feature is added to the configurator. This feature, in turn, allows generating an execution log file containing all operations (i.e., all modeling steps) of respective subjects. We further record snapshots (i.e., images) of the data collection instrument modeled by a subject after each operation in order to allow for a graphic evaluation as well. The errors made are classified manually based on the submitted model and are weighted accordingly. Finally, hypotheses were derived (cf. Table 2).

3.2   Experimental Design

To be able to quickly react to possible malfunctions, the study is conducted as an offline experiment in a controlled environment. For this scenario, the computer lab of the Institute of Databases and Information Systems at Ulm University is prepared accordingly. The lab provides 10 workstations, each comparable with respect to hardware resources (e.g., RAM or CPU cores). Each workstation is equipped with one monitor using a common screen resolution. Before the

Table 2. Derived Hypotheses

|	
H_a0	Novices are not slower when solving advanced tasks compared to easy tasks
H_a1	Novices are significantly slower when solving advanced tasks compared to easy tasks
H_b0	Experts are not slower when solving advanced tasks compared to easy tasks
H_b1	Experts are significantly slower when solving advanced tasks compared to easy tasks
H_c0	Novices do not make more errors when solving advanced tasks compared to easy tasks
H_c1	Novices make significantly more errors when solving advanced tasks compared to easy tasks
H_d0	Experts do not make more errors when solving advanced tasks compared to easy tasks
H_d1	Experts make significantly more errors when solving advanced tasks compared to easy tasks
H_e0	Novices are not slower than experts when solving tasks
H_e1	Novices are significantly slower than experts when solving tasks
H_f0	Novices do not make more errors than experts when solving tasks
H_f1	Novices make significantly more errors than experts when solving tasks
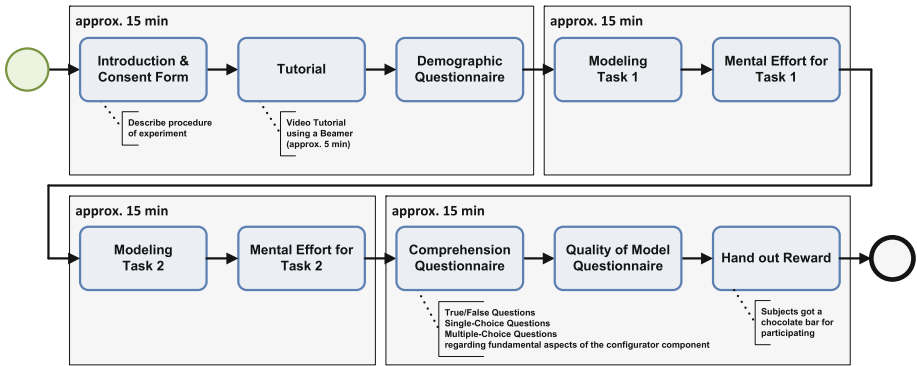
Fig. 4. Experiment design

experiment is performed, respective workstations are prepared carefully. This includes re-installing the configurator component and placing the consent form, task descriptions, and mental effort questionnaires beside each workstation.

The procedure of the experiment is outlined in Fig. 4: The experiment starts with welcoming the subjects. Afterwards, the goal of the study is described and the overall procedure is introduced. Then, the subjects are asked to sign an informed consent form. Next, we provide a 5 min live tutorial to demonstrate the most important features of the configurator component. Up to this point, the subjects may ask questions. Following this short introduction, the subjects are asked to fill in a demographic questionnaire that collects personal information. Afterwards, subjects have to model their first data collection instrument using the configurator, followed by filling in questions regarding their mental effort when handling respective task. Then, subjects have to model a second instrument (with increasing difficulty) and answer mental effort questions again. Thereby, subjects need to answer comprehension questions with respect to fundamental aspects of the developed configurator component. In the following, one final questionnaire dealing with the quality of the modeled data collection instruments has to be answered. Altogether, the experiment took about 60 min in total[1].

