Workflow Management versus Case Handling: Results from a Controlled Software Experiment

Bela Mutschler
Daimler AG
Group Research
Ulm, Germany
bela.mutschler@daimler.com

Barbara Weber
Dept. of Computer Science
University of Innsbruck
Austria
barbara.weber@uibk.ac.at

Manfred Reichert
Information Systems Group
University of Twente
The Netherlands
m.u.reichert@utwente.nl

ABSTRACT

Business Process Management (BPM) technology has become an important instrument for improving process performance. When considering its use, however, enterprises typically have to rely on vendor promises or qualitative reports. What is still missing and what is also demanded by IT decision makers are quantitative evaluations based on empirical and experimental research. This paper picks up this demand and illustrates how experimental research can be applied in the BPM field. The conducted experiment compares efforts for implementing a sample business process either based on standard workflow technology or on a case handling system. We motivate and describe the experiment design, discuss threats for the validity of experiment results (as well as risk mitigations), and present experiment results. In general, more experimental research is needed in order to obtain more valid data on the various aspects and effects of BPM technology and tools.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Office Automation—*Workflow Management, Case Handling*

1. INTRODUCTION

Providing effective IT support for business processes has become crucial for enterprises to stay competitive in their market [1]. In response to this need numerous process support paradigms (e.g., workflow management, service flow management, case handling), process specification standards (e.g., WS-BPEL, BPML), and BPM tools (e.g., ARIS Toolset, Tibco Staffware, FLOWer) have emerged [4].

When evaluating suitability of existing BPM technology for a particular project or when arguing about its strengths and weaknesses, typically, it becomes necessary to rely on qualitative criteria. As one example consider *workflow patterns* [19], which can be used to evaluate the expressiveness of the workflow modeling language provided by a particular BPM tool. As another example consider process change patterns [22]. What has been neglected so far are more profound evaluations of BPM technology based on empirical or experimental research. This is surprising as the benefits of these research methods have been demonstrated in areas like software engineering (e.g., in the context of software development processes or code reviews [10, 7]) for a long time [17]. From the introduction of experimental research to BPM as well as to the development of process-aware information systems, we expect more valid, quantitative data on costs and benefits of BPM technology. This, in turn, becomes increasingly important for IT managers and project leaders [8].

Picking up this demand, this paper illustrates how experimental research can be applied in the BPM context. For this purpose we have conducted a controlled software experiment with 48 participants. Exemplarily, this experiment investigates efforts related to the implementation and change of business processes either using a conventional workflow system [20] or case handling technology [21]. More precisely, we have used Tibco *Staffware* [18] as representative of workflow technology and *FLOWer* [2] as representative of case handling systems. We describe our experiment design, give a mathematical model of the experiment, and discuss potential threats for the validity of experiment results. The results of our experiment help to better understand the complex efforts caused by using BPM technology.

Section 2 motivates the need for experimentation in BPM and provides background information needed for understanding our experiment. Section 3 describes our experimental framework. Section 4 deals with the performance and results of our experiment. Finally, Section 5 discusses related work and Section 6 concludes with a summary.

2. BACKGROUNDS

Assume that a business process for refunding traveling expenses shall be supported by a *Process-Aware Information System* (PAIS) realized on top of BPM technology. This *eTravel business process* distinguishes between four roles (cf. Fig. 1). The *traveler* initiates the refunding of his expenses. For this purpose, he has to summarize the travel data in a *travel expense report*. This report is then forwarded either to a *travel expense responsible* (in case of a national business trip) or to a *verification center* (in case of an international

SAC'08 March 16-20, 2008, Fortaleza, Ceará, Brazil

business trip). Both the *travel expense responsible* and the *verification center* fulfill the same task, i.e., they verify a received travel expense report. "Verification" means that the declared travel data is checked for correctness and plausibility (e.g., regarding accordance with receipts). An incorrect travel expense report will be send back to the traveler (for correction). If it is correct, it will be forwarded to the *travel supervisor* for final approval. The supervisor role may be filled, for example, by the line manager of the traveler. If a travel expense report is approved by the supervisor, the refunding will be initiated. Otherwise, it will be send back to either the travel expense responsible (national trip) or the verification center (international trip). Note that this is a characteristic (yet simplified) process as it can be found in many organizations.

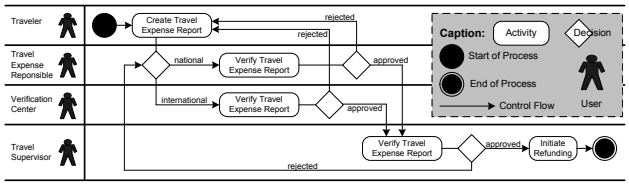


Figure 1: The eTravel Business Process.

When realizing a PAIS supporting this process, one challenge is to select the most adequate BPM technology for this. Currently, there exist different BPM paradigms, which can be applied in the given context. Among them are *workflow management* and *case handling*.

Workflow Management. Contemporary *workflow management system* (WfMS) enable the modeling, execution, and monitoring of business processes. When working on a particular process step (i.e., activity), typically, in WfMS-based applications, only data needed for executing this activity is visible to respective actors, but no other workflow data. This is also known as "context tunneling". WfMSs coordinate activity execution based on routing rules, which are described by process definitions and which are strictly separated from processed data. If an activity is completed, subsequent activities will become active. Accompanying to this the worklists of potential actors will be updated accordingly. Finally, electronic forms are typically used to implement activities and to present data being processed.

