

**Debriefing Research Games
Context, Substance and Method**

van den Hoogen, Jop; Lo, Julia; Meijer, Sebastiaan

DOI

[10.1177/1046878116651023](https://doi.org/10.1177/1046878116651023)

Publication date

2016

Document Version

Final published version

Published in

Simulation & Gaming: an international journal of theory, design and research

Citation (APA)

van den Hoogen, J., Lo, J., & Meijer, S. (2016). Debriefing Research Games: Context, Substance and Method. *Simulation & Gaming: an international journal of theory, design and research*, 47(3).
<https://doi.org/10.1177/1046878116651023>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

Debriefing Research Games: Context, Substance and Method

Simulation & Gaming
2016, Vol. 47(3) 368–388
© The Author(s) 2016
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1046878116651023
sag.sagepub.com



Jop van den Hoogen¹, Julia Lo¹, and Sebastiaan Meijer²

Abstract

Background. **Debriefing** is an intrinsic part of games for learning and proper **debriefing** can also be beneficial to **research games**. However, the literature on how to debrief research games is sparse and only provides the professional with an abstract **topic guide**.

Aim. The purpose of this study was to design a **framework** for the **debriefing** of **research games** that are used in ongoing **innovation** processes.

Method. We used the literature on debriefing and experimental research and our experience as game designers to build a **framework** that tackles the **context**, **substance** and **method** of debriefing research games.

Results. Our framework provides three contributions. First, it shows how the context in which a **research game** is applied sometimes impacts the functionality of the game in negative ways. This can be helped by designing both the game and the **debriefing** together. Second, we operationalize **validity** to a greater extent, as this is the core of a good **research game**. Third, we provide a **methodology** for debriefing professionals that opens up the black box of the **gaming simulation** session.

Conclusion. The **debriefing** framework provides a **method** to collectively assess the validity, reliability and robustness of the causal claims associated with the research conducted.

¹Delft University of Technology, The Netherlands

²KTH Royal Institute of Technology, Sweden

Corresponding Author:

Jop van den Hoogen, Faculty of Technology, Policy and Management, Delft University of Technology, Jaffalaan 5, Delft, 2628 BX, The Netherlands.

Email: j.vandehoogen@tudelft.nl

Keywords

complex sociotechnical system, debriefing, event-structure analysis, gaming simulation, innovation, railway system, research game, sensitivity, validity

Introduction

The topic of debriefing is not solely of interest to gaming simulation researchers as the activity merely refers to a collective discussion of events that happened before the debriefing. Debriefing is in fact used in many more instances, for example following military operations and traumatic events or after deceptive psychological experiments (Lederman, 1992). Debriefing can be defined as: “the process in which people who have had an experience are led through a purposive discussion of that experience” (Lederman, 1992, p. 146). In a real-life event, debriefing is clearly distinguishable from the event itself. In a gaming simulation, the debriefing plays a more intrinsic role. For the debriefing of educational gaming simulations in particular, learning comes from the debriefing rather than from the game itself (Crookall, 2010).

Many scholars have pointed to the crucial importance of debriefing in realizing the overall value of gaming simulations, also known as simulation games or serious games (Crookall, 2010; Decker et al., 2013; Lederman, 1992). Games are devices that allow experiential learning to be practiced, yet effective learning only comes with reflection (Decker et al., 2013). Debriefing allows experience and reflection to be integrated in the learning process, so it is striking how little attention is paid to this crucial element of gaming simulation even though scholars have consistently called for more attention (Crookall, 2010; Dennehy, Sims, & Collins, 1998; Fanning & Gaba, 2007; Lederman, 1992).

Just as we design games differently depending on whether we intend to use them for learning, policy making or research, the way we debrief should also be in line with the game’s purpose (Peters & Vissers, 2004). For instance, compared to training games, research games do not focus on players’ knowledge creation or adaptation, but instead allow researchers to investigate elements, such as actors and processes, in a controlled environment. In the debriefing of research games, the validity and reliability of the gaming situation are the key topics rather than the learning process. The aim of this article is to enrich the existing debriefing methodology by introducing a framework for debriefing research games.

We start by examining the use of gaming simulation, especially from the perspective of the researcher, rather than the game designer. We build on previous work that remained on a rather abstract level in order to provide a complete framework for the debriefing of research games (Lederman & Stewart, 1986; Van Ments, 1983). Our framework is presented by combining insights from existing literature on debriefing, empirical work on the context in which our gaming simulations are applied, and our own experience in conducting and debriefing gaming simulations. The framework has both a structural and a methodological component and we provide a topic guide that

shows which topics a debriefing should cover and a methodology for approaching them.

Gaming Simulation for Research

According to Peters, Vissers, and Heijne (1998) gaming simulation can be used for education and training, for designing policies, and for research. The gaming simulations we designed for the Dutch railway sector are mostly research games which enable stakeholders to test hypotheses about the value of their innovations in a safe and experimental environment. Research games differ from learning games and policy games in that the transfer of knowledge is not from the game to players or between players, but rather from the game, including the players, to some outside observer (Peters et al., 1998). This observer can use the simulation to study processes holistically and dynamically, as well as provide the simulated system with stimuli. The gaming part of the simulation increases the simulation's validity since human behavior is an intrinsic part of a railway system. This is especially true of innovations that focus on changing both the technical and the social make-up of a system. The redesign of a station layout combined with traffic control procedures to allow for more capacity is an example of this innovation. When experimenting with such an innovation, a dynamic model needs to incorporate human game players and this creates the need for gaming simulation as a research tool. At first sight, such an application of gaming simulation needs little to no debriefing. A simple pretest and posttest setup, logging agreed-upon performance values and measuring the difference should suffice. However, the peculiarities of innovating in complex sociotechnical systems prohibit the researcher from taking this simplistic approach, which is why research games specifically used for innovation purposes, as addressed in this article are needed.

