Chapter 6
Examining Designed Experiences: A Walkthrough for Understanding Video Games as Performance Assessments

Michael P. McCreery, P. G. Schrader, S. Kathleen Krach, Jeffrey R. Laferriere, Catherine A. Bacos, and Joseph P. Fiorentini

6.1 Introduction

Empirical investigations of video games follow a few primary approaches. Typically, they examine: (1) consequences of gaming (e.g., learning from games; De Freitas, 2018), (2) interactions with games (e.g., from a human–computer interaction perspective; Fortes Tondello et al., 2018), or (3) learning within games as a situated context (Jabbari & Eslami, 2019). Broadly, the majority of learning-related video game literature tends to fall into one of four general categories: intervention studies (Stefanidis, Psaltis, Apostolakis, Dimitropoulos, & Daras, 2019), addiction studies (Mancini, Imperato, & Sibilla, 2019), learning studies (Wouters, Van Nimwegen, Van Oostendorp, & Van der Spek, 2013), or social interaction research (McCreery, Vallett, & Clark, 2015).

Although the breadth of work associated with learning and video games continues to develop, there is a dearth of examples on how to extract complex, dynamic, and emergent data using video game contexts. Similarly, there are limited examples that outline strategies and tools for interpreting game-based data. As such, the main purpose of this paper is to outline one possible process to use the complex environ-

M. P. McCreery (✉) · P. G. Schrader · C. A. Bacos · J. P. Fiorentini
University of Nevada, Las Vegas, Las Vegas, NV, USA
e-mail: michael.mccreery@unlv.edu; pg.schrader@unlv.edu; bacosc@unlv.nevada.edu; fiorej2@unlv.nevada.edu

S. K. Krach
Florida State University, Tallahassee, FL, USA
e-mail: skrach@fsu.edu

J. R. Laferriere
Lebanon Valley College, Annville, PA, USA
e-mail: laferrie@lvc.edu

D. Ifenthaler, Y. J. Kim (eds.), *Game-Based Assessment Revisited*, Advances in Game-Based Learning, https://doi.org/10.1007/978-3-030-15569-8_6

ment of a video game as a data collection tool. Readers should expect to exact a greater understanding of how data captured from observing video gameplay can be used in conjunction with path analytic techniques to elucidate the process of learning. Fundamentally, this work exposes strategies to leverage existing off-the-shelf video games as contexts for performance assessment.

6.2 Performance Assessments

There has been substantive effort to evaluate performance in video games as spaces for experiential learning (i.e., how game experiences impact learning; Anetta, Minogue, Holmes, & Cheng, 2009; Harvianinen, Lainema, & Saarinen, 2014; Shaffer, Squire, Halverson, & Gee, 2005; Squire, 2011). However, less research has been conducted on leveraging video games as encapsulated, performance assessments (i.e., how interconnected gameplay experiences influence outcomes). At their core, performance assessments are grounded in the principle that learning occurs within a situated or sociocultural context (Wang, Shute, & Moore, 2015). From this perspective, learners develop mental representations (i.e., schemata, scripts) as they interact with the world. Subsequently, those representations are called upon as heuristics to aid in decision-making processes (Govaerts, Van Der Vleuten, Schuwirth, & Muijtjens, 2007). Accordingly, the best way to assess performance learning is to ask the learner to demonstrate higher-order thinking and apply their conceptual understanding of the world in novel situations (Shavelson, Baxter, & Gao, 1993).

Typically, performance assessments are designed in ways that position the learner to: (a) perform a goal-oriented exercise that demonstrates success on a summative task, and (b) demonstrate understanding of the process or steps associated with its successful completion (Shavelson et al., 1993). This dual-oriented emphasis (i.e., goal-oriented performance from a process-oriented lens) serves to reveal the connection between higher-order thinking and conceptual understanding in novel situations. Consequently, performance assessments differ substantially from most traditional assessments, particularly multiple-choice tests. For example, items on multiple choice tests are generally designed to be independent of one another; items can be arranged in any order, and success on one item does not influence the success on subsequent items (Yen, 1993).

In contrast, performance assessments are defined in terms of item interdependence. In most cases, a setting (e.g., narrative) is first established and learners must make decisions within that narrative (Yen, 1993). Each decision has predefined and intentional dependencies that are linked to previous choices and early choices have implications for subsequent decisions. For example, some decisions may expose new options or limit choices. As such, specific decisions may be examined formatively; while collectively, the sum of those activities can be examined in the context of summative outcomes to provide meaningful insight into the overall process, degree, and nature of learning (Shute, Leighton, Jang, & Chu, 2016).