Case Handling. An alternative BPM paradigm is provided by case handling [21]. A *case handling system* (CHS) aims at more flexible process execution by avoiding restrictions known from (conventional) workflow technology (cf. Fig. 2)). Examples of such restrictions include rigid control flow and the aforementioned context tunneling. The central concepts behind a CHS are the case and its data as opposed to the activities and routing rules being characteristic for WfMSs. Usually, CHSs present all data about a case at any time to the user (assuming proper authorization), i.e., context tunneling as known from WfMSs is avoided. Furthermore, CHSs orchestrate the execution of activities based on the data assigned to a case. Thereby, different kinds of data objects are distinguished (cf. Fig. 2). *Free data objects* are not explicitly associated with a particular activity and can be changed at any time during a case execution (e.g., Data Object 3 in Fig. 2). *Mandatory* and *restricted data objects*, in turn, are explicitly linked to one or more activities. If a data object is mandatory for an activity, a value will have to be assigned to it before the activity can be completed (e.g., Data Object 5 in Fig. 2). If a data object is restricted for an activity, this activity needs to be active in order to assign a value to the data object (e.g., Data Object 6 in Fig. 2). Like in WfMSs, forms linked to activities are used to provide context-specific views on case data. Thereby, a CHS does not only allow to assign an *execution role* to activities, but a *redo role* (to undo an executed activity) and a *skip role* as well (to omit the execution of activities). User 2 in Fig. 2, for example, may execute Activities 3, 4, 5 and 6, and redo Activities 2 and 3.

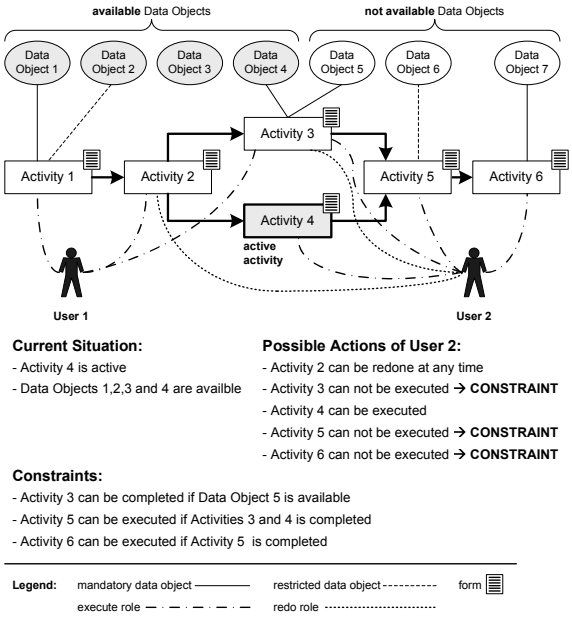


Figure 2: Data-driven Case Handling.

Despite conceptual differences, both paradigms can be used for implementing processes in general and our eTravel process in particular. Usually, the selection of "the most adequate" BPM technology depends on project-specific requirements. While some IT managers will consider BPM technology as adequate if best practices are available, others will take into account more specific selection criteria like the support of a sufficient degree of process flexibility. Likewise, IT managers can be interested in value-based considerations as well. In practice, for example, a frequently asked question is whether *there is a difference in the efforts for implementing a business process either with BPM technology A or BPM technology B* and if "yes" *how strong this difference is*. Demanding for such considerations, IT managers typically have to rely on vendor data (e.g., about the return-on-investment of their products) and qualitative experience reports. What has been not available so far are precise quantitative data about the use of BPM technology and PAIS (e.g., concerning efforts for implementing processes).

To generate quantitative data, controlled software experiments offer promising perspectives. In the following, we pick up this idea and describe the results of an experiment in which we investigate efforts related to the implementation of business processes using either a WfMS or a CHS.

We use Tibco *Staffware* [18] (Version 10.1) as typical representative of workflow technology. Its build-time tools in-

clude, among other components, a visual process modeling tool and a graphical form editor. The used CHS, in turn, is *FLOWer* [2] (Version 3.1), the most widely used commercial CHS. Like Staffware, FLOWer provides a visual process modeling tool and a form editor.

3. EXPERIMENTAL FRAMEWORK

This section describes the experimental framework underlying our experiment. Section 3.1 discusses general issues to be considered when designing an experiment. Section 3.2 describes the specific design underlying our experiment. Section 3.3 discusses factors threatening the validity of experiment results as well as potential mitigations.

3.1 Basic Issues

Literature about software experiments [3, 5, 6, 17, 23] provides various design guidelines for setting up an experiment. *First*, an experiment design should allow to collect as much data as possible with respect to the major goals of the experiment. *Second*, collected data should be unambiguous. *Third*, the experiment must be feasible within the given setting (e.g., within the planned time period). Meeting these design criteria is not trivial. Often, an experiment cannot be accomplished as planned due to its complex design or an insufficient number of participants [17].

Considering these major criteria, we accomplished our experiment as a *balanced single factor experiment* with *repeated measurement* (cf. Fig. 3). This design is particularly suitable for comparing software development technologies [6]. Specifically, single factor experiments investigate the effects of one *factor*[1] (e.g., a particular software development technology) on a common *response variable* (e.g., implementation efforts). This design also allows to analyze variations of a factor (e.g., two alternative tools for software development). Generally, these variations are called *factor levels*. The response variable is determined when the participants of the experiment (who are also called *subjects*) apply the factor or factor levels to an *object* (e.g., a specification to be implemented, based on a set of requirements).