Klabbers (2003) stated that designing effective gaming simulations is an interplay between designing the game itself, i.e. design in the small (DIS), and the intended effects of the game on the design of the referent system, i.e. design in the large (DIL). If we wish to structure a debriefing, and thus make the debriefing a design consideration as proposed by Crookall (2010), we should inform this process with the peculiarities of the context in which the simulation is employed. As Klabbers (2006) stated, the goal of a gaming simulation (DIS) should serve the meta-goal of DIL processes. Kriz and Hense (2006) sought to combine these two design processes by linking common applications of gaming simulation to Greif and Kurtz' (1996) model of organizational development. This model identifies four types of gaming simulation (or simulation games, SG), which are often sequentially applied to support organizational development:

1. Present state SG
2. Future state SG
3. Test scenario SG
4. Training SG

These gaming simulation typologies help us to distinguish different functionalities of gaming simulation, such as diagnosis, design, testing and training. Since research is defined by the generation (present state SG) and testing (test scenario SG) of hypotheses, we characterize game types 1 and 3 as games for research. In game type 2 (future state SG) gaming simulation is used for the in-game design of artifacts, policies and strategies, and we see similarities with policy games. The fourth game type (training SG) is well known and takes the form of educational games and games for learning. This distinction closely follows the categorization by Peters et al. (1998) of games for research, policy and learning.

Like Kriz and Hense (2006), we sought to combine innovation processes with the use of gaming simulation. However, clear phases are hard to distinguish in these chaotic processes (Anderson & Joglekar, 2012; Cheng & Van de Ven, 1996). It may be chaotic because the railways—as complex sociotechnical systems—are hard to grasp using linear models. That is, the innovation processes that intend to change these systems are erratic, rather than linear. The complexity of the system we wish to model and simulate leads to many validity issues (Lo, Van den Hoogen, & Meijer, 2013), for which the game itself cannot control. We propose that debriefing plays a significant role in adjusting for the flaws of gaming simulation as a research tool (Van den Hoogen et al., 2014a). In addition, the same gaming simulation might perform different functions depending on the observer. For one organizational entity it might serve as a way to rigorously test hypotheses, while for another entity, it might provide an ideal opportunity to observe a system holistically and perform a diagnosis. For operators, the game is a way to gain influence in the innovation process or a threat to their autonomy. Potentially, all of these conflicting expectations have two detrimental consequences: the research game might not be able to validly answer one specific research question and converge on a single final design, and the game might create negative effects outside of the game. For both purposes we feel that a debriefing is a valuable, even necessary, addition to the design of a research game.

The Role of Debriefing in Gaming Simulation

In general, debriefing is the collective assessment of in-game events and the discussion with game participants about the events' relation to the real world. While such processes are highly valuable for learning purposes, we argue that the same mechanism will also improve research games. Even without the need to allow for game player learning, the assessment of in-game events, their significance and their relation to the outside world are enormously relevant for research games as a methodology. For this reason a considerable part of the theoretical background for our framework is based on existing work on debriefing games for learning.

The notion that debriefing should be an intrinsic component of gaming simulation design is supported by the fact that experiential learning is a matter of experiencing an event and reflecting on this experience. While the gaming simulation is designed in such a way to provide the player with a realistic experience, the debriefing allows for reflection. Historically, many debriefing frameworks for gaming simulation have

focused solely on games for learning and applied Kolb's cyclical model of experiential learning (Kolb, 1984) as the foundation (Decker et al., 2013; Dennehy et al., 1998; Van der Meij, Leemkuil, & Li, 2013). This cyclical model portrays experiential learning as moving from experimentation, via experiencing and reflection, to conceptualization. Debriefing usually involves two parts: a collective assessment of what has happened and a discussion on the implications of these events outside of the game (Kriz, 2003). Debriefing ensures better task performance and allows players to learn more about a decision domain and develop heuristics to significantly reduce the time between observation and decision-making (Qudrat-Ullah, 2007).

Topics of a Debriefing

Most frameworks focus on the phases, or topics, that a debriefing should have. In the realm of games for learning, Sims (2002), Thiagarajan (1993), and Lederman (1992) provide insightful frameworks. However, frameworks for games for research, the topic of this article, are less developed.

Peters and Vissers (2004) are among the few gaming simulation scholars who specifically target the debriefing of research games. According to them, debriefing of research games has three functions:

1. Providing an opportunity for participants to cool down
2. Protecting the instrument of gaming simulation
3. Validating the researcher's interpretation of simulation outcomes.

At first sight, validation seems the most obvious of the three phases of debriefing. Gaming simulations are artificial environments in which the simulation is open due to the involvement of human game players. This creates internal and external validity issues. The researcher's interpretation of the simulation outcomes should therefore be validated using feedback from participants in the gaming simulation, although the first two functions are also important. As we strive for high levels of immersion when we want game players to portray realistic behavior in a game, we ask game players to enter into a reactive mode, dealing solely with the decisions presented to them by the game model and not reflecting on the model itself. In a debriefing, we ask them to reflect on what happened, and possibly also ask them to question the model. The transition between these two modes, from reactive to reflexive, does not happen automatically and is facilitated by cooling-down phase. Willing game players tend to be scarce, particularly in organizational settings where game players are also employees responsible for day-to-day operations, so successive participation or participation by their colleagues is key. We also need to ensure that controversial issues, such as contested innovations tested in the game or conflicts between game players, stay within the realms of the game. As games for research do not primarily look for interventionist effects, what happens in the game should not have any immediate impact outside of the game. The debriefing is the ideal means of controlling these factors.

Table 1. Phases in a Debriefing.

Phase	Topic	Explanation
1	How did you feel?	Cooling down of the participants
2	What happened?	Data collection
3	How are the game and reality connected?	External validity
4	What did you/we learn?	Reaching conclusions
5	What would happen if...?	Testing replicability/sensitivity
6	How do we proceed from here?	Planning for action

Source. Kriz (2010: 669-671).

A Systems Perspective on Debriefing

Kriz (2010) was one of the first to apply a systems perspective to the debriefing process where games are intended to say something about referent systems or designs in the large scale. The systems perspective pervades the framework as it acknowledges complex features of systems, including its multi-interpretability and path dependence. To do justice to these properties of both the referent system and the game model, Kriz (2010) recommended using six distinct phases in the debriefing process. A key component of this debriefing framework is that gaming simulations allow for the collective and holistic study of complex systems. This collectiveness and holism requires researchers to combine many insights from players and observers and converge the results towards valid propositions concerning the main causal mechanisms that drive the simulation outcomes. An overview of these phases is provided in Table 1.