4   Evaluation

A total of 44 subjects participated in the experiment. Prior to analyzing the results, data is validated. [23] states that it has to be ensured that all subjects understand the tasks as well as the forms to be processed. Furthermore, invalid data (e.g., due to non-serious participation) has to be detected and removed. Two datasets need to be excluded due to invalidity (one participant aborts the study during Task 2) and doubts regarding the correctness of demographic information

---

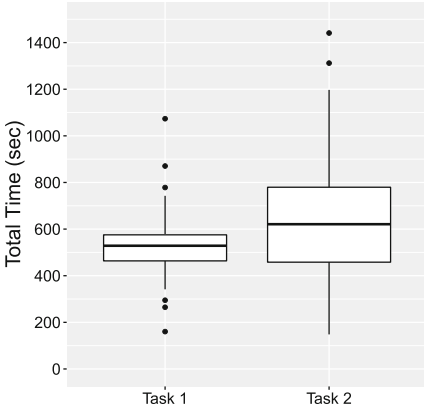[1] The dataset can be found at https://www.dropbox.com/s/tjte18zfu1j4bfk/dataset.zip.

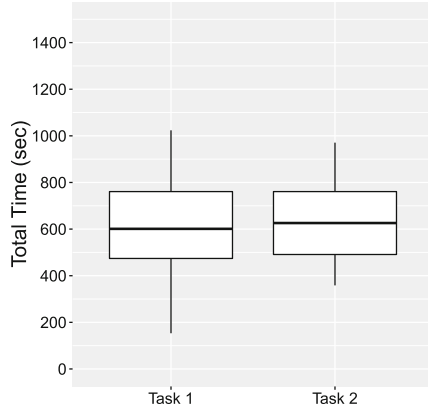**Fig. 5.** Total time (novices)

Fig. 6. Total time (experts)

(> 20 years of process modeling experience). After excluding these datasets, the final sample comprises 42 subjects. Based on their prior experience in *process modeling*, subjects are divided into two groups. Applying our criterion (*have read no more than 20 process models or have created less than 10 process models within the last 12 months*) finally results in 24 *novices* and 18 *experts*. Most of the subjects receive between 15 and 19 years of education up to this point. As no special knowledge is required for participating (besides prior experience in *process modeling* to count as *expert*), we consider the collected data as valid with respect to the goal of the study.

First, the *total time* (sec) subjects need to complete both modeling tasks is evaluated (cf. Table 3). Overall, novices need less time than experts to complete respective tasks. This may be explained by the fact that novices are not as conscientious as experts. Possibly, novices do not focus on all details needed to create data collection instruments. Next, the difference in the median is approximately 80 sec. for Task 1. The time to complete Task 2, however, barely differs for both groups. Furthermore, both groups need less time for modeling Task 1. Given the fact that Task 2 is more complex than the first one, this can be explained as well. Figs. 5 and 6 present boxplots for the total time needed. Note that the plot for novices indicates outliers in both directions. All outliers are carefully analyzed to check whether they need to be removed from the dataset. However,

Table 3. Total time and number of errors when handling tasks (median values)

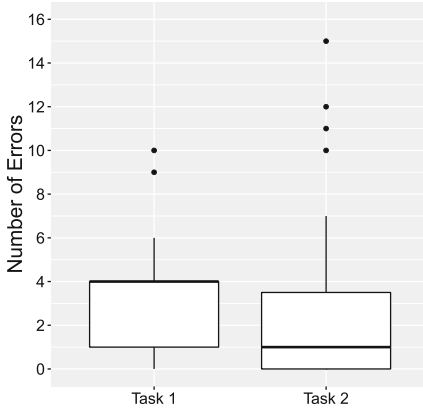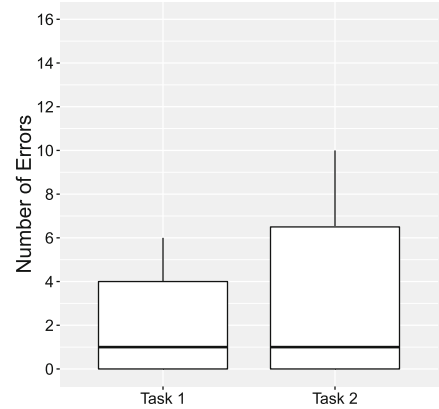|	Total time (sec)		Number of errors	
Group	Task 1	Task 2	Task 1	Task 2
Novices	528.61	620.97	4	1
Experts	601.08	625.98	1	1

Fig. 7. Number of errors (novices)

Fig. 8. Number of errors (experts)

when considering other aspects (e.g., the number of errors), it can be shown that the outliers represent valid datasets and, therefore, must not be removed.