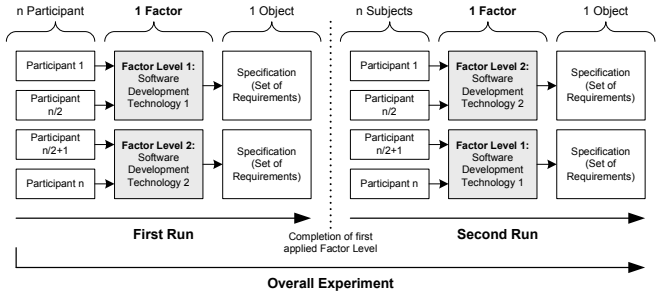


Figure 3: Single Factor Experiment.

We denote a single factor experiment as *balanced* if all factor levels are used by all participants of the experiment. This enables *repeated measurements* and the collection of more precise data as every subject generates data for every treated factor level. Generally, repeated measurements can be realized in different ways. Fig. 3 shows a frequently applied variant which is based on two subsequent *runs*. During the *first run* half of the subjects apply "Software Development Technology 1" to the treated object, while the other half uses "Software Development Technology 2". After having completed the first run, the *second run* begins. During this run each subject applies that factor level to the object not been treated so far.

[1] *Multi factor experiments*, by contrast, investigate the effects of factor combinations on a common response variable, e.g., of a software development technology and a software development process on implementation efforts. Despite such experiments can improve the validity of experiment results, they are rarely applied in practice [12].

3.2 Experiment Design

Considering the generic experiment design from Section 3.1, our specific design comprises the following elements:

- ***Subjects***: Subjects are 48 students of a combined Bachelor/Master Computer Science course at the University of Innsbruck. These 48 students are divided into 4 *main groups* each consisting of 4 *teams* with 3 students (cf. Fig. 4). This results in an overall number of 16 teams. The students are randomly assigned to the teams prior to the start of the experiment.

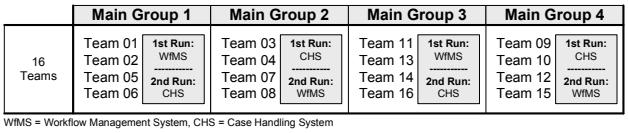


Figure 4: Main Groups and Teams.

- ***Object***: The object to be implemented is the *eTravel business process* (cf. Section 2). Its specification comprises two parts: an initial "Base Implementation" (*Part I*) and an additional "Change Implementation" (*Part II*). While the first part deals with the realization of process support for the refunding of national business trips, the second part specifies a process change, namely, additional support for refunding international business trips. Both parts describe the elements to be implemented: the process logic, user roles, and the data to be presented using simple electronic forms. Note that this specification does not only enable us to investigate efforts for (initially) implementing a business process, but also to examine efforts for subsequent process changes. In our experiment, with "process change" we mean the adaptation of the implemented business process. After having realized such a process change new process instances are based on the new process model. We do not investigate the migration of changing process instances to the new process schema in this context.

- ***Factor & Factor Levels***: In our experiment, *BPM technology* is the considered factor with factor levels "WfMS" (Staffware) and "CHS" (FLOWer).

- ***Response Variable***: In our experiment the response variable is the *time* the subjects (i.e., the students) need for implementing the given object (i.e., the eTravel specification) with each of the factor levels (WfMS and CHS). All effort values related to the Staffware implementation are denoted as "WfMS Sample". All effort values related to the FLOWer implementation are denoted as "CHS Sample".

Besides, the following issues are important:

- ***Instrumentation***: To precisely measure the response variable, we have developed an application called *TimeCatcher*. This "stop watch" allows to log time in six typical "effort categories" related to the development of a process-oriented application: (1) *process modeling*, (2) *user/role management*, (3) *form design*, (4) *data modeling*, (5) *test*, and (6) *miscellaneous efforts*. To collect qualitative feedback as well (e.g., concerning the maturity or usability of the applied WfMS and CHS), we use a structured *questionnaire*.

- ***Data Collection Procedure***: The TimeCatcher tool is used by the students during the experiment. The aforementioned questionnaire is filled out by the students after completing the experiment.

- ***Data Analysis Procedure***: For data analysis well-established statistical methods and standard metrics are applied (cf. Section 4.3 for details).

The mathematical model of our experiment can be summarized as follows: n subjects $S_1, ..., S_n$ (n $\in \mathbb{N}$) divided into m teams $T_1, ..., T_m$ ($m \in \mathbb{N}, m \geq 2$, m even) have to implement the eTravel business process specification. This specification describes a "Base Implementation" O_1 (corresponding to the "national case" of the eTravel process) and a "Change Implementation" O_2 (additionally introducing the "international case"). During the experiment one half of the teams ($T_1, ..., T_{m/2}$) implements the complete specification (i.e., base and change implementation) using a WfMS (PMS_1, Staffware), while the other half ($T_{m/2+1}, ..., T_m$) does this using a CHS (PMS_2, FLOWer). After finishing the implementation with the first factor level (i.e., the *first run*), each team has to implement the eTravel process using the second factor level in a *second run* (i.e., the development technologies are switched). The response variable "Effort[Time] of T_m implementing O using PMS_j" is logged with the TimeCatcher tool.

