Bus Inf Syst Eng 62(2):87–101 (2020)
https://doi.org/10.1007/s12599-019-00579-2

RESEARCH PAPER

Developing Serious Games with Integrated Debriefing

Findings from a Business Intelligence Context

Christian Karl Grund · Michael Schelkle

Received: 5 December 2017 / Accepted: 8 November 2018 / Published online: 21 January 2019
© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2019

Abstract Serious games (SG) are recognized in several domains as a promising instructional approach. When it comes to the field of Information Systems, however, they are not yet broadly investigated. Especially in business intelligence and analytics, a literature review indicates the absence of SG for proper report design. Such games, however, seem beneficial since many business reports suffer from poor business information visualization (BIV). To address this issue, the scope of the study is twofold: first, the paper presents a SG that aims to foster learning about BIV. Second, it evaluates this SG in a laboratory experiment, comparing it to a more conventional instructional approach (i.e., presentation) and testing two different versions of the game: One version integrates debriefing into the game itself, whereas the other version uses classical post hoc debriefing. Results indicate that it is favorable to integrate debriefing into the game in terms of motivation and learning outcomes. In the vein of design science research, the authors thus intend to contribute a useful artifact as well as a novel design principle for this instructional approach: Integrating debriefing into SG.

Keywords Serious games · Debriefing · Business information visualization · Business intelligence · Design science research · Laboratory experiment · Evaluation

Accepted after one revision by Matthias Jarke.

C. K. Grund (✉) · M. Schelkle
Faculty of Business and Economics, University of Augsburg, Augsburg, Germany
e-mail: christian.grund@wiwi.uni-augsburg.de

M. Schelkle
e-mail: michael.schelkle@wiwi.uni-augsburg.de

1 Introduction

Serious games (SG) are recognized in several domains as a promising instructional approach (Connolly et al. 2012). Examples include health care (Basole et al. 2013), computer science (Papastergiou 2009), and business (Faria et al. 2009). Among the desired and often realized outcomes of these games are increased motivation and learning (Connolly et al. 2012; Grund 2015; Wouters et al. 2009). Despite its popularity in other domains, the field of information systems (IS) has not yet broadly investigated this instructional approach, although technology-related learning plays an important role for instance in digital transformation processes in organizations (Matt et al. 2015; Legner et al. 2017). While there are some studies about SG in the field of IS, they are seemingly not yet discussed in publications following the design science research (DSR) paradigm (Grund and Meier 2016). Hence, there is still a major opportunity for the field of IS to gain insights about how to design effective SG that help organizations to train their employees in IS-related skills.

One of the most prominent IS-related capabilities for future employees is handling the ever increasing amount of information (Chen et al. 2012). This includes analytical skills, business and domain knowledge as well as communication skills (Chen et al. 2012). Especially the latter often seems to be not prominently investigated in the domain of business intelligence and analytics (BI&A). This domain instead focuses mostly on analytical aspects like how to mine big data and not how the resulting findings are best presented to target audiences (Chen et al. 2012). Not surprisingly, many business reports (i.e., where results are communicated) suffer from poor business information visualization (BIV) (Beattie and Jones 2008). Since decision makers relying on these flawed reports may be misled, it appears beneficial to develop SG with this focus to equip employees with appropriate reporting skills. Although the BI&A domain already provides some studies about SG, none of these games focus on report design and BIV yet (Grund and Meier 2016).

To fill this gap, we set out to develop a SG that aims to increase BIV capabilities (namely being able to identify inadequate BIV and being able to suggest reasonable improvements) among players by letting them compete across several minigames (Grund and Schelkle 2016). Each minigame confronts players with insufficient BIV, which they are supposed to avoid when designing reports. While prior research focused mainly on describing the development and architecture of this SG (Grund and Schelkle 2016; Grund et al. 2017), the current study emphasizes its thorough evaluation. In particular, we are interested in the differences between learners playing our SG, and learners in a more conventional training condition (i.e., a presentation about the same BIV guidelines). Hence, we pose our first research question:

> RQ1: Which effects on motivation and learning outcomes has using serious games for business information visualization compared to presentations?

One of the most important concerns of DSR is to generate knowledge about how an artifact is best designed to fulfill its purpose, which often includes designing different alternatives of an artifact (Hevner et al. 2004). For the development of SG, there are several possible design choices that may be investigated, including which game elements to use (Blohm and Leimeister 2013), how to connect educational content with game content (Charsky 2010) as well as how to facilitate the reflection on experiences after the game (Lederman 1992). This last design aspect, which is often referred to as "debriefing", is considered an essential part of any SG, where instructors discuss the learning content of the game after the experience to ensure *learning outcomes* (Garris et al. 2002). Many scholars even consider this the most crucial part of SG (Lederman 1992; Crookall 1992), since experiential learning has to be accompanied by appropriate learner support for effective learning to happen (Garris et al. 2002; Kolb 1984). Despite its importance for learning in SG, this design aspect is often not prominently investigated or even ignored by SG scholars (Crookall 2010). In addition, the conventional approach of conducting debriefing after the game experience may be costly and time-consuming, since it requires participants of SG to be spatially and/or temporally synchronized with an instructor or so-called "debriefer" (Lederman 1992). To overcome this drawback, integrating the debriefing into the game itself may be a viable solution. However, prior research has thus far not directly compared integrating debriefing into the game with conducting it in an often advocated post hoc manner. We therefore pose our second research question to investigate this design principle:

> RQ2: Which effects on motivation and learning outcomes has integrated debriefing in comparison to post hoc debriefing as a design principle for Serious Games?

To address these research questions, we developed a SG for BIV and evaluated it in a multivariate 1 × 3 between-group laboratory experiment at a German University. Two groups played different versions of the game and one group was attending a presentation about the same learning content, which represented a more conventional training method. In this paper, we present and discuss the results of this experimental evaluation. Hence, this article is structured as follows: first, we describe our terminology and related work in Sect. 2. Second, the theoretical background alongside hypotheses for the evaluation are presented in Sect. 3. Section 4 provides a brief description of the developed artifact which is evaluated in Sect. 5. The paper closes with a discussion and conclusion as well as an outlook on future research in Sects. 6 and 7.

2 Terminology and Related Work

In the following, we describe the terminology as well as related work for both SG that foster BIV skills and debriefing in SG.

2.1 Serious Games for Business Information Visualization

To investigate whether there are similar approaches to our proposed SG, we aim to characterize the state of the art of BIV as a learning goal or a learning outcome in SG. In this context, information visualization is defined as using computer-supported, interactive graphical representations of abstract data to amplify cognition (Card et al. 1999). When information visualization technologies are used to depict business information (e.g., with tables or column charts) it is referred to as BIV (Tegarden 1999). SG may be characterized as games that have an "explicit and carefully thought-out educational purpose and are not intended to be played primarily for amusement" (Abt 1987). In contrast to gamification, where game design elements are used in a non-game context, SG are full-fledged games that serve an educational purpose (Deterding et al. 2011). In our case, we thus intend to identify SG that incorporate BIV capabilities as their educational purpose.

In a basic overview of SG, Susi et al. (2007) find that communication skills (i.e., effectively presenting ideas when speaking, writing, etc.) are important for employees in corporations. Although this might include BIV, this learning goal is not explicitly stated. Connolly et al. (2012) investigate empirical evidence on the *learning outcomes* of computer games and SG in a systematic literature review. Out of the 129 publications they identified, 17 higher quality studies report knowledge acquisition and content understanding outcomes. However, none of these studies mention BIV as a learning outcome. Another literature review about the *learning outcomes* of SG conducted by Wouters et al. (2009) concludes that cognitive *learning outcomes* (i.e., knowledge and cognitive skills) can be observed in 12 out of the 28 empirical studies investigated. Although they argue that SG seem to be effective when it comes to cognitive *learning outcomes*, BIV was again not a learning goal in any of the studies. In a recent literature review about using SG to improve the decision process, Grund and Meier (2016) show that BIV is not addressed in their sample of SG that include business reporting. In summary, according to the investigations mentioned above, SG that specifically focus on improving BIV skills seem to be still missing. We intend to fill this gap with the SG described in Sect. 4.

