Evaluating Methodologies for Business Process Architecture Design – A Pilot Study

Fernanda Gonzalez-Lopez and Guillermo Bustos

Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile
m.fernanda.gonzalez@gmail.com, guillermo.bustos@pucv.cl

Abstract. A pilot user study was conducted for evaluating methodologies for business process architecture design. We adapted the Method Evaluation Model for assessing these kind of methodologies in terms of perceived and actual usefulness and ease of use, as well as intention to use. The study allowed to compare two methodologies and evaluate them individually. Also, we found that usefulness was the most relevant aspect when predicting their use in practice.

Keywords: Business process model, methodology, experimental evaluation

1 Introduction

Process oriented organizations handle large collections of business process models [16]. A Business Process Architecture (BPA) provides an overview of such a collection [4]. Though multiple methodologies have been proposed for designing BPAs, in most cases they are accompanied by an ilustrative example of their feasibility but lack further evaluation [8]. Green and Ould [9] propose one of the first steps towards a more in-depth assessment of BPA design methodologies: a framework composed by a set of questions regarding the form, content, purpose, and lifecycle perspectives. Some later studies offer empirical evaluations, but these evaluations present some limitations. For example, the study by Dijkman et al. [4] compares BPA approaches in terms of their use, usefulness, and popularity. A limitation of this study is that it has an exploratory scope and, therefore, its findings cannot be generalized. Another example is the empirical assessment of a BPA approach proposed by Eid-Sabbagh [5]. Though this study offers generalizable results, it does not focus on the methodology itself, but rather on the understandability of its outcomes (i.e. the BPA models).

The goal of the work reported in this paper was to contribute to filling the aforementioned gap. In particular, it aims to provide a means for empirically assessing BPA design methodologies in a replicable and generalizable way. Such a tool will allow validating and comparing available as well as new approaches. For this purpose, we designed a user study based on the Method Evaluation Model (MEM) by Moody [14]. The MEM is a specilization of the Technology Acceptance Model (TAM) [3]. The TAM suggests that the acceptance of a new technology depends on how its potential users perceive its usefulness and ease of

S. Kolb, C. Sturm (Eds.): 11th ZEUS Workshop, ZEUS 2019, Bayreuth, Germany, 14-15 February 2019, published at http://ceur-ws.org/Vol-2339

use. The MEM, on the other hand, suggests that the acceptance of a new method depends not only on these variables, but also on the actual usefulness and ease of use of the method. We chose to use the MEM for a number of reasons: (i) it has a sound rationale behind it; (ii) it has proven to deliver valid and reliable results for evaluating Information System design methods [14]; and (iii) it has been successfully adapted for specific types of methods in the past, e.g. [15]. In order to use the MEM for BPA design methodologies, we needed to adapt it. This paper reports on a pilot study conducted in the process of adapting the MEM. Pilot studies are useful means for detecting issues that can be improved for future versions of a study [11]. We followed the guidelines for empirical evaluation by Shull et al. [17] for both the design and report of the pilot study.

The remainder of the paper is structured as follows: Sect. 2 describes planning of the experiment and adaptation of the MEM and its questionnaire, Sect. 3 reports on the execution of the study and the necessary tunning of the questionnaire (taking into account validity and reliability), and, finally, Sect. 4 provides the concluding remarks. In the following, and provided it is clear from the context, we will use the term *methodology* when refering to a *BPA design methodology*.

2 Planning

2.1 Variables

The *methodology* was the single *independent variable* of the experiment. We opted to use two levels for such factor: the *methodology by BPTrends Associates* [10] and the *domain-based methodology* [7]. The decision to use these two lies in the fact that they differ regarding their origins and, more importantly, regarding their perspectives on the problem of designing a BPA. On one hand, the *methodology by BPTrends Associates* – which has been used in consultancy for several years – derives processes based on a recurrent decomposition of previously identified value chains [10]. On the other hand, the *domain-based methodology* – recently developed in the academia – derives processes and their relationships from the analysis of the domain-relevant data [7].