Second, the *number of errors* in the resulting models are evaluated (cf. Table 3). As expected, experts make fewer errors than novices in the context of Task 1. Considering the results for the time needed, one can observe that novices are faster, but produce more errors than experts when accomplishing Task 1. When modeling Task 2, however, both groups can be considered the same. This may be explained by the fact that experts have prior knowledge with respect to process modeling. Furthermore, it is conceivable that some kind of *learning effect* has taken place during Task 1 as novices make fewer errors when performing the second one. Boxplots in Figs. 7 and 8 show results for each task. Again, outliers can be observed in the group of novices.

Third, *mental effort* and *comprehension* questionnaires are evaluated with respect to the previously mentioned variables. Recall that each subject has to fill in a short questionnaire after handling a certain task (cf. Table 4, top part). Figures 9 and 10 show respective medians. The calculated score (median value)

Table 4. Mental effort questionnaires

| Question	Answers
The mental effort for creating the questionnaire model was considerably high	7 Point Likert-Scale
The mental effort for changing elements was considerably high	7 Point Likert-Scale
I was able to successfully solve the task	7 Point Likert-Scale
Do your models represent the questionnaires in the given tasks?	7 Point Likert-Scale
Are there significant aspects that are missing in your models?	7 Point Likert-Scale
Do your models represent the logic of the given questionnaires exactly?	7 Point Likert-Scale
Are there any significant errors in your models?	7 Point Likert-Scale
Would you change your models if you were allowed to?	7 Point Likert-Scale
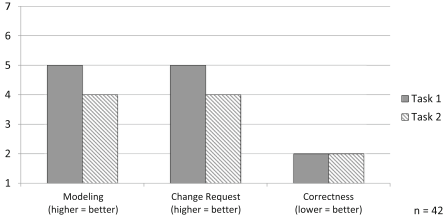
Fig. 9. Mental effort (novices)
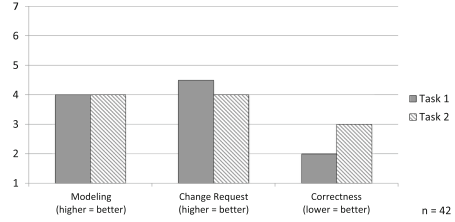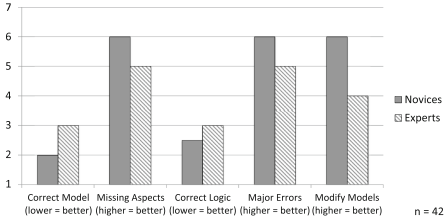

Fig. 10. Mental effort (experts)

Fig. 11. Quality of models

Table 5. Comprehension questionnaire

Group	Score (median)
Novices	20.5 out of 25
Experts	21.5 out of 25

for the comprehension questionnaire is shown in Table 5. We consider the results for both the mental effort and comprehension questionnaire as reasonable. Table 4 (bottom part) shows the questions for rating the model quality when completing the experiment (cf. Fig. 11). When combining answers of the subjects (e.g., how satisfied they are with their own models) with the analysis of the errors made, results are convincing. Interestingly, novices rate their models better compared to experts. Note that from 84 data collection instruments in total (Task 1 and Task 2 combined), 43 models (21 models from novices and 22 from experts) have zero or one error. The results regarding mental effort, comprehension, and model quality questionnaires as well as the submitted instrument models do not differ largely among the two groups. Therefore, our results indicate that the modeling concept of the developed configurator component is intuitive and end-users with relatively low prior process modeling experience are able to use the configurator.

The collected data is further checked for its normal distribution (cf. Fig. 12). The first graph shows a quantile-quantile (Q-Q) graph plotting the quantiles of the sample against the ones of a theoretical distribution (i.e., normal distribution). The second graph presents a histogram of probability densities including the normal distribution (i.e., blue) and density curve (i.e., red line).