2.2 Debriefing in Serious Games

As mentioned above, debriefing plays a crucial role when it comes to SG. In an experiential learning context, debriefing may be defined as a process that allows participants to process meaningful experiences that happened during an activity, thus facilitating learning (Lederman 1992). It is important to note that in this definition, debriefing takes place after learners have engaged in a learning activity, often in a guided discussion. This is also reflected in prior research on debriefing in SG.

In a special issue in 1992, the journal Simulation & Gaming called for research articles focusing on debriefing, since this topic seemed to be neglected by too many scholars (Crookall 1992). Following this call, researchers contributed definitions of debriefing (Lederman 1992), practical recommendations (e.g., Steinwachs 1992), and technologies for debriefing (Thiagarajan 1992). Ever since, research on debriefing in SG discussed how to design debriefing sessions and what makes debriefing effective (Kriz 2010; Rudolph et al. 2008; Pavlov et al. 2015; Qudrat-Ullah 2007; Der Sahakian et al. 2015). In an effort to provide a structure for the reflection phase in debriefing, Kriz (2010) lays out several parameters that may be taken into account, including the role of debriefers, the use of media, oral vs. written debriefing, etc. However, whether debriefing is integrated into the activity is not among these parameters. Instead, he only mentions that when the game is too lengthy, several small rounds of debriefing may be performed after each game round. This is, however, not an integration of the reflection into the game itself as debriefing and the gaming activity are still separated. Rudolph et al. (2008) propose that debriefing might be conducted as formative assessment. In contrast to summative assessment, where feedback is given after the activity, formative assessment immediately addresses shortcomings of participants (Rudolph et al. 2008). Although this approach seems similar to integrating debriefing into the learning activity, it focuses on giving feedback to increase participants' performance during the activity, rather than fostering reflection about the meaning of the activity. The literature reviewed above shows that while the importance of debriefing is undisputed in the field of SG, studies explicitly investigating the differences between integrated debriefing and post hoc debriefing seem to remain elusive. Hence, we examine this matter by utilizing two different versions of our SG. To lay out our reasoning as to why we expect differences between these two approaches, the theoretical background of this study is described below.

3 Theoretical Background and Hypothesis Development

Since SG are concerned with improving player capabilities as well as providing an entertaining experience, both learning and motivation theories are used in literature to explain the benefits of SG (Grund 2015; Wu et al. 2012; Ryan et al. 2006). To explain the motivational effects of our SG, we draw on self-determination theory (Deci and Ryan 1985). One of its central assumptions is that *intrinsic motivation* [i.e., when individuals engage in behavior for the pleasure and satisfaction that they inherently experience with participation (Deci and Ryan 1985)] requires the satisfaction of three basic psychological needs: Competence, relatedness, and autonomy. Findings in the context of self-determination theory show that video games in general foster *intrinsic motivation* by fulfilling these needs (Ryan et al. 2006).

In our case, perceived competence may be fostered by players succeeding in the different minigames and earning points for doing so. Since players in a competition are unlikely to form meaningful social bonds, relatedness as it is described in self-determination theory may not directly be established by our SG. However, by having players compete with each other and using a leaderboard that allows for comparisons with other players, they might get a feeling of each other's social presence, which may be regarded a prerequisite for relatedness. Last, a sense of autonomy may be achieved by players being able to choose their own approaches of how to succeed in the minigames.

In contrast, participants who only attend a presentation are not expected to experience competence, since they are only passively consuming (i.e., not receiving any performance feedback). Furthermore, we expect participants only attending a presentation to experience less social presence, because they are not supposed to interact with each other. Last, perceived autonomy is expected to be below the participants in a SG setting, since only attending a presentation does not include influencing the course of actions. Resulting from these anticipated differences, we expect that participants in any SG condition will perceive higher *intrinsic motivation* than participants not playing the SG, since fulfilling these psychological needs fosters *intrinsic motivation* (Sheldon and Filak 2008; Ryan and Deci 2000). Often accompanied by increased *intrinsic motivation* is an increase in the perceived task value (Ryan 1982). In our case, this task value refers to whether participants deem the learning activity as important and adequate for learning about BIV. Hence, we propose that participants who play any version of the SG show increased motivational outcomes compared to participants in a presentation setting according to self-determination theory. This leads to our first group of hypotheses:

H1a Participants who play any version of the serious game will experience higher autonomy than participants only attending a presentation.

H1b Participants who play any version of the serious game will experience higher competence than participants only attending a presentation.

H1c Participants who play any version of the serious game will experience higher social presence than participants only attending a presentation.

H1d Participants who play any version of the serious game will experience higher intrinsic motivation than participants only attending a presentation.

H1e Participants who play any version of the serious game will experience higher task value than participants only attending a presentation.

When it comes to expected differences between the two versions of our SG, the basic psychological needs described in self-determination theory may be used to provide possible explanations. As mentioned above, the first version of our SG includes debriefing during the gameplay, whereas the second version uses debriefing after the game ("post hoc debriefing"). Hence, in both versions, players still solve the same tasks and compete identically, which is why we do not expect differences in either perceived competence or social presence. However, we do expect a difference in perceived autonomy. The reason for this is that players who receive a debriefing after the game may perceive a shift in their locus of control, meaning that they no longer control what is going on after playing. Instead, either the debriefer or a debriefing video determines all following events. In contrast, when the meaning of the exercise is presented during the game, players may still opt to simply close this description after reading it, thus still being able to control what is being displayed and for how long. Since a change in any of the psychological needs may have an impact on *intrinsic motivation* (Sheldon and Filak 2008), we further expect the *intrinsic motivation* of the integrated debriefing group to be higher due to a higher feeling of autonomy. Again, this may also positively impact the perceived task value of the group with integrated debriefing. Hence, we derive our second group of hypotheses:

H2a Participants who play the serious game with integrated debriefing will experience higher autonomy than participants who play the game with post hoc debriefing.

H2b Participants who play the serious game with integrated debriefing will experience higher intrinsic motivation than participants who play the game with post hoc debriefing.

H2c Participants who play the serious game with integrated debriefing will experience higher task value than participants who play the game with post hoc debriefing.

Regarding the desired *learning outcomes*, prior studies suggest that participants who engage in experiential learning (e.g., playing SG) rather than only attending a presentation, show higher observed *learning outcomes* (Connolly et al. 2012; Wouters et al. 2009). The theoretical underpinning of this increased learning success is rooted in experiential learning theory (Kolb 1984). Its main rationale is that individuals learn most effectively when they reflect on concrete experiences and actively experiment based on the resulting conceptualizations (Kolb 1984). Since SG allow players to go through all stages of the so-called learning cycle, we expect participants engaging in our SG to show higher observed *learning outcomes* than participants only attending a presentation. However, this is not the only reason for possible differences between the groups. The anticipated differences in *intrinsic motivation* may also lead to differences in observed *learning outcomes*, since several studies suggest a positive relationship between *intrinsic motivation* and learning (e.g., Kusurkar et al. 2013; Taylor et al. 2014). Based on the anticipated differences in *intrinsic motivation* described above, we thus propose our third group of hypotheses:

H3a Participants who play any version of the serious game will show higher learning outcomes than participants only attending a presentation.

H3b Participants who play the serious game with integrated debriefing will show higher learning outcomes than participants who play the game with post hoc debriefing.

To investigate these hypotheses, we will evaluate our SG after briefly describing it in the following section.

4 Artifact: Dashboard Tournament

To develop the Dashboard Tournament, we employed the human-centred design process (see Grund and Schelkle 2016 for details). For its implementation, we used the game engine unity with C# as its programming language. An overview of the game's technical architecture is provided by Grund et al. (2017). In the following, we briefly describe the game's educational purpose as well as its structure (for a more detailed description see Grund and Schelkle 2016, 2017).