6.3 Video Games and Assessment

For decades, researchers have asserted that video games are rich tools and environments for the study of learning and related mechanisms (de Freitas, 2018). However, in recent years this work has expanded its focus to include the examination of process-oriented data (Schrader, McCreery, & Vallett, 2017). From this perspective, games provide access to behavioral and learning data that are dynamic, emergent, and complex. Researchers have argued that these process-oriented data have great potential to yield insight into learning as it evolves through gameplay. For example, Vallett (2016) described the dynamic process of acting and adjusting behavior to the environment as situated learning via "soft failure" (e.g., dying and restarting a level). Here, gameplay experiences act as a performance tuning mechanism (Schrader et al., 2017; Vallett, 2016). Each interaction within the system provides information and a potential source of data. Players must discern what information is useful and adjust their behavior accordingly. Failure is inevitable and when it occurs, the situation provides the player an opportunity to reevaluate the usefulness of the information, problem solve, and reattempt the action (Schrader et al., 2017; Vallett, 2016). Collectively, these data provide evidence of patterns of behavior during the learning process. As assessments, games offer more than a mechanism to examine performance through outcomes. Games provide new opportunities for researchers to collect, analyze, and interpret data during these experiences (Schrader et al., 2017).

Although it is often difficult to capture process-oriented data, games regularly monitor interactions within the environments and commonly collect data on player performance (Shute, Ke, & Wang, 2017). While these data are typically used to provide feedback and cues for players, the same data may be captured and used by researchers to provide unique and additional insights into variables associated with processes (Schrader et al., 2017). It follows from this perspective that although *summative* evaluation of performance is useful for many questions, the development of a meaningful *formative* understanding of learning through systematic observation and analysis of behaviors within a video game (e.g., game's cues and player's actions) adds numerous options to researchers' repertoire (Schrader et al., 2017).

By leveraging games as performance assessments that capture process data (i.e., data that are complex, dynamic, and emerge over time), researchers can look beyond the gameplay as a singular or aggregated experience to be observed. This subtle, yet important, shift augments the research perspective in a fundamental way by moving the focus from assessments characterized by success or failure, to understanding how higher-order thinking and the learner's conceptual understanding of the world informs connected outcomes (Schrader et al., 2017; Shute et al., 2017). With respect to games that provide a finite number of choices, the game structure is similar to a nested multiple-choice decision tree or flowchart. In this example, each decision relies on the previous one, and taken as a whole, performance can be characterized by the path that player takes coupled with the outcome (e.g., Tic Tac Toe, Othello, or a Moral Choice game). As noted earlier, each gameplay decision is interdependent with other decisions. By extension, play serves as an opportunity to document and capture dynamic, in-game interactions, link those interactions to formative activities, and then examine the ways in which those activities influence the overall goal.

With these ideas in mind, and because games differ significantly in their structure, affordances, and capabilities, we first outline the factors involved with evaluating a game's suitability (Schrader & McCreery, 2012). In particular, we focus on games that function as complex systems and produce data that are aligned to a process-oriented perspective (Schrader & McCreery, 2012). Second, we establish a heuristic for identifying data and their coding. Third, we explore analytic techniques that are appropriate to process-oriented data. In this case, we describe path analysis and its potential to elucidate how player interactions are tied to learning as an emergent, dynamic process. Throughout, methods for capturing, coding, and analyzing within-game data are described pursuant to this goal.

6.4 Game Selection

Researchers have described various reasons for selecting the specific video game contexts they study. In some cases, the environments are constructed as part of broader work (e.g., Quest Atlantis, River City, or Whyville). In others, selection criteria and rationale focus on game popularity or interesting interactions within the system (see Schrader & McCreery, 2008). Whatever the reason, game selection is a vital component of the research process. The game governs the types of affordances that are available to players, shapes the research questions, informs the types of data that can be collected, and impacts researchers' choice of designs and methods. When a dual-oriented emphasis (i.e., goal-oriented performance from a process-oriented lens) is adopted, game selection is even more important.

In general, all players' choices within games can be represented or mapped in some manner. For example, actions within open-world games, although vast and overwhelming, can be observed as classes, categories, or groups of actions that are based on the constraints and affordances of the game being investigated. By contrast, player decisions within moral-choice games (i.e., *The Deed*) are finite and can be mapped more easily. When represented visually, the decision structure is similar to a flowchart, in which each fork represents a choice or interaction within the game. Similar to a performance assessment, each fork provides the player with an opportunity to select an optimal or a suboptimal solution (i.e., correct or incorrect choice). As a result, these actions serve as isolated error checks, as well as a more holistic performance assessment that is readily quantified and analyzed. In this way, the format of the game provides an ideal platform to evaluate gameplay performance methodology; specifically, concrete data that are specific to the player's decision-making processes at every stage of gameplay.

