Made available by Hasselt University Library in https://documentserver.uhasselt.be

The influence of business process representation on performance of different task types

Peer-reviewed author version

Ritchi, Hamzah; JANS, Mieke; Mendling, Jan & Reijers, Hajo (2020) The influence of business process representation on performance of different task types. In: Journal of information systems, 34 (1), p. 167-194.

DOI: 10.2308/isys-52385
Handle: http://hdl.handle.net/1942/28115

The influence of business process representation on performance of different task types

Hamzah Ritchi
Padjadjaran University, Jl. Dipati Ukur No.35,Bandung 40132, Indonesia,
hamzah.ritchi@unpad.ac.id

Mieke Jans
Hasselt University, Martelarenlaan 42, 3500 Hasselt, Belgium,
mieke.jans@uhasselt.be

Jan Mendling
Wirtschaftsuniversität Wien, Welthandelsplatz 1, 1020 Wien, Austria,
jan.mendling@wu.ac.at (Corresponding Author)

Hajo Reijers
Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands,
h.a.reijers@vu.nl

The influence of business process representation on performance of different task types

Abstract

The analysis of business processes is an integral part of risk assessment procedures and audit methodology. In both auditing research and process modeling research, there is an ongoing debate on which representation format might be best suited to support the analysis task. Most important in this context is the question whether business process models as visual representation might be superior to textual narratives. This paper investigates the affinity of different tasks with two process representational formats: textual narratives and visual diagrams (BPMN process models). Our findings demonstrate that the representation format has an impact on task performance and that the direction of this impact depends upon the affinity of the tasks type with the representation format. This implies that auditors are best provided with different process representations, depending on the task they are performing at that moment. These findings have important implications for research on auditing tasks, and more broadly also for software engineering and information systems research.

Keywords: Visual representation format, textual representation format, BPMN, process model, task type

INTRODUCTION

Business process analysis using models plays an important role for improving time, cost and quality of business operations (Dumas et al. 2013). Business process models are not only useful for performance analysis, but also for assessing the effectiveness of internal controls in financial reporting (Boritz, Borthick, and Presslee 2012). While already Bell et al. (1997) emphasize their potential for auditing, it was the Sarbanes-Oxley Act of 2002 that triggered a broader uptake. Likewise, the ISA 315 of the International Auditing and Assurance Standards Board (IAASB) and the Auditing Standard No. 12 of Public Company Accounting Oversight Board (PCAOB) explicitly state the need to analyze business processes for auditing purposes. The recent uptake of business process models is reflected by various contributions to audit on investigating the impact of a process focus (O'Donnell and Schultz Jr 2003), combining strategic and process-level analysis for audit (Ballou, Earley, and Rich 2004), comparing modelling techniques (Carnaghan 2006) and their effectiveness for auditing (Bradford, Richtermeyer, and Roberts 2007) and for testing internal controls (Bierstaker, Hunton, and Thibodeau 2009).

Research in Accounting Information Systems (AIS) has long recognized the importance of information representation (Kelton, Pennington, and Tuttle 2010, Dunn, Gerard, and Grabski 2017). The representation of business processes plays an important role in this context, since auditors have the choice between various types of textual and visual representations (Boritz, Borthick, and Presslee 2012). Business processes can be described using text or using visual models like the flowchart standard Business Process Model and Notation (BPMN). This raises the question which type of representation should be preferred. For instance, Carnaghan (2006) states that visual representations have some advantages over text, but a combination of both visual and textual representations might be superior to either approach alone. Some studies in AIS researched this question using experimental designs. Dunn and Gerard (2001) investigated the effect of the representation format of database schemas on three types of tasks (search, recognition, and inference tasks). Their results indicated that participants find the visual representation easier to use, used it faster, and they were more satisfied with it (no impact on accuracy). Boritz, Borthick, and Presslee (2012) compared the effect of providing either a visual or a textual representation of a business process on the accuracy and efficiency of students' risk assessment. They found that the textual representation resulted in a higher efficiency (no impact on accuracy), which seems to contradict the findings of Dunn and Gerard (2001) at a first glance. Indeed, also studies in software engineering on representation formats have provided partially inconclusive results (Whitley 1997).

Boritz, Borthick, and Presslee (2012) emphasize that the instrumentation of a different task might lead to different results regarding the mutual benefits of visual and textual representations. This paper builds on this observation and systematically investigates the affinity of the representation format with different task types. Our theoretical argument is grounded in cognitive fit theory, emphasizing that a representation should be less assessed from the perspective of being *better* than another in general terms, but rather providing a better performance *relative* to a particular type of task. In recent work in the accounting domain, also Dunn, Gerard, and Grabski (2017) experimented with representations that had a high or low fit with the task to be performed. Following this line of thought, we identify tasks from prior research as having *visual or textual affinity*. This characteristic complements the common distinction of schema-based and non-schema-based task types, with the former being tasks that can be completed with information of the presented schema; the latter being tasks where the user is required to go beyond the information that is provided in the schema, requiring a deeper level of understanding (Gemino and Wand 2003, Khatri et al. 2006). To test our hypothesis of visual and textual task affinity, we conducted an experiment on the fit between the types of process representation on the one hand with four different task types on the other hand. Our results highlight that there is no format that is generally superior, as we anticipated. For search and recognition and partially for inference tasks, visual business process models appear to be better suited than textual models (in terms of efficiency and effectiveness, both for experts and for novices). For recall tasks (for experts and novices), and partially for problem-solving tasks (only for the experts), textual models provide better results than visual models. These findings support our theoretical concept of affinity.

The remainder of the paper is organized as follows. The next section summarizes prior research and the theoretical background upon which we define our hypotheses. Then, we present an experimental design as our research method. The subsequent section presents the results of our experiment. After that, we discuss the results and their implications for research and practice. We conclude with a summary and reflections on limitations.

RELATED LITERATURE AND HYPOTHESIS DEVELOPMENT

Cognitive theory is a solid foundation for hypothesizing the mutual strengths and weaknesses of visual and textual representations of business processes. More specifically, this study builds on the theory of cognitive fit, which we will shortly introduce hereafter. Further, the dimensions of both task type and process representations will be elaborated on to distill our hypotheses from.

Cognitive fit

The debate on the mutual strengths and weaknesses of different representation formats dates back to the 1970s when proponents of procedural and declarative computer programming argued for either paradigm being superior. It were Gilmore and Green (1984) who provided the compelling argument that general superiority is not the right perspective, but that different types of tasks might benefit more or less from different representations of program logic. Larkin and Simon (1987) formulated the question whether a visual diagram is worth 10,000 words. They found empirical support for their argument that representations can be informationally equivalent—they provide the same information—while not being computationally equivalent—the information is easier to access from a particular type of representation. Consider a business process represented as a visual BPMN model and an informationally equivalent text. In the model, we can very quickly spot how many activities the process contains. This information is also provided in the text, but we have to read it sequentially from start to end. Apparently, the model is more computationally efficient for identifying the number of activities.

Identifying the number of activities is only one type of task with very specific characteristics. The theory of cognitive fit (Vessey 1991) further develops the observation that tasks are diverse (Kelton, Pennington, and Tuttle 2010). More specifically, it posits that problem-solving performance depends upon the internal mental representation of the problem, which itself depends upon the external representation format of the problem (Vessey 1991). A fit between the problem representation and the task at hand improves the task performance. An extended version of this theory integrates the interaction of the problem-solving task and the mental representation of the task solution (Shaft and Vessey 2006). According to cognitive fit theory, a mismatch between the external representation and the problem-solving task leads to additional mental effort in reorganizing the information to meet the requirements of the task. Cognitive fit theory can be reformulated from the perspective of tasks. Indeed, Vessey and Galletta (1991) find that symbolic (text, numeric) tasks are better supported by tables, and spatial or visualization tasks are better supported by graphs.

The importance of cognitive theory is recognized in various auditing studies. Bierstaker, Bedard, and Biggs (1999), for instance, find that a shift in the mental problem representation leads auditors to better insights into possible explanations of observed discrepancies. Dunn and Grabski (2001) were the first to integrate the concept of localization (Larkin and Simon 1987) with the theory of cognitive fit in the domain of accounting models: the more localized the information that is needed to complete a given task, the better the performance. Dunn, Gerard, and Grabski (2017) investigated the interaction and the effect of users' schemas (debit-credit-accounting and Resource-Event-Agent) and cognitive fit on task performance. We expand upon these works and approach the research problem from the perspective of the respective type of cognitive task.

Task types

The performance of completing certain tasks does not only depend upon their type. Libby and Luft (1993) emphasize that accounting tasks depend upon experience and knowledge as distinct theoretical constructs. The authors emphasize the need to distinguish different task types when researching performance. Kelton, Pennington, and Tuttle (2010) identify task type, task complexity, and task environment as factors influencing cognitive fit. In this context, task types are originally categorized as spatial or symbolic by Vessey (1991). Task complexity covers a variety of aspects including the number of elements, their relationships, and their ability to change over time. Task environment has been related to time constraints, task demands, practice, and task interruptions (Kelton, Pennington, and Tuttle 2010).

-------- Insert Table 1 – Overview of studies on the benefits of visual versus textual representations showing the inconsistencies on this topic -----

Table 1 shows that various studies investigate the benefits of visual and textual representations with inconclusive results. Some of these works are summarized in a comprehensive survey by Whitley (1997). Most of the mentioned studies focus on one specific type of task, in particular the older ones. Although cognitive fit theory is acknowledged in the more recent studies, hardly any of them provides insights into the effects of considering different task types. Among the latter are Dunn and Gerard (2001) who examined the difference between diagrammatic and linguistic conceptual model representations. They utilize three different types of tasks: search, recognition, and inference. In this regard, their study is unique even though all tasks yield consistent results. Later studies by Jones, Tsay, and Griggs (2002), Boritz, Borthick, and Presslee (2012), and Boritz, Carnaghan, and Alencar (2014) study representation formats, but do not focus on task types. However, they speculate that the cognitive fit between the type of task and the representation format is presumably important to understand when and why a certain format might be superior. Boritz, Borthick, and Presslee (2012) emphasize that the “use of a different task could lead to different results”.

The three types of tasks that Dunn and Gerard (2001) use are based on Larkin and Simon (1987), who describe problem solving as a cognitive process that involves search, recognition, and inference. Larkin and Simon (1987) suggest that search and recognition time will increase when a textual representation format is used, because it forces search into a sequential processing mode following the flow of the text. They also expect that inference tasks are less susceptible to representation effects, because they build on semantic processing once search and recognition are completed. This positive relationship between visual representations and efficiency was confirmed in the study of Dunn and Gerard (2001). In terms of accuracy, Larkin and Simon (1987) expected no difference between two informationally equivalent representations. This was also confirmed in the study of Dunn and Gerard (2001), although hypothesized differently by the authors. One of the possible explanations of this unexpected outcome is the set of task categories in the conducted experiment. Note that also performance measures differ (Gemino and Wand 2003). Here, we will focus on efficiency in terms of task duration and effectiveness in terms of accuracy of the task output.

Prior literature distinguishes different categories of tasks. Several of them are generic, while others are domain-specific. The three *generic* task types search, recognition, and inference of Larkin and Simon (1987) all require explicitly represented information to complete them, and nothing further. We call these generic tasks *schema-based* tasks. Some tasks require information that goes beyond directly represented information (Gemino and Wand 2003). We call these tasks *non schema-based* tasks. Agarwal, Sinha, and Tanniru (1996) classify tasks as *object-oriented* versus *process-oriented*. Bodart et al. (2001) link the type of tasks to the *level of understanding* that is required of the user: a *surface-level* or a *deep-level* understanding of the domain. Beyond these generic task types, there are also taxonomies of *domain-specific tasks* in the context of auditing. A comprehensive list of 332 audit tasks is identified in a survey by Abdolmohammadi (1999), organized in the four phases of orientation, control structure, substantive tests, and forming an opinion. Many of these tasks such as, e.g., OR41 (assessment of complexity of transaction flows), CS7 (evaluation of procedures to safeguard assets, TC5 (determination of adequacy of procedures to reconcile records), and ST41 (tracing sales from sales register to shipping records) are concerned with business processes. In practice, these and other audit tasks are supported by textual and by visual representations of the underlying business processes. In the following, we investigate *generic* task types and representational effects on task performance in the context of auditing. We will build on general insights from cognitive fit theory for theorizing.