Considering the presented results, several statistical methods are used to test the hypotheses described in Sect. 3.1 (with p-value $\leq \alpha$ (0.05)). For normally distributed datasets, *t-Tests* are applied. Non-normally distributed datasets are tested with *One-Tailed Wilcoxon(-Mann-Whitney) Tests* [23]. When applying the tests, H_a showed significant results (p-value $= 0.046$). The other tests, however, show non-significant results (with p-value > 0.05) and the corresponding null hypotheses are accepted. Besides the hypothesis that novices are
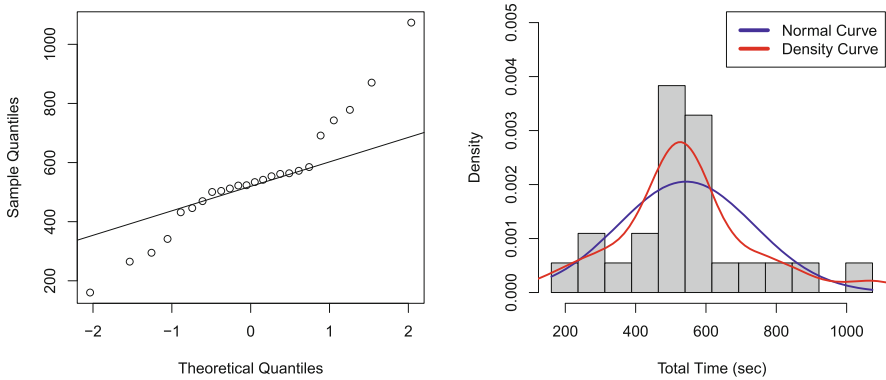
Fig. 12. Distribution of total time for Task 1 (novices) (Color figure online)

significantly slower in solving more advanced tasks, all other alternative hypotheses have to be rejected. In particular, the one stating that experts are faster than novices (i.e., hypothesis H_e1) cannot be confirmed. Considering the errors in the context of Task 1, however, novices make more errors. This may be explained by the fact that subjects having no prior experience in process modeling are not as conscientious as subjects with more experience. Novices, in turn, possibly not focus on details needed to model data collection instruments properly. The latter may be addressed by conducting an *eye-tracking* experiment with respective subjects. Furthermore, the assumption that experts make fewer errors than novices (i.e., hypothesis H_f1) cannot be confirmed. Although there is a difference in the descriptive statistics in Task 1, the difference does not attain statistical significance. In summary, results indicate that the prior experience of a subject does not affect the modeling of data collection instruments. In particular, the experiment shows that users without prior experience may gain sufficient knowledge within approximately 60 min (total time of the experiment) to model data collection applications themselves. Moreover, the learning effect between the first and second task have to be addressed more specifically in a future experiment.

To conclude, the results indicate the feasibility of the modeling concept. Overall, 43 out of 84 created instruments have been completed with zero or only one error. Given the fact that none of the subjects had ever used the application before, this relatively low number of errors confirms that the application can be easily used by novices. Hence, the QuestionSys configurator is suited to enable domain experts create mobile data collection applications themselves.

Threats to Validity. First of all, *external*, *internal*, *construct* and *conclusion validity*, as proposed in [7], were carefully considered. However, any experiment bears risks that might affect its results. Thus, its levels of validity need to be checked and limitations be discussed. The selection of involved subjects is a possible risk. First, in the experiment, solely subjects from Computer Science (34) and Business Science (8) participated. Second, 36 participants have already

worked with process models. Concerning these two risks, in future experiments we will particularly involve psychologists and medical doctors (1) being experienced with creating paper-based questionnaires and (2) having no experiences with process modeling. Third, the categorization of the subjects to the groups of novices and experts regarding their prior experience in process modeling is a possible risk. It is debatable whether an individual, who has read more than 20 process models or created more than 10 process models within the last 12 months, can be considered as an expert. A broader distinguishing, for example, between novices, intermediates, and experts (with long-term practical experience) could be evaluated as well. The questionnaires used for the modeling task of the experiment constitute an additional risk. For example, if subjects feel more familiar with the underlying scenario of the questionnaire, this might positively affect the modeling of the data collection instrument. Furthermore, the given tasks might have been too simple regarding the low number of modeling errors. Hence, additional experiments should take the influence of the used questionnaires as well as their complexity into account. In addition, we address the potential learning effect when modeling data collection instruments in more detail. Finally, another limitation of the present study is the relatively small sample size of $N = 42$ participants. However, the sample is large enough to run meaningful inferential statistical tests, though their results can only be seen as preliminary with limited external validity. Therefore, we will run another experiment to evaluate the configurator component with a larger and more heterogeneous sample.