4.1 Educational Purpose

As mentioned earlier, the Dashboard Tournament aims at improving BIV skills of players. A possible approach to improve these skills is conveying visualization guidelines that inform report design decisions. Although several guidelines for information visualization exist (e.g., Ware 2012), only few focus on elements used specifically in business reports. One framework that highlights the design of business reports and presentations is called International Business Communication Standards (IBCS) (Hichert and Faisst 2015). This framework comprises specific guidelines that showcase examples of poor BIV alongside their proposed corrections. We hence incorporated these guidelines in our SG to enable players to identify inadequate BIV and to suggest reasonable improvements. These two skills, namely being able to identify inadequate BIV and being able to suggest reasonable improvements, are what we refer to as BIV skills in this study. The specific guidelines included in our SG are described in the following alongside the structure of the game.

4.2 Game Structure

The Dashboard Tournament is a multiplayer SG featuring a competition across four minigames (Grund and Schelkle 2016). Each minigame addresses one specific guideline for adequate BIV from different perceptual IBCS rule sets (Hichert and Faisst 2015). For each minigame, there is only limited time for players to fulfill their task, which induces a sense of urgency but also helps to ensure that there are no long waiting times emerging for fast players. They can score between 0 and 100 points per minigame that are displayed in a global leaderboard after finishing (for details see Grund and Schelkle 2017). The game ends when every minigame is finished and the overall results (i.e., a leaderboard featuring points and ranks) are announced. An overview of all four minigames implemented in the Dashboard Tournament is provided in Fig. 1.

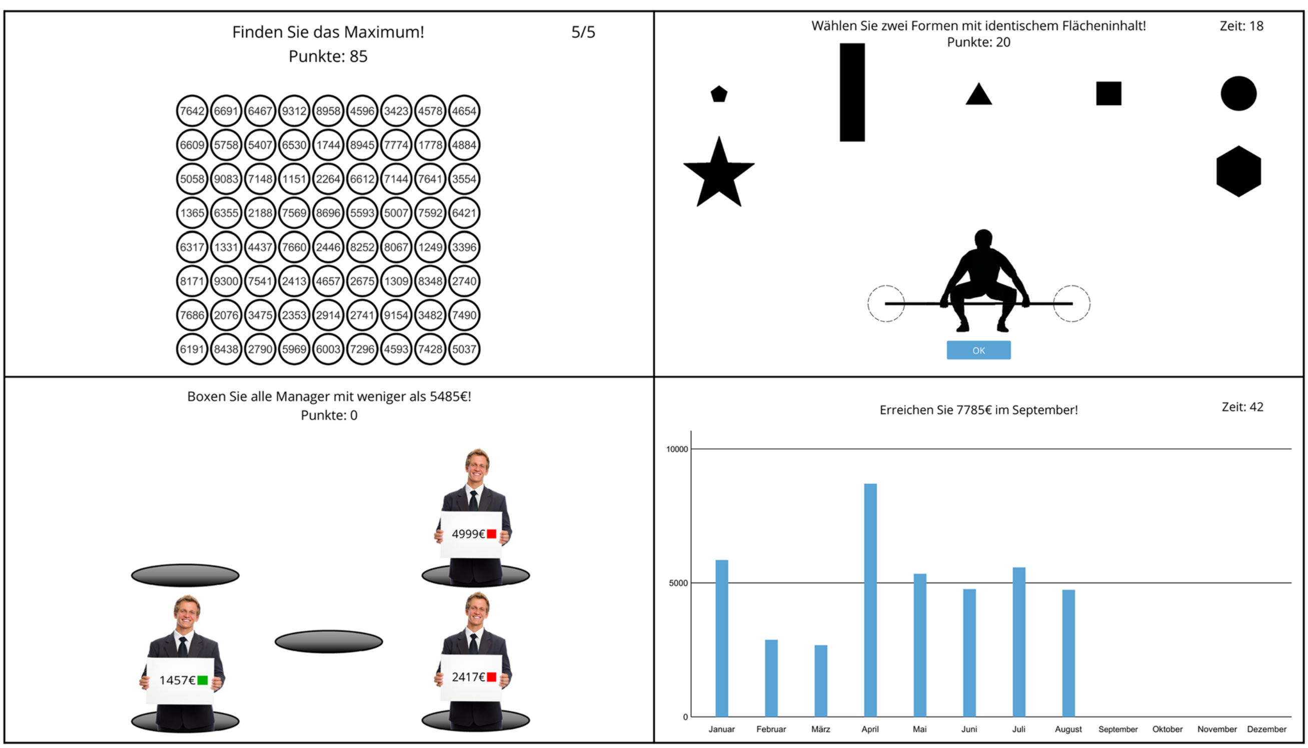

Fig. 1 Minigames implemented in the Dashboard Tournament (screenshots from the software used in the experiment)

The first minigame (upper left image in Fig. 1) addresses the guideline CO 4.4 (Hichert and Faisst 2015). This guideline recommends using graphical elements in tables to easily identify differences in size between numbers. The basic layout of the minigame is a grid of targets with numbers (similar to a table) without graphical support. Players only have limited time to identify the maximum value and lose points the longer they need to accomplish the goal. This is to demonstrate that having no graphical support in tables and the resulting high cognitive effort is slowing them down in what they are trying to achieve. To show that this problem is amplified with more numbers, there are five rounds in this minigame: each round adds more targets with numbers that are potentially the maximum value.

In the second minigame (upper right image in Fig. 1), the guideline CH 3.1 is covered (Hichert and Faisst 2015). This guideline advises against using area comparisons in reports (which is the case for example with pie charts) and instead suggests using length comparisons. To experience the difficulty of correctly comparing area sizes, players have to select two shapes with identical areas out of several different shapes and attach them to a weightlifting bar. As with the previous minigame, there are five rounds with increasing difficulty by adding more shapes to choose from in each round. Although there is a time limit in each round (20 s), scores only depend on correct solutions. The time limit serves mainly as a means to reduce waiting times for fast players.

The next minigame (lower left image in Fig. 1) is concerned with the guideline EX 2.5 (Hichert and Faisst 2015). This guideline disadvises from using traffic light indicators in reports, since they distract from comprehending the actual numbers. To show this effect, players have to hit all managers holding numbers below a given threshold in a "Whac-A-Mole"-style minigame. Hitting the correct targets increases the score while hitting the wrong ones decreases it. Inconsistencies between the traffic light colors and the numbers lead to wrong decisions when players blindly trust the traffic light indicators.

The last minigame (lower right image in Fig. 1) addresses the guideline SI 3.1 (Hichert and Faisst 2015). This guideline recommends replacing value axes in column charts with data labels. Players are given a target value and hold a key to "grow" a column with the corresponding height. With ongoing progression, gridlines start to disappear, gradually increasing the difficulty of the minigame. Although there is a time limit of 60 s for the whole minigame, the score only depends on the deviations from the correct values. Through this, players experience difficulties when estimating the exact height given only a value axis and gridlines.

The experienced difficulties in all four minigames lay the foundation for debriefing, where experiences may be reflected upon (Lederman 1992). As mentioned in Sect. 2, literature in the domain of SG suggests conducting a debriefing session after the learning activity took place (i.e., after all minigames are completed). To investigate the differences between this approach and integrating debriefing into the game itself, we developed two versions of the game: The first version shows participants the corresponding IBCS guideline after each minigame, explaining why several kinds of BIV should be avoided in business reports ("integrated debriefing"). In the second version, these explanations are missing and participants only play the minigames. Therefore, in the second version of the game, a conventional debriefing is required after the game for learning to take place ("post hoc debriefing"). These two versions of the game are used in the experimental evaluation of our artifact which is described below.

5 Evaluation

To evaluate our artifact, we conducted a laboratory experiment. In the following, we describe the study setup, the development of the measurement instrument, as well as the results of this experimental evaluation.