The experiment considered all five *dependent latent variables* used in the MEM [14]: (a) *Actual Ease of Use* (AEOU), i.e. effort required to apply a method, which can be evaluated by input measures such as modeling time; (b) *Actual Usefulness* (AU), i.e. degree to which a method achieves its objectives, which can be evaluated by output measures such as model quality; (c) *Perceived Ease of Use* (PEOU), i.e. degree to which a person believes that using a particular method would be free of effort in aspects such as learning and using; (d) *Perceived Usefulness* (PU), i.e. degree to which a person believes that a particular method will be effective in achieving its intended objectives which might include usefulness, relevance to modeling purpose, and model comprensibility; and (e) *Intention To Use* (ITU), i.e. extent to which a person intends to use a particular method.

Observable variables allow to measure the previously described latent ones (e.g. $PEOU_1 ... PEOU_6$ for PEOU). Table 1 lists these observable variables showing their operationalization together with the respective data gathering technique, all of which was adapted from the MEM. For AU_1, data were gathered by assesing the treatment output. For the rest of the variables, data were collected using a post-task questionnaire (see Sect. 2.4) consisting of: (i) an open question for AU_1, and (ii) 5-point Likert scale questions for the remaining variables, ranging from 1 for *strongly disagree* to 5 for *strongly agree*.

Table 1. Items of the adapted instrument: variables, their operationalization, and data gathering technique

Variable	Operationalization	Data
$AEOU_1$	Total time in hours to apply the methodology.	OQ
AU_1	Total number of mistakes in models.	OA
$PEOU_1$	I find learning the methodology for BPA design is easy.	CQ
$PEOU_2$	I find it would be easy for me to become skillful at using the BPA design methodology.	CQ
$PEOU_3$	I find using the method in the BPA design task required a lot of mental effort.	CQ
$PEOU_4$	I find using the method in the BPA design task required a lot of time.	CQ
$PEOU_5$	Overall, I find the BPA design methodology easy to use.	CQ
$PEOU_6$	I find I am now competent to use this method in practice.	CQ
PU_1	I find the methodology provides an effective solution for designing BPAs.	CQ
PU_2	Overall, I find the methodology useful for designing BPAs.	CQ
PU_3	I find the methodology useful for identifying BP relations.	CQ
PU_4	I find the BP relations in the resulting model to be expressive.	CQ
PU_5	I find the BPA model resulting from using the methodology to be understandable.	CQ
PU_6	I find others, provided a quick explanation, would easily understand BP relations in the BPA resulting model.	CQ
ITU_1	If I am given the task of designing a BPA in the future, my intention would be to use this methodology.	CQ
ITU_2	If someone asks me to recommend a BPA design methodology for clearly identifying BP relations, I predict I would recommend this one.	CQ

OA : output analysis, CQ : closed question, OQ : open question

2.2 Subjects and Treatments

The subjects of the study were 25 Industrial Engineering senior undergraduate students of the *Pontificia Universidad Católica de Valparaíso*. These subjects were chosen because they were easily accessible and — most importantly — represent the original context to some extent [13] because of: (i) having a background in conceptual modeling due to a previous *System and Business Process Modeling* course; (ii) being familiarized with Business Process Management (BPM) in the context of their current enrollment in a *Business Engineering* course; and (iii) being likely to soon become BPM practitioners. All subjects were assigned to two treatments that we called A and B corresponding to applying the *BPTrends Associates methodology*, and the *domain-based methodology*, respectively, for designing the BPA for a Higher Education Institution (HEI) business scenario.

2.3 Hypotheses

The study considers 24 hypotheses. We used nonparametric statistical methods to test them. Unlike parametric methods, nonparametric methods do not require that observations are defined in an interval or ratio scale and, also, are independent from the underlying data distribution [1]. Both of these aspects fit the data collected in our study since we have a small set of, mainly, ordinal-scaled observations and no assumption of normality could be made (see Fig. 1).

Individual evaluation. We used six hypotheses to evaluate the perceptions of the users about the methodologies, individually. Hypotheses H1 to H3 assess the perceptions of subjects about the *methodology by BPTrends Associates* regarding ease of use, usefulness, and intention to use. Hypotheses H4 to H6 are the analogous for the *domain-based methodology.* Testing each of these six hypotheses required checking whether the observed median score of the respective variable significantly differed from the zero point of the 5-point Likert scale (median_0=3). We used one-sample Wilcoxon signed rank tests [12]: the null hypothesis states that observed median scores are smaller or equal than the zero point (H_0: $\text{median}_{\text{variable}} \leq \text{median}_0$) and the alternative hypothesis states that observed median scores are greater than the zero point (H_A: $\text{median}_{\text{variable}} > \text{median}_0$).