3.3 Risk Analysis and Mitigations

When accomplishing experimental research and generating results, related risks have to be taken into account. Generally, there exist factors that threaten both the *internal validity* ("Are the claims we made about our measurements correct?") and the *external validity* ("Can we generalize the claims we made?") of an experiment. In our context, threats to internal validity are as follows:

- ***People***: Participating students differ in their skills and their productivity for two reasons: (i) general experience with software development and (ii) experience with BPM technology. The first issue can only be balanced by conducting the experiment with a sufficiently large and representative set of students. The number of 48 students promises to achieve such a balance. The second issue can be mitigated by using BPM tools unknown to every student. Only three of the participating students have rudimental (and thus negligible) workflow knowledge. As we cannot exclude that this knowledge influences our experiment results, we have assigned those three students to different teams in order to minimize potential effects as far as possible.

- ***Data collection process***: Data collection is one of the most critical threats. To mitigate it, we need to continuously control data collection in the context of the experiment. We further have to ensure that students understand which TimeCatcher categories have to be selected during the experiment.

- ***Time for optimizing an implementation***: The specification to be implemented does not include any guideline concerning the number of electronic forms or their layout. This implies the danger that some teams spend more time for implementing a "nice" user interface than others do. To minimize such effects, we explicitly indicate to the students that the development of a "nice" user interface is not a goal of our experiment. To ensure that the implemented solutions are similar across different teams, we accomplish acceptance tests.

Besides, there are threats to the external validity:

- ***Students instead of professionals***: Involving students instead of IT professionals may be critical. However, it has been shown before that the results of student experiments are transferable and can provide valuable insights into an analyzed problem domain [15]. Also note that the use of professional software developers is hardly possible in practice as no profit-oriented organization will simultaneously implement a business process twice using two different BPM technologies.

- ***Investigation of tools instead of concepts***: In our experiment, BPM tools are used as representatives for the analyzed concepts (i.e., workflow management and case handling). Investigating the concepts therefore always depends on the quality of the used tools. To mitigate this risk, the used BPM technologies should be representative for state-of-the-art technologies in practice (which is the case as both selected BPM tools are market leader in their domain).

- ***Choice of object***: To avoid that the chosen business process setting strongly supports the goals of our experiment, we have picked a business process that can be found in many organizations, i.e., the eTravel process (cf. Section 2).

4. PERFORMING THE EXPERIMENT

This section deals with the preparation, execution and analysis of our experiment. This also includes the presentation of experiment results.

4.1 Experiment Preparation

In the run-up of the experiment, we prepare a technical specification of the eTravel process. This specification comprises UML activity diagrams[2], an entity relationship diagram describing the generic data structure of a travel expense report, and system-specific data models for the considered tools (Staffware, FLOWer). In order to ensure that

[2]One may argue that the use of UML activity diagrams can undermine the validity of the experiment as these diagrams are very similar to the explicit, flow-driven notation of Staffware process models, but different from the implicit, more data-driven FLOWer process models. However, in practice, UML activity diagrams are widely used to describe standard business processes. Thus, the use of UML activity diagrams can even improve internal validity as a typical practical scenario is investigated.

the specification is self-explanatory and correct, two student assistants are involved in its development.

Before the experiment, the same two students implement the specification with each of the utilized BPM technologies. This allows us to ensure the feasibility of our general experiment setup and to identify critical issues with respect to the performance of the experiment. This pre-test also provides us with feedback that helps to further improve the comprehensibility of our specification. Finally, we compile a "starter kit" for each participating team. It includes original tool documentation, additional documentation created by us when preparing the experiment (and which can be considered as a compressed summary from the original documentation), and the technical process specification.

4.2 Experiment Execution

Due to infrastructure limitations, we split up the experiment in two events. While the first one took place in October 2006, the second one was conducted in January 2007. Each event lasted 5 days, involved 24 participants (i.e., students), and was based on the following procedure: Prior to the start of the experiment, all students have to attend an introductory lecture. We introduce to them basic notions of workflow management and case handling. We further inform them about the goals and rules of the experiment. Afterwards, each team receives its "starter kit". Then, the students have to implement the given eTravel business process specification (with both considered factor levels). After having implemented the eTravel specification with a factor level, an acceptance test is accomplished by us in order to ensure that the developed solution corresponds to the specification. After finishing their work on the experiment, students have to fill out the aforementioned questionnaire.

We further optimize experiment results by applying *Action Research* [13]. Action Research is characterized by an intensive communication between researchers and subjects. At an early stage, we optimize the data collection process by assisting and guiding the students in using the TimeCatcher data collection tool properly (which is critical with respect to the quality of the gathered data). Besides, we document emotional reactions of the students regarding their way of working. This helps us to design the questionnaire. Note that Action Research does not imply any advice for the students on how to implement the eTravel process.

4.3 Data Analysis Procedure

Data analysis comprises three steps: an initial validation of the collected data (*Step 1*), data analysis itself (*Step 2*), and analysis of the questionnaire results (*Step 3*).

Step 1: Data Validation. We validate the collected data regarding its *consistency* ("Is all expected data available?") and *plausibility* ("Is all available data meaningful?"):

- ***Data Consistency***: We discard the data of two teams as their data is flawed. Both have made mistakes using the TimeCatcher tool. Hence, the data provided by 14 teams is finally included in data analysis.
- ***Data Plausibility***: We analyze data plausibility based on *box-whisker-plot diagrams*. Such diagrams visualize the distribution of a sample and particularly show outliers. A low number of outliers indicates plausible data [12]. Fig. 5A, for example, shows a box-whisker-plot diagram which illustrates the distributions of the base implementation efforts in our experiment. The diagram takes the form of a box that spans the distance between the 25% quantile and the 75% quantile (the so called *interquantile range*) surrounding the median which splits the box into two parts. The "whiskers" are straight lines extending from the ends of the box to the maximum and minimum values. Outliers are defined as data points beyond the interquantile range, i.e., beyond the edge of the box. As can be seen in Fig. 5A, there are no outliers, i.e., all data from these samples lie within the boxed area. Moreover, there exists only one (negligible) outlier in the distribution of the change implementation efforts (cf. Fig. 5B), and no outliers regarding the distribution of the overall implementation efforts (cf. Fig. 5C).