Although not specifically targeted at games for research, this framework provides a good direction for the debriefing of research games that involve the study of complex systems. In summary, a properly structured debriefing should contain distinct phases: cooling down, data collection, validity and reliability analysis, planning for action and protecting the instrument.

Missing Links

To summarize the literature: frameworks for debriefing seem well developed for gaming simulation for learning, whereas debriefing for research games deserves further attention. We have seen three key phenomena that create a need for a more fully developed framework. First, the context in which gaming is applied pervades only slightly in the debriefing framework. Second, validity is not operationalized in enough detail to serve as a structuring force on debriefing frameworks suited for research games. The fact that research games are not solely about the design of the game but also about the design of the experiment is particularly overlooked. There is a need to incorporate matters such as internal validity in the debriefing. Third, there is no clear methodology for how to tackle the topics. A topic guide alone does not help the debriefing professional to actually assess all the topics; it merely points to those topics that require further attention.

Framework for Debriefing Games for Research

We have developed a debriefing framework based on our experience in designing, facilitating, and debriefing games for research, including incorporating the conclusions we have drawn and the lessons learned over the years. We were involved in the Railway Gaming Suite (RGS) from 2009, designing ad hoc low-tech tabletop gaming simulations for the Dutch railway sector. Being designed on an ad hoc basis, there were significant differences in the specific research questions. However, a common factor in all the games was the simulation of operational processes of railway systems (trains running according to a schedule, a realistic depiction of the infrastructure, and operators dealing with scenarios such as major disruptions around railway stations). An example of a typical question on which our games were intended to shed light is: does the punctuality of train traffic around the central node of the network increase if we separate two heavily used corridors by removing railway switches (points)? The fact that these questions involved a unit of analysis at system level, and contained both technical and social elements, created the need to use gaming simulation to test such measures in a safe environment. For a more in-depth analysis, we refer the reader to Meijer (2012), Lo et al. (2013), Van den Hoogen, Lo, and Meijer (2014b) and Van den Hoogen, Lo, and Meijer (2014a). The framework presented here is a distillation of all of this work.

The framework tackles precisely those problems we encountered in the current literature on debriefing when applied solely to gaming simulations intended to test hypotheses. First, it takes into account contextual influences on the ability of a game to test a hypothesis solely by running a simulation (and disregarding the debriefing). Sometimes gaming simulation seems to have a undesirable, natural tendency to allow for exploration rather than explanation, caused by contextual influences. Our debriefing framework helps to counter this tendency. Secondly, it uses a topic guide that operationalizes validity in more detail. This enables our framework to alleviate many of the inherent validity threats of using a method that lingers between field observations and classical experiments. Our framework also provides a specific methodology that enables the debriefing to open up the *black box* of the simulation run. This last contribution is significant in that a topic guide alone barely helps when actually debriefing a research game. For example, a topic guide may direct us to assess ecological validity, but does not provide us with a method for actually doing so. In addition, our framework truly coalesces the debriefing process to make it an intrinsic part of the discipline of gaming simulation. This is because the gaming simulation and the debriefing mutually reinforce one other. Conversely, a carefully designed gaming simulation helps to improve the debriefing, but when our framework is employed, the game and the debriefing become whole.

Context

The gaming simulations we have designed, employed and debriefed up to now were used as applied experiments that organizations could explore or test innovations [See

Meijer (2012) and Lo et al. (2013) for an overview]. Gaming simulation is not an isolated phenomenon, but is embedded in ongoing technological, social and institutional processes over time. The design of a gaming simulation needs input from the environment, for example, innovations to be tested, data, models, game players, and feeding back the results of the simulation. Two parameters seem particularly relevant in this case for both the innovation and the function of a game:

1. Innovation processes can be either *stable* or *volatile*, or move from one to the other over time. Volatility entails rapid changes in the design of the innovation, rapid entrance and exits of designers and decision makers and fluid and flexible institutions that govern these activities.
2. Gaming simulation can either create *convergence* or *divergence*. Divergence is the exploration of a multitude of designs, the opening up of the arena of designers and the exploration of viable institutional arrangements. Convergence is the opposite, where designers and decision makers become more fixed, increasingly focusing on a single design as the final option under increasingly stable institutional arrangements (Van den Hoogen & Meijer, 2015).

This conceptualization of the value of gaming simulation in light of the context in which the method is applied led us to study the practical value of using so-called explanatory research games. We term these explanatory for their ability to test hypotheses rather than generate them. However, to stay in line with the literature we will continue to use the more general term of research games.

A research game diminishes volatility because the design of the game and the experiment is such that it should allow researchers observing the game to focus on the acceptance or rejection of a single hypothesis. In an applied context this would mean testing an innovation. After the experiment, the researcher would be able to say for example, *my innovation caused a 10% increase in system performance compared to the base scenario without the innovation*. We have seen that this function of gaming simulation has often led to the method being employed in times when the innovation process is highly volatile. In such times, when many designs, ideas and innovations float through the organization and many new designers, decision makers and other stakeholders enter the decision-making arena, stakeholders view gaming simulation as a valid tool to alleviate this volatility. However, contextual influences of this volatility have a pervasive effect on the ability of gaming simulation to actually create convergence (Van den Hoogen & Meijer, 2015).

In highly volatile times, rapid changes occur in the set of feasible design alternatives. On many occasions we had to make last-minute changes to the game model to incorporate alterations in the innovation or in other relevant parameters. Since there is always a lag between the input for the game design, the design of the experiment and the output in the form of results, there is a chance that the game answers questions are deemed no longer relevant by the organization.