Comparative evaluation. We used three hypotheses to evaluate the perceptions of the users about the *methodology by BPTrends Associates* and the *domain-based methodology* in a comparative way. Hypotheses H7 to H9 compare the methodologies in terms of their perceived ease of use, usefulness, and intention to use. This required contrasting the observed median scores for each variable between two different treatments as reported by the same subjects. We used the paired Wilcoxon signed rank test [2] : the null hypothesis states that observed median scores of treatment A are equal or smaller than those for B (H_0: $\text{median}_{\text{variable\_A}} \leq \text{median}_{\text{variable\_B}}$) and the alternative hypothesis states that observed median scores of treatment A are greater than those for B (H_A: $\text{median}_{\text{variable\_A}} > \text{median}_{\text{variable\_B}}$).

Likelihood of adoption in practice. We used fifteen hypotheses to seek relationships that allow to predict the likelihood of use in practice of BPA design methodologies in general. Hypotheses H10 to H24 test the existence of a relationship between every pair of variables in the experiment plus an additional variable called *Modeling Skills*. We wanted to check whether modeling skills of the subjects influenced their performance and/or perceptions. The value of *Modeling Skills* corresponded to the final grade of the *System and Business Process Modeling* course. To assess relations among variables, we used Spearman rank-order correlation test [12]: the null hypothesis states that there is no relationship (H_0: $\rho = 0$) and the alternative hypothesis states that there is a relationship (H_A: $\rho \neq 0$).

2.4 Materials and Steps

The materials for the experiment include: (i) training media for each methodology; (ii) HEI business scenario specification; (iii) task execution instructions; (iv) task outcome assessment guidelines; and (v) post-task questionnaire (see Table 1). The questionnaire played the role of collecting experiences, perceptions, and intentions of the subjects after using each methodology. Our questionnaire operationalized the variables from the MEM sligthly differently, mainly by using questions addressing the particularities of BPA design methodologies, and written in the native language of the subjects, i.e. Spanish. As in the MEM, we used an opposing format Likert scale for closed questions and also arranged the questions randomly to reduce the potential ceiling effect that could induce monotonous responses to question measuring the same latent variable. Additionally, we considered guidelines from Fowler [6], e.g. questionnaire formatting and 20-word limit per statement. Table 1 – leaving aside the first row – provides the English-speaking version of the items of the questionnaire.

The sequence of steps for the experiment were: (i) treatment *A* training; (ii) treatment *A* execution (two weeks deadline, groupal outcome) and respective post-task survey; (iii) treatment *B* training; (iv) treatment *B* execution (three weeks deadline, individual outcome) and respective post-task survey; (v) data consolidation; (vi) instrument evaluation; and (vii) data analysis. Training material, as well as task execution instructions and questionnaire, were available for self-administration via the course website.

3 Execution

3.1 Data Consolidation and other Considerations

In our study, we ruled out AEOU and AU data for treatment *A*, since such treatment was not applied individually. Also, and since considering this does not bias statistical inference, we decided to reject observations of subjects producing incomplete questionnaires, which lead to a total of 23 observations.

We inverted scores for questions formulated in a negative form, namely $PEOU_3$ and $PEOU_4$. Finally, for each subject, we calculated the value for each latent variable as the average of the scores in the respective observable variables.

Statistical analysis in this study considered a 95% confidence.

3.2 Instrument Evaluation

A key step for a proper building or adaptation of an instrument is the evaluation of its *validity* and *reliability* [11]. While validity is concerned with how well an instrument measures what it is supposed to measure, reliability is concerned with reproducibility of the data obtained using the instrument [11]. We conducted such an evaluation for the previously described questionare by analyzing the latent variables that are operationalized via multiple observable variables, namely PEOU, PU, and ITU.

