Those are not the Stories you are Looking For: Using Text Prototypes to Evaluate Game Narratives Early

Jason T. Bowey and Regan L. Mandryk
Department of Computer Science, University of Saskatchewan
Saskatoon, SK, Canada, S7N 5C9
{firstname.lastname}@usask.ca

ABSTRACT
In game design, evaluation is important to do early and often; however, evaluating game narratives early in development is an open problem. We don’t know how the evaluation of a game narrative will be affected when it is experienced outside of the context of the game’s mechanics and graphics. In this paper, we test the plausibility of using a text-based narrative prototype, by evaluating player experience and