A second point, and far more significantly if one intends to create convergence with a gaming simulation, is that gaming simulation serves as a window of opportunity to

test other innovations as well. This may be particularly relevant in the capital-intensive and safety-critical industries in which we operate as there are opportunities to test innovations. When organizational entities other than the primary client of the game become aware of the possibility to test their innovation in a gaming environment, we see an influx of additional research questions. The usual way of going about adhering to these questions would be to increase the factorial design of an experiment. However, pure experimental research often demands full factorial designs (making all possible combinations of innovations), resulting in exponential increases in the numbers of runs. Due to time constraints, real-life operators are usually only available as game players for a limited time. The choice is to either not test them all or to make the simulation more abstract and omit real-time play, thus risking lower levels of game player immersion. These are design choices that endanger internal and external validity, respectively. Finally, immersion is sometimes a problem in itself, as game players are usually operators who enjoy a certain degree of autonomy in their daily work. This creates both a desirable and undesirable distance between those who carry out the work (and are part of the game) and those who design the overall system in which the operators are placed (and observe the game). When this distance is removed by employing a gaming simulation, two phenomena can occur: game players either feel under heightened scrutiny and start behaving differently than in real life, or they feel heard and desire a dialogue with the designers of the innovation during the game. Both phenomena create immersion issues since we want game players to behave just as they would in real life where there are no designers observing or able to communicate with them.

Making the debriefing an intrinsic part of the game design significantly helps to alleviate the aforementioned context-based problems. The debriefing can serve as a way of testing all innovations while keeping the number of runs relatively low. This is valuable since a low number of runs enables the game designers to use real-time play, a design parameter often, but not always, associated with high levels of immersion. In the design of the experiment, game designers and the innovation managers involved can decide together which innovations truly need to be tested in the game run and which innovations can be assessed in the debriefing. Taking into account the possibilities of assessing additional innovations in the debriefing helps make the gaming simulation more adaptable to last-minute changes in the innovation. To achieve this, a robustness analysis should be included in the debriefing. Game players and observers can concertedly assess the extent to which simulation outcomes will differ if either the innovation changes later on or additional innovations are introduced. This is an important part of the debriefing since one cannot expect the innovation being tested to be exactly the same when it is implemented, especially in volatile times. Furthermore, the debriefing can be used to postpone the inherent tendency of gaming simulation to lead to a dialogue between game players and observers. This means that incorporating the debriefing allows the facilitator to better manage expectations. The facilitator could demand that dialogue be non-existent during the game run, thereby increasing immersion, and in return promise that the desired dialogue will take place during the debriefing.

Substance

Hypothesis-testing research is in essence an experiment in which one or more independent variables are manipulated to investigate their effects on a dependent variable (Zechmeister, Zechmeister, & Shaughnessy, 2001). Two streams can be identified for experimental research: the first is a classical linear perspective on causality and the second is a complexity perspective. The classical linear perspective sees experimental objects as trivial machines, which implies that the same treatment given to the same participant will always have a similar outcome. The complexity perspective takes non-triviality into account, which implies that systems with dynamic feedback show path-dependent and chaotic behavior. In line with this perspective, units of analyses are therefore respectively regarded as black boxes or as a collection of interacting elements. However, two critical concepts are key to determining the quality of both streams of experimental research: reliability and validity (Lo et al., 2013).

Reliability

Measurement reliability. Measurement reliability is the extent to which a research method or measurement tool provides a similar value if the measurement is repeated (Messick, 1975). In quantitative terms, the reliability of the measurement tool can be expressed as a margin of error. For instance, if a thermometer should be measuring a temperature of 39 degrees Celsius, but indicates a value of 38 degrees half of the time and a value of 40 degrees for the remaining measurements, the margin of error of the measurement tool is $1/39$.

Sensitivity. The sensitivity of the experiment is often determined in computer simulation experiments, in which the researcher determines whether similar causal relationships are found when the experiment is repeated with exactly the same sample and setup. This complexity perspective on reliability follows from experiments with dynamic feedback systems. Because dynamic feedback systems inherit stochastic and sometimes chaotic properties, different results can be found when experiments are repeated with the same or almost the same starting conditions. An indication of the sensitivity of an experiment is useful in order to assess whether the results are sensitive to the initial conditions or to critical decisions by game players.

Validity. Internal, external and measurement validity are the core validity types in experimental research, in which external and internal validity play a dominant role in determining the quality of the experiment (Zechmeister et al., 2001).

Internal validity. In establishing a causal relationship, the research needs to meet the conditions of co-variation, time-order relationships and elimination of plausible alternative causes (Zechmeister et al., 2001). Co-variation is the first step in establishing a causal inference which can be fulfilled by finding a relationship between the independent and dependent variable. Identifying the cause and effect for the independent and dependent variable enables a time-order relationship to be established. And finally,

confounding variables need to be isolated to eliminate plausible alternative causes.

External validity. External validity has multiple definitions that are subject to conflicting interpretations (Morton & Williams, 2010). We distinguish external validity in terms of generalizability, i.e. results that can be transferred from the current sample to the population, versus ecological validity, from the simulated environment to a real-world setting, which would be in line with the *fieldness* of the experiment (Harrison & List, 2004). Selection of a representative sample ensures the generalizability of the results as a reflection of the population. Parallel resemblances can be drawn for ecological validity with the three gaming simulation validity types defined by Raser (1969). Gaming simulation validity can be broken down into structural validity, process validity and psychological reality. The simulated gaming model may be rather abstract or simplified in terms of processes, interactions, and contextual and physical cues in comparison to the reference system. As such, the omitted characteristics of the reference system may endanger the transfer of causal claims made within the gaming simulation to the real world. Applying a sensitivity analysis could support the assessment of this type of external validity by focusing on whether parameter sensitivity, tipping points and critical decisions by game players could be a resemblance of events in the reference system.

Measurement validity. Measurement validity, also known as test validity, refers to the validity of the measurement tool or instrument itself. Psychometric researchers have predominantly focused on the different typologies involved in the use of measurement instruments, often questionnaires. The American Psychological Association, American Educational Research Association, and National Council on Measurement in Education have set Joint Standards (1966) in which construct, criterion, and content validity are distinguished as the three main measurement validity categories.