5.1 Method, Participants, and Design

Following the DSR paradigm, this study aims to evaluate our developed artifact in order to generate design knowledge (Hevner et al. 2004). The purpose of this evaluation is twofold: First, we aim to evaluate an instantiation of our designed artifact to establish its utility and efficacy for achieving its stated purpose (Venable et al. 2012), namely increasing motivation and learning. Second, we intend to evaluate our designed artifact in comparison to other designed artifacts' ability to achieve a similar purpose (Venable et al. 2012), as we seek to compare our SG featuring integrated debriefing with our SG using post hoc debriefing. Since an artificial evaluation environment provides the benefit of controlling for possibly confounding circumstances and since the artifact has already been developed ("ex post evaluation"), we chose to conduct a laboratory experiment using a multivariate 1 × 3 between-group design, as suggested by Venable et al. (2012). Participants were recruited at a German University and comprised different fields of study. Since our SG targets laypersons in report design and since BIV is relevant in many professional domains, the sample was not limited to business students. Every participant received a monetary compensation for being included in the study. The

Table 1 Demographics

	Group 1	Group 2	Group 3	Total
Gender				
Male	7 (37%)	10 (59%)	9 (56%)	26 (50%)
Female	12 (63%)	7 (41%)	7 (44%)	26 (50%)
Age				
18–24	14 (74%)	14 (82%)	11 (69%)	39 (75%)
25–34	5 (26%)	3 (18%)	5 (31%)	13 (25%)
Field				
Business/economics	11 (58%)	8 (47%)	8 (50%)	27 (51%)
Industrial engineering	3 (15%)	2 (12%)	2 (13%)	7 (13%)
Law	2 (11%)	1 (6%)	3 (18%)	6 (12%)
Education	2 (11%)	2 (12%)	2 (13%)	6 (12%)
Others/missing	1 (5%)	4 (23%)	1 (6%)	6 (12%)
Education				
High school degree	11 (58%)	12 (71%)	9 (56%)	32 (62%)
University degree	8 (42%)	5 (29%)	7 (44%)	20 (38%)

demographics of participants are depicted in Table 1, grouped by the treatments described in the following.

Participants have been randomly assigned to one of three groups: the first group played the Dashboard Tournament with integrated debriefing (i.e., corresponding guidelines were shown after each minigame). The second group played an identical game without the guidelines being shown and with a post hoc debriefing afterwards. Last, there was a control group only attending a presentation about the same BIV guidelines. All groups received information about the same BIV guidelines, including the identical pictures of the respective guidelines provided by the IBCS Association (Hichert and Faisst 2015). The duration of each treatment was also similar and reached from approximately 10–15 min per group. To ensure that the debriefing was delivered identically in groups 2 and 3, we used a video of a presentation as debriefing. This presentation included, alongside a general introduction to the topic, the identical guidelines that were delivered in the SG conditions. It was projected onto the front of a classroom to make the experience as close as possible to an actual presentation held by an instructor. Although literature usually suggests that debriefing should be personalized to the learners and include active discussions (Lederman 1992), there are also findings indicating that video-assisted self-debriefing is on par with instructor-guided debriefing (Boet et al. 2011). Since competition and changing leaderboards may confound independency of observations, every participant was shown their own score alongside fictional competitor scores after playing. To assess the motivational effects of each treatment, participants in every group filled out post-experience questionnaires regarding motivational outcomes. For assessing *learning outcomes*, pre-and posttests addressed participants' BIV capabilities. To see whether these acquired capabilities are sustainable, posttests have been conducted one week after the treatment. A summary of this design is presented in Table 2.

The measurement instrument utilized for post-experience questionnaires as well as for pre-and posttests is described in the following.

5.2 Development and Validation of the Measurement Instrument

The measurement instrument for post-experience questionnaires was mainly based on the intrinsic motivation inventory (IMI) that has been used in many studies to measure basic psychological needs as well as *intrinsic motivation* after an experience (Ryan 1982). We included the subscales Interest/Enjoyment (i.e., *intrinsic motivation*), Competence, Autonomy, and Task Value. Changes have been made to the Autonomy subscale, which has been adjusted to express the amount of control and influence participants felt (Grund and Tulis 2017). As described

Table 2 Experimental design of the evaluation

Group (N)	Pretest	Treatment	Post-experience	Posttest
1 (19)	BIV skills	Integrated debriefing	Motivation	BIV skills
2 (17)	BIV skills	Post-hoc debriefing	Motivation	BIV skills
3 (16)	BIV skills	Presentation	Motivation	BIV skills

earlier, we did not measure relatedness of participants but rather social presence as a potential prerequisite for relatedness. For this, we drew from the Behavioral Engagement subscale of the "social presence in gaming questionnaire (SPGQ)" developed by de Kort et al. (2007). To measure participants' overall appreciation of video games, which may arguably confound their motivational outcomes in the treatments with our SG, we used the "Usefulness, Importance, and Interest" subscale from Wigfield and Eccles (2000). In our study, we refer to it as "Game Value", since it expresses how each participant values video games in general. All items adapted and derived from other instruments were modified to relate to the context and translated into German. Items were assessed using a 6-point scale, ranging from 1 = *not at all true* to 6 = *very true*, and were randomized across all subscales. In addition to the questionnaire items, students were provided with space for leaving any comments or suggestions.

To validate the psychometric properties of the resulting instrument and to examine the overall model fit of our measurement model, we conducted a confirmatory factor analysis. This type of analysis is performed to determine whether items used in the questionnaire belonged to the factors anticipated (i.e., convergent validity), and whether factors were distinguishable from each other (i.e., discriminant validity). After minor modifications (e.g., correlated errors, for an overview see Brown 2015), our measurement model reached a satisfactory model fit according to generally accepted thresholds (Hu and Bentler 1999). The ratio between χ^2 and *df* was 1.23, which is below the maximum threshold of 3. The root mean standard error of approximation (RMSEA) was .068 and therefore within the range of acceptable model fit of .08. Last, both comparative fit index (CFI) and Tucker–Lewis index (TLI) are above their common suggested minimum value of .90 (CFI = .92, TLI = .91). We may hence conclude that our measurement instrument achieved a satisfactory model fit, which means that the data is in line with the proposed measurement model. In addition, we accounted for reliability of the scales by computing Cronbach's α, which ranges from .82 to .96 and is hence above the desired minimum of .70 (Krippendorff 2004). To account for discriminant validity, we investigated the square root of the average variance extracted (AVE) of each construct in combination with the correlations between constructs (Fornell and Larcker 1981; Gefen and Straub 2005). As shown in Table 3, each inter-construct correlation lies below the square root of AVE of each construct, hence discriminant validity is demonstrated.

To ensure convergent validity, standardized factor loadings (λ) are investigated for each construct. They range from .55 to .98 and are thus above the recommended minimum of .45 for a fair rating (Tabachnick and Fidell 2013). Overall, construct validity is shown by confirming both discriminant and convergent validity. Table 4 summarizes our measurement model in the post-experience questionnaire and shows its psychometric properties.

Table 3 Square root of AVE (bold) and inter-construct correlations

	IMOT	COMP	AUTO	SOP	TASKV	GAMV
IMOT	**.74**					
COMP	.14	**.76**				
AUTO	.33	.12	**.80**			
SOP	− .37	− .10	.48	**.71**		
TASKV	.47	.18	.31	.09	**.80**	
GAMV	− .07	.42	.10	− .01	.08	**.88**

Learning outcomes have been assessed by comparing participants' initial knowledge of the IBCS guidelines included in our SG with their knowledge about these guidelines after the experiment. For this purpose, participants were provided with different examples of business reports and requested to suggest improvements. The provided reports suffered from inadequate BIV that is addressed by the guidelines covered in the different treatments. To keep participants from simply guessing, we also included obvious other mistakes that were not related to the IBCS guidelines addressed. We could hence check whether improvements suggested by participants complied with the BIV guidelines included in the treatment. If a participant did not suggest an improvement consistent with the IBCS guideline in the pretest but managed to do so in the posttest, we considered this an observed learning outcome of the participant. The flawed business reports presented to participants are shown in Fig. 2.

5.3 Results

As a first analysis, we were interested in whether the perceived game value (GAMV) affects motivational outcomes (e.g., *intrinsic motivation*) among participants in SG conditions. To see potential influences of this variable, we investigated bivariate correlations between GAMV and the dependent variables in our first group of hypotheses (H1a–H1e). These correlations are presented in Table 5.