The main aspect of validity is *construct validity* which, due to our small sample size and ordinal data, was analyzed via the Spearman coefficient $\rho \in$ [-1,1] for *inter-item correlation*. In this analysis, the fact that variables theoretically measuring the same construct hold a significant negative correlation ($\rho \leq 0$) suggests that such variables might actually not be measuring the same construct, and thus they are candidates for removal from the instrument. In our questionnaire, we removed $PEOU_6$ for eliminating significant negative correlations.

Reliability of the questionnaire was assessed using the Chronbach's alpha coefficient $\alpha \in$ [0,1] due to its suitability for summated scales such as the one used in our study. Chronbach's alpha evaluates internal consistency of a set of variables aiming to measure a single construct by calculating the influence of error in the measure. The ideal value of Chronbach's alpha is 1 (measure is error-free), but values over 0.7 are considered adequate. After removing $PEOU_6$, the coefficient for PEOU, PU, ITU, as well as for their combination, lie arround this lower limit of adequate reliability.

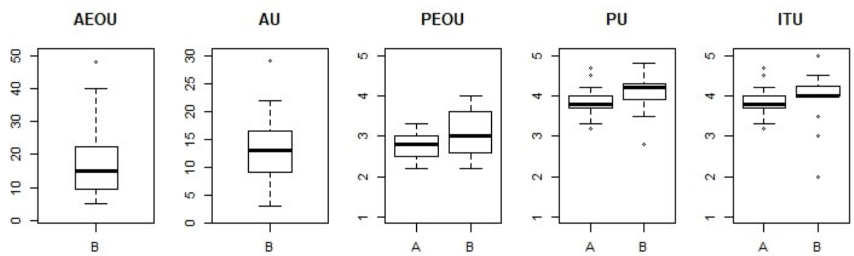

Fig. 1. Summary of the data (A: *BPTrends* and B: *domain-based*)

3.3 Data Analysis

The data gathered in the study (after removing $PEOU_6$) are summarized in Fig. 1. They show that the time used for applying the domain-based methodology ranged from 5 to 48 h, with an average value of 17.6 h. For the same treatment, the number of mistakes ranged between 3 and 29, with an average value of 13.2. It was not possible to compare the methodologies in these aspects due to data considerations described in Sect. 3.1. For both methodologies, PEOU, PU, and ITU were analyzed via the hypotheses described in Sect. 2.3 and whose analysis is detailed in the following.

Individual evaluation. According to the obtained p-values for H1 to H6, there was significant evidence to reject $H2_0$ (p=0.0000), $H3_0$ (p=0.0002), $H5_0$ (p=0.0000), and $H6_0$ (p=0.0000). This provides evidence supporting that both methodologies were perceived as useful and there is intention to use them, but they are not perceived as easy to use.

Comparative evaluation. According to the obtained p-values for H7 to H9, there was no significant evidence to reject any null hypotheses. Hence, the data weakly support the thesis that the *domain-based* methodology outperforms the *BPTrends Associates* methodology in terms of perceived ease of use, perceived usefulness, and intention to use.

Likelihood of adoption in practice. According to the obtained p-values for H10 to H24, there is only significant evidence to reject that AU does not hold an inverse relation with PU ($\rho_{\mathrm{AU,PU}}$ = -0.52, p=0.0114), and that PU does not hold a direct relation with ITU ($\rho_{\mathrm{PU,ITU}} = 0.53$, p=0.0001). While the latter agrees with the MEM, the former does not and makes no sense conceptually either. The remaining relations identified in the MEM are missing, namely between AEOU and PEOU, PEOU and ITU, PEOU and PU, and AU and PU. Finally, it was possible to verify that modeling skills were unrelated to the variables measured in this study; so it was possible to rule it out as a confounding variable threatening internal validity of the experiment.

4 Conclusion and Outlook

We created an instrument for evaluating BPA design methodologies that has proven valid and reliable in the aspects of perceived ease of use, perceived usefulness, and intention to use. This allowed to conduct an individual as well as a comparative evaluation of two methodologies, namely the *BPTrends Associates* and *domain-based* methodology. We found that both methodologies were perceived as useful but not easy to use, and there was intention to use them in practice. Also, results suggested that the *domain-based* methodology outperforms the *BPTrends Associates* but, since the data were inconclusive, further research is needed to confirm this.