Topic guide. Based on the literature review in the previous sections, we identified eight phases that need to be addressed in a debriefing session of a research game, in which a large overlap exists with existing literature by Kriz (2010) and Peters and Vissers (2004). However, this paper recognizes the gap in the existing literature regarding the specific topics that need to be addressed within the validity and reliability analysis phase. In order to incorporate the context of volatile innovation processes, the topic guide introduces a robustness analysis to determine to what extent the outcomes are robust against slight changes in the innovation. Table 2 summarizes the findings from the previous sections and integrates the different debriefing phases with the topics and the ideal participants involved for each phase.

Method

We used gaming simulation to find causal patterns between an innovation, the independent variable, and a performance measure, the dependent variable. Unlike classical medical and psychological experiments, however, we apply the treatment to a system,

Table 2. Research Game Framework of Phases, Topics Addressed and Participant Involvement.

Phase	Description	Topics	Participant Involvement
Cooling down	Change game player's mental state from immersion to retrospection.	Experience Emotions	Facilitator Game players
Data collection	Additional qualitative data from players, observers and facilitators.	Measurement reliability Validity	All participants
Reliability	Assess whether repetition would result in similar outcomes.	Sensitivity	Game players Observers
Internal validity	Can we state with confidence that the experienced causal claim holds within the game situation?	Potential confounding variables	Game players Observers
External validity	Assess whether causal claim holds in real life (ecological) and for different samples (generalizability).	Game artificiality Impact of omissions in game model Sample-specific behavior	Game players Observers
Robustness	Do variations of the tested innovation, or the introduction of additional innovations, create strikingly different outcomes?	Longevity of the relevance of outcomes if innovation processes persist	Game players Observers
Planning for action	Determine what follow-up questions need to be answered. Determine what concrete actions need to be taken and by whom.	Future research questions and actions	All participants
Protect the instrument	Evaluate gaming simulation session. Determine what outcomes may be shared. Ensure a durable relationship with game players.	Experience Emotions	Facilitator Game players

which is the game model, rather than to a single atomistic entity. This system comprises many interdependent elements in a web of complex causal relationships and adaptable human game players. This difference is best explained using the notions of

trivial machines (TMs) and non-trivial machines (NTMs) by Von Foerster (1984), and it subsequently impacts how we can claim any causality after experimenting with systems. In general, opening the *black box of the game* run, and systematically assessing what has actually happened during the game, helps to find causality in NTMs such as our gaming simulations.

Trivial and non-trivial machines. In traditional experiments, researchers assume that some conceptual device transform the input x into output y and that this transformation is both linear and independent of context, time and history. In those instances when the relationship between x and y is established and the researcher is solely interested in prediction, there is no need to open up the black box of this device. How x causes y is irrelevant. In contrast to these trivial machines, non-trivial machines bring about causality in a far more complex manner. NTMs are devices in which the transformation of x into y is highly dependent on history, time and context and in which the device itself changes as a result of x . Social systems, consisting of adaptable and interdependent human beings, are perfect examples of NTMs (Klabbers, 2006). Here, how x is transformed into y , becomes highly relevant and the researcher thus needs to open up the black box (Von Foerster, 1984). Because we assume that the systems we manipulate in a gaming environment are like NTMs, we cannot simply perform a pretest and posttest with and without an innovation, as is customary in classical psychological and medical experiments.

An ontology of events and processes. Researching non-trivial machines in which potential causality is brought about by an interplay of complexity, path dependence, chaos and interdependence on multiple levels of analysis is common in the more qualitatively oriented fields of the historical and sociological sciences (Griffin, 1993; Hedström & Bearman, 2009). Here researchers rely heavily on narrative explanations that allow them to better describe what is actually going on and also to better incorporate the highly relevant context. This explanation is based on event sequences rather than relationships of variables (Abbott, 2001; Geels, 2011). For example, a usual description is that the Great Depression in the 1930s, Event A, partly triggered the Second World War, Event B. According to Weber (1949), most events are too complex to state any causal generalization about them, so claiming that economic decline and the likelihood of war are always causally related becomes infeasible. In contrast to linear causal models, narratives allow the researcher to gain insight into the complex interplay between social structure and human agency over time (Giddens, 1979; Griffin, 1993; Sewell, 1992). Narrative style explanations also gained more popularity in the fields of management sciences as topics became ontologically more complex and linear models failed to acknowledge this. Examples of this can be found in innovation management and organizational theory research (Langley, 2007; Tsoukas & Hatch, 2001; Van de Ven, Angle, & Poole, 2000) and in research on transitions of sociotechnical systems (Geels, 2011). Since what happens in a gaming simulation is really a sequence of events rather than a link of variables, their methodologies could support our

debriefing. Our games are more like discrete-event simulations than system dynamics models so observing them needs to acknowledge the *eventness* of the simulation.

Methodologies. Of all the methodologies applied by historians and sociologists, *event-structure analysis* seems to be the most developed (Heise, 1989). Event-structured analysis enables the researcher to structure events and portray how accumulations of past actions constrain or instigate future events. For a better overview of these methodologies we refer to Manzo (2010).

The event-structure analysis starts by drawing a timeline of the events that have occurred. In other words, the events have a specific temporal ordering. After this, one must determine the extent to which an event causally triggered the next event or another event later on. Key elements of this assessment are counterfactuals, which are negations or modifications of a specific event and basically involve asking *what if* questions (Griffin, 1993). If Event A1 occurred, could Event A2 also have occurred? Having done this, we must determine whether these counterfactuals are objective possibilities (Weber, 1949). This means that the counterfactual is in itself realistic and remains conceptually close to the real past. If the hypothetical negation or modification of the event would have caused a completely different unfolding of events later on, this event is a causal triggering for all subsequent events (Griffin, 1993). To assess this counterfactual world, researchers can either use other cases as a benchmark or theoretically deduce how the story would unfold. In Table 3, we briefly summarize the steps commonly found in narrative analyses that focus on causality.