Step 2: Data Analysis. The main goal of the experiment is to investigate whether there is a significant difference between the efforts of implementing a business process with a WfMS and the efforts of an implementation using case handling technology. Hence, the 0-hypothesis to be analyzed is as follows: "*Using workflow technology yields no significant difference in implementation efforts when compared to case handling technology*". We analyze this 0-hypothesis based on a *two-sided t-test* [12] (respectively an additional sign test if the t-test fails). Doing so, we are able to assess whether the means of the WfMS sample and the CHS sample are statistically different from each other. A successful t-test (with $|T| > t_0$) rejects our 0-hypothesis. Specifically, the following steps have to be executed in order to accomplish a t-test (with $\alpha = 0.05$ as the level of significance):

1. ***Paired Comparison***: The t-test is combined with a *paired comparison* [12], i.e., we analyze "pairs of effort values". Each pair comprises one effort value from the WfMS sample and one from the CHS sample. Note that we compose pairs according to the performance of the teams, i.e., effort values of "good" teams are not combined with effort values of "bad" teams (cf. [6]).
2. ***Standardized Comparison Variable***: For each pair, a *standardized comparison variable* X_j is derived. It is calculated by dividing the difference of the two compared effort values by the first one:

$$X_j := \frac{EFFORT_{j+m/2} - EFFORT_j}{EFFORT_{j+m/2}} \cdot 100\%$$

   In other words, X_j denotes how much effort team T_j saves using workflow technology when compared to team $T_{j+m/2}$ which uses case handling technology. Together, all X_j constitute a standardized comparison sample $x = (X_{11}, ..., X_{1m/2})$ used as basis when performing the t-test.
3. ***Statistical Measures***: For the standardized comparison sample x we calculate the *median* (m), the *interquantile range* (IQR), the *expected value* (μ), the *standard deviation* (σ), and the *skewness* (sk).
4. ***Two-sided t-Test***: Finally, we apply the t-test to x. Note that the t-test will be only possible if x emanates a *normal distribution* and if the WfMS and CHS sample have *same variance*. The first condition can be

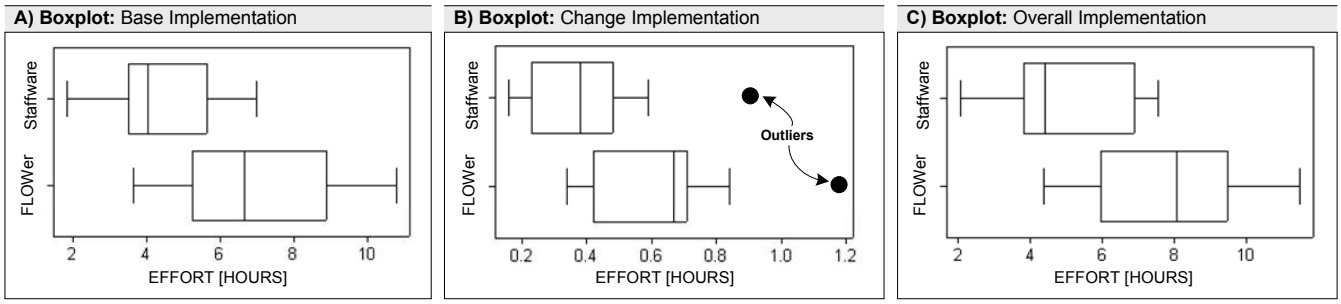


Figure 5: Data Distribution (Box-Whisker-Plot Diagrams).

tested using the *Kolmogoroff/Smirnov test* [16]. In particular, the result of the Kolmogoroff/Smirnov test has to be smaller than K_0 (with K_0 a predefined value depending on the size of x and α). The second condition can be tested based on the *test for identical variance* [16]. The variance of the WfMS and CHS sample will be identical, if the result of this test is smaller than F_0 (with F_0 a predefined value depending on the size of the samples and α). Only one violated precondition is sufficient to avoid the accomplishment of the t-test.

Step 3: Questionnaire Analysis. We analyze the data collected based on the questionnaire each student has to fill out. As one of the students became ill at the last day of the January 2007 event, only 47 students participate.

4.4 Experiment Results

Fig. 6A shows the results for the *overall implementation efforts*. When studying the efforts for the workflow implementation, we can see that they are lower than the efforts for the case handling implementation. This difference is confirmed by the results of the (successful) t-tests for both the first and the second run, i.e., our 0-hypothesis is to be rejected. In the first run, the use of workflow technology has resulted in effort savings of 43.04% (fluctuating between 27.51% and 50.81%) when compared to the efforts for using case handling technology. In the second run, the use of workflow technology has still resulted in savings of 28.29% (fluctuating between 11.48% and 53.16%).

Fig. 6A also shows that efforts for the first run are generally higher than those for the second run. Regardless which technology is used first, all teams reduce their efforts in the second run. This can be explained either through learning effects on the used BPM technologies or an increasing process knowledge gathered during the experiment. Based on questionnaire results (see below), we assume that this effect is not necessarily related to learning effects concerning the used BPM technologies (i.e., tool knowledge), but to increasing process knowledge (which, in turn, reduces the risk of comparing tools instead of concepts).