5   Related Work

Several experiments measuring mental efforts in the context of process modeling are described in literature. Common to them is their focus on the resulting process model. For example, [13] evaluates the process of modeling processes itself. Furthermore, [22] identifies a set of fixation patterns with eye tracking for acquiring a better understanding of factors that influence the way process models are created by individuals. The different steps a process modeler undertakes when modeling processes are visually presented in [6]. However, in our study the process models represent data collection instruments. Therefore, additional aspects have to be modeled that are normally not important for process models (e.g., different versions of elements). On the other hand, these aspects may increase overall mental efforts during modeling. Consequently, our experiment differs from the ones conducted in the discussed approaches.

Various approaches supporting non-programmers with creating software have proven their feasibility in a multitude of studies. For example, [10] provides an environment allowing system administrators to visually model script applications. An experiment revealed the applicability of the proposed approach. In turn, [3] introduces a graphical programming language, representing each function of a computer program as a block.

Regarding the systematic evaluation of configurators that enable domain experts to create flexible questionnaires on their own, only few literature exists.

For example, [1] evaluates a *web-based configurator for ambulatory assessments* against *movisensXS*. More precisely, two studies are described. On one hand, the configurator component is assessed by two experts. On the other, 10 subjects evaluate the respective client component capable of enacting the configured assessment. Both studies, however, rely on standardized user-experience questionnaires (e.g., *System Usability Scale* [4]) to obtain feedback. The results are limited due to the low number of subjects. In [18], a web-based application to create and coordinate *interactive information retrieval* (IIR) experiments is presented. The authors evaluate their application in two ways: First, usability analyses are performed for the application backend by a human computer interaction researcher and a student. Both confirm an easy-to-use user interface. Second, the frontend is evaluated by performing an IIR experiment with 48 participants. Thereby, the time to complete tasks is measured by the application and participants are asked to provide feedback on how they rate their performance. Though these studies focus on the usability of the developed applications, our study pursues a different approach as it evaluates the configurator by observing correctness aspects when solving specific tasks. To the best of our knowledge, when using a configurator application for modeling data collection instruments, no similar approaches are available so far.

6   Summary and Outlook

This paper investigated the questionnaire configurator of the QuestionSys framework with respect to its usability. The configurator, in turn, shall enable domain experts to create mobile data collection applications on their own. To address the usability of the configurator, a controlled experiment with 44 participants was conducted. For the experiment, the participants were separated into two groups, based on their background knowledge and experience in process modeling. To evaluate differences between both groups, we focused on the total time needed to solve specific tasks as well as the number of errors in the submitted models. We showed that user experience in process modeling has minimal effects on the overall understanding of the configurator. Furthermore, the subjects gained respective knowledge to use the configurator in adequate time. One could argue that a learning effect took place. However, contrary to our expectations, the study showed there are no significant differences in working with the configurator regarding the experience the user has with process modeling. In order to evaluate the results with respect to domain differences, we plan a large-scale study with subjects from multiple domains. Currently, we are recruiting subjects from Psychology and Business Science. Furthermore, we address the learning effect observed and, therefore, rerun respective studies multiple times with the same subjects. The results obtained in this study confirm the intuitiveness and improve the overall user-experience of the developed configurator component. Altogether, the QuestionSys approach will significantly influence the way data is collected in large-scale studies (e.g., clinical trials).