Using this methodology helps to tackle all the topics in the aforementioned topic guide in a more systematic manner. The collective determination of events improves the data collection phase of the debriefing. In the years we have spent designing gaming simulations for the railway sector, we have found the tool to be an ideal method for enabling multiple stakeholders to holistically observe processes that would otherwise be separated in space and time. In addition to the more quantitative data usually logged during gameplay, more qualitative observations are possible. Qualitative data is valuable for two reasons: first, it is better able to capture the complex nature of the dynamics that occur during gameplay. Second, it requires less operationalization beforehand. This increases the possibility of testing innovations for which the performance measures are still being debated or hard to quantify. To fully benefit from this in the debriefing, there must be considerable attention for data collection during the design of the game and the experiment. Observers, most often designers of the innovation and subject matter experts, need to be present during the game and given instructions. Although retroactive accounts of game players cannot be identified beforehand, it is possible to determine what observers should look for in advance. For instance, observers could be provided with a topic guide. During the debriefing, all shared observations form a common picture of what occurred during the game. This serves the purpose of calibrating the observations and improving the measurement reliability, and of concertedly creating a chain of crucial events.

In addition to discussing how variables changed during gameplay, we discuss the event chains that caused these dynamics in the variables. In other words, we open up

Table 3. Event-Structure Analysis.

Step	Action	Description
1	Determine events	Map all game player decisions, changes in game parameters and context.
2	Determine counterfactuals	Map for the potential counterfactual events for every event.
3	Assess realism of counterfactual	Determine whether the counterfactual is close to the real past and is realistic in real life.
4	Determine counterfactual world	Assess to what extent the different event would trigger different subsequent events.

Source. Based on Griffin (1993).

the black box. In this phase of the debriefing, we map all key events that occurred during the game. We use the musical staff as a metaphor, with each line representing an element of the system, for instance: game player, train and station. The notes are events instigated by the element. The story is the temporal progression of events. For instance, a train might break down as Event A, which invokes a response by a traffic controller as Event B, and so on. Figure 1 presents a graphical example of this, with four elements and the green path representing the actual events and the blue path representing the possible alternative decisions.

However, it should be noted that the level of detail we use here is merely for didactic purposes. The level of detail we usually apply is much lower, focusing on around 10 events that best describe the gameplay. Some of the questions we use to draw up such an event chain are: What happened? What was crucial for the experienced gameplay? What processes did you observe?. The role of the facilitator is to combine all of these insights, assess their congruence and juxtapose contradictory observations.

The event chain becomes especially valuable for the systematic assessment of the validity and reliability of gaming simulation outcomes. In addition, it provides a good method to collectively discuss the impacts of inherent internal and external validity issues. To start, the internal validity of the causal claim is increased by determining *how* the innovation brought about changes in variables rather than simply stating *that* the innovation did so (George & Bennett, 2005). Following this, if we want to determine to severity of validity threats, both internal and external, we can use the event chain analysis to assess whether simulation outcomes are highly dependent on certain validity-threatening phenomena. For instance, we can assess whether certain events are triggered by omissions in the game model (to test its ecological validity), whether one decision by a game player could have just been another decision resulting in a different unfolding of events (to test its sensitivity), or whether other game players who were not involved in the game would have decided something different for a certain event (to test its generalizability).

The event-structure analysis would enable the facilitator to study the sensitivity of the simulation outcomes to validity issues. Critical decisions can be assessed to

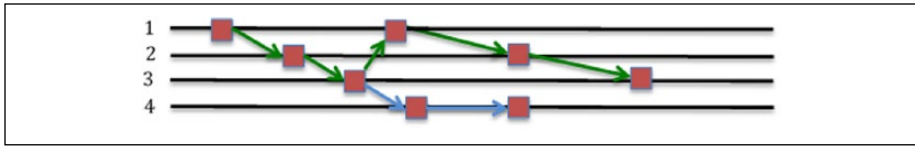


Figure 1. A four-element event chain system.

Source: Van den Hoogen et al. (2014a: 3511).

determine whether the player could have just as well have decided something else, and to what extent this would have caused a completely different unfolding of events, as represented by the blue path in Figure 1. Both game players and observers usually determine which decisions were critical. Immersed game players usually cannot recall all the events that took place during a game, but do know the significance of the decisions they have made. Observers are less aware of the significance but are more likely to recall decisions, especially when are tasked to do so. For this collective imagining of a different unfolding of events, we rely on a mental simulation of the changed game. An advantage of this is that low-tech gaming simulations are easy to re-use and so can serve to support this analysis. If the mental simulation places too much cognitive strain on the game players, we can use the game that is still available to quickly replay a few events.

Synthesis

In Table 4 we present a brief overview of a possible debriefing of a research game. For a more in-depth look at how we applied part of this framework in specific cases in the Dutch railway sector, we refer to Van den Hoogen et al. (2014a). The framework incorporates existing notions from the literature (Peters & Vissers, 2004; Kriz, 2010) and adds a more thorough operationalization of validity and a methodology by which to actually study validity in the debriefing. The table also shows the ideal roles of each participant.

Conclusion

Although the methodological literature on research games is slowly growing, the literature on debriefing this specific type of gaming simulation is scarce. For this reason this article provides a useful framework for debriefing research games. In particular, the paper focused on research games used for innovation. Discussing robustness and planning for action are especially important for this specific context of use, whereas the other phases are also valuable for games used for fundamental research.

We used our framework to tackle the context, substance and method of debriefing research games. The debriefing framework enables a gaming simulation to do justice to the volatile context of innovation processes. By making debriefing intrinsic to the design considerations of a research game, the game designer is better able to cope with this volatility. As far as substance is concerned, the framework delves deeper into the

Table 4. Reflection of Validity and Reliability Issues During Debriefing.