Hypotheses

We observe that previous research has provided partially inconclusive results on the mutual strengths and weaknesses of visual and textual representations. For instance, while Dunn and Gerard (2001) observe faster problem solving supported by a visual representation of a database schema, Boritz, Borthick, and Presslee (2012) find students to be faster in solving risk assessment problems with a textual business process representation. The studies shown in Table 1 are difficult to reconcile a posteriori, because they use different types of tasks supported by different types of information. In our study we make a distinction between task types based on two dimensions: schema-based versus not schema-based and surface-level understanding versus deep-level understanding. In Figure 1 we visualize these dimensions, along with the task types we selected to operationalize these classes of task types.

	Surface-level understanding	Deep-level understanding
Schema-based	Strong visual affinity: Search and recognition	Weak visual affinity: Inference
Not schema-based	Strong textual affinity: Recall	Weak textual affinity: Problem solving

Figure 1- Classification of task types along two dimensions

What we believe is crucial for interpreting the diverging results in previous studies is to assess the *affinity* of the task at hand with the representation of the informational material that is provided. This notion of affinity is grounded in cognitive fit theory and further develops the dichotomy of spatial versus symbolic tasks by Vessey (1991). Most importantly, affinity defines a spectrum and not a dichotomy.

Strong Visual Affinity Tasks

We call those tasks that have a strong affinity with visual representation formats *strong visual affinity tasks*. In an audit context, this type of task is relevant when an auditor for example reads a business process diagram and identifies the different stakeholders involved in a process. Tasks of *strong visual affinity* are characterized by the fact that they strongly benefit from multidirectional spatial search capabilities of humans and corresponding visual cognition for a given visual representation of information. The provision of the equivalent information using a textual representation typically leads to a decline in task performance, because sequential search in text and symbolic processing is slow. Tasks from prior literature that are likely to be found in this category are those that are schema-based and require surface understanding. The search and recognition tasks defined by Dunn and Gerard (2001) fulfil these criteria. They can also be described as spatial tasks according to Vessey (1991), since they requiring finding the right piece of information in a visual representation. The processing of a visual representation will be faster due to the visual cognition. This will be accompanied with a more efficient utilization of cognitive resources, and as a consequence, less errors and better results. For these reasons, we formulate the following hypotheses.

H1.a: For search and recognition tasks, a visual representation is more *efficient* than a textual representation.

H1.b: For search and recognition tasks, a visual representation is more *effective* than a textual representation.

Weak Visual Affinity Tasks

We refer to those tasks that have some affinity with visual representation formats *weak visual affinity tasks*. In an audit context, this type of task is relevant when an auditor for example reads a business process diagram and identifies the paths that lead to the cancellation of an order. Tasks of *weak visual affinity* are characterized by the fact that they partially benefit from multidirectional spatial search capabilities of humans, but these are not sufficient for a successful task completion alone. The reason is that the provided information also needs to be integrated with prior knowledge. The provision of the equivalent information using a textual representation has a negative effect on visual cognition, but does not fully affect the semantic processing and information integration. Tasks from prior literature that are likely to be found in this category are those that are schema-based, but require deep understanding of a specific problem domain (Bodart et al. 2001). The inference tasks used by Dunn and Grabski (2001), Reijers and Mendling (2011) or Trkman, Mendling, and Krisper (2016) meet these criteria. The processing supported by a visual representation will be faster due to the advantages of human visual cognition. This will be accompanied with a more efficient utilization of cognitive resources, and as a consequence, less errors and better results. For these reasons, we formulate the following hypotheses.

H2.a: For inference tasks, a visual representation is more *efficient* than a textual representation.

H2.b: For inference tasks, a visual representation is more *effective* than a textual representation.

Weak Textual Affinity Tasks

Those tasks that have some affinity with textual representation formats are referred to as *weak textual affinity tasks* here. In an audit context, this type of task is relevant when an auditor reads documents that describe policies and work instructions, and then identifies the potential risks and suitable controls. Tasks of *weak textual affinity* have in common that they partially benefit from a sequential processing of information, while this not being sufficient for a successful completion of the task alone. These tasks also require a semantic processing of the information and an integration with prior knowledge. They cannot be fully solved based on the provided schema (Gemino and Wand 2004).

Newell and Simon (1972) and Simon and Newell (1971) also refer to them as problem-solving tasks that require both searching and understanding. The absence of prior knowledge typically results in a linear search down the entire text, although experts often make use of selective search by choosing good moves and then conducting a limited search to test the choice of moves (Chi, Glaser, and Rees 1981). Tasks from prior literature that are likely to be found in this category are those that are not schema-based and require a deep understanding of the problem domain. Textual representations might benefit from an easier establishment of associative connections between words with the corresponding entities within the verbal cognitive system. Visual representations require an additional establishment of referential connections between words and visual elements and vice versa (Paivio 1991). Therefore, we might see a degrading performance for visual representations (due to the split-attention effect), when the working memory is overwhelmed by constructing these connections (Sweller 1988, Chandler and Sweller 1991). The problem-solving tasks used by Boritz, Borthick, and Presslee (2012) meet these criteria. For reasons of split attention, cognitive processing using a textual representation will be faster. Since less cognitive resources are used, also the results will be better.

> **H3.a:** For problem-solving tasks, a textual representation is more *efficient* than a visual representation.
>
> **H3.b:** For problem-solving tasks, a textual representation is more *effective* than a visual representation.

Strong Textual Affinity Tasks

We call those tasks that have a strong affinity with textual representation formats *strong textual affinity tasks*. In an audit context, this type of task is relevant when an auditor has to memorize policies and procedures, and then cross-checking their interaction with information on the effectiveness of internal controls. Tasks of *strong textual affinity* generally benefit from a sequential processing of information, potentially only requiring a surface understanding. Typically they are also not schema-based and might partially benefit from prior knowledge and semantic processing. Tasks from prior literature that are likely to be found in this category are fill-in-blanks (also called Cloze) recall tasks (Gemino and Wand 2004). Textual representations might more easily facilitate the establishment of associative connections between words of the input and the corresponding entities of the verbal cognitive system (Schiefele and Krapp 1996). Also a sequential access of memorized textual information might be helpful for these tasks (Schiefele and Krapp 1996). Similar to the split-attention effect (Chandler and Sweller 1991, Sweller 1988), it will likely be a burden to cognitively retrieve and reformulate a visual representation into a text for working on a recall task. For these reasons, this type of task should benefit more from a textual representation in terms of faster and better answers. Accordingly, we formulate the following hypotheses.

H4.a: For recall tasks, a textual representation is more *efficient* than a visual representation.

H4.b: For recall tasks, a textual representation is more *effective* than a visual representation.

METHOD

We tested our hypotheses using an experimental research design. More specifically, we use a between-subjects experiment in which participants worked on two cases with four task types for each case (so each participant solved eight tasks in total). The tasks related to risk assessments and internal control analysis with one case being supported by a *visual* business process representation and the other with a *textual* business process representation. This study employs a 2 x 2 factorial design with two categorical independent factors: the process representation format with two levels (BPMN and Text) and the expertise in process analysis with two levels (Expert and Novice). The effects of two other factors were controlled for: task type and task complexity. Our hypotheses and the corresponding statistical analysis compares the effects of expertise and representation on performance for each task type and for each of the two cases separately.

Participants

For our experimental study, we were specifically interested in recruiting participants who can be characterized as experts or novices. To this end, we implemented an interactive website and distributed its link. To recruit experts, we approached practitioner networks that have a focus on business process analysis and risk assessment on LinkedIn and Xing. Experts were incentivized by the offer to take part in a draw to win one of three business process management textbooks and to receive the final report of the research findings. To recruit novices, we collaborated with colleagues from universities in Australia, Austria, Germany, Indonesia and the Netherlands who advertised it to students who had completed courses on business process analysis, audit risk assessment or accounting information system. In addition to obtaining a report and participation in the book draw, novices could obtain bonus points for their final grades. With this procedure, we recruited 345 participants who completed the experiment on the website. Data cleansing is specifically important for internet-based research (Birnbaum 2000, Rodgers et al. 2003). Checking data consistency (Rodgers et al. 2003), minimal duration and statistical outliers (Verardi and Dehon 2010) led to the exclusion of spurious data points. We excluded outliers below 30 minutes and greater than 120 minutes duration. Also signs of interruption were used to reduce the number of data points. Finally, we used robust Mahalanobis distance (Verardi and Dehon 2010), which is the distance of all variables from the centroid in a multivariate space to exclude outliers. All these procedures reduced the data set to 167 participants, with 83 being experts and 84 being novices.

These 167 participants stem from different countries: 68% were from Asian countries and Australia, while 32% were from European (mostly Germany and the Netherlands) countries, Brazil and North America. Among the Asian respondents, Indonesia made up the majority (63%), in which 60% of them were third and fourth year undergraduate accounting and information system students. These Indonesian students received the material in Indonesian while all other participants worked with the material in English. The gender distribution was fairly even across all educational levels, with men accounting for 51% and women for 49% of the data points. Participants had to report whether they consider themselves as novices or experts with regards to the focus of the experiment. This self-report was highly consistent with them having a bachelor degree or not, which provides prima facie validity to the self-assessment.

Table 3 summarizes the demographic statistics of the participants. Participants were randomly assigned to the experimental tasks. The assignment to the expert or novice group was based on the self-assessment of the participant. It can be observed that novices appear to cluster in the undergraduate cohort. Both the novice-expert distribution and the prior working areas show only marginal differences between the text and BPMN treatment groups.

-------- Insert Table 3- Demographic statistics participants------

Experimental Tasks

The experimental tasks refer to two separate cases of typical business processes: a Goods receipt handling process (GHP) and a Procure-to-pay process (PPP). The GHP case was partially adopted from a study that compared two informationally equivalent process models (Recker and Dreiling 2011). The PPP case was an adaptation of the case developed by Borthick, Schneider, and Vance (2012). It is a typical business process found in an organization's accounting cycle, which starts from the purchase of vendors' goods, until the clearing of account payables. Adhering to Larkin and Simon's (1987) proposition, we had to ensure informational equivalence of the alternative representation formats in order to provide a fair comparison of studying computational advantages. Therefore, we first created a BPMN diagram for each case and then constructed a corresponding text according to the formally defined transformation rules of Leopold, Mendling, and Polyvyanyy (2014). Furthermore, we used a manipulation check which confirmed that there was no statistical difference between the number of activities identified by the participants (t-test, $p=0.38$ for GHP, $p=0.15$ for PPP). The PPP case is slightly more complex than the GHP, because it had more activities, organizational units and business objects. For the GHP case, the total number of elements (pools, lanes, events, gateways and activities) in the BPMN model is 24. The text has a total of 136 words. For the PPP case, the BPMN model contains 31 elements, whereas the PPP text is expressed in 255 words. Subjects were randomly assigned and counter balanced. The complete experimental material is in the appendix.

For each case, the participants had to work on four types of tasks, each corresponding to one of the hypotheses.