References

1. Bachmann, A., Zetzsche, R., Schankin, A., Riedel, T., Beigl, M., Reichert, M., Santangelo, P., Ebner-Priemer, U.: ESMAC: a web-based configurator for context-aware experience sampling apps in ambulatory assessment. In: 5th International Conference on Wireless Mobile Communication and Healthcare, pp. 15–18 (2015)
2. Basili, V.R.: Software Modeling and Measurement: The Goal/Question/Metric Paradigm (1992)
3. Begel, A., Klopfer, E.: Starlogo TNG: an introduction to game development. J. E-Learn. (2007)
4. Brooke, J., et al.: SUS - a quick and dirty usability scale. Usability Eval. Ind. **189**(194), 4–7 (1996)
5. Carlbring, P., Brunt, S., Bohman, S., Austin, D., Richards, J., Öst, L.G., Andersson, G.: Internet vs. paper and pencil administration of questionnaires commonly used in panic/agoraphobia research. Comput. Hum. Behav. **23**(3), 1421–1434 (2007)
6. Claes, J., Vanderfeesten, I., Pinggera, J., Reijers, H.A., Weber, B., Poels, G.: A visual analysis of the process of process modeling. Inf. Syst. e-Bus. Manag. **13**(1), 147–190 (2015)
7. Cook, T.D., Campbell, D.T., Day, A.: Quasi-Experimentation: Design and Analysis Issues for Field Settings, vol. 351. Houghton Mifflin, Boston (1979)
8. Fernandez-Ballesteros, R.: Self-report questionnaires. Compr. Handb. Psychol. Assess. **3**, 194–221 (2004)
9. Höst, M., Regnell, B., Wohlin, C.: Using students as subjects - a comparative study of students and professionals in lead-time impact assessment. Empirical Softw. Eng. **5**(3), 201–214 (2000)
10. Kandogan, E., Haber, E., Barrett, R., Cypher, A., Maglio, P., Zhao, H.: A1: end-user programming for web-based system administration. In: Proceedings of 18th ACM Symposium on User Interface Software and Technology. ACM (2005)
11. Lane, S.J., Heddle, N.M., Arnold, E., Walker, I.: A review of randomized controlled trials comparing the effectiveness of hand held computers with paper methods for data collection. BMC Med. Inform. Decis. Mak. **6**(1), 1 (2006)
12. Marcano Belisario, J.S., Jamsek, J., Huckvale, K., O'Donoghue, J., Morrison, C.P., Car, J.: Comparison of self-administered survey questionnaire responses collected using mobile apps versus other methods. The Cochrane Library (2015)
13. Martini, M., Pinggera, J., Neurauter, M., Sachse, P., Furtner, M.R., Weber, B.: The impact of working memory and the process of process modelling on model quality: investigating experienced versus inexperienced modellers. Sci. Rep. **6**, 1–11 (2016)
14. Palermo, T.M., Valenzuela, D., Stork, P.P.: A randomized trial of electronic versus paper pain diaries in children: impact on compliance, accuracy, and acceptability. Pain **107**(3), 213–219 (2004)
15. Pavlović, I., Kern, T., Miklavčič, D.: Comparison of paper-based and electronic data collection process in clinical trials: costs simulation study. Contemp. Clin. Trials **30**(4), 300–316 (2009)
16. Rafique, I., Lew, P., Abbasi, M.Q., Li, Z.: Information quality evaluation framework: extending ISO 25012 data quality model. World Acad. Sci. Eng. Technol. **65**, 523–528 (2012)
17. Reichert, M., Weber, B.: Enabling Flexibility in Process-Aware Information Systems: Challenges, Methods, Technologies. Springer, Heidelberg (2012)

18. Renaud, G., Azzopardi, L.: SCAMP: a tool for conducting interactive information retrieval experiments. In: IIiX, pp. 286–289 (2012)
19. Schobel, J., Pryss, R., Schickler, M., Reichert, M.: A configurator component for end-user defined mobile data collection processes. In: Demo Track of the 14th International Conference on Service Oriented Computing, October 2016
20. Schobel, J., Pryss, R., Schickler, M., Reichert, M.: A lightweight process engine for enabling advanced mobile applications. In: Debruyne, C., et al. (eds.) OTM 2016. LNCS, vol. 10033, pp. 552–569. Springer, Cham (2016). doi:10.1007/978-3-319-48472-3_33
21. Schobel, J., Pryss, R., Schickler, M., Ruf-Leuschner, M., Elbert, T., Reichert, M.: End-user programming of mobile services: empowering domain experts to implement mobile data collection applications. In: IEEE 5th International Conference on Mobile Services. IEEE Computer Society Press, June 2016
22. Weber, B., Pinggera, J., Neurauter, M., Zugal, S., Martini, M., Furtner, M., Sachse, P., Schnitzer, D.: Fixation patterns during process model creation: initial steps toward neuro-adaptive process modeling environments. In: 2016 49th Hawaii International Conference on System Sciences (HICSS), pp. 600–609. IEEE (2016)
23. Wohlin, C., Runeson, P., Höst, M., Ohlsson, M.C., Regnell, B., Wesslén, A.: Experimentation in Software Engineering. Springer Science & Business Media, Heidelberg (2012)