In most games, the structures, models, algorithms, and rules within these systems are implicit. As a result, the deconstruction of the game model begins with an inductive process associated with extensive play or game experience (Schrader, Deniz, & Keilty, 2016). Essentially, researchers are encouraged to observe the various options for action and the constraints on action, particularly as they relate to the agency of: (a) players, (b) developers, and (c) researchers. Although there may be some overlap, the agency for players is often different than the agency for developers or researchers. For example, the ability to access command line input may be available to developers, but unavailable to players because they are intended to rely more heavily on visual stimuli. Collectively, player and developer affordances inform everything from the type of questions that are appropriate to opportunities for data collection. It should be noted that this process is focused on the potential for action and the constraints imposed on the system rather than the intentions behind either. For these reasons, the deconstruction of the game model is both reasonable and necessary; it provides a means to evaluate key design characteristics and affordances (e.g., narrative and gameplay mechanics) in relation to research suitability. This typically happens prior to game selection, but certainly before any empirical study commences.

Often, environments are selected because they are popular and/or have a set of features that give rise to interesting studies or player interactions. This means that research frequently involves commercial and publicly available software. Unfortunately, researchers do not usually have access to the design principles, guidelines, or gameplay diagrams. Similarly, it is very difficult to capture clickstream data, process data, or the "under-the-hood" mechanics due to the proprietary nature of commercial games. For researchers, this is a common scenario and often requires a labor-intensive scheme to extract and code data from the system. In this case, researchers identified, catalogued, and mapped all available actions within the game. This is a necessary step in quantifying key data for analysis.

6.5 Selecting the Deed

In the current example, *The Deed* (Grab the Games, 2015) was selected because of its structure, compelling story and plot, and alignment with guidelines for performance assessments (see Shute et al., 2017). The process of selecting *The Deed* followed the same approach identified above. Members of the research team identified the game as a potential candidate for research based on reviews and game descriptions. Subsequently, they played the game multiple times with an intent to identify the key elements of agency in the game based on what players might be able to accomplish through their experience, what developers intended, and how those two perspectives might inform research. Briefly, *The Deed* is a moral-choice role-playing murder mystery video game in which players' in-game decisions are limited in ways that are like a choose-your-own-adventure novel. There is a compelling social narrative that contextualizes a complex, puzzle-oriented game that focuses on the players' ability to reverse traditional moral roles. Unlike many other murder-mystery games, the objective of *The Deed* is to commit the act of murder (i.e., "the deed") and secure the family inheritance, rather than solve a crime that has been committed. The plot involves murdering the main character's own sister, framing another character for the murder, and ensuring that the main character avoids conviction for the crime. The plot helps shape players' decisions and social interactions, all of which result in a finite number of outcomes. More importantly, the social interactions with characters in the game allow players to unravel the clues to the social puzzle they are attempting to solve (e.g., interacting with characters, and the various weapon and evidence choices).

Similar to a play, the narrative of *The Deed* can be divided into five experiences: The Introduction and Four Acts. These acts include: (1) the homecoming (2) the dinner (3) the deed, and (4) the murder investigation and verdict. At the start of the game, the player has an opportunity to read the Introduction. This is the first learning opportunity for the player. If the player chooses to read the Introduction, they receive critical information that includes how to experience the game narrative, the importance of weapon and evidence selection (i.e., formative activities), and how planting evidence will impact the outcome (i.e., the summative outcome). Act One immediately follows the Introduction. Throughout this act, the player is given numerous learning opportunities to interact with characters (i.e., maid, butler, mother, father, and sister) and objects (i.e., weapons, evidence items, and story flashbacks). These interactions are intended to help players gain critical information to better develop problem-solving strategies. Moreover, they inform a set of formative tasks, including the successful (or not) selection of a weapon and an item of evidence that will be used to commit the deed and scapegoat another character for the murder. The player is given the choice to engage in these learning opportunities or to pass on them. However, in order for the player to move on to the second act, two items must be selected (i.e., a weapon and piece of evidence [correct response], two weapons, or two pieces of evidence [incorrect response]).