Fig. 6B and Fig. 6C show results for the *base implementation* and the *change implementation*. Again, our results allow to reject the 0-hypothesis (the failed t-test can be compensated with a successful *sign test*). Using workflow technology results in effort savings of 44.11% for the change implementation in the first run (fluctuating between 16.29% and 56.45%). In the second run, the use of workflow technology results in effort savings of 40.46% when compared case handling efforts.

4.5 Questionnaire Results

Fig. 6D shows that the methodical soundness of using process management technology is easier to understand in the case of workflow technology, i.e., using case handling technology is considered as being more difficult. Fig. 6E illustrates what we have already mentioned above, i.e., process knowledge gained during the first run significantly simplifies the second run. By contrast, Fig. 6F shows that the increased efficiency during the second run cannot be related to a gained tool knowledge. Finally, Fig. 6G deals with the usability of the applied process management systems. As can be seen, there remains a lot of space for improvement from the students' viewpoint.

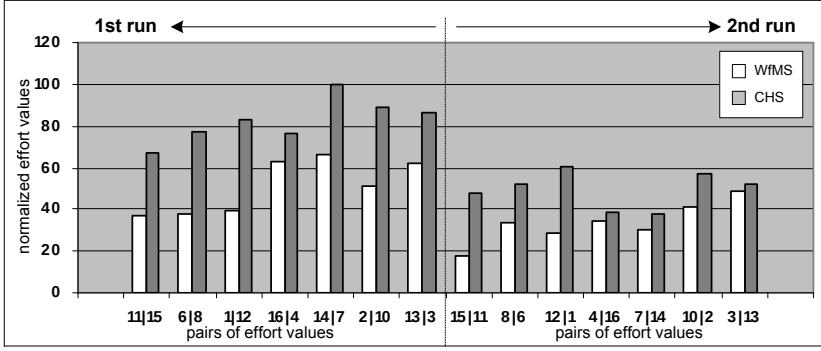
4.6 Discussion

Our results indicate that process implementations based on workflow technology generate lower efforts when compared to implementations based on case handling technology. Moreover, our results show that initial implementations of processes generate significantly higher efforts than subsequent process changes (cf. Fig. 7). This is particularly important for policy makers, who often focus on short-term costs (e.g., for purchasing BPM technology and initially implementing business processes) rather than on long-term benefits (e.g., low costs for realizing process changes).

Finally, our data indicates that increasing knowledge about the processes to be implemented results in increased productivity of software developers. Regardless which BPM technology is used first, all teams reduce their efforts in the second run. Questionnaire results further indicate that this effect is not necessarily related to an increasing knowledge about the used BPM technologies. This also emphasizes the need to involve domain experts with high process knowledge when applying BPM technology.

Considering our experiment design, it is inevitable to acknowledge that our experiment results are influenced by the quality of the used BPM tools. However, by selecting leading commercial BPM tools as representatives for the analyzed concepts (i.e., workflow management and case handling), we can reduce the impact of the tool quality. Yet, based on this single experiment, results cannot be generalized, i.e., substantial conclusions regarding the strengths and weaknesses of workflow management and case handling cannot be derived. For this purpose, additional experiments with different experiment designs and more specific research questions will be necessary. As one example consider the comparison of conventional WfMS, adaptive WfMS, and CHS regarding the effectiveness of realizing process changes.

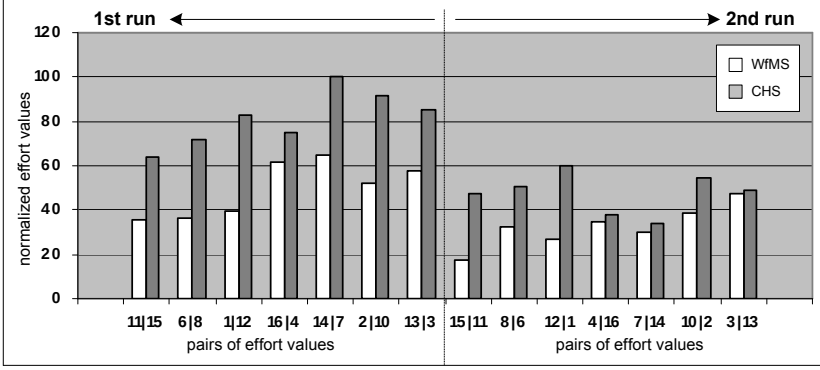
A) EXPERIMENT RESULTS - Paired Comparison (Overall Efforts)



Statistical Data

first run	second run
m = 43.048	m = 28.2913
IQR = [27.51 ; 50.81]	IQR = [11.48 ; 53.16]
μ = 38.6896	μ = 31.2358
σ = 12.7703	σ = 20.6792
sk = -0.6309	sk = 0.4490
K = 0.135 (K_0 = 0.349)	K = 0.129 (K_0 = 0.349)
F = 0.661 (F_0 = 4.284)	F = 0.76 (F_0 = 4.284)
T = -5.059 (t_0 = 2.179)	T = -3.294 (t_0 = 2.179)