Phase	Participant Debriefing Roles		
	Player/Operator	Observer/SME	Facilitator
Cooling down	Taking a break, discussing game experiences	Summarizing observations	Leading discussions on game experience
Data Collection	Establishing event chains	Establishing event chains	Juxtaposing statements; assessing measurement validity and reliability
Sensitivity	Determining counterfactuals and their effects on subsequent events (based on experience)	Determining counterfactuals and their effects on subsequent events (based on theory, rules, etc.)	Asking players and observers about crucial events and objective possibilities
Internal Validity	Determining how treatment impacted the events; determining effect of confounding variables	Determining how treatment impacted the event chain; determining effect of confounding variables	Identifying potential confounding variables due to experimental context
Generalizability	Comparing own decisions with probable decisions made by peers; comparing sensitivity of decisions to changes in other dimensions of the sample: different timetable, etc.	Identifying differences between the sample and the population	Linking differences found by observers with players' comparisons
Ecological Validity	Determining perceived realism and effect of omissions of elements and processes of referent system on event chains	Determining the effect of omissions of processes and structural properties of referent system on event chains in game	Discussing what omissions were applied during game design
Robustness	Determining effects of changes in innovation and introduction of additional innovations on event chains	Determining in what ways the innovation might change later on	Introducing the agreed-upon leftover category of innovations not tested in the game run
Planning for action	Determining to what extent other operators are able to handle the innovation once implemented, and if additional training is needed	Determining what follow-up research is needed and how concrete actions will be coordinated with all stakeholders	Summarizing findings of the previous discussions to start up this phase
Protect the instrument	Discussing to what extent the innovation or the game was controversial and what can and cannot be fed back into real world	Discussing to what extent the innovation or the game was controversial and what can and cannot be fed back into real world	

specific topics a debriefing should address: data collection, sensitivity, internal validity, generalizability, ecological validity and robustness. Event-structure analysis, a method used in the qualitative historical and sociological sciences, allows for a more thorough and rigorous analysis of causality and validity, thereby opening the black box of the game run.

The limitations of this framework are twofold. First, the framework was distilled from the many experiences we gained in designing and debriefing a multitude of different games, yet, the applicability of the framework in its entirety has still to be tested. Future research should look at the feasibility of rigorously applying this entire framework in a debriefing. Such a study could also examine whether or not the framework improves the gaming simulation, by whatever metric. For the practitioner, it could result in a set of exemplary questions that operationalize the dimensions and phases mentioned in this article. Secondly, our proposed method requires game players to mentally simulate the answer to *what if* questions. Given that the focal point is a complex system, the extent to which game players are able to do this is still debatable. However, the player's ability largely determines the validity of the claims we make on basis of the debriefing. Are the results really robust or is the game player simply unable to perceive that a slight change in the innovation will bring about radical changes in the dynamics of the system? Future research could look at the players' cognitive capacities as well as methods to improve the collective assessment of alternate courses of gameplay. Nevertheless, we feel that debriefing is an intrinsic part of designing games for research as well as games for learning. We have seen how debriefing has become more and more intertwined with the designing of models and simulations. With this framework, we intend to improve this cross-fertilization between gaming and debriefing.

Author Contributions

J.H, J.L and S.M. designed the gaming simulations. S.M. facilitated the gaming sessions. S.M. conducted the debriefing sessions. J.H. and J.L. analyzed the debriefing and designed the framework. J.H. and J.L. wrote the manuscript. S.M. reviewed the manuscript.

Acknowledgments

This article is an augmented and enhanced version of a paper presented at the 45th annual international conference of the International Simulation and Gaming Association (ISAGA), Dornbirn, Austria, July 7-11, 2014. The authors would like to thank the project teams involved in the gaming sessions, both at ProRail and Delft University of Technology. Special thanks go out to Gert Jan Stolk and Emdzad Sehic. Additionally, we would like to thank the anonymous reviewers for their valuable comments.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is funded by the Railway Gaming Suite (RGS), a joint project of Delft University of Technology and ProRail.

References

- Abbott, A. (2001). *Time matters: On theory and method*. Chicago, IL: University of Chicago Press.
- American Psychological Association, American Educational Research Association, and National Council on Measurement in Education. (1966). *Standards for educational and psychological tests and manuals*. Washington, DC: American Psychological Association.
- Anderson, E. G., & Joglekar, N. R. (2012). *The innovation butterfly*. New York, NY: Springer.
- Cheng, Y. T., & Van de Ven, A. H. (1996). Learning the innovation journey: Order out of chaos? *Organization Science*, 7, 593-614.
- Crookall, D. (2010). Serious games, debriefing, and simulation/gaming as a discipline. *Simulation & Gaming*, 41, 898-920.
- Decker, S., Fey, M., Sideras, S., Caballero, S., Rockstraw, L., Boese, T., . . . Borum, J. C. (2013). Standards of best practice: Simulation standard VI: The debriefing process. *Clinical Simulation in Nursing*, 9(6), S26-S29.
- Dennehy, R. F., Sims, R. R., & Collins, H. E. (1998). Debriefing experiential learning exercises: A theoretical and practical guide for success. *Journal of Management Education*, 22, 9-25.
- Fanning, R. M., & Gaba, D. M. (2007). The role of debriefing in simulation-based learning. *Simulation in Healthcare*, 2, 115-125.
- Geels, F. W. (2011). The multi-level perspective on sustainability transitions: Responses to seven criticisms. *Environmental Innovation and Societal Transitions*, 1, 24-40.
- George, A. L., & Bennett, A. (2005). *Case studies and theory development in the social sciences*. Cambridge, MA: MIT Press.
- Giddens, A. (1979). *Central problems in social theory: Action, structure, and contradiction in social analysis*. Berkeley: University of California Press.
- Greif, S., & Kurtz, H. J. (1996). *Handbuch selbstorganisiertes lernen* [Manual self-organized learning]. Göttingen, Germany: Verlag für Angewandte Psychologie.
- Griffin, L. J. (1993). Narrative, event-structure analysis, and causal interpretation in historical sociology. *American Journal of Sociology*, 98, 1094-1133.
- Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 42, 1009-1055.
- Hedström, P., & Bearman, P. (2009). *The Oxford handbook of analytical sociology*. Oxford, UK: Oxford University Press.
- Heise, D. R. (1989). Modeling event structures. *Journal of Mathematical Sociology*, 14, 139-169.
- Klabbers, J. H. (2003). Gaming and simulation: Principles of a science of design. *Simulation & Gaming*, 34, 569-591.
- Klabbers, J. H. (2006). A framework for artifact assessment and theory testing. *Simulation & Gaming*, 37, 155-173.
- Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice Hall.
- Kriz, W. C. (2003). Creating effective learning environments and learning organizations through gaming simulation design. *Simulation & Gaming*, 34, 495-511.