Act Two consists of a dinner celebrating the father's birthday. The player is seated at a table while interacting with other characters through a series of response options to statements made during the dinner conversation. Act Three is when the deed is committed; during this act, gameplay includes the formative tasks of successfully planting the evidence selected and using the weapon selected in Act One. The player has the option to forgo planting evidence and advance to committing the deed. However, not planting evidence is the only option if the player decided not to select an item of evidence during Act One (i.e., selected two weapons). Conversely, if the player decided not to select a weapon in Act One (i.e., selected two items of evidence), the only option is to commit the deed using the character's bare hands. Finally, in Act Four the murder investigation takes place. The player faces an investigator who has been called to the house. During the interview with the investigator, the player is questioned in relation to their prior decisions. In order to achieve a successful summative outcome (i.e., not going to prison), the player must succeed at each of the formative tasks presented throughout the narrative.

Ultimately, *The Deed* was determined to: (1) be a contextualized experience (i.e., social narrative); (2) provide clear linkages between choices (i.e., formative activities); and (3) be a goal-oriented exercise (i.e., summative outcome). In total, this game can take up to an hour to complete. For the purpose of research and assessment, this short time period is crucial (see Schrader et al., 2017). It may be unreasonable to use a game where players have different levels of expertise (McCreery, Schrader, & Krach, 2011), or that are overly time consuming given the purpose of the assessment (Kline, 2005). Collectively, these characteristics, evident in *The Deed*, provided researchers with access to, and the ability to assess, transactional learning experiences during gameplay in a situation that meets the added constraints (e.g., time, setting, replicability) that researchers often impose on design. In other words, learning experiences within *The Deed* are grounded in the interplay among the learner (i.e., player), context (i.e., narrative), and content (i.e., plot) (Moore, 1993).

Essentially, the game selected for this study was reverse engineered to understand the behind-the-scenes game mechanics that afford the range of player actions and outcomes in the game. Because *The Deed* involved a finite number of choices, the act of defining game elements and choices was somewhat straightforward. The selection and deconstruction process resulted in a *data dictionary* and *behavioral observation protocol* through which all gameplay data could be collected and analyzed.

6.6 Creating a Data Dictionary

Once the researcher has played the game, consumed other details and media, and deconstructed its mechanics, the next step is to define pertinent game elements. In some cases, this means observing general trends of players' interactions. For example, McCreery et al. (2015) created a matrix of observable behaviors that was based on Whiteside's model of social presence (Whiteside & Garrett Dikkers, 2012). The researchers then addressed questions related to players' interactions within a complex, dynamic, and emergent game (i.e., *World of Warcraft*) through cataloging observed behaviors in the game. By contrast to the open-endedness of the *World of Warcraft*, as well as many other games, *The Deed* includes a finite number of choices. Although there is no set pattern or pre-scripted path through the game, researchers were able to identify and define all game content. As a result, each opportunity for action and all player interactions were able to be tracked and analyzed. In this case, a detailed inventory of actions and interactions was appropriate because of the specific type of game originally selected. Below are the suggested steps of a game deconstruction process:

1. Identify all potential outcomes: go to prison (failure); get away with murder but no inheritance (partial success); get away with murder and gain inheritance (full success).
2. Identify the formative activities that must be accomplished in order to achieve a successful outcome: weapon selection, evidence selection, evidence planting.
3. Identify broad categories of in-game affordances that players can interact with in order to gain information necessary for problem-solving: non-player characters (i.e., computer controlled), weapons, evidence, flashback objects (e.g., painting on a wall that when interacted with provides narrative clues).
4. Identify all individual in-game affordances within each broad category (i.e., each character; weapon; piece of evidence; and flashback object).

The sum of all this information resulted in a *data dictionary*. In this example, a data dictionary outlined and defined key concepts, terms, ideas, and behaviors that were known to exist in the game. The data dictionary was created to provide the entire research team with consistent and shared understanding of game elements, features, mechanics, and play. Further, the data dictionary allowed the team to organize and categorize each of the game elements based on the constructs being analyzed and the variables being measured.

6.7 The Behavioral Observation Protocol and Coding Data

Once the essential elements of a game are defined and, in this case, categorized in a dictionary, the next step involves creating a resource for coding. For this example, a behavioral observation protocol was developed that included an array of important, observable player exhibited behaviors (i.e., it happened or it didn't) in order to limit qualitative inference. These behaviors were organized in ways that address the research question and its underlying theoretical framework. Moreover, whether researchers are mapping the game space in its entirety or a targeted set of behaviors (see McCreery, Krach, Schrader, & Boone, 2012 for an example), a behavioral

Evidence Interations

UID	E-LP	E-LT	E-MD	E-UG	Evidence Selected	Evidence Planted	Total
000	0	2	1	1	E-LT	0	4

LP (Love Poem); LT (Leather Tawse); MD (Mother's Diary); UG(Undergarments)