According to Table 5, there have been significant correlations between GAMV and COMP in both groups. This seems reasonable, since individuals who value video games are more likely to have higher skills in them, thus assessing their own competence in a game-based activity as higher. However, this does not seem to influence other motivational outcomes, especially *intrinsic motivation* does not seem to be affected by GAMV. This might be a first indicator that aversion towards video games in general does not erode the motivational outcomes of the SG.

Table 4 Measurement instrument (post-experience questionnaire)

Factor	Item	M	SD	λ	α
Intrinsic motivation	The session has been fun	4.94	.85	.89	.82
(IMOT)	I thought the session was boring. (R)	5.21	.78	.72	
(Ryan 1982)	I thought the session was quite enjoyable	4.92	.84	.56	
	I enjoyed attending this session very much	5.27	.66	.76	
Perceived competence	I think I was pretty good in this session	4.08	.95	.85	.84
(COMP)	I think I did pretty well in this session, compared to other students	4.06	.94	.71	
(Ryan 1982)	I am satisfied with my performance in this session	4.63	.86	.74	
	I was pretty skilled in this session	3.96	1.00	.71	
Perceived autonomy	In this session I could choose what to do	2.19	1.21	.70	.86
(AUTO)	In this session I had the feeling to be able to co-determine	1.98	1.02	.92	
(Ryan 1982;	I had the feeling to be able to influence the session	2.25	1.27	.66	
Grund and Tulis 2017)	I had the impression to be able to co-determine what happens	2.04	1.08	.91	
Social	During the session, I felt close to the other students	2.06	.94	.55	.86
presence	During the session, I sensed the presence of the other students	2.58	1.29	.91	
(SOP)	During the session, I sensed the attendance of the other students	3.02	1.31	.76	
(de Kort et al. 2007)	During the session, I thought of the other students	2.33	1.28	.65	
	During the session, I was wondering how the other students are doing	3.02	1.61	.70	
	During the session, I was wondering how easy the task might be for the other students	3.44	1.78	.63	
Task value	I believe this session was of value to me	4.19	1.16	.89	.91
(TASKV)	I think this session was well-suited for learning	4.27	1.27	.80	
(Ryan 1982)	I think this session was important to learn something about its content	4.54	1.15	.87	
	I believe this session has helped me gain a better understanding	4.19	1.21	.71	
	I believe that this session was beneficial to me	4.52	1.08	.86	
	I think this session was important	3.90	1.12	.60	
Game value	Video games are interesting to me	3.67	1.75	.98	.96
(GAMV)	Engaging with video games provides fun to me	4.17	1.53	.88	
(Wigfield and Eccles 2000)	Video games have a personal utility for me	2.94	1.59	.87	
	Video games are beneficial to me	2.48	1.32	.87	
	Being good at video games is important to me	2.71	1.46	.77	
	Video games are important to me personally	2.56	1.61	.91	

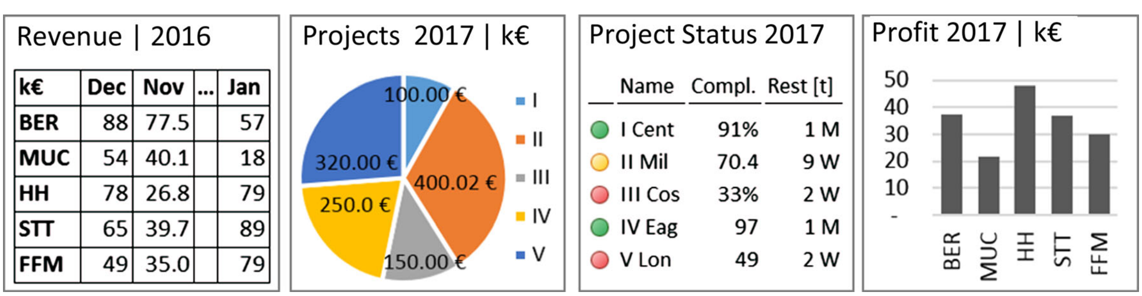

Fig. 2 Flawed business reports (pre-and posttest of BIV skills)

To investigate differences in motivation between our three experimental groups, we conducted a one-way MANCOVA with planned contrasts to test our hypotheses. This method of analysis is specifically useful when intercorrelations between dependent variables are expected (Tabachnick and Fidell 2013), which is the case with our variables measuring different aspects of *intrinsic motivation*. Regarding the requirements for this analysis method, we first checked whether covariance matrices are equal among groups. This is the case, since Box’s M test turned

Table 5 Bivariate correlations with the control variable

	COMP	AUTO	SOP	IMOT	TASKV
GAMV (Group 1)	.69**	.17	.09	.30	.41
GAMV (Group 2)	.57*	.47	.28	− .23	− .03
GAMV (Group 1 + 2)	.59***	.19	.21	− .10	.19

(*$p < .05$, **$p < .01$, ***$p < .001$)

out non-significant ($p = .45$). Next, we used Levene's test for equality of error variances across groups, which turned out to be non-significant for all dependent variables except for perceived autonomy ($p = .046$). Hence, we adjusted the level of significance for this variable to $p = .025$ as suggested by Tabachnick and Fidell (2013). After checking for the requirements, we may proceed with our one-way MANCOVA. To account for the possible differences due to GAMV (see Table 5), we included it as a covariate in our group comparison. As dependent variables, we included all motivational outcomes described in our first group of hypotheses (H1a–H1e). The result of this analysis shows that the treatment led to significant differences between groups with Wilk's $\Lambda = .63$, $p = .016$, and partial $\eta^2 = .207$. Our covariate, namely GAMV, also had a significant impact on group differences with Wilk's $\Lambda = .74$, $p = .020$, and partial $\eta^2 = .256$. To investigate the nature of these differences, we used planned contrasts in line with our hypotheses.

In a first contrast analysis, we aimed at testing our first group of hypotheses (H1a–H1e), namely whether participants in any SG condition show increased motivational outcomes compared to participants in a presentation. Hence, we used simple contrasts comparing the means of the two SG groups with the control group. The results of this analysis are shown in Table 6.

Table 6 shows that, despite theoretically expected differences, there are no significant differences in terms of *intrinsic motivation* (H1d) and satisfaction of basic psychological needs (H1a–H1c) between the SG conditions and the control group. Surprisingly, H1e was supported in the opposite direction, indicating that participants in the control group found the presentation more important and appropriate for learning. Regarding our control variable GAMV, there was a significant impact on COMP ($p < .001$, partial $\eta^2 = .232$). In other words, participants who valued games higher, felt higher competence.

To test our second group of hypotheses, a simple contrast between the two SG groups was used to investigate mean differences. Table 7 shows the results of this analysis.

Although perceived autonomy did not differ significantly between the two groups, the group with integrated debriefing reported significantly higher *intrinsic motivation*. This is interesting, since there is no significant difference in any of *intrinsic motivation*'s antecedents proposed by self-determination theory. In addition, there was no significant difference in perceived task value.

Regarding the *learning outcomes*, we were interested in whether participants were able to increase their knowledge about BIV guidelines in each group. As described earlier, an observed learning outcome shows when participants were not able to make a suggestion in accordance with the IBCS guideline in the pretest, but were able to do so in the posttest. Since this kind of comparison is essentially a within-subject analysis, we used dependent t-tests to observe increases in BIV knowledge for each group. Table 8 shows the results of this analysis.