- Kriz, W. C. (2010). A systemic-constructivist approach to the facilitation and debriefing of simulations and games. *Simulation & Gaming, 41*, 663-680.
- Kriz, W. C., & Hense, J. U. (2006). Theory-oriented evaluation for the design of and research in gaming and simulation. *Simulation & Gaming, 37*, 268-283.
- Langley, A. (2007). Process thinking in strategic organization. *Strategic Organization, 5*, 271-282.
- Lederman, L. C. (1992). Debriefing: Toward a systematic assessment of theory and practice. *Simulation & Gaming, 23*, 145-160.
- Lederman, L. C., & Stewart, L. P. (1986). *Instruction manual for THE MARBLE COMPANY: A simulation board game*. New Brunswick, NJ: Rutgers University.
- Lo, J. C., Van den Hoogen, J., & Meijer, S. A. (2013). Using gaming simulation experiments to test railway innovations: Implications for validity. In R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, & M. E. Kuhl (Eds.), *Proceedings of the 2013 Winter Simulation Conference* (pp. 1766-1777). Piscataway, NJ: Institute of Electrical and Electronics Engineers.
- Manzo, G. (2010). Analytical sociology and its critics. *European Journal of Sociology, 51*, 129-170.
- Meijer, S. A. (2012). Gaming simulations for railways: Lessons learned from modeling six games for the Dutch infrastructure management. In X. Perpinya (Ed.), *Infrastructure design, signaling and security in railway* (pp. 275-294). Rijeka, Croatia: IntechOpen.
- Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist, 30*, 955-966.
- Morton, R. B., & Williams, K. C. (2010). *Experimental political science and the study of causality: From nature to the lab*. New York, NY: Cambridge University Press.
- Peters, V., & Vissers, G. (2004). A simple classification model for debriefing simulation games. *Simulation & Gaming, 35*, 70-84.
- Peters, V., Vissers, G., & Heijne, G. (1998). The validity of games. *Simulation & Gaming, 29*, 20-30.
- Qudrat-Ullah, H. (2007). Debriefing can reduce misperceptions of feedback: The case of renewable resource management. *Simulation & Gaming, 38*, 382-397.
- Raser, J. C. (1969). *Simulation and society: An exploration of scientific gaming*. Boston, MA: Allyn & Bacon.
- Sewell, W. H. (1992). A theory of structure: Duality, agency, and transformation. *American Journal of Sociology, 98*, 1-29.
- Sims, R. R. (2002). Debriefing experiential learning exercises in ethics education. *Teaching Business Ethics, 6*, 179-197.
- Thiagarajan, S. (1993). How to maximize transfer from simulation games through systematic debriefing. In F. Percival, S. Lodge, & D. Saunders (Eds.), *The Simulation and Gaming Yearbook 1993* (pp. 45-52). London, England: Kogan Page.
- Tsoukas, H., & Hatch, M. J. (2001). Complex thinking, complex practice: The case for a narrative approach to organizational complexity. *Human Relations, 54*, 979-1013.
- Van de Ven, A. H., Angle, H. L., & Poole, M. S. (2000). *Research on the management of innovation: The Minnesota studies*. New York, NY: Oxford University Press.
- Van den Hoogen, J., Lo, J. C., & Meijer, S. A. (2014a). Debriefing in gaming simulation for research: Opening the black box of the non-trivial machine to assess validity and reliability. In A. Tolk, S. D. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, & J. A. Miller (Eds.), *Proceedings of the 2014 Winter Simulation Conference* (pp. 3503-3516). Piscataway, NJ: Institute of Electrical and Electronics Engineers.

- Van den Hoogen, J., Lo, J. C., & Meijer, S. A. (2014b). The debriefing of research games: A structured approach for the validation of gaming simulation outcomes. In W. C. Kriz (Ed.), *The shift from teaching to learning* (pp. 88-100), Bielefeld, Germany: W. Bertelsmann.
- Van den Hoogen, J., & Meijer, S. A. (2015). Lessons on the design of gaming simulations for convergence and divergence in volatile innovation environments. In L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, & M. D. Rossetti (Eds.), *Proceedings of the 2015 Winter Simulation Conference* (pp. 1044-1055). Piscataway, NJ: Institute of Electrical and Electronics Engineers.
- Van der Meij, H., Leemkuil, H., & Li, J. L. (2013). Does individual or collaborative self-debriefing better enhance learning from games? *Computers in Human Behavior*, 29, 2471-2479.
- Van Ments, M. (1983). *The effective use of role-play: A handbook for teachers and trainers*. London, England: NP Cogan Page.
- Von Foerster, H. (1984). Principles of self-organization - in a socio-managerial context. In H. Ulrich & G. J. B. Probst (Eds.), *Self-organization and management of social systems* (pp. 2-24). Berlin, Germany: Springer.
- Weber, M. (1949). *The methodology of the social sciences*. New York, NY: Free Press.
- Zechmeister, J. S., Zechmeister, E. B., & Shaughnessy, J. J. (2001). *Essentials of research methods in psychology*. New York, NY: McGraw-Hill.

Author Biographies

Jop van den Hoogen is a PhD candidate at Delft University of Technology. His research focuses on decision-making patterns during systemic innovation processes in the railway sector. Specifically, he studies the extent to which simulation and gaming can influence these patterns.

Contact: j.vandehoogen@tudelft.nl

Julia Lo is a PhD candidate at Delft University of Technology. With a background in Psychology and Communication Science, her research focuses on studying (team) situation awareness of operators in the railway sector through the use of gaming simulation methods.

Contact: j.c.lo@tudelft.nl

Sebastiaan Meijer is a Professor of Healthcare Logistics at KTH Royal Institute of Technology and a part-time Associate Professor at Delft University of Technology. He specializes in gaming simulation and other interactive methods to involve the operational level of organizations in innovation processes.

Contact: sebastiaan.meijer@sth.kth.se