Character Interactions

UID	Maid	Butler	Mother	Father	Sister	Total
000	2	1	2	1	2	8

Weapon Interactions

UID	W-BH	W-BR	W-CS	W-FS	W-GL	W-KN	W-RO	W-RP	W-Q	W-SG	Weapon Selected	Weapon Used	Total
000	0	0	0	1	0	0	0	1	0	1	W-GL	W-GL	3

BH (Bare Hands); BR (Broom Handle); CS (Candlestick); FS (Fencing Sword); GL (Shard of Glass); KN (Knife); RO (Rope); Q (Pool Cue); SG (Shotgun)

Flashbacks

UID	F-CH-MO	F-GC-FA	F-MI-SI	Total
000	0	1	0	1

Trigger Object - Character Involved, CH-MO (Chair-Mother); GC-FA (Class Cabinet-Father); MI-SI (Mirror-Sister)

Fig. 6.1 Behavioral observation protocol example

observation protocol provides boundary conditions on the behavior that must be recorded and those that are not pertinent to the questions being answered (Alevizos, DeRisi, Liberman, Eckman, & Callahan, 1978; Milne, 2015).

The development of a behavioral observation protocol is an applied psychological approach to data collection that in the context of a video game entails two major steps. First, researchers begin by translating the elements of the data dictionary into a spreadsheet(s) that will become a comprehensive record of relevant player behaviors. This spreadsheet becomes a scorecard on which to record (i.e., tally) all of the observable behaviors, formative activities, and summative outcomes for each player. Behaviors must be operationally defined (e.g., specific, quantifiable, observable, concrete action) in order to ensure content validity and interrater reliability (Tapp, Wehby, & Ellis, 1995). Second, the protocol template is then generated for each player and distributed to the coders. The template then serves as a checklist for each coder to observe and record player behavior. For example, in Fig. 6.1, four types of interactions (i.e., evidence, character, weapon, and flashback) as defined during the creation of the data dictionary were translated into the behavioral observation protocol. Additionally, more specific interactions associated with interaction type (e.g., E-LP = evidence, love poem) are also defined. The coder can then record every time a player (represented by UID or user identification in the example) interacts with that specific element of the game.

The behavioral observation protocol was created to account for each of the possible interactions in *The Deed*. In Act One, the following player behaviors were recorded based on elements defined in the data dictionary: watching the introduction, dialogue with characters, story flashbacks viewed (i.e., objects in the story setting that when selected trigger a story flashback revealing more information about the other characters), weapons viewed and selected, and items of evidence viewed and selected. In Act Two, the dialogue with characters during dinner is coded in the same format as conversations in Act One. The dialogue checklist for the coder provides a listing of all the character statements and response choices to those statements. While viewing the video recording of the player's gameplay, the coder checks a box indicating the character interaction (e.g., spoke with the mother) and the response selected among the possible options listed for that character interaction (e.g., response choice 1, 2, or 3).

In Act Three, coders used a checklist to mark whether the player planted evidence selected in Act One, where the evidence was planted, and finally, what weapon was used to commit the murder. In Act Four, coders used a checklist to indicate responses to the crime investigator's interview questions. A checklist was also provided to coders to indicate one of the following outcomes: (1) the player was convicted of murder and sent to prison, (2) the player was not convicted of murder, or (3) another character was convicted of the murder because of the evidence planted against them, and the player received the inheritance.

6.8 Analytics of Gameplay

Once all the data from the player's gameplay is recorded, additional spreadsheets can be created for each of the constructs and related variables being measured as defined in the data dictionary. Further, because the nature of the data is a count (i.e., it happened or it didn't) interrater agreement in its true form, consistency of subject ratings is not needed (McHugh, 2012). However, for the sake of accuracy interrater data should be collected. In the present example, the coded spreadsheets for *The Deed* noted each interaction (exogenous variables) with weapons, story flashbacks, characters, and evidence items. The coded spreadsheet also noted the successful completion of each linked outcome (endogenous variables) across the game. Specifically, the variables coded as formative outcomes included: successful selection of a weapon and evidence item (Item Selection); successful planting of the evidence (Evidence Planted); and finally, the summative outcome, successfully get away with murder (Successful Outcome).

6.9 Analytic Techniques to Understand Player Experience

Using this process, data that are extracted from observations of players' behavior within *The Deed* are dynamic, emergent, and complex. It is common practice in low-dimensional, independent systems to test for significance using techniques like, t-test, ANOVA, MANOVA, etc. By contrast, complex systems involve increasing degrees of emergence and higher levels of dimensionality; this ilk of analyses is not very informative or useful. Fortunately, there exists a variety of analytic techniques that have the potential to expose patterns in data extracted from video games. For example, time series techniques, analysis of spline equations, structural equation modeling, and path analysis have been used with this class of data. It should be noted that each approach has distinct assumptions and each address different types of questions. For more details, please refer to Little, Bovaird, and Slegers (2006).