A limitation of this pilot study was that, due to the different way the treatments were applied, it was not possible to compare the methodologies in regard to their ease of use and usefulness. To overcome this issue, future versions of the study will ask the subjects to produce individual outcomes for both treatments. Also, we will change the way treatments were applied into a crossover design, since this would provide better data by minimizing learning and fatigue effect. After the training, it may be adequate to add a little test at the end to ensure an even learning level.

As an additional means to improve the experiment, its future versions will consider some issues reported by the subjects, namely the needs to clarify some terminology used in the survey (e.g. symbol), to provide more detail on the goal of the survey, and to review the wording of some questions.

Finally, the experiment also provided insights into the aspects that allow predicting the adoption in practice of methodologies. Though ease of use and usefulness seem to be predictors of the use in practice of a given IT artifact [14], we found that for the assessed BPA design methodologies only perceived usefulness seemed to be a predictor of future use.

Acknowledgments This paper was written during a research stay of the first author with the *BPT-Group* at the *Hasso Platner Institute* in *Potsdam University*. However, the study was conducted at the *Pontificia Universidad Católica de Valparaíso*.

References

1. Canavos, G.C.: Probabilidad y Estadística: Aplicaciones y Métodos. Mc Graw-Hill (1998)
2. Corder, G.W., Foreman, D.I.: Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach. John Wiley & Sons, Inc., Hoboken, New Jersey (2009)
3. Davis, F.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly pp. 319–340 (1989)
4. Dijkman, R., Vanderfeesten, I., Reijers, H.A.: Business Process Architectures: Overview, Comparison and Framework. Enterprise Information Systems 12(2), 129–158 (2016)
5. Eid-Sabbagh, R.H.: Business Process Architectures: Concepts, Formalism, and Analysis. Ph.D. thesis, Hasso Plattner Institute, University of Postdam, Germany (2015), `https://publishup.uni-potsdam.de/files/7971/eid-sabbagh{_}diss.pdf`
6. Fowler, F.: Survey Research Methods. Sage Publications, Boston, 5th edn. (2014)
7. Gonzalez-Lopez, F., Bustos, G.: Business process architecture baselines from domain models. In: Business Process Management Workshops. BPM 2016. LNBIP, vol. 281, pp. 118–130. Springer (2017)
8. Gonzalez-Lopez, F., Bustos, G.: Business process architecture design methodologies – a literature review. Business Process Management Journal (2019), `https://doi.org/10.1108/BPMJ-09-2017-0258`
9. Green, S., Ould, M.A.: A Framework for Classifying and Evaluating Process Architecture Methods. Software Process Improvement and Practice 10(4), 415–425 (2005), `http://dx.doi.org/10.1002/spip.244`
10. Harmon, P.: Business process change: a business process management guide for managers and process professionals. Elsevier, Amsterdam, 3^{rd} edn. (2014)
11. Hernández, R., Fernandez, C., Baptista, P.: Metodología de la Investigación. Mexico, D.F., 3rd edn. (2006)
12. Hollander, M., Wolfe, D.A., Chicken, E.: Nonparametric Statistical Methods. John Wiley & Sons, Inc., Hoboken, New Jersey, 3rd edn. (2014)
13. Höst, M., Regnell, B., Wohlin, C.: Using Students as Subjects—A Comparative Study of Students and Professionals in Lead-Time Impact Assessment. Empirical Software Engineering 5(3), 201–214 (2000)
14. Moody, D.L.: The Method Evaluation Model : A Theoretical Model for Validating Information Systems Design Methods. Information Systems Journal pp. 1327–1336 (2003)
15. Reijers, H., Vanderfeesten, I., Plomp, M., Van Gorp, P., Fahland, D., van der Crommert, W., Diaz Garcia, H.: Evaluating data-centric process approaches: Does the human factor factor in? Software & Systems Modeling 16(3), 649–622 (2017)
16. Rosemann, M.: Potential Pitfalls of Process Modeling : Part A. Business Process Management Journal 12(2), 249–254 (2006)
17. Shull, F., Singer, J., Sjoberg, D.: Guide to Advanced Empirical Software Engineering. Springer, London, 1st edn. (2008)