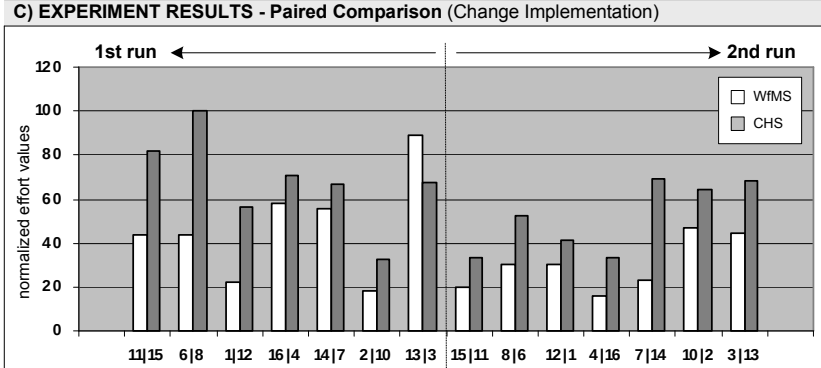
B) EXPERIMENT RESULTS - Paired Comparison (Base Implementation)



Statistical Data

first run	second run
m = 43.0116	m = 28.5209
IQR = [32.03 ; 50.06]	IQR = [8.11 ; 54.90]
μ = 39.2788	μ = 29.3332
σ = 11.9261	σ = 23.5067
sk = -09141	sk = 0.4401
K = 0.138 (K_0 = 0,349)	K = 0.198 (K_0 = 0,349)
F = 0.972 (F_0 = 4,284)	F = 0.895 (F_0 = 4,284)
T = -4.816 (t_0 = 2,179)	T = -3.024 (t_0 = 2,179)

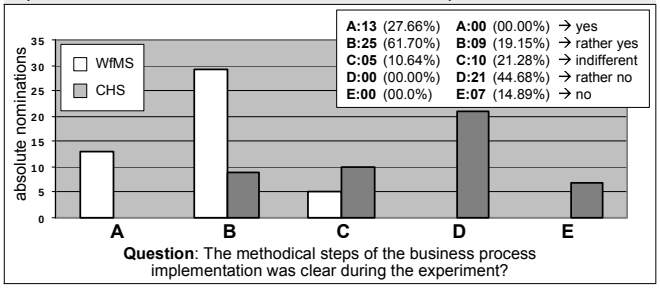
C) EXPERIMENT RESULTS - Paired Comparison (Change Implementation)



Statistical Data

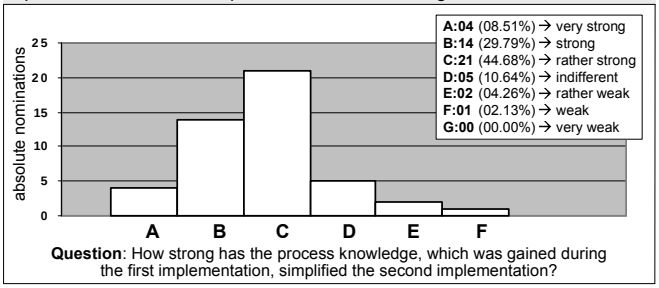
first run	second run
m = 44.1152	m = 40.4666
IQR = [16.29 ; 56.45]	IQR = [26,63 ; 52,20]
μ = 29.9807	μ = 41.4368
σ = 32.4034	σ = 14.4722
sk = -1.3034	sk = 0.8501
K = 0.172 (K_0 = 0.349)	K = 0.198 (K_0 = 0.349)
F = 0.752 (F_0 = 4.284)	F = 1.784 (F_0 = 4.284)
T = -1.724 (t_0 = 2.179)	T = -2.884 (t_0 = 2.179)
→ **t-test failed**	

D) QUESTIONNAIRE - Methodical Soundness of Implementation



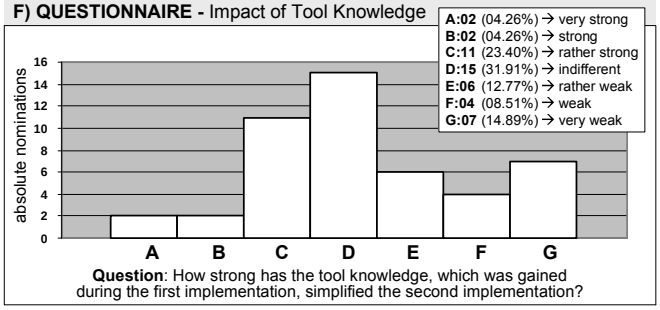
Question: The methodical steps of the business process implementation was clear during the experiment?

E) QUESTIONNAIRE - Impact of Process Knowledge



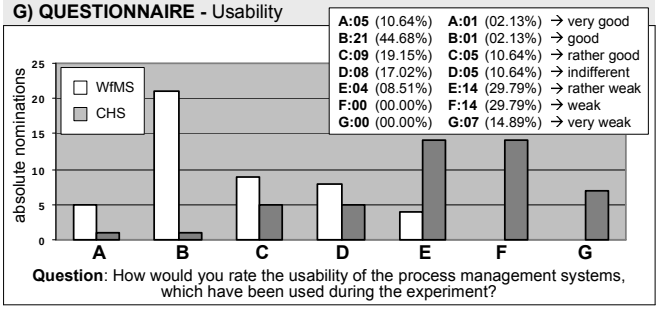
Question: How strong has the process knowledge, which was gained during the first implementation, simplified the second implementation?

F) QUESTIONNAIRE - Impact of Tool Knowledge



Question: How strong has the tool knowledge, which was gained during the first implementation, simplified the second implementation?

G) QUESTIONNAIRE - Usability



Question: How would you rate the usability of the process management systems, which have been used during the experiment?

Figure 6: Results of our Experiment.

We apply these experiment results in the EcoPOST project [9]. This project aims at the development of an approach to investigate complex causal dependencies and related cost effects in PAIS engineering projects. In particular, our results enable us to quantify causal dependencies in PAIS engineering projects. As an example consider the impact of process knowledge on the productivity of process implementation.