- *Search and recognition tasks:* First, participants worked on a set of schema-based search and recognition tasks with multiple-choice answers adapted from process model comprehension experiments (Mendling, Strembeck, and Recker 2012). Questions in the search part included "What is the next task after setting invoice as payable?" These questions required participants to scan or locate a piece of information. Recognition questions were formulated in an open way such as "List *only* six activities which are observable …. when a delivery note without purchase order is approved *from the point when* the delivery is identified until the goods are placed in stock". Here, participants have to recognize the sequence of activities in the process description.
- *Inference tasks:* Second, we used schema-based inference tasks adapted from Recker and Dreiling (2007a). Participants were required to integrate prior knowledge with the process description to provide the right answer to a multiple-choice question as "Yes", "No", and "Unknown". There were questions for which "Unknown" was the correct answer when essential information to answer the question was missing in the material.
- *Problem-solving tasks:* Third, the participants had to work on a problem-solving task adapted from Mayer (2005). The participants did not have the case description available anymore, which means that they had to rely on their cognitive model they developed while working on the schema-based tasks. The problem-solving task was a risk assessment of the specific case, for which the participants had to identify as many potential risks that have to be considered for the case.
- *Recall tasks:* Fourth, we used a fill-in-the-blank (Cloze) test as proposed by Gemino and Wand (2004). The Cloze test asked participants to fill in the missing words in the text representation of the process description. The Cloze test is suitable to test recall in long-term memory. Participants were also allowed to enter synonyms of the missing original word. Respectively, 10 and 17 words were left blank for the GHP case and the PPP case.

The same sequence of tasks was presented afterwards for the second case (in the other representation format than the first case was presented). Finally, a post test was used to assess cognitive effort. It checked the participants' perceived ease of understanding (PEOU) for alternative process representations. The measurement was adopted from the scale initially developed by Moore and Benbasat (1991). At the end of the experiment, participants could express their interest in the final report and the book draw.

Dependent Variables

Answers to the tasks allowed us to calculate different dependent variables as summarized in Table 2. We operationalized *effectiveness* as the overall score of given correct answers on a specific type of task. We operationalized *efficiency* as the time elapsed for working on a specific type of task. Duration was automatically measured by the web system based on the answer time of the participants.

-------- Insert Table 2- Dependent variables ------

The search and recognition tasks and inference tasks were automatically evaluated using the correct multiple-choice options. A total score for *search and recognition tasks* was calculated as the sum of two parts: multiple-choice questions and one question about the sequence pattern. As there were eight closed questions, it ranges from 0 – 8 points. The sequence question required the participants to recognize a sequence of six activities that match a specific condition. This question used binary scores based on the following consideration: 1) whether the listed activity in fact existed in the relevant area (identification) and 2) whether an activity was placed in the correct sequence (quality of sequencing). This makes a maximum of two points for each of the six activities plus eight such that the overall range of the variable is 0 – 20 points. The *inference task* score was calculated based on ten closed questions, so it ranges from 0 – 10 points. To score the *problem-solving task*, we use the procedure by Recker and Dreiling (2007b) and Bodart et al. (2001) who distinguish three types of answers: (a) the number of plausible answers based on information inferable from the model, (b) the number of plausible answers that showed knowledge beyond the information provided in the model, and (c) the number of implausible or missing answers. We also followed Mayer (1989) and created a set of acceptable responses for problem solving. The total score was created by summing up the number of acceptable answers across two problem-solving questions in each case. Two independent raters using the predefined scoring scheme evaluated each response. The number of acceptable responses ranged between 5 and 25. The *Cloze score* was evaluated based on the correct words or synonyms filled in by participants. This yielded a total maximum score for the Cloze test of 10 and 17, respectively.

RESULTS

This section presents the results of the statistical analysis. We first provide descriptive statistics and then the results of the MANOVA analysis.

Descriptive Statistics

We report the descriptive statistics of the effectiveness and efficiency variables, captured by the number of correct answers and the duration (in minutes) of working on the corresponding tasks, respectively. Table 4 to Table 7 summarize the means, standard deviations and numbers of observations. In general, the expert cohort appears to outperform the novice cohort. For some tasks, the BPMN treatment yields the better performance, namely for tests related to search and recognition and for inference tasks, while the text treatment mostly leads to better results for problem-solving and recall tasks. We follow up on these observations below in the section on statistical hypothesis tests using MANCOVA.

-------- Insert Table 4 to Table 7----------------------

We also tested whether the treatment groups were balanced as validity checks. There were no demographic differences between the BPMN-first and text-first cohorts. Regarding experts and novices, expected differences in terms of familiarity, intensity, competency and confidence are observed (Table 8). Furthermore, we checked for potential learning and fatigue effects (Table 9). It seems that there might be a learning effect in terms of efficiency. The cohort that first received the PPP case took significantly more time for the four types of tasks on that case than the cohort that first solved the other case. The other way around, for two of the task types, the PPP-first cohort was significantly faster on the GHP case. In terms of effectiveness, however, no indications of learning or fatigue effects could be observed.

-------- Insert Table 8- Validity checks between treatments ------

-------- Insert Table 9- Validity checks between cases, where 'Score' measures effectiveness and 'Time' efficiency------

MANCOVA results

Before conducting the MANCOVA analysis, we checked potential *multicollinearity.* Lawley's χ^2 (119, 465.73) and Bartlett test of sphericity's χ^2 (120, 668.32) were both significant at $p<0.001$. This indicates a low degree of multicollinearity, such that the power of main analysis of MANOVA will not be compromised. Regarding *normal distribution*, we observe the Shapiro-Wilk's *z*-scores ranged between $0.05 < p < 0.95$. Omnibus Doornik-Hansen's multivariate p-values for χ^2 however indicate that most pooled dependent variables across four cells follow normal distribution ($0.12 < p < 0.69$). This indicates that the dependent variables for interpretational accuracy are normally distributed, while the dependent variables associated with interpretational timeliness deviate from the normal distribution. Regarding *homogeneity*, Box's *M* test is significant with F (408, 54298.2) = 3.14, $p<0.000$, which indicates that there is heterogeneity of the within-cell variance and their covariance. Furthermore, the univariate Levene's test points to variability in scores across cells. Heterogeneity is problematic when group sizes across factors are unbalanced (largest size/smallest size >1.5) and there are significant differences of variance of each group (group with largest variance – group with smallest variance > 4). Given that we have approximately balanced group sizes (1.01 times) and the differences between group variances are modest (< 4), we conclude that the assumption is fairly satisfied and heterogeneity does not compromise our findings.

After testing the assumptions that adhere to MANCOVA analyses, we conducted the multivariate analyses with representation and expertise as factors and all covariates of Table 8. Since none of these covariates appeared to be significant, they are omitted in the results that we present next. Table 10 and Table 11 summarize the results for effectiveness and efficiency for the two cases and the two factors, representation and expertise. The F-statistic provides the basis for determining the significance level of each factor. Eta^2 indicates small, medium and large effects for the thresholds of greater than 0.01, 0.06, 0.14, respectively (Field 2013, Kirk 1996). Table 12 and Table 13 show the differential plots for effectiveness and efficiency.

-------- Insert Table 10 to Table 13----------------------

Strong Visual Affinity Tasks

We hypothesized that tasks that strongly benefit from the multidirectional spatial search capabilities of humans have a strong affinity with visual representations. More precisely, we expect these tasks to be conducted more efficiently and more effectively when starting from a visual representation, compared to starting from a textual representation. The search and recognition task is used in our experiment to test this hypothesis.

In terms of efficiency, the representation factor was significant in the PPP case, where the visual representation was more efficient than the textual. On average, an expert (novice) needed only 7.78 (5.96) minutes to solve the search and recognition task of the PPP case with the BPMN representation, against 10.71 (8.04) minutes with the textual representation. This difference was significant (p=0.000) with a medium effect size of 0.123. No significant difference was found for the other case.

In terms of effectiveness, solving the search and recognition task with the visual representation led on average to higher scores for both the GHP and the PPP case, for both experts and novices (GHP: 18.16 (17.03) with BPMN versus 14.87 (14.71) with text; PPP: 18.30 (16.70) with BPMN versus 11.79 (10.29) with text). The representation factor is significant with a large effect size (0.181 and 0.598).

These results support our first hypotheses that search and recognition tasks can be viewed as tasks with a strong visual affinity, where both efficiency and effectiveness will benefit from a visual presentation over a textual presentation. Our results provide more support for the positive effect of a visual presentation on effectiveness (large effect in both cases) than on efficiency (medium effect in one case).

The multivariate analysis also allows us to assess the effect of expertise. Expertise had a significant impact on efficiency for both cases with a medium effect size (Eta^2= 0.109 and 0.101). The impact on effectiveness was only significant for the PPP case with a medium effect (Eta^2=0.079).

Weak Visual Affinity Tasks

We hypothesized that tasks that partially benefit from multidirectional spatial search capabilities of humans have a weak affinity with visual representations. More precisely, we expect these tasks, like the tasks with a strong visual affinity, to be conducted more efficiently and more effectively when starting from a visual representation. The inference task is used in our experiment to test this hypothesis.

The analysis shows a small effect (Eta^2=0.047) of representation on efficiency in the PPP case. On average, an expert (novice) needed 3.64 (2.37) minutes to solve the inference task of the PPP case with the BPMN representation, against 2.71 (1.78) minutes with the textual representation. This indicates that in the PPP case, the textual representation outperformed the visual for the inference task. This is contrary to what we expected. In the GHP case, no significant impact was found.

In terms of effectiveness, the results partially support our hypothesis. Solving the inference task with the visual representation led on average to higher scores for the PPP case, for both experts and novices (7.73 (6.85) with BPMN versus 6.39 (6.09) with text). The representation factor is significant (p=0.000) with a medium effect size of 0.113 for this case. No significant effect is found for the other case.

In sum, these results suggest a rejection of hypothesis H2.a that visual representations positively impact efficiency when conducting inference tasks. Partial support was found for hypothesis H2.b that visual representations positively impact effectiveness of inference tasks. With regards to expertise, a significant impact was measured on both efficiency and effectiveness and for both cases with a small to medium effect size (Eta^2= 0.093 and 0.094 for efficiency and 0.086 and 0.039 (small) for effectiveness).

Weak Textual Affinity Tasks

We hypothesized that tasks that partially benefit from a sequential processing of information, while this not being sufficient for a successful completion of the task alone, are tasks with a weak textual affinity. These tasks would benefit from textual representations over visual representation in terms of efficiency and effectiveness. We tested this hypothesis by selecting problem-solving tasks as tasks with a weak textual affinity.

The representation factor was not significant for the problem-solving tasks, neither for efficiency and effectiveness, nor for any of the two cases. We therefore reject the hypotheses on tasks with a weak textual affinity. With regards to expertise, again a medium-sized significant effect was measured on both efficiency and effectiveness and for both cases (Eta^2= 0.084 and 0.107 for efficiency, and 0.085 and 0.050 (small) for effectiveness).

Strong Textual Affinity Tasks

We hypothesized that tasks that generally benefit from a sequential processing of information, potentially only requiring a surface understanding, would benefit from textual representations over visual representations. We called these tasks strong textual affinity tasks.

In terms of efficiency, the representation factor was significant in the GHP case, where the textual representation indeed was more efficient than the visual. On average, an expert (novice) needed 6.15 (4.22) minutes to solve the recall task of the GHP case with the text representation, against 6.54 (5.93) minutes with the visual representation. This difference was significant (p=0.003) with a small effect size of 0.052. No effect was found for the PPP case.

In terms of effectiveness, more explicit evidence that supports our hypothesis is found. Solving the recall task with the textual representation led on average to higher scores for both the GHP and the PPP case, for both experts and novices (GHP: 9.11 (8.91) with text versus 8.48 (7.73) with BPMN; PPP: 13.76 (12.24) with text versus 10.32 (8.83) with BPMN). The representation factor is significant (p=0.000) with a medium effect size in the GHP case (Eta^2=0.077) and a large effect size in the PPP case (Eta^2=0.223).

The results support our hypothesis concerning tasks with a strong textual affinity: for recall tasks, a textual representation is more efficient and effective than a visual representation. There is full support for the impact of the textual representation on effectiveness, and partial support for the impact on efficiency. With regards to expertise, a small effect was measured in both cases on efficiency (Eta^2= 0.024 and 0.029) and a small effect on effectiveness for the PPP case (Eta^2=0.052).

DISCUSSION OF THE RESULTS, LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH

This study investigates whether different task types have a different affinity towards a visual or a textual process representation. Table 14 summarizes our findings. We now highlight our contributions and discuss their implications for theory and practice.