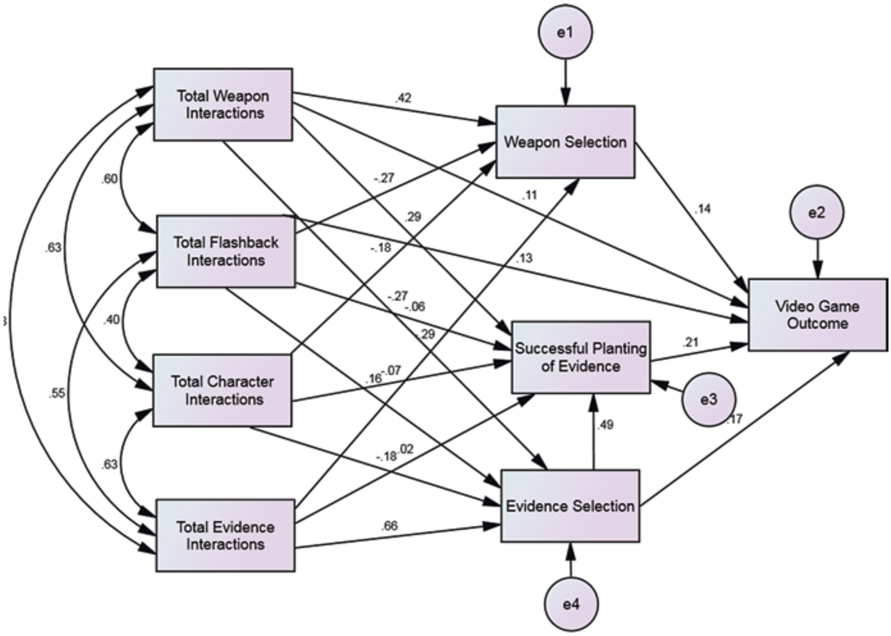

Fig. 6.2 Example path model

In this example, data were coded based on an event-dependent sample (as opposed to a time-dependent sample). Researchers employed path analysis to demonstrate causal effects among constructs in the game model: knowledge interaction, formative activities, and the summative outcomes. This form of analysis allows the researchers to link in-game *observable* information activities (emphasis added) directly with both formative and summative outcomes to better understand the process of learning. This process yielded a viable model (see Fig. 6.2) based on the relationships between the game constructs. While the details for this study are presented elsewhere (see McCreery, Laferriere, Bacos, & Krach, 2018), what should be noted is that the model illustrates that player outcomes are specifically related to the information acquired through interaction in the game space. For example, the more a player interacts with the available evidence (i.e., Total Evidence Interaction), the better is the understanding they appear to have in terms of the required Evidence Selection necessary to win the game. Alternatively, as a player increases their interaction with weapons (i.e., Total Weapon Interaction), the more likely those interactions become a distractor in terms of Evidence Selection necessary to win the game.

6.10 Discussion and Implications

The current work demonstrates the potential for video games to serve as unique and useful data-collection methods. By following the steps outlined in this chapter, researchers can extract data from complex contexts, in which players' choices can be represented or mapped. In the most general terms, researchers should plan carefully when deciding on the appropriate game to choose, how the game context allows for data collection of constructs of interest, and how the data can be collected in a psychometrically sound manner. Researchers are encouraged to plan for data collection in games from multiple lenses, perspectives, and levels. This includes whether it is appropriate to capture behavioral data. Moreover, if behavioral data are deemed appropriate, examine whether it is feasible to map the game space (e.g., *The Deed*) or does emergent gameplay (e.g., *World of Warcraft*) require a more targeted approach. Answers to these questions are critical as they will provide insight into the underlying mechanics and encapsulating contexts of games, and promote an increased understanding for the purpose of hypothesis generation, study design, data collection, data coding, and analytic approaches.

The example employed in this chapter (i.e., *The Deed*) is best characterized as a moral-choice game. By design, players are forced to make decisions in an attempt to achieve the game's main objective. From a limited point of view, the game is a finite collection of mappable choices that are either beneficial (right) or not (wrong). From this perspective, *The Deed* is structured in the same way as any performance assessment including: a contextualized narrative, goal-oriented summative outcome, and clearly linked formative activities. Moreover, unlike traditional multiple-choice tests, where each item is independent of one another and evaluated individually, in choice-based games, each decision is necessarily dependent upon the previous response. This suggests that there is an opportunity to examine choices at a discreet, individual level and also collectively as a whole. As a result, path analysis is the logical procedure to examine performance in these systems when overall performance, defined here to be the sum of all items is dependent upon one another.