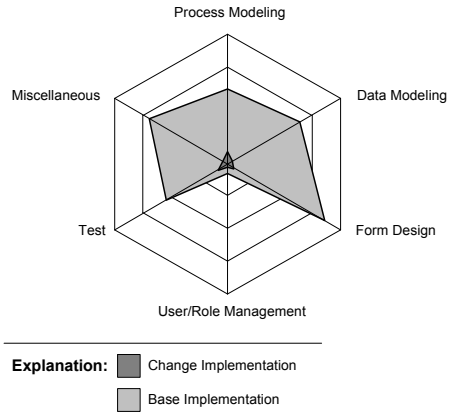


Figure 7: Base versus Change Implementation.

5. RELATED WORK

The most similar experiment design when compared to our own is provided by [6] which investigates the impact of workflow technology on software development and software maintenance. Generally, only few data is available on the effects of workflow technology (regarding case handling, no data is available at all). Oba et al. [11], for example, analyze the introduction of WfMS and particularly focus on the identification of factors influencing work efficiency, processing time, and business process standardization. A mathematical model is provided for predicting the reduction rate of processing times. An extension of this work is [14] where simulation is used to compare pre- and post-implementations of information systems relying on workflow technology. Focus of this work is on analyzing process performance based on criteria such as lead time, waiting time, service time, and utilization of resources.

6. SUMMARY

This paper presents the results of a controlled BPM software experiment with 48 students. Our results indicate that business process implementation based on workflow technology generates lower efforts than using case handling technology. Thereby, initial process implementations result in higher efforts than subsequent process changes. Our data can help enterprises – which crave for quantitative data completing their qualitative decision criteria – to better understand the efforts of using BPM technology.

7. REFERENCES

[1] Y. L. Antonucci. *Using Workflow Technologies to improve Organizational Competitiveness*. Int'l. J. of Management, 14(1), pp.117-126, 1997.

[2] P. Athena. *Case Handling with FLOWer: Beyond Workflow*. 2002.

[3] V. R. Basili, R. W. Selby, and D. H. Hutchens. *Experimentation in Software Engineering*. IEEE Trans. in SW Engin., 12(7), pp.733-743, 1986.

[4] M. Dumas, W. M. P. van der Aalst, and A. ter Hofstede. *Process-aware IS*. Wiley, 2005.

[5] N. Juristo and A. M. Moreno. *Basics of Software Engineering Experimentation*. 2001.

[6] N. Kleiner. *Can Business Process Changes Be Cheaper Implemented with Workflow-Management-Systems?* IRMA 2004, pp.529-532.

[7] C. M. Lott and H. D. Rombach. *Repeatable Software Engineering Experiments for Comparing Defect-Detection Techniques*. Empirical Software Engineering, 1(3), pp. 241-277, 1996.

[8] B. Mutschler, M. Reichert, and J. Bumiller. *Unleashing the Effectiveness of Process-oriented Information Systems: Problem Analysis, Critical Success Factors, Implications*. IEEE Transactions on Systems, Man, and Cybernetics - Part C: Application and Reviews, 2008 (accepted for publication).

[9] B. Mutschler, M. Reichert, and S. Rinderle. *Analyzing the Dynamic Cost Factors of Process-aware IS: A Model-based Approach*. CAiSE 2007.

[10] G. J. Myers. *A controlled Experiment in Program Testing and Code Walkthroughs/Inspections*. Comm. of the ACM, 21(9), pp. 760-768., 1978.

[11] M. Oba, S. Onoda, and N. Komoda. *Evaluating the Quantitative Effects of Workflow Systems based on Real Cases*. HICSS 2000.

[12] L. Prechelt. *Controlled Experiments in Software Engineering (in German)*. Springer, 2001.

[13] P. Reason and H. Bradbury. *Handbook of Action Research*. 2001.

[14] H. A. Reijers and W. M. P. van der Aalst. *The Effectiveness of Workflow Management Systems - Predictions and Lessons Learned*. Int'l. J. of Inf. Manag., 25(5), pp.457-471, 2005.

[15] P. Runeson. *Using Students as Experiment Subjects*. EASE 2003.

[16] D. J. Sheskin. *Handbook of Parametric and Nonparametric Statistical Procedures*. 2000.

[17] D. I. K. Sjoberg, J. E. Hannay, O. Hansen, V. B. Kampenes, A. Karahasanovic, N.-K. Liborg, and A. C. Rekdal. *A Survey of Controlled Experiments in Software Engineering*. IEEE Trans. in SW Engin., 31(9), pp.733-753, 2005.

[18] Tibco. *Staffware Process Suite*. User Manual, 2005.

[19] W. van der Aalst, A. ter Hofstede, B. Kiepuszewski, and A. Barros. *Workflow Patterns*. Distributed and Parallel Databases, 14(3), pp.5-51, 2003.

[20] W. M. P. van der Aalst and K. van Hee. *Workflow Management*. MIT Press, 2004.

[21] W. M. P. van der Aalst, M. Weske, and D. Grunbauer. *Case Handling: A New Paradigm for Business Process Support*. DKE, 53(2), pp.129-162, 2005.

[22] B. Weber, S. Rinderle, and M. Reichert. Change patterns and change support features in process-aware is. CAiSE 2007.

[23] M. V. Zelkowitz and D. R. Wallace. *Experimental Models for Validating Technology*. IEEE Computer, 31(5), pp.23-31, 1998.