-------- Insert Table 14- Overview of findings --------

First, we developed a theoretical argument that task performance depends on the match between the representation format and the task type. We refer to this mutual dependency as task representation affinity and describe a spectrum from strong visual affinity towards strong textual affinity. This theoretical argument builds on cognitive fit theory (Vessey 1991) and its extension by Shaft and Vessey (2006). Our findings support this theoretical argument at the extreme ends of the spectrum with medium to large effect sizes for both efficiency and effectiveness. All corresponding differential plots, search and recognition on the one end and recall at the other end, show the expected gradient and all hypotheses are at least partially supported. These tasks at the end of the spectrum have in common that they require surface level understanding. In an audit engagement, tasks like scanning documents to determine whether a control exists or not (search and recognition) or gaining a general understanding of the environment (recall) are associated with these ends of the spectrum. This suggests that auditors, when examining process documentation, might benefit from a textual representation in the audit planning phase, while a visual representation might be preferred to assess controls.

Second, the two tasks for which we posited weak affinity with either visual or textual representation (inference and problem-solving tasks) have in common that they require a deep understanding of the task and integration of prior knowledge on model semantics and domain semantics. Our results are not conclusive for these tasks, with only one of four related hypotheses being partially supported with a medium effect size. Performing analytical procedures (inference) and tracing potential root causes of material errors (problem solving) are example tasks that would fall in this category. For these types of tasks, it seems that no overall affinity for one representation might exist. These results are consistent with the findings of Boritz, Borthick, and Presslee (2012) who also find no advantage of visual representations for problem-solving and Dunn and Gerard (2001) who find better performance of visual affinity tasks for visual representations in terms of efficiency. It is the contribution of our study to provide and test a theoretical argument to reconcile these separate findings. Also experiments by Trkman, Mendling, and Krisper (2016), Boritz et al 2014, Ottensooser et al 2012, Bierstaker et al 2009, and Dunn & Grabski 2001 appear to be at least partially consistent with our argument. Therefore, our findings are not only relevant for accounting information systems, but also for software engineering (Whitley 1997) and information systems research (Figl 2017), which have largely focused on a distinction between sequential and circumstantial tasks so far (Fahland et al. 2009a,b, Pichler et al. 2011).

Our study has also limitations and we discuss how they can be addressed by future research. The first limitation relates to external validity. Our findings emphasize the need for further investigating how representations affect deep-understanding tasks without strong representational affinity. Clearly, conducting an audit involves several tasks with different characteristics. Therefore, different tasks in an auditing assignment might need different process representations to obtain the best results. Different strategies can be combined to address this research problem. Methods such as hierarchical task analysis (Stanton 2006) can be an important aid to study these complex tasks in more detail. Likely, this will reveal that complex tasks are composed of several more fine-granular tasks that partially have visual and partially textual affinity. External validity of experiments in this area can be increased by focusing on more complex tasks that play a key role in the audit process. The taxonomy of audit tasks described by Abdolmohammadi (1999) is an excellent reference in this regard. Furthermore, the actual interaction between the experimental participants and the task material can be studied in more detail using eye tracking. The potential of this technology is apparent from recent studies (Petrusel, Mendling, and Reijers 2016, 2017).

The second limitation relates to content validity. Our research emphasizes the need for further investigating the spectrum from visual to textual representations. On the one hand, most visual representations including BPMN models or other types of flowcharts contain at least a small amount of text, which has the effect that not all information can be spatially searched. On the other hand, text is often visually structured, ranging from the use of paragraphs and sections to bullet lists and tables, which does facilitate spatial search. A starting point for future research in this area might be the multimedia learning theory by Mayer (2005), which provides several guidelines on how visual and textual representations can be integrated to achieve good understanding performance. Recent studies including the one by Ottensooser et al (2012) already found benefits of a combination of different representations, but research that explicitly connects variants of such combined representations with task types is missing so far.

The third limitation relates to conclusion validity. We identify a need for large-scale experiments. The number of factors for audit task performance is long as evidenced by Kelton, Pennington, and Tuttle (2010), and the number of potential interactions is even larger. Large-scale experiments would permit to vary various factors in one experiment. The statistical analysis of such experiments would allow us to understand the relative importance of the different factors.

Nevertheless, our findings demonstrate that the representation format has an impact on comprehension performance and that the size and direction of this impact is relative to the type of task. In this way, our research provides a foundation for future research into the mutual benefits of representation formats for different types of auditing tasks.

References

Abdolmohammadi, Mohammad J. 1999. "A comprehensive taxonomy of audit task structure, professional rank and decision aids for behavioral research " *Behavioral Research in Accounting* 11:51-92.

Agarwal, Ritu, Atish P. Sinha, and Mohan Tanniru. 1996. "Cognitive fit in requirements modeling: A study of object and process methodologies." *Journal of Management Information Systems: JMIS; Armonk* 13 (2):137-162.

Ballou, Brian, Christine E. Earley, and Jay S. Rich. 2004. "The Impact of Strategic-Positioning Information on Auditor Judgments about Business-Process Performance." *Auditing: A Journal of Practice & Theory* 23 (2):71-88.

Bell, Thimothy, Frank Marrs, Ira Solomon, and Howard Thomas. 1997. *Auditing Organizations Through a Strategic-Systems Lens: The KPMG Business Measurement Process*. Montvale, NJ: KPMG Peat Marwick LLP.

Bierstaker, James L., Jean C. Bedard, and Stanley F. Biggs. 1999. "The Role of Problem Representation Shifts in Auditor Decision Processes in Analytical Procedures." *Auditing: A Journal of Practice & Theory* 18 (1):18-36.

Bierstaker, James Lloyd, James E. Hunton, and Jay C. Thibodeau. 2009. "Do Client-Prepared Internal Control Documentation and Business Process Flowcharts Help or Hinder an Auditor's Ability to Identify Missing Controls?" *Auditing: A Journal of Practice & Theory* 28 (1):79-94.

Birnbaum, M.H. 2000. *Psychological Experiments on the Internet*. San Diego: Elsevier: Academic Press.

Bodart, François, Arvind Patel, Marc Sim, and Ron Weber. 2001. "Should optional properties be used in conceptual modelling? A theory and three empirical tests." *Information Systems Research* 12 (4):384-405.

Boritz, J. Efrim, A. Faye Borthick, and Adam Presslee. 2012. "The Effect of Business Process Representation Type on Assessment of Business and Control Risks: Diagrams versus Narratives." *Issues in Accounting Education* 27 (4):895-915.

Boritz, J. Efrim, Carla Carnaghan, and Paulo S. Alencar. 2014. "Business Modeling to Improve Auditor Risk Assessment: An Investigation of Alternative Representations." *Journal of Information Systems* 28 (2):231-256.

Borthick, A. Faye, Gary P. Schneider, and Anthony Vance. 2012. "Using Graphical Representations of Business Processes in Evaluating Internal Control." *Issues in Accounting Education* 27 (1):123-140.

Bradford, Marianne, Sandra B. Richtermeyer, and Douglas F. Roberts. 2007. "System Diagramming Techniques: An Analysis of Methods Used in Accounting Education and Practice." *Journal of Information Systems* 21 (1):173-212.

Carnaghan, Carla. 2006. "Business process modeling approaches in the context of process level audit risk assessment: An analysis and comparison." *International Journal of Accounting Information Systems* 7 (2):170-204.

Chandler, P., and J. Sweller. 1991. "Cognitive load theory and the format of instruction." *Cognition and Instruction* 8 (4):293-332.

Chi, Michelene T. H., Robert Glaser, and Ernest Rees. 1981. Expertise in problem solving. Pittsburgh univ PA learning research and development center.

Cunniff, Nancy, and Robert P. Taylor. 1987. "Graphical vs. Textual Representation: An Empirical Study of Novices' Program Comprehension."

Curtis, Bill, Sylvia B. Sheppard, Elizabeth Kruesi-Bailey, John Bailey, and Deborah A. Boehm-Davis. 1989. "Experimental evaluation of software documentation formats." *Journal of Systems and Software* 9 (2):167-207.

Day, Ruth S. 1988. "Alternative representations." In *Psychology of learning and motivation*, 261-305. Elsevier.
Dumas, Marlon, Marcello La Rosa, Jan Mendling, and Hajo A. Reijers. 2013. *Fundamentals of business process management*: Springer.
Dunn, Cheryl, and Severin Grabski. 2001. "An Investigation of Localization as an Element of Cognitive Fit in Accounting Model Representations." *Decision Sciences* 32 (1):55-94.
Dunn, Cheryl L, Gregory J Gerard, and Severin V Grabski. 2017. "The combined effects of user schemas and degree of cognitive fit on data retrieval performance." *International Journal of Accounting Information Systems* 26:46-67.
Dunn, Cheryl L., and Gregory J. Gerard. 2001. "Auditor efficiency and effectiveness with diagrammatic and linguistic conceptual model representations." *International Journal of Accounting Information Systems* 2 (4):223-248.
Field, Andy. 2013. *Discovering statistics using IBM SPSS statistics*: SAGE.
Figl, Kathrin. 2017. "Comprehension of Procedural Visual Business Process Models – A Literature Review." *Business & Information Systems Engineering (BISE)* 59 (1):41-67.
Gemino, Andrew, and Yair Wand. 2003. "Evaluating Modeling Techniques Based on Models of Learning." *Commun. ACM* 46 (10):79-84.
Gemino, Andrew, and Yair Wand. 2004. "A framework for empirical evaluation of conceptual modeling techniques." *Requirements Engineering* 9 (4):248-260.
Gilmore, D. J., and T. R. G. Green. 1984. "Comprehension and recall of miniature programs." *International Journal of Man-Machine Studies* 21 (1):31-48.
Green, Thomas R. G., Marian Petre, and R. K. E. Bellamy. 1991. "Comprehensibility of visual and textual programs: A test of superlativism against the'match-mismatch'conjecture." *Empirical Studies of Programmers: 4th Workshop* 91 (743):121-146.
International Auditing and Assurance Standards Board (IAASB). 2010. Understanding the Entity and Its Environment and Assessing the Risks of Material Misstatement (and related appendices). In *ISA 315*. New York: International Federation of Accountants.
Jones, Roberta Ann, Jeffrey J Tsay, and Kenneth Griggs. 2002. "An Empirical Investigation of the Cognitive Fit of Selected Process Model Diagramming Techniques." *The Review of Business Information System* 6:1-10.
Kelton, Andrea Seaton, Robin R. Pennington, and Brad M. Tuttle. 2010. "The Effects of Information Presentation Format on Judgment and Decision Making: A Review of the Information Systems Research." *Journal of Information Systems; Sarasota* 24 (2):79-105.
Khatri, Vijay, Iris Vessey, V. Ramesh, Paul Clay, and Park Sung-Jin. 2006. "Understanding Conceptual Schemas: Exploring the Role of Application and IS Domain Knowledge." *Information Systems Research; Linthicum* 17 (1):81-102.
Kirk, Roger E. 1996. "Practical significance: A concept whose time has come." *Educational and psychological measurement* 56 (5):746-759.
Larkin, Jill H, and Herbert A Simon. 1987. "Why a diagram is (sometimes) worth ten thousand words." *Cognitive science* 11 (1):65-100.
Leopold, Henrik, Jan Mendling, and Artem Polyvyanyy. 2014. "Supporting Process Model Validation through Natural Language Generation." *IEEE Transactions on Software Engineering* 40 (8):818-840.
Libby, Robert, and Joan Luft. 1993. "Determinants of judgment performance in accounting settings: Ability, knowledge, motivation, and environment." *Accounting, Organizations and Society* 18 (5):425-450.
Mayer, Richard E. 1989. "Models for understanding." *Review of educational research* 59 (1):43-64.
Mayer, Richard E. 2005. *The Cambridge Handbook of Multimedia Learning*, *Cambridge Handbooks in Psychology*: Cambridge University Press.
McGuinness, Carol. 1986. "Problem representation: The effects of spatial arrays." *Memory & Cognition* 14 (3):270-280.