Using this logic, information can be presented as a hint to aid the player or as distractor to lead them astray. Further, some choices could be considered correct answers (e.g., Evidence Selection), which are conducive to increased success. By contrast, distractor or error choices correspond with diminished success (e.g., the longer you examine your weapons choices, or Total Weapon Interactions, the less likely you are to experience success at the game). Ultimately, designers of *The Deed* presented information in three key ways: (a) there is information that is critical to success (e.g., information gained from interacting with pieces of evidence predicts the selection of evidence); (b) there is information that contributes to the atmosphere or narrative, but is not germane to the solution (e.g., interactions with flashback objects do not influence the selection of evidence); (c), there is information that is intended to distract and test your problem-solving ability.

Collectively, the manner in which the information is presented to the player and the heuristics that must be employed shift the focus of the experience away from a recall task to a situated performance assessment. Moreover, the fundamental structure of choice-based games and this process approach to capturing data, raise exciting possibilities for new forms of assessment. Future assessments could be designed to capture process data, rather than after the fact as presented here. There are several significant benefits to such a design: (a) it would provide researchers with a clearer understanding of how design elements impact the assessment (e.g., usability and psychometrics); (b) integrated data capture tools would limit resource expenditures (e.g., time coding data); and (c) provide a clearer manner in which to evaluate learning process discrepancies between actual and target learning.

Although the first two points are obviously important, the last one warrants additional discussion. Since the days of Dewey (1899), researchers and theorist alike have argued the importance of understanding learning as a process rather than solely an outcome. It is within the process that one can tease out misunderstanding, ineffectual problem-solving strategies, and misplaced heuristics. Game-based performance assessments may provide new opportunities to better understand how these issues arise. Specifically, a players' individual process model can be evaluated against the successful solution(s) in order to better understand where additional help should be given. This not only provides both teacher and learner with a more detailed understanding of where a problem(s) has emerged, but also discussion points to better understand both the *how and why* (emphasis added) choices were made.

References

Alevizos, P., DeRisi, W., Liberman, R., Eckman, T., & Callahan, E. (1978). The behavior observation instrument: A method of direct observation for program evaluation. *Journal of Applied Behavior Analysis, 11*(2), 243–257. https://doi.org/10.1901/jaba.1978.11-243

Anetta, L. A., Minogue, J., Holmes, S. Y., & Cheng, M. (2009). Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers & Education, 53*, 74–85.

de Freitas, S. (2018). Are games effective learning tools? A review of educational games. *Journal of Educational Technology & Society, 21*(2), 74–84. Retrieved from http://www.jstor.org/stable/26388380

Dewey, J. (1899). *The school and society*. Chicago, IL: University of Chicago Press.

Fortes Tondello, G., Valtchanov, D., Reetz, A., Wehbe, R. R., Orji, R., & Nacke, L. E. (2018). Towards a trait model of video game preferences. *International Journal of Human–Computer Interaction, 34*(8), 732–748.

Govaerts, M. J. B., Van der Vleuten, C. P. M., Schuwirth, L. W. T., & Muijtjens, A. M. M. (2007). Broadening perspectives on clinical performance assessment: Rethinking the nature of in-training assessment. *Advances in Health Sciences Education, 12*(2), 239–260. https://doi.org/10.1007/s10459-006-9043-1

Grab the Games. (2015). *The deed* [Computer software]. Retrieved from http://store.steampowered.com/app/420740/The_Deed/

Harvianinen, J. T., Lainema, T., & Saarinen, E. (2014). Player-reported impediments to game-based learning. *Transactions of the Digital Games Research Association, 1*(2), 1–11. Retrieved June 14, 2018, from http://todigra.org/index.php/todigra/article/view/14/21

Jabbari, N., & Eslami, Z. R. (2019). Second language learning in the context of massively multi-player online games: A scoping review. *ReCALL, 31*(1), 92–113.

Kline, T. (2005). *Psychological testing: A practical approach to design and evaluation*. Thousand Oaks, CA: Sage.

Little, T. D., Bovaird, J. A., & Slegers, D. W. (2006). Methods for the analysis of change. In D. K. Mroczek & T. D. Little (Eds.), *Handbook of personality development* (pp. 181–211). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Mancini, T., Imperato, C., & Sibilla, F. (2019). Does avatar's character and emotional bond expose to gaming addiction? Two studies on virtual self-discrepancy, avatar identification and gaming addiction in massively multiplayer online role-playing game players. *Computers in Human Behavior, 92*, 297–305.