As can be seen in Table 8, participants who played the SG with integrated debriefing were able to significantly increase their knowledge about all four BIV guidelines. For instance, 32% of the participants in this group were already familiar with the guideline CO 4.4 in the pretest. In the posttest, 68% of the participants were able to make the correct suggestion. This increase of 37 percentage points was statistically significant at the $p < .01$ level. Looking at the *learning outcomes* in the SG group with post hoc debriefing, we find that only knowledge about half of the guidelines presented could be significantly increased (namely CO 4.4 and EX 2.5). Last, in the control group, knowledge about three out of the four guidelines could be significantly increased. These findings indicate that integrating debriefing into SG may yield the highest *learning outcomes*. Using SG with post hoc debriefing, however, seems to be even inferior to conventional presentations. This means that, with regard to hypothesis H3a, we did not find support that using any version of our SG yields higher *learning outcomes* than providing only a presentation: It is

Table 6 MANCOVA results for control group comparisons (*$p < .05$)

H	Construct	M_{G1}	M_{G2}	M_{G3}	$M_{G1} - M_{G3}$	$M_{G2} - M_{G3}$	Support
H1a	COMP	3.78	3.56	3.48	.30	.08	Not supported
H1b	AUTO	2.64	2.11	2.35	.29	− .24	Not supported
H1c	SOP	2.27	2.58	2.15	.12	.43	Not supported
H1d	IMOT	3.41	3.04	3.30	.11	− .26	Not supported
H1e	TASKV	2.88	2.82	3.32	− .44*	− .50*	Supported (opposite)

Table 7 MANCOVA results for comparisons between SG groups (*$p < .05$)

H	Construct	M_{G1}	M_{G2}	$M_{G1} - M_{G2}$	Support
H2a	AUTO	2.64	2.11	.53	Not supported
H2b	IMOT	3.41	3.04	.37*	Supported
H2c	TASKV	2.88	2.82	.06	Not supported

important how the debriefing is integrated into the learning activity. Regarding hypothesis H3b, we found that integrating debriefing into the SG seems superior to conducting it in a classical post hoc manner, since knowledge about twice as many guidelines could be significantly increased.

Regarding participants' comments on their experiences, we conducted a summative qualitative content analysis (Hsieh and Shannon 2005). We investigated two different open questions: First, what did participants like about the session? And second, what should be changed about the session? Answers were manually assigned to categories by the authors in a consensual procedure for each of the SG groups. Only comments about the SG and debriefing were analyzed (not, for instance, comments on the questionnaires used). Table 9 shows which aspects have been mentioned by participants.

As can be seen in Table 9, participants in the SG group with integrated debriefing most often mentioned the debriefing as their favorite part of the game, followed by statements that referred to the game itself as a positive experience (without further differentiation). In the group with post hoc debriefing, however, debriefing was only mentioned by two participants as something they liked about the session. In this group, the game itself received the most positive remarks. This indicates that debriefing was more popular in the group with integrated debriefing. The game overall, however, was apparently appreciated in both groups. Recommendations for improving the game are scattered and span from longer gameplay to improved instructions in the game (i.e., tutorials). They do not indicate a single major issue with the game in both groups. These and other aspects of our results will be discussed in the following section.

6 Discussion

Looking at the results described above, there are several unexpected findings. First and foremost, contrary to what we expected from prior literature (e.g., Ryan et al. 2006), we found no differences in *intrinsic motivation* and

Table 8 Learning outcomes per group (*$p < .05$, **$p < .01$, ***$p < .001$)

Guideline	Integrated debriefing (N = 19)			Post-hoc debriefing (N = 17)			Control group (N = 16)		
	M_{PRE}	M_{POST}	ΔM	M_{PRE}	M_{POST}	ΔM	M_{PRE}	M_{POST}	ΔM
CO 4.4	.32	.68	.37**	.12	.53	.41**	.44	.56	.12
CH 3.1	.16	.63	.47**	.24	.35	.12	.19	.69	.50**
EX 2.5	.05	.42	.37*	.00	.35	.35**	.06	.75	.69***
SI 3.1	.26	.74	.47**	.24	.41	.17	.13	.69	.56**

Table 9 Results of the summative qualitative content analysis

Group	Participants liked	#	Participants wished for	#
Integrated debriefing	Debriefing	7	Longer game	1
	Game overall	5	Longer tutorials	1
	Competition	2	More precise tutorials	1
	Tutorials	2	Less waiting time	1
	Interactivity	1	More comparisons	1
	Feeling of success	1	Longer display of results	1
	Variety	1		
Post-hoc debriefing	Game overall	8	Better video quality	1
	Debriefing	2	Less waiting time	1
	Variety	2	Slower presentation	1
	Competition	1		
	Interactivity	1		
	Tutorials	1		

satisfaction of basic psychological needs in the SG groups compared to the group only attending a presentation. Although particularly the group with integrated debriefing showed higher means in these variables, none of these differences turned out to be significant. In addition, the control group reported significantly higher task value than both SG groups. In other words, participants attending a presentation rated it more appropriate for learning about BIV guidelines than both SG groups. A possible explanation for this might be that students are used to presentations as a prevalent method of knowledge distribution. Hence, when they attend an apparently interesting presentation, they rate it as highly appropriate for learning. In contrast, students are usually not used to play games for learning, they may thus be more hesitant to rate them as a very useful activity. Regarding the lack of motivational differences, the effect size of using SG on the basic psychological needs as well as *intrinsic motivation* may be too small for the present sample size in this study. The effect size of integrating debriefing versus conducting it in a post hoc manner, however, seems to be higher. This is shown by a significant difference in *intrinsic motivation* between these two groups. Participants who played our SG with integrated debriefing enjoyed the experience more than participants who played it with post hoc debriefing. Interestingly, however, this difference may not be explained with the hypothesized difference in perceived autonomy, since it did not turn out to be significant. This finding, alongside the lack of significant differences in satisfaction of basic psychological needs between the SG groups and the control group, may indicate that an additional theoretical lens for describing motivational differences might be beneficial in future studies. Among the potential theories for explaining motivational differences between such groups in future research are for instance flow theory (Csikszentmihalyi 1991) or goal-setting theory (Locke and Latham 2002) which are also used in recent literature to investigate motivational aspects of game-based approaches (Grund 2015).

Differences in *learning outcomes* show that integrating debriefing into SG may not only lead to higher *intrinsic motivation*, but also to increased *learning outcomes*. More specifically, participants who played the game with integrated debriefing were able to significantly increase their knowledge about twice as many BIV guidelines compared to participants in the post hoc debriefing group. This is in line with our expectation that increased motivation in the integrated debriefing group may foster *learning outcomes*. When compared to the control group, participants in the integrated debriefing group showed slightly higher *learning outcomes* and participants in the post hoc debriefing group showed slightly lower *learning outcomes*. This may indicate that when using SG with post hoc debriefing, participants may actually learn less than in a regular presentation. A possible reason for this is the temporal proximity of reflection on the activity. This is in line with research on the role of the immediacy of feedback on learning, stating that timely feedback may improve learning performance (de Freitas et al. 2017; Arbel et al. 2017; Nadolski and Hummel 2017; Gee 2003; Kickmeier-Rust and Albert 2010). While participants in the integrated debriefing group are asked to reflect about each minigame immediately after they played it, participants with post hoc debriefing are forced to remember their experiences in each minigame. Although this does not seem like a daunting task, given that only four minigames are played, this form of debriefing apparently leads to less learning. This confirms findings from educational research that hint at forgetfulness of students in many higher education programs (e.g., Lindsey et al. 2014). Interestingly, although participants in the control group deemed the session as more important and appropriate for learning, they seem to have fewer *learning outcomes* than participants in the integrated debriefing group. This indicates that while SG seem to be able to increase *learning outcomes* compared to conventional training methods, they are not yet recognized as "serious" enough. Regarding the qualitative comments of participants, we also find support for integrating debriefing into SG. While most participants in the group with integrated debriefing mentioned this very debriefing as a positive aspect of the session, only two participants in the group with post hoc debriefing explicitly mentioned the debriefing as something they liked. This finding is similar to what has been reported in studies investigating formative feedback in SG. Here, authors showed that if the feedback is timely and unobtrusive, meaning not hampering the game flow, SG are received better by participants (Nadolski and Hummel 2017).

Regarding the findings discussed above, this study provides several contributions customary to DSR (Briggs and Schwabe 2011). The first mode of inquiry we employed is applied research and engineering, which leads to instances of generalizable solutions, proof-of-concept prototypes, and evidence that solutions are useful and generalizable (Briggs and Schwabe 2011). In our case, we developed and evaluated the (according to our literature review) first SG about BIV, thus contributing a novel artifact to the domain of BI&A. In a laboratory experiment, we showed that this SG is useful for increasing knowledge about BIV guidelines and is appreciated by participants judging by their qualitative comments. When compared to a more conventional instructional approach (i.e., a presentation), we did not find significant differences in motivation from the theoretical lens of self-determination theory. However, providing the SG with integrated debriefing indicates higher *learning outcomes* than a conventional presentation.