Mendling, Jan, Mark Strembeck, and Jan Recker. 2012. "Factors of process model comprehension—Findings from a series of experiments." *Decision Support Systems* 53 (1):195-206.
Moher, Thomas G., D. Mak, B. Blumenthal, and L. Levanthal. 1993. "Comparing the comprehensibility of textual and graphical programs." 1993.
Moore, Gary C, and Izak Benbasat. 1991. "Development of an instrument to measure the perceptions of adopting an information technology innovation." *Information systems research* 2 (3):192-222.
Newell, Allen, and Herbert Alexander Simon. 1972. *Human problem solving*. Vol. 104: Prentice-Hall Englewood Cliffs, NJ.
O'Donnell, Ed, and Joseph J. Schultz Jr. 2003. "The Influence of Business-Process-Focused Audit Support Software on Analytical Procedures Judgments." *Auditing: A Journal of Practice & Theory* 22 (2):265-279.
Ottensooser, Avner, Alan Fekete, Hajo A. Reijers, Jan Mendling, and Con Menictas. 2012. "Making sense of business process descriptions: An experimental comparison of graphical and textual notations." *Journal of Systems and Software* 85 (3):596-606.
Paivio, Allan. 1991. "Dual coding theory: Retrospect and current status." *Canadian Journal of Psychology/Revue canadienne de psychologie* 45 (3):255.
PCAOB. 2010. Auditing Standard No. 12 - Identifying and Assessing Risks of Material Misstatement. edited by Public Company Accounting Oversight Board (PCAOB).
Petrusel, Razvan, Jan Mendling, and Hajo A. Reijers. 2016. "Task-specific visual cues for improving process model understanding." *Information and Software Technology* 79:63-78.
Petrusel, Razvan, Jan Mendling, and Hajo A. Reijers. 2017. "How visual cognition influences process model comprehension." *Decision Support Systems* 96 (Supplement C):1-16.
Ramsey, H. Rudy, Michael E. Atwood, and James R. Van Doren. 1983. "Flowcharts versus program design languages: an experimental comparison." *Communications of the ACM* 26 (6):445-449.
Recker, Jan, and Alexander Dreiling. 2007a. "Does It Matter Which Process Modelling Language We Teach or Use? An Experimental Study on Understanding Process Modelling Languages without Formal Education." Australasian Conference on Information Systems Toowomba.
Recker, Jan, and Alexander Dreiling. 2007b. "Does it matter which process modelling language we teach or use? an experimental study on understanding process modelling languages without formal education."
Recker, Jan, and Alexander Dreiling. 2011. "The effects of content presentation format and user characteristics on novice developers'understanding of process models." *Communications of the Association for Information Systems* 28:65-84.
Reijers, Hajo A., and Jan Mendling. 2011. "A Study Into the Factors That Influence the Understandability of Business Process Models." *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* 41 (3):449-462.
Rodgers, Jacqui, Tom Buchanan, Andrew B Scholey, Thomas M Heffernan, Jonathan Ling, and Andrew C Parrott. 2003. "Patterns of drug use and the influence of gender on self-reports of memory ability in ecstasy users: a web-based study." *Journal of Psychopharmacology* 17 (4):389-396.
Scanlan, David A. 1989. "Structured flowcharts outperform pseudocode: An experimental comparison." *IEEE software* 6 (5):28-36.
Schiefele, Ulrich, and Andreas Krapp. 1996. "Topic interest and free recall of expository text." *Learning and Individual Differences* 8 (2):141-160.
Schwartz, Steven M., and Daniel L. Fattaleh. 1972. "Representation in deductive problem-solving: The matrix." *Journal of Experimental Psychology* 95 (2):343-348.
Shaft, Teresa M., and Iris Vessey. 2006. "The role of cognitive fit in the relationship between software comprehension and modification." *MIS Quarterly* 30 (1):29-55.

Shneiderman, Ben, Richard Mayer, Don McKay, and Peter Heller. 1977. "Experimental investigations of the utility of detailed flowcharts in programming." *Communications of the ACM* 20 (6):373-381.

Simon, Herbert A., and Allen Newell. 1971. "Human problem solving: The state of the theory in 1970." *American Psychologist* 26 (2):145.

Stanton, Neville A. 2006. "Hierarchical task analysis: Developments, applications, and extensions." *Applied ergonomics* 37 (1):55-79.

Sweller, John. 1988. "Cognitive Load During Problem Solving: Effects on Learning." *Cognitive Science* 12 (2):257-285.

Trkman, Marina, Jan Mendling, and Marjan Krisper. 2016. "Using business process models to better understand the dependencies among user stories." *Information and Software Technology* 71:58-76.

Verardi, Vincenzo, and Catherine Dehon. 2010. "Multivariate outlier detection in Stata." *The Stata Journal* 10 (2):259-266.

Vessey, Iris. 1991. "Cognitive Fit: A Theory-Based Analysis of the Graphs Versus Tables Literature." *Decision sciences* 22 (2):219-240.

Vessey, Iris, and Dennis Galletta. 1991. "Cognitive Fit: An Empirical Study of Information Acquisition." *Information Systems Research* 2 (1):63-84.

Vessey, Iris, and Ron Weber. 1986. "Structured tools and conditional logic: An empirical investigation." *Communications of the ACM* 29 (1):48-57.

Whitley, K. N. 1997. "Visual Programming Languages and the Empirical Evidence For and Against." *Journal of Visual Languages & Computing* 8 (1):109-142.

TABLES

Table 1 – Overview of studies on the benefits of visual versus textual representations showing the inconsistencies on this topic

Paper	**Text Representation**	**Visual Representation**	**Task**	**Effectiveness**	**Efficiency**	**Favoring**
Schwartz and Fattaleh (1972)	Sentence	Tree	Deductive Inference	Visual better		Visual
Shneiderman et al. (1977)	Fortran	Fortran + Flowchart	Design	Inconclusive		Inconclusive
Ramsey, Atwood, and Van Doren (1983)	PDL	Flowchart	Design	Text better		Text
Vessey and Weber (1986)	Pseudocode	Decision Tree	Conditional Logic Inference	Visual better	Visual faster	Visual
McGuinness (1986)	Matrix	Family Tree	Family Relationship Inference	Inconclusive		Visual
Cunniff and Taylor (1987)	Pascal	Flowchart	Code Recognition	Visual better		Visual
Day (1988)	List	Spatial Map	Cursor Movement	Visual better		Visual
Curtis et al. (1989)	Pseudocode	Flowchart	Design	Text better		Text
Scanlan (1989)	Pseudocode	Flowchart	Conditional Logic Inference	Visual better	Visual faster	Visual
Green, Petre, and Bellamy (1991)	Pseudocode	LabVIEW	Conditional Logic Inference	Inconclusive	Text faster	Text
Moher et al. (1993)	Pseudocode	Petri Net	Conditional Logic Inference	Text better	Matched faster	Text
Dunn and Gerard (2001)	BNF Grammar	Entity Relationship	Schema Search, Recognition, Inference	Inconclusive	Visual faster	Visual
Dunn and Grabski (2001)	DCA	REA	Inference and Problem Solving	Visual better	Text faster	
Bierstaker, Hunton, and Thibodeau (2009)	Documentation	Flowchart	Missing Controls Problem Solving	Visual+Text better		Visual+Text>Text

Paper	Text Representation	Visual Representation	Task	Effectiveness		Efficiency
Ottensooser et al. (2012)	Use Case	BPMN	Requirements Inference	Visual+Text better		Text
Boritz, Borthick, and Presslee (2012)	Narrative	BPMN	Risk Assessment	Inconclusive	Text faster	Visual+Text>Text
Boritz, Carnaghan, and Alencar (2014)	Tabular	Business Model	Risk Assessment	Visual better	Visual better	Visual
Trkman, Mendling, and Krisper (2016)	Use Case	BPMN + Use Case	Requirements Inference	Visual better	Visual faster	

Table 2- Dependent variables and ranges

	Goods receipt handling (GHP)	Procure-to-pay process (PPP)
Effectiveness	GHP Search and Recognition Score [0-28]	PPP Search and Recognition Score [0-28]
	GHP Inference Score [0-10]	PPP Inference Score [0-10]
	GHP Problem Solving Score [0-43]	PPP Problem Solving Score [0-26]
	GHP Cloze Score [0-10]	PPP Cloze Score [0-17]
Efficiency	GHP Search and Recognition Time [0-28] / duration	PPP Search and Recognition Time [0-28] / duration
	GHP Inference Time [0-10] / duration	PPP Inference Time [0-10] / duration
	GHP Problem Solving Time [0-43] / duration	PPP Problem Solving Time [0-26] / duration
	GHP Cloze Time [0-10] / duration	PPP Cloze Time [0-17] / duration

Table 3- Demographic statistics participants

		Text		BPMN	
		Expert	Novice	Expert	Novice
		(self-assessment of Expert versus Novice)			
Male [Female]		23 [15]	20 [25]	26 [18]	17 [23]
English as first language		5	1	3	2
Age – Mean (SD)		36.05 (6.84)	20.16 (0.80)	35.11 (8.82)	20.65 (2.30)
Education	Undergraduate	1	41	2	37
	Graduate	37	4	42	3
Prior Working Area[1]	IT Audit	9	8	10	8
	Financial Audit	9	10	11	5
	Risk and Control Advisory	7	1	7	1
	IT Governance	4	-	4	-
	Business Process Analysis	8	6	21	2
	Others	15	5	13	4

[1] Multiple areas could have been selected per participant.

Table 4- Descriptive statistics of effectiveness for the Goods Handling case, measured by score performance

	Task *M* (*SD*) [*n*]			
	Search and Recognition	Inference	Problem-solving	Recall
Panel A: BPMN				
Expert	18.16 (1.778) [44]	7.73 (1.169) [44]	6.55 (3.701) [44]	8.48 (1.772) [44]
Novice	17.03 (2.616) [40]	6.93 (1.457) [40]	5.4 (2.073) [40]	7.73 (1.71) [40]
Panel B: text				
Expert	14.87 (3.974) [38]	7.37 (1.46) [38]	7.29 (2.967) [38]	9.11 (1.41) [38]
Novice	14.71 (3.341) [45]	6.44 (1.575) [45]	5.07 (2.093) [45]	8.91 (1.395) [45]
Panel C: total				
Expert	16.63 (3.408) [82]	7.56 (1.316) [82]	6.89 (3.381) [82]	8.77 (1.635) [82]
Novice	15.8 (3.221) [85]	6.67 (1.531) [85]	5.22 (2.078) [85]	8.35 (1.653) [85]
Total	16.21 (3.331) [167]	7.11 (1.493) [167]	6.04 (2.909) [167]	8.56 (1.652) [167]

Table 5- Descriptive statistics of efficiency for the Goods Handling case, measured by time

	Task *M* (*SD*) [*n*]			
	Search and Recognition	Inference	Problem-solving	Recall
Panel A: BPMN				
Expert	8.88 (3.294) [44]	2.48 (0.983) [44]	7.72 (7.598) [44]	6.54 (3.941) [44]
Novice	6.45 (1.728) [40]	1.92 (0.704) [40]	4.39 (2.381) [40]	5.93 (1.931) [40]
Panel B: text				
Expert	9.29 (5.713) [38]	2.58 (2.143) [38]	7.46 (6.122) [38]	6.15 (6.011) [38]
Novice	6.93 (1.967) [45]	1.59 (0.509) [45]	4.56 (2.633) [45]	4.22 (1.237) [45]
Panel C: total				
Expert	9.07 (4.551) [82]	2.53 (1.616) [82]	7.6 (6.913) [82]	6.36 (4.979) [82]
Novice	6.7 (1.863) [85]	1.75 (0.627) [85]	4.48 (2.504) [85]	5.03 (1.808) [85]
Total	7.87 (3.644) [167]	2.13 (1.275) [167]	6.01 (5.38) [167]	5.68 (3.767) [167]

Table 6- Descriptive statistics of effectiveness for the Procure-to-Pay case, measured by score performance