McCreery, M. P., Krach, S. K., Schrader, P. G., & Boone, R. (2012). Defining the virtual self: Personality, behavior and the psychology of embodiment. *Computers in Human Behavior, 28*(3), 976–983. https://doi.org/10.1016/j.chb.2011.12.019

McCreery, M. P., Laferriere, J., Bacos, C. A., & Krach, S. K. (2018). Video games as a model for assessing learning behaviors. In *American Educational Research Association Annual Conference*, New York, NY.

McCreery, M. P., Schrader, P. G., & Krach, S. K. (2011). Navigating massively multiplayer online games: Evaluating 21st century skills for learning within virtual environments. *Journal of Educational Computing Research, 44*(4), 473–493. https://doi.org/10.2190/EC.44.4.f

McCreery, M. P., Vallett, D., & Clark, C. (2015). Social interaction in a virtual environment: Examining sociospatial interactivity and social presence using behavioral analytics. *Computers in Human Behavior, 51*, 203–206. https://doi.org/10.1016/j.chb.2015.04.044

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282. Retrieved from https://hrcak.srce.hr/89395

Milne, D. (Ed.). (2015). *Evaluating mental health practice (psychology revivals): Methods and applications*. London, UK: Routledge.

Moore, M. G. (1993). Theory of transactional distance. In *Theoretical principles of distance education* (Vol. 1, pp. 22–38). London, UK: Routledge.

Schrader, P. G., Deniz, H., & Keilty, J. (2016). Breaking SPORE: Building instructional value in science education using a commercial, off-the shelf game. *Journal of Learning and Teaching in Digital Age, 1*(1), 32–46. Retrieved from http://joltida.org/index.php/joltida/article/view/6/54

Schrader, P. G., & McCreery, M. (2008). The acquisition of skill and expertise in massively multiplayer online games. *Educational Technology Research & Development, 56*(5–6), 557–574.

Schrader, P. G., & McCreery, M. (2012). Are all games the same? Examining three frameworks for assessing learning from, with, and in games. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in game-based learning: Foundations, innovations, and perspectives* (pp. 11–28). New York, NY: Springer.

Schrader, P. G., McCreery, M. P., & Vallett, D. (2017). Performance in situ. In Y. K. Baek (Ed.), *Game-based learning: Theory, strategies and performance outcomes* (pp. 421–438). Hauppauge, NY: Nova Science Publishers.

Shaffer, D. W., Squire, K. R., Halverson, R., & Gee, J. P. (2005). Video games and the future of learning. *Phi Delta Kappan, 87*(2), 104–111. https://doi.org/10.1177/003172170508700205

Shavelson, R., Baxter, G., & Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement, 30*(3), 215–232. Retrieved from http://www.jstor.org/stable/1435044

Shute, V. J., Ke, F., & Wang, L. (2017). Assessment and adaptation in games. In P. Wouters & H. van Oostendorp (Eds.), *Instructional techniques to facilitate learning and motivation of serious games*. Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-39298-1_4

Shute, V. J., Leighton, J. P., Jang, E. E., & Chu, M. W. (2016). Advances in the science of assessment. *Educational Assessment, 21*(1), 34–59.

Squire, K. D. (2011). *Video games and learning: Teaching and participatory culture in the digital age*. New York, NY: Teachers College Press.

Stefanidis, K., Psaltis, A., Apostolakis, K. C., Dimitropoulos, K., & Daras, P. (2019). Learning prosocial skills through multiadaptive games: A case study. *Journal of Computers in Education, 6*, 167–190.

Tapp, J., Wehby, J., & Ellis, D. (1995). A multiple option observation system for experimental studies: MOOSES. *Behavior Research Methods, Instruments, & Computers, 27*(1), 25–31.

Vallett, D. (2016). Soft failure in educational and learning games. In R. Lamb & D. McMahon (Eds.), *Educational and learning games: New research*. New York, NY: Nova Science.

Wang, L., Shute, V., & Moore, G. R. (2015). Lessons learned and best practices of stealth assessment. *International Journal of Gaming and Computer-Mediated Simulations, 7*(4), 66–87.

Whiteside, A. L., & Garrett Dikkers, A. (2012). Using the Social Presence Model to maximize interactions in online environments. In K. S. Amant & S. Kelsey (Eds.), *Computer-mediated communication across cultures: International interactions in online environments*. Hershey, PA: IGI Global.

Wouters, P., Van Nimwegen, C., Van Oostendorp, H., & Van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology, 105*(2), 249–265. https://doi.org/10.1037/a0031311

Yen, W. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement, 30*(3), 187–213. https://doi.org/10.1111/j.1745-3984.1993.tb00423.x