Concerning our first research question (i.e., *Which effects on motivation and learning outcomes has using Serious Games for Business Information Visualization compared to more conventional presentations?*), we may thus conclude that while not necessarily leading to increased motivation, SG may improve *learning outcomes* compared to conventional training methods.

The second mode of inquiry leading to DSR contributions used in this study is experimental research (Briggs and Schwabe 2011). This mode of inquiry leads to hypotheses, experimental designs, and analyzed data sets (Briggs and Schwabe 2011). With these contributions, DSR aims to measure the degree to which design objectives have been achieved. In this study, hypotheses have been derived from self-determination theory, which served as the kernel theory for artifact construction. As an important contribution, we developed a measurement instrument that may be used in future studies about SG in the IS domain. Using this measurement instrument, we were able to show that one of the most important dependent variables, namely *intrinsic motivation*, significantly differed between the groups with integrated and post hoc debriefing. In addition, *learning outcomes* seem to be higher when debriefing is integrated into the SG. Being the (according to our literature review) first study that deliberately investigates the differences between integrated and post hoc debriefing by implementing two different versions of a SG (although it has been proposed earlier, for instance by Kriz 2010), we contribute to the design of effective SG. Thus, with regard to our second research question (i.e., *Which effects on motivation and learning outcomes has integrated debriefing in comparison to* post hoc *debriefing as a design principle for Serious Games?*), we may conclude that integrating debriefing into SG may yield beneficial outcomes in terms of learning and motivation compared to post hoc debriefing, thus being a promising design principle for SG.

However, this study also comes with some limitations. Regarding the generalizability of the findings, it is important to note that the presentation in the control group was not varied (i.e., we only investigated one specific presentation). To thoroughly compare SG with conventional training methods, we also must alter different aspects of presentations (e.g., length or quality of visual support). Another limitation of these findings might be the way that debriefing was conducted in the group with post hoc debriefing. Although there are studies indicating that video-assisted self-debriefing is on par with discussion-based debriefing with an instructor (Boet et al. 2011), this was not investigated in this study. Hence, future research should deliberately examine whether our findings about integrated debriefing may be replicated when compared to discussion-based post hoc debriefing. Another potential limitation lies in the way of measuring *learning outcomes*. Although we showed that participants increased their ability to identify inappropriate BIV and make reasonable improvement suggestions, we did not measure whether they would also be able to incorporate these guidelines when creating novel reports instead of correcting existing ones. In addition, although we included different fields of study in our sample, it mainly consisted of business students and may thus only be valid for this field of study. Future research may thus aim to replicate our study with a more diverse set of participants to investigate possible differences due to fields of study.

7 Conclusion

This study set out to evaluate a SG about BIV, which likely constitutes a novel artifact in the domain of BI&A. In addition, we investigated the role of integrated debriefing in SG, which has thus far not been deliberately examined. Our findings indicate that SG are able to increase BIV skills and are acknowledged by participants. We also found that integrating debriefing into SG may yield significant benefits: It leads to higher motivation and *learning outcomes* compared to SG with post hoc debriefing. This might be an important finding, especially since SG still heavily rely on this post hoc debriefing. In addition, findings indicate that SG with integrated debriefing may enhance learning compared to conventional presentations. SG with post hoc debriefing, however, seem inferior to these presentations. We thus found evidence that simply using SG will not necessarily increase learning and motivation compared to conventional training methods. Instead, it is important to thoroughly investigate design principles for SG in order to harness their potential. In addition to the specific findings provided, this study invites the field of IS to examine how technology-supported learning in IS may be improved using DSR in future studies. This is a question that may become ever more important, as required skill sets of employees might change more frequently due to ever more frequent technological innovations. As learning is required in many different domains (e.g., business process management and knowledge management), this research may help to support ongoing learning processes in organizations facing the challenges of digital transformation.

References

Abt CC (1987) Serious games. University Press of America, Lanham

Arbel Y, Hong L, Baker TE, Holroyd CB (2017) It's all about timing: an electrophysiological examination of feedback-based learning with immediate and delayed feedback. Neuropsychologia

99:179–186. https://doi.org/10.1016/j.neuropsychologia.2017.03.003
Basole RC, Bodner DA, Rouse WB (2013) Healthcare management through organizational simulation. Decis Support Syst 55(2):552–563. https://doi.org/10.1016/j.dss.2012.10.012
Beattie V, Jones MJ (2008) Corporate reporting using graphs: a review and synthesis. J Acc Lit 27:71–110
Blohm I, Leimeister JM (2013) Gamification. Wirtschaftsinformatik 55(4):275–278. https://doi.org/10.1007/s11576-013-0368-0
Boet S, Bould MD, Bruppacher HR, Desjardins F, Chandra DB, Naik VN (2011) Looking in the mirror: self-debriefing versus instructor debriefing for simulated crises. Crit Care Med 39(6):1377–1381. https://doi.org/10.1097/ccm.0b013e31820eb8be
Briggs RO, Schwabe G (2011) On expanding the scope of design science in IS research. In: Jain H, Sinha AP, Vitharana P (eds) Service-oriented perspectives in design science research: DESRIST 2011. Springer, Heidelberg, pp 92–106
Brown TA (2015) Confirmatory factor analysis for applied research, 2nd edn. Methodology in the social sciences, Guilford
Card SK, Mackinlay JD, Shneiderman B (1999) Readings in information visualization: using vision to think. Morgan Kaufmann, Burlington
Charsky D (2010) From edutainment to serious games: a change in the use of game characteristics. Games Cult 5(2):177–198. https://doi.org/10.1177/1555412009354727
Chen H, Chiang RHL, Storey VC (2012) Business intelligence and analytics: from big data to big impact. MIS Q 36(4):1165–1188
Connolly TM, Boyle EA, MacArthur E, Hainey T, Boyle JM (2012) A systematic literature review of empirical evidence on computer games and serious games. Comput Educ 59(2):661–686
Crookall D (1992) Editorial: debriefing. Simul Gaming 23(2):141–142. https://doi.org/10.1177/1046878192232001
Crookall D (2010) Serious games, debriefing, and simulation/gaming as a discipline. Simul Gaming 41(6):898–920. https://doi.org/10.1177/1046878110390784
Csikszentmihalyi M (1991) Flow: the psychology of optimal experience, vol 41. Harper Perennial, New York
de Freitas S, Gibson D, Alvarez V, Irving L, Star K, Charleer S, Verbert K (2017) How to use gamified dashboards and learning analytics for providing immediate student feedback and performance tracking in higher education. In: Barrett R, Cummings R (eds) Proceedings of the 26th international conference on world wide web companion, Geneva, pp 429–434
de Kort YAW, IJsselsteijn WA, Poels K (2007) Digital games as social presence technology: development of the social presence in gaming questionnaire (SPGQ). In: Moreno L (ed) Proceedings of the 10th annual international workshop on presence, October 25–27, 2007, Starlab, Barcelona, Spain, pp 195–203
Deci EL, Ryan RM (1985) Intrinsic motivation and self-determination in human behavior. Plenum, New York
Der Sahakian G, Alinier G, Savoldelli G, Oriot D, Jaffrelot M, Lecomte F (2015) Setting conditions for productive debriefing. Simul Gaming 46(2):197–208. https://doi.org/10.1177/1046878115576105
Deterding S, Dixon D, Khaled R, Nacke L (2011) From game design elements to gamefulness: defining "gamification". In: Lugmayr A, Franssila H, Safran C, Hammouda I (eds) MindTrek'11. ACM, New York, pp 9–15
Faria AJ, Hutchinson D, Wellington WJ, Gold S (2009) Developments in business gaming: a review of the past 40 years. Simul Gaming 40(4):464–487
Fornell C, Larcker DF (1981) Evaluating structural equation models with unobservable variables and measurement error. J Market Res 18(1):39–50
Garris R, Ahlers R, Driskell JE (2002) Games, motivation, and learning: a research and practice model. Simul Gaming 33(4):441–467. https://doi.org/10.1177/1046878102238607
Gee JP (2003) What video games have to teach us about learning and literacy. Comput Entertain 1(1):20. https://doi.org/10.1145/950566.950595
Gefen D, Straub D (2005) A practical guide to factorial validity using PLS-graph: tutorial and annotated example. Commun Assoc Inf Syst 16(5):90–110
Grund CK (2015) How games and game elements facilitate learning and motivation: a literature review. In: Cunningham D, Hofstedt P, Meer K, Schmitt I (eds) Informatik 2015. Ges. für Informatik, Bonn, pp 1279–1293
Grund CK, Meier MC (2016) Towards game-based management decision support: using serious games to improve the decision process. In: Stelzer D, Nissen V, Straßburger S (eds) Proceedings of the Multikonferenz Wirtschaftsinformatik (MKWI) 2016, pp 155–166
Grund CK, Schelkle M (2016) Developing a serious game for business information visualization. In: Proceedings of the 22nd Americas conference on information systems (AMCIS), pp 1–4
Grund CK, Schelkle M (2017) Visualisieren spielend erlernen - Ein Serious Game zur Verbesserung von Managementberichten. In: Strahringer S, Leyh C (eds) Gamification und Serious Games: Grundlagen, Vorgehen und Anwendungen. Springer, Wiesbaden, pp 167–181
Grund CK, Tulis M (2017) Using elected elements in large-scale information systems lectures. In: Leimeister JM, Brenner W, Peters C (eds) Proceedings of the 13th international conference on Wirtschaftsinformatik, pp 776–787
Grund CK, Schelkle M, Hurm M (2017) Architecture and evaluation design of a prototypical serious game for business information visualization. In: Leimeister JM, Brenner W, Peters C (eds): Proceedings of the 13th international conference on Wirtschaftsinformatik, pp 1271–1274
Hevner AR, March ST, Park J, Ram S (2004) Design science in information systems research. MIS Q 28(1):75–105
Hichert R, Faisst J (2015) International business communication standards: IBCS Version 1.0. Proposals for the conceptual, perceptual, and semantic design of comprehensible business reports and presentations.https://www.ibcs.com/standards/. Accessed 21 Dec 2018
Hsieh H-F, Shannon SE (2005) Three approaches to qualitative content analysis. Qual Health Res 15(9):1277–1288. https://doi.org/10.1177/1049732305276687
Hu L-T, Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model Multidiscip J 6(1):1–55. https://doi.org/10.1080/10705519909540118
Kickmeier-Rust MD, Albert D (2010) Micro-adaptivity: protecting immersion in didactically adaptive digital educational games. J Comput Assist Learn 26(2):95–105. https://doi.org/10.1111/j.1365-2729.2009.00332.x
Kolb DA (1984) Experiential learning: experience as the source of learning and development. Prentice-Hall, Englewood Cliffs
Krippendorff K (2004) Reliability in content analysis: some common misconceptions and recommendations. Hum Commun Res 30(3):411–433. https://doi.org/10.1111/j.1468-2958.2004.tb00738.x
Kriz WC (2010) A systemic-constructivist approach to the facilitation and debriefing of simulations and games. Simul Gaming 41(5):663–680. https://doi.org/10.1177/1046878108319867
Kusurkar RA, Ten Cate TJ, Vos CMP, Westers P, Croiset G (2013) How motivation affects academic performance: a structural equation modelling analysis. Adv Health Sci Educ 18(1):57–69. https://doi.org/10.1007/s10459-012-9354-3
Lederman LC (1992) Debriefing: toward a systematic assessment of theory and practice. Simul Gaming 23(2):145–160. https://doi.org/10.1177/1046878192232003