	Task *M* (*SD*) [*n*]			
	Search and Recognition	Inference	Problem-solving	Recall
Panel A: BPMN				
Expert	18.3 (1.995) [44]	7.73 (1.208) [44]	6.91 (4.258) [44]	10.32 (2.939) [44]
Novice	16.7 (3.09) [40]	6.85 (1.642) [40]	6.4 (3.601) [40]	8.83 (3.327) [40]
Panel B: text				
Expert	11.79 (2.877) [38]	6.39 (1.636) [38]	9.37 (4.334) [38]	13.76 (2.645) [38]
Novice	10.29 (2.677) [45]	6.09 (1.443) [45]	6.31 (3.522) [45]	12.24 (3.821) [45]
Panel C: total				
Expert	15.28 (4.068) [82]	7.11 (1.564) [82]	8.05 (4.441) [82]	11.91 (3.282) [82]
Novice	13.31 (4.307) [85]	6.45 (1.577) [85]	6.35 (3.538) [85]	10.64 (3.967) [85]
Total	14.28 (4.294) [167]	6.77 (1.601) [167]	7.19 (4.084) [167]	11.26 (3.692) [167]

Table 7- Descriptive statistics of efficiency for the Procure-to-Pay case, measured by time

		Task *M* (*SD*) [*n*]			
		Search and Recognition	Inference	Problem-solving	Recall
Panel A: BPMN					
	Expert	7.78 (3.694) [44]	3.64 (2.891) [44]	6.55 (5.19) [44]	6.54 (3.941) [44]
	Novice	5.96 (1.604) [40]	2.37 (0.809) [40]	4.31 (2.208) [40]	5.93 (1.931) [40]
Panel B: text					
	Expert	10.71 (5.007) [38]	2.71 (1.466) [38]	7.74 (5.532) [38]	6.15 (6.011) [38]
	Novice	8.04 (2.393) [45]	1.78 (0.694) [45]	4.37 (2.336) [45]	4.22 (1.237) [45]
Panel C: total					
	Expert	9.14 (4.566) [82]	3.21 (2.374) [82]	7.1 (5.351) [82]	6.36 (4.979) [82]
	Novice	7.06 (2.3) [85]	2.06 (0.803) [85]	4.34 (2.264) [85]	5.03 (1.808) [85]
	Total	8.08 (3.732) [167]	2.62 (1.846) [167]	5.7 (4.299) [167]	5.68 (3.767) [167]

Table 8- Validity Checks between treatments

Variables	Representation			Expertise		
	Text (*n=83)*	**BPMN** (*n=84)*	**Sig.** *(t–test)*	**Expert** (*n=82)*	**Novice** *(n=85)*	**Sig.** *(t–test)*
Prior use of text to document business processes (yes = 1, no = 2)	1.05	1.08	0.36	1.09	1.05	0.32
Prior use of diagram to document business process (yes = 1, no = 2)	1.08	1.07	0.76	1.02	1.13	**0.01**
Overall familiarity with text	5.35	5.08	0.19	5.04	5.39	**0.09**
Overall familiarity with diagrams	5.41	5.26	0.44	5.74	4.94	**0.00**
Perceived intensity of using text	4.53	4.29	0.31	4.27	4.54	0.26
Perceived intensity of using diagrams	4.76	4.55	0.35	4.95	4.36	**0.01**
Perceived competency with text	4.89	4.74	0.46	4.87	4.76	0.63
Perceived competency with diagram	5.06	4.83	0.27	5.33	4.58	**0.00**
Perceived confidence with text	5.1	5.05	0.82	4.96	5.18	0.31
Perceived confidence with diagram	5.45	5.15	0.13	5.6	5.01	**0.00**
Perceived knowledge of goods handling	4.66	4.57	0.66	4.43	4.8	**0.07**
Perceived knowledge of procure-to-pay	4.67	4.46	0.33	4.49	4.65	0.46
Perceived involvement with business process	5.33	5.25	0.6	5.63	4.95	**0.00**
Perceived involvement with risk and audit	4.42	4.26	0.45	4.54	4.15	**0.07**
Perceived involvement with internal control	4.72	4.55	0.4	4.65	4.62	0.91
Perceived involvement of acc. information system	4.49	4.6	0.64	4.5	4.59	0.69

Table 9- Validity checks between cases, where 'Score' measures effectiveness and 'Time' efficiency

	Case Order		
Variables	**GHP first** *(n=86)*	**PPP first** *(n=81)*	**Sig.** *(t-test)*
GHP Search and Recognition Score	16.48	15.93	0.29
GHP Inference Score	7.3	6.9	0.08
GHP Problem Solving Score	6.56	5.49	**0.02**
GHP Cloze Score	8.45	8.67	0.41
GHP Search and Recognition Time	9.18	6.47	**0.00**
GHP Inference Time	2.29	1.96	0.1
GHP Problem Solving Time	7.81	4.11	**0.00**
GHP Cloze Time	3.49	2.79	0.22
PPP Search and Recognition Score	14.31	14.23	0.91
PPP Inference Score	6.7	6.85	0.53
PPP Problem Solving Score	6.88	7.51	0.33
PPP Cloze Score	11.4	11.12	0.63
PPP Search and Recognition Time	7.31	8.9	**0.01**
PPP Inference Time	2.3	2.97	**0.02**
PPP Problem Solving Time	4.72	6.74	**0.00**
PPP Cloze Time	4.94	6.46	**0.01**

Table 10- MANCOVA results for efficiency (measured in time), based on representation and expertise

Hypothesis:	Strong visual affinity			Weak visual affinity			Weak textual affinity			Strong textual affinity		
Efficiency	Search & Recognition			Inference			Problem Solving			Cloze		
Goods Handling Process	F	Sig.	Eta^2	F	Sig.	Eta^2	F	Sig.	Eta^2	F	Sig.	Eta^2
Representation	0.679	0.411	0.004	0.384	0.536	0.002	0.004	0.951	0.000	8.986	**0.003**	0.052
Expertise	19.885	**0.000**	0.109	16.731	**0.000**	0.093	14.972	**0.000**	0.084	4.058	**0.046**	0.024
Representation*Expertise	0.004	0.949	0.000	1.253	0.265	0.008	0.069	0.793	0.000	3.365	0.068	0.020
	R Squared = .110 (Adjusted = .094)			R Squared = .103 (Adjusted = .087)			R Squared = .085 (Adjusted = .068)			R Squared = .095 (Adjusted = .078)		

Efficiency	Search & Recognition			Inference			Problem Solving			Cloze		
Procure-to-Pay	**F**	Sig.	Eta^2	F	Sig.	Eta^2	F	Sig.	Eta^2	F	Sig.	Eta^2
Representation	22.762	**0.000**	0.123	8.037	**0.005**	0.047	0.975	0.325	0.006	3.373	0.068	0.020
Expertise	18.239	**0.000**	0.101	16.872	**0.000**	0.094	19.611	**0.000**	0.107	4.890	**0.028**	0.029
Representation*Expertise	0.644	0.423	0.004	0.382	0.537	0.002	0.794	0.374	0.005	1.320	0.252	0.008
	R Squared = .193 (Adjusted = .178)			R Squared = .142 (Adjusted = .126)			R Squared = .113 (Adjusted = .097)			R Squared = .059 (Adjusted = .042)		

Table 11- MANCOVA results for effectiveness (measured by score), based on representation and expertise

Hypothesis:	Strong visual affinity			Weak visual affinity			Weak textual affinity			Strong textual affinity		
Effectiveness	Search & Recognition			Inference			Problem Solving			Cloze		
Goods Handling Process	F	Sig.	Eta^2	F	Sig.	Eta^2	F	Sig.	Eta^2	F	Sig.	Eta^2
Representation	35.974	**0.000**	0.181	3.621	0.059	0.022	0.224	0.637	0.001	13.652	**0.000**	0.077
Expertise	1.910	0.169	0.012	15.315	**0.000**	0.086	15.065	**0.000**	0.085	3.716	0.056	0.022
Representation*Expertise	1.093	0.297	0.007	0.076	0.783	0.000	1.541	0.216	0.009	1.292	0.257	0.008
	R Squared = .197 (Adjusted = .182)			R Squared = .110 (Adjusted = .093)			R Squared = .092 (Adjusted = .076)			R Squared = .099 (Adjusted = .083)		

Effectiveness	Search & Recognition			Inference			Problem Solving			Cloze		
Procure-to-Pay	F	Sig.	Eta^2	F	Sig.	Eta^2	F	Sig.	Eta^2	F	Sig.	Eta^2
Representation	242.481	**0.000**	0.598	20.700	**0.000**	0.113	3.767	0.054	0.023	46.809	**0.000**	0.223
Expertise	13.930	**0.000**	0.079	6.610	**0.011**	0.039	8.527	**0.004**	0.050	9.012	**0.003**	0.052
Representation*Expertise	0.013	0.909	0.000	1.542	0.216	0.009	4.353	0.038	0.026	0.001	0.980	0.000
	R Squared = .619 (Adjusted = .612)			R Squared = .157 (Adjusted = .142)			R Squared = .088 (Adjusted = .071)			R Squared = .247 (Adjusted = .233)		

Table 12- MANCOVA differential plots for effectiveness (dark line: experts, light line: novices)

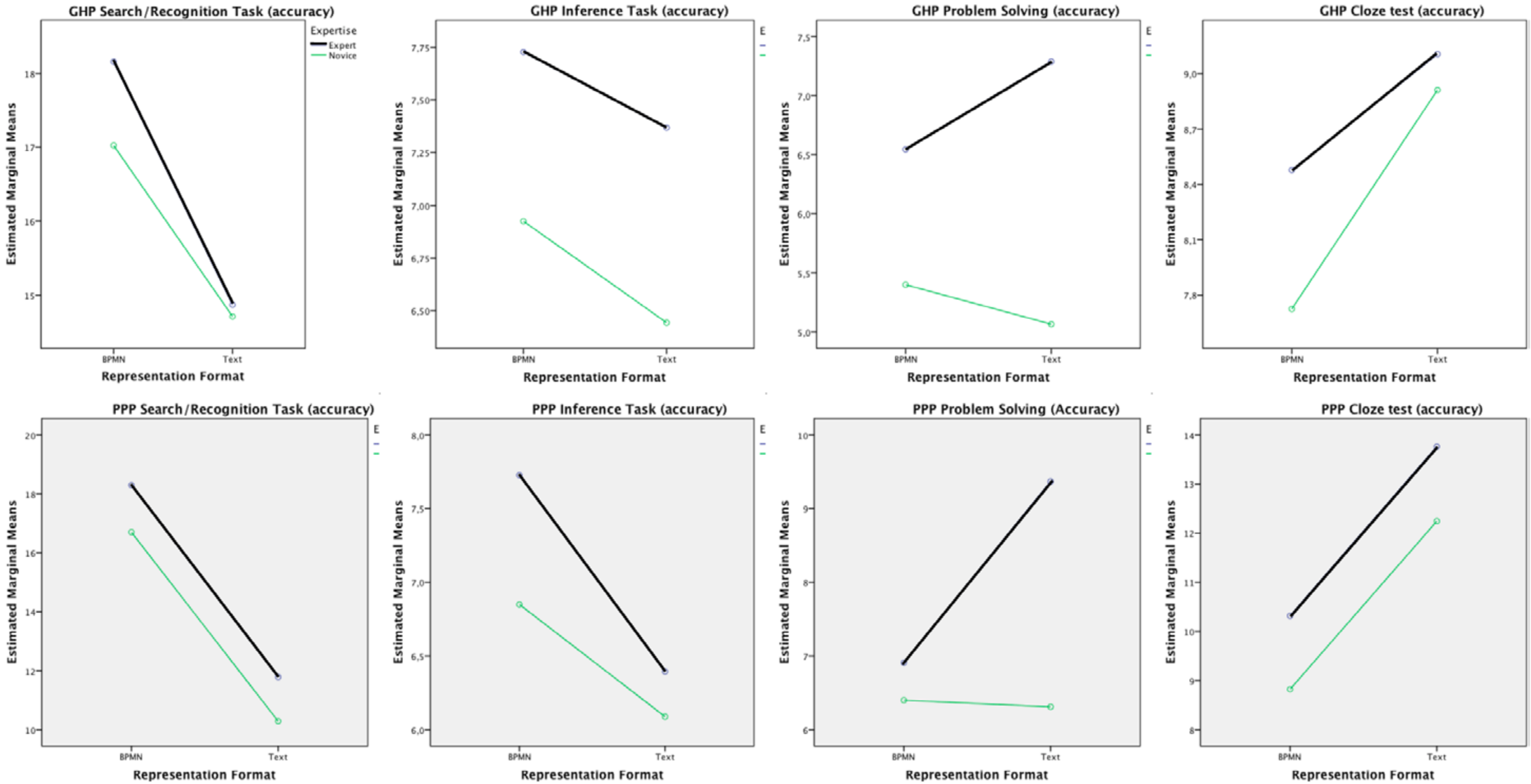


Table 13- MANCOVA differential plots for efficiency (dark line: experts, light line: novices)

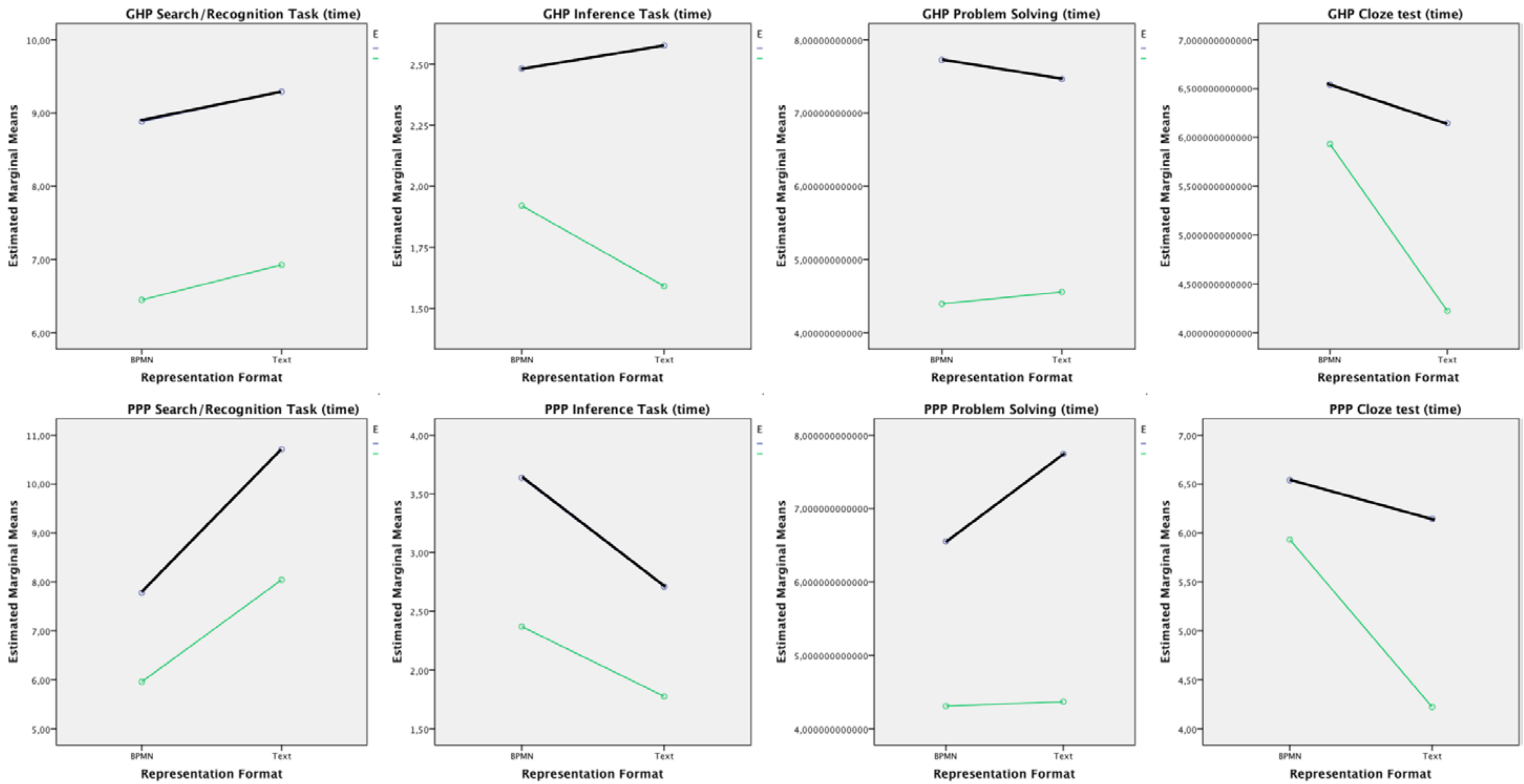


Table 14- Overview of findings

Task type	Experimental task		Hypothesis	GHP case	PPP case
Strong visual affinity task	Search and recognition tasks	H1.a	visual is more efficient than textual	no effect	**medium effect**
		H1.b	visual is more effective than textual	**large effect**	**large effect**
Weak visual affinity task	Inference tasks	H2.a	visual is more efficient than textual	no effect	small opposite effect
		H2.b	visual is more effective than textual	no effect	**medium effect**
Weak textual affinity task	Problem solving tasks	H3.a	textual is more efficient than visual	no effect	no effect
		H3.b	textual is more effective than visual	no effect	no effect
Strong textual affinity task	Recall task	H4.a	textual is more efficient than visual	**small effect**	no effect
		H4.b	textual is more effective than visual	**medium effect**	**large effect**

APPENDICES

The participants were randomly assigned to one of the four variants of the experimental material. The sequence of these variants are:

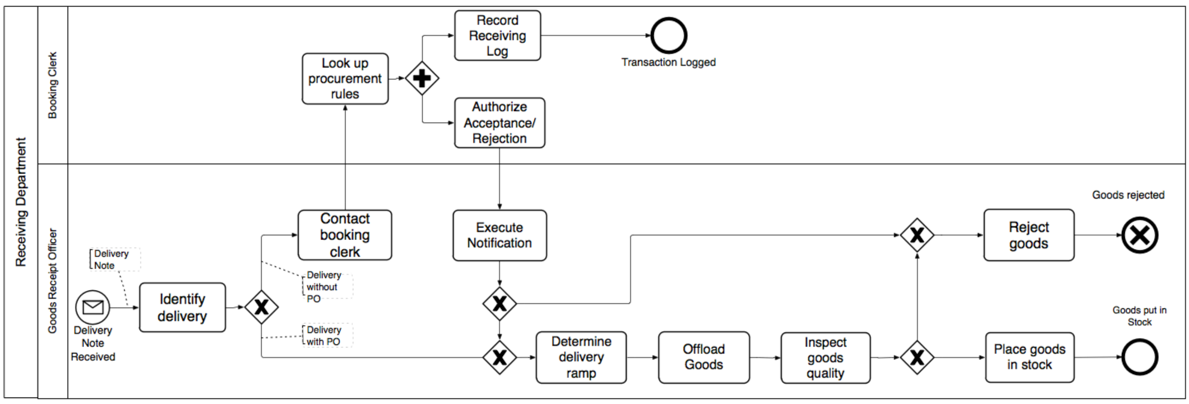
Variant 1	Demographic Questions	Tasks 1-4 for Case 1 with visual representation	Tasks 1-4 for Case 2 with textual representation	Closing Questions
Variant 2	Demographic Questions	Tasks 1-4 for Case 1 with textual representation	Tasks 1-4 for Case 2 with visual representation	Closing Questions
Variant 3	Demographic Questions	Tasks 1-4 for Case 2 with visual representation	Tasks 1-4 for Case 1 with textual representation	Closing Questions
Variant 4	Demographic Questions	Tasks 1-4 for Case 2 with textual representation	Tasks 1-4 for Case 1 with visual representation	Closing Questions
		Between-subject comparison	Between-subject comparison	

Appendix A – Case descriptions

Goods Handling Process – Textual Representation

"A truck driver registers at the goods receiving department with a delivery note to a good receipt officer. The officer identifies the delivery type. In his case, it is a delivery related to a purchase order. In case of deliveries without a purchase order, a booking clerk has to be contacted. The booking clerk shall look up the procurement rules before authorizing the delivery to be accepted or not. When the decision has been made, the booking clerk notifies the goods reception officer to execute the acceptance or rejection and records the receiving transaction in file. Following the assignment of a delivery ramp to the truck driver, the goods are inspected after offloading them. Since the goods inspection proceeds without complaints, the goods are placed into stock. In case of inspection complaints, the goods would have been rejected."

Goods Handling Process – BPMN Representation



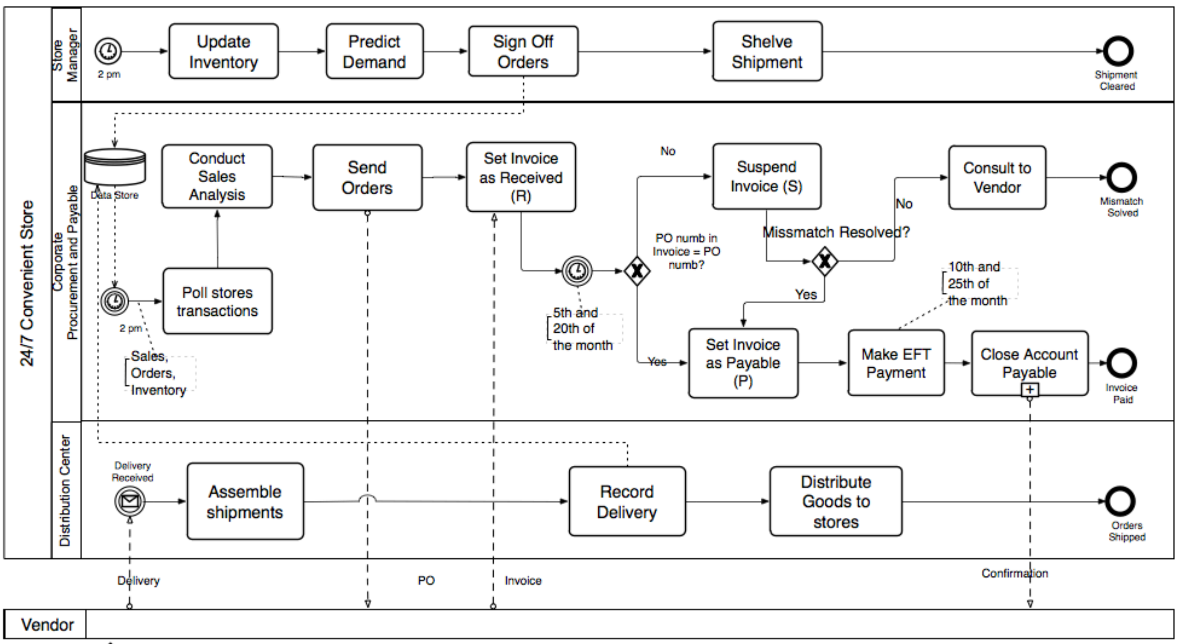
Remark: **Exclusive Gateway**. There can be only one path leaving the gateway/decision point.

Parallel Gateway. All available paths are executed in parallel after leaving the gateway/decision point.

Procure-to-Pay – Textual Representation

Daily between noon and 2 pm, each store manager updates the inventory with a handheld computer that synchronizes the inventory of the in-store computer and predicts what the store needs for the next day, before signing-off. The store manager signs off on the store order by 2 pm to corporate database. At the same time, 24-Seven Company Corporate polls stores electronically for sales, inventory, and orders from the updated corporate database, and analyzes sales data by store, region, and overall. Based on the sales analysis, Corporate prepares store orders and posts purchase orders (PO) for vendors. The corporate's distribution center, after receiving delivery notification from all vendors, assembles shipments for stores, records all deliveries and distributes goods to all stores. Each store clerk shelves the shipment of inventory. Monthly, 24-Seven's corporate procurement and payable staffs sets the status of the invoices received from vendor to 'R' for 'received.' On the 5th and 20th of the month, the system matches all unprocessed invoices to purchase orders. PO number and amounts on the invoice is matched with PO number on the PO. When not matched, invoice status is set to 'S' for 'suspended', and payables staff resolves the mismatch. When finally matched, the invoice status is set to 'P' for payable. If still not matching, the payable clerk will consult the vendors. On the 10th and 25th of the month, electronic funds transfer (EFT) payment are made. To close the prior month, the accounts payable is closed and confirmation of payment is sent to the respective vendor."