Legner C, Eymann T, Hess T, Matt C, Böhmann T, Drews P, Mädche A, Urbach N, Ahlemann F (2017) Digitalization: opportunity and challenge for the business and information systems engineering community. Bus Inf Syst Eng 59(4):301–308. https://doi.org/10.1007/s12599-017-0484-2

Lindsey RV, Shroyer JD, Pashler H, Mozer MC (2014) Improving students' long-term knowledge retention through personalized review. Psychol Sci 25(3):639–647. https://doi.org/10.1177/0956797613504302

Locke EA, Latham GP (2002) Building a practically useful theory of goal setting and task motivation: a 35-year odyssey. Am Psychol 57(9):705–717. https://doi.org/10.1037/0003-066x.57.9.705

Matt C, Hess T, Benlian A (2015) Digital transformation strategies. Bus Inf Syst Eng 57(5):339–343. https://doi.org/10.1007/s12599-015-0401-5

Nadolski RJ, Hummel HGK (2017) Retrospective cognitive feedback for progress monitoring in serious games. Br J Educ Technol 48(6):1368–1379. https://doi.org/10.1111/bjet.12503

Papastergiou M (2009) Digital game-based learning in high school computer science education: impact on educational effectiveness and student motivation. Comput Educ 52(1):1–12

Pavlov OV, Saeed K, Robinson LW (2015) Improving instructional simulation with structural debriefing. Simul Gaming 46(3–4):383–403. https://doi.org/10.1177/1046878114567980

Qudrat-Ullah H (2007) Debriefing can reduce misperceptions of feedback: the case of renewable resource management. Simul Gaming 38(3):382–397. https://doi.org/10.1177/1046878107300669

Rudolph JW, Simon R, Raemer DB, Eppich WJ (2008) Debriefing as formative assessment: closing performance gaps in medical education. Acad Emerg Med 15(11):1010–1016. https://doi.org/10.1111/j.1553-2712.2008.00248.x

Ryan RM (1982) Control and information in the intrapersonal sphere: an extension of cognitive evaluation theory. J Personal Soc Psychol 43(3):450–461

Ryan RM, Deci EL (2000) Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 55(1):68

Ryan RM, Rigby CS, Przybylski A (2006) The motivational pull of video games: a self-determination theory approach. Motiv Emot 30(4):344–360

Sheldon KM, Filak V (2008) Manipulating autonomy, competence, and relatedness support in a game-learning context: new evidence that all three needs matter. Br J Soc Psychol 47(Pt 2):267–283. https://doi.org/10.1348/014466607x238797

Steinwachs B (1992) How to facilitate a debriefing. Simul Gaming 23(2):186–195. https://doi.org/10.1177/1046878192232006

Susi T, Johannesson M, Backlund P (2007) Serious games: an overview. Technical Report HS-IKI-TR-07-001, Sweden

Tabachnick BG, Fidell LS (2013) Using multivariate statistics, 6th edn. Always Learning, Pearson

Taylor G, Jungert T, Mageau GA, Schattke K, Dedic H, Rosenfield S, Koestner R (2014) A self-determination theory approach to predicting school achievement over time: the unique role of intrinsic motivation. Centemp Educ Psychol 39(4):342–358

Tegarden DP (1999) Business information visualization. Commun Assoc Inf Syst 1(1):1–38

Thiagarajan S (1992) Using games for debriefing. Simul Gaming 23(2):161–173. https://doi.org/10.1177/1046878192232004

Venable J, Pries-Heje J, Baskerville R (2012) A comprehensive framework for evaluation in design science research. In: Peffers K, Rothenberger M, Kuechler B (eds) Design science research in information systems: advances in theory and practice. Springer, Berlin, pp 423–438

Ware C (2012) Information visualization: perception for design, 3rd edn. Elsevier, Amsterdam

Wigfield A, Eccles JS (2000) Expectancy—value theory of achievement motivation. Centemp Educ Psychol 25(1):68–81. https://doi.org/10.1006/ceps.1999.1015

Wouters P, van der Spek ED, van Oostendorp H (2009) Current practices in serious game research: a review from a learning outcomes perspective. In: Connolly T, Stansfield M, Boyle L (eds) Games-based learning advancements for multi-sensory human computer interfaces. Hershey, Derry Township, pp 232–250

Wu W-H, Hsiao H-C, Wu P-L, Lin C-H, Huang S-H (2012) Investigating the learning-theory foundations of game-based learning: a meta-analysis. J Comput Assist Learn 28(3):265–279. https://doi.org/10.1111/j.1365-2729.2011.00437.x