Procure-to-Pay – BPMN Representation



Remark: **Exclusive Gateway**.There can be only one path leaving the gateway/decision point.

Parallel Gateway. All available paths are executed in parallel after leaving the gateway/decision point.

Appendix B – Demographic questions

Please answer each question as precisely as possible, if you are unsure please make an estimate.

----PAGE 1----

1. Have you ever used Text to document a business process? [Yes or No]
2. Have you ever used Diagrams to document a business process? [Yes or No]
3. Overall, I am very familiar with the following method to document a process.
   [1=not at all familiar ------ 7=Very familiar]
4. Estimate how INTENSIVELY you have worked with the following methods in the last 4 (four) years.
   [1=not at all intensive ------ 7=Very intensive]
5. Estimate the level of COMPETENCE that you have attained in using the following methods for process documentation.
   [1=not at all competent ------ 7=Very competent]
6. Estimate the level of CONFIDENCE that you have attained in understanding process documentation with the following methods.
   [1=not at all confident ------ 7=Very confident]

----PAGE 2----

Knowledge of risk, business process, internal control, accounting information system

1. Please rate your level of knowledge of business processes.
   [1=no knowledge at all ------ 7=Very high knowledge]
2. Please rate your level of knowledge of audit and risk.
   [1=no knowledge at all ------ 7=Very high knowledge]
3. Please rate your level of knowledge of internal control.
   [1=no knowledge at all ------ 7=Very high knowledge]
4. Please rate your level of knowledge of accounting information systems.
   [1=no knowledge at all ------ 7=Very high knowledge]

Knowledge of business case (accounting cycles)

5. Indicate your level of knowledge of the following business processes:
   a. A Trading Company Goods Receipt/Handling Process
   b. A Retailer Purchase-to-Pay Process

[1=not knowledge at all ------ 7=Very high knowledge]

6. Check or place a mark on the boxes of the activities listed below that you have been involved in:

• Operated a goods handling activity	• Worked at a retailing activity
• Received a delivery note	• Verified Invoices
• Reviewed a purchase order	• Placed a purchase order
• Inspected goods received	• Managed an invoice payment
• Made decision about goods delivery	• Recorded account payable transaction

Appendix C – Tasks on Goods Handling Process case

----PAGE 1 Manipulation Checks ---- Multiple-choice

1. How many activities does the process have?
2. Does the documentation depict a manufacturing process?
3. Does this process start with rejection of goods?
4. Does this process only allow for delivery with purchase order?

----PAGE 2 Search/Recognition Questions ---- Multiple-choice

1. What is the subsequent task for looking up procurement rules?
2. Which task handles the notification of acceptance/rejection authorization from the booking clerk?
3. Which party handles the approval for a delivery without purchase order?
4. Which task is performed before inspecting goods quality?
5. Which party initiates the goods handling process?
6. How many decisions have to be taken by a delivery note without purchase order to be placed in the stock?
7. What type of delivery does the good receipt officer require in order to contact the booking clerk?
8. Which activities should the booking clerk perform after reviewing the procurement rules?
9. List six activities that are observable from the process documentation in a case when a delivery note without purchase order (PO) is approved **from the point when** the delivery is identified as 'without PO' until the goods are placed in stock. List the activities in the correct order (1 to 6).

----PAGE 3 Inference Questions ---- Yes/No/Unknown

1. Are deliveries without purchase order automatically rejected?
2. Can the goods quality be inspected before a delivery ramp is determined?
3. Is the booking clerk responsible for acceptance decisions of goods without purchase orders?
4. Can goods be rejected for multiple reasons?
5. Does the booking clerk notify the goods receipt officer via a text message?
6. Can goods be placed into multiple warehouses?
7. Can goods be rejected immediately after they have been assigned a delivery ramp?
8. Can goods still be accepted without a delivery note?
9. Can goods be rejected after they have been inspected?
10. Have all goods quality placed in stock been inspected?

----PAGE 4 Problem-solving questions --- Open

1. A delivery had been successfully inspected and was placed in stock. It was found out the next day that it was a delivery without accompanying purchase order in which no approval was given by the booking clerk.
    - List any identifiable risks that may affect financial accounts and the company operation as the result of the above situation.
2. The procurement manager wants to reduce the rejected goods rate. The high number of rejected goods deliveries reduces the inventory available to sell and creates too many backorders (out-of stock orders), which end up being cancelled.
    - What control procedures need to be enforced to help the manager to sort out the problem?

----PAGE 5 Recall questions --- Cloze test

A truck driver registers at the goods receiving department with a ______________ note to a good receipt officer. The officer identifies the delivery type. In his case, it is a delivery related to a ______________ order. In case of deliveries without a purchase order, a ______________ clerk has to be contacted. The booking clerk shall look up the ______________ rules before ______________ the delivery to be accepted or not. When the decision has been made, the booking clerk ______________ the goods ______________ officer to execute the acceptance or rejection and ______________ the receiving transaction in file. Following the assignment of a delivery ramp to the truck driver, the goods are ______________ after offloading them. Since the goods inspection proceeds without complaints, the goods are ______________ into stock. In case of inspection complaints, the goods would have been rejected.

Acceptable answers for the problem-solving task

1. List any identifiable risks that may affect financial accounts and the company operation as the result of the above situation.

Financial Account Risks

Effect on Cash and Expenses

Understated Cash

Unnecessary shipping tax and costs

Fake payment

Illegitimate disbursement of cash

Negative cash flow

Cash loss

Supplier/Vendor

Fake supplier

Unwanted order

Incomplete order

Account Payable

Incorrect AP amount is posted

Incorrect AP balance

More invoices unnecessarily paid

Overstated Payable

Inventory

Incorrect / Overstated inventory record

All accepted inventories are not made to the supplier master file

No segregation of duty between the custody, recording and cash holders

Overstated COGS

Matching principle on COGS and Revenue was violated

Operational Risks

Purchase or Procurement

Unauthorized goods receipt

Goods receipts are not recorded in transaction log

Non-existent purchase

Goods receipts are not placed for approved requisitions, undetected defect

Unqualified goods quality

Collusion of vendors and warehouse clerks go unnoticed

Bogus transactions

The company may pay twice for the same order.

Fictive order ;Inventory theft

The goods are not recorded in the system

Goods already received have different specification with company's inventory policy

Warehouse overcapacity

Unmatched goods

Unused stock, stock obsolence

2. Which control procedures need to be enforced to help the manager to sort out the problem?

   Shipments should be recorded

   Enforce minimum inventory control

   Constant monitoring over inspection process

   Maintain examination over rejection rate by vendor

   Consistent review on supplier performance

   Pre-delivery sample

   Improve workforce and quality assurance on the offloading and inspection proces

   Pre-delivery inspection in vendor site

   Vendor evaluation

   Improve warehouse - shop window visibility

   Minimum buffer for stock

Appendix D – Tasks on Purchase-to-Pay case

----PAGE 1 Manipulation Checks---- Multiple-choice

1. How many activities does the process have?
2. Does the documentation depict a purchasing process?
3. Does this process start with updating inventory?
4. Does this process deliver orders to a distribution center?

----PAGE 2 Search/Recognition Questions ---- Multiple-choice

1. What is the next task after setting invoice as payable?
2. What activity follows the EFT payment?
3. Which division handles the approval for a matched invoice number?
4. Which task is performed before recording the delivery?
5. Which party firstly initiates the procure-to-pay process?
6. How many decisions have to be taken for a suspended invoice to be set as Payable from the point when it is received from vendors?
7. Which type of condition requires an invoice to be suspended (S)?
8. Which activity follows when the mismatch goes unresolved on suspended invoices?
9. List six activities that are performed by the corporate payable staffs **excluding** activities that handle unmatched and suspended invoices. List the activities in the correct order (1 to 6).

----PAGE 3 Inference Questions ---- Yes/No/Unknown

1. Is the matching of POs and Invoices always performed every day?
2. Are vendors' invoices always approved for payment?
3. Is the corporate procurement staff responsible for distributing delivered goods by vendors to store?
4. Can store managers immediately receive goods after they sign off the order?
5. Does the store manager make demand predictions at a particular time?
6. Can goods be delivered into multiple warehouses?
7. Can goods be delivered directly to each store?
8. Can a payment be made for unmatched PO numbers and invoices?
9. Are suspended invoices stored in corporate files manually?
10. Is it possible to have a PO number that is not matched with the invoice?

----PAGE 4 Problem-solving questions --- Open

1. When matching invoice amount with the associated purchase order (PO), there are invoices whose number correspond to different POs. Those invoices actually have been paid two months ago as evidenced by prior PO number stored in the PO spreadsheet.
   - List any identifiable risks that may affect financial accounts and the company operation as the result of the above situation.
2. A store manager has submitted the order for tomorrow's shipment. But when the shipment takes place at the distribution center the next day, he can only find 75% of his orders to be correct, while the remaining have not been received yet or even considered unordered.
   - Which control procedures need to be enforced to help the manager to sort out the problem?

----PAGE 5 Recall questions --- Cloze test

Daily between noon and 2 pm, each ____________manager updates the ____________ with a handheld computer that synchronizes the inventory of the in-store computer and predicts what the store needs for the next day, before signing-off. The store manager signs off on the store order by 2 pm to ____________ database. At the same time, 24-Seven Company Corporate ____________ stores electronically for sales, inventory, and orders from the updated corporate database, and analyzes sales data by ____________, region, and overall. Based on the ____________ analysis, Corporate prepares store orders and posts purchase orders (PO) for vendors. The corporate's ____________ center, after receiving delivery notification from all vendors, assembles shipments for stores, ____________ all deliveries and distributes goods to all stores. Each store clerk shelves the shipment of inventory. Monthly, 24-Seven's corporate procurement and payable staffs sets the status of the invoices received from vendor to 'R' for 'received.' On the ____________and 20th of the month, the system ____________ all unprocessed invoices to purchase orders. PO number and amounts on the ____________ is matched with PO number on the ____________. When not ____________, ____________status is set to 'S' for 'suspended', and payables staff resolves the mismatch. When finally matched, the invoice status is set to 'P' for ____________. If still not matching, the payable clerk will consult the vendors. On the 10th and ____________ of the month, electronic ____________ transfer (EFT) payment are made. To close the prior month, the accounts payable is closed and confirmation of payment is sent to the respective vendor.

Acceptable answers for the problem-solving task

1. List any identifiable risks that may affect financial accounts and the company operation as the result of the above situation.

Financial Account Risks

Incorrect cut-off of account payable

Incorrect account payable amount is posted

Double disbursement to the same vendors

Overstated inventory record

Payment made to non-existent vendors

Unauthorized goods receipts

Collusion between vendors and warehouse clerks go unnoticed

Unqualified goods quality per spec

Bogus transactions hidden under different PO no

Cash loss

Overstated account payable posted

Payment failure from original PO transaction

Negative impact on cash budget

Untrustworthy vendors

Failed cash management

Incorrect bank transfer/payment

Outstanding balance account payable due to failed redemption

Incorrect invoice number

2. Which control procedures need to be enforced to help the manager to sort out the problem?

POs are placed only for approved requisitions and invoices.

POs are accurately and completely entered.

All POs issued are input and processed.

Account Payable (A/P) are adjusted only for valid reason.

The A/P payments should be checked against the open-only POs.

Ensuring that the closed PO has already flagged in the system.

Disbursement are made only for goods received.

Segregation of duty between PO entry clerk and the A/P clerk with supported by shared database.

---------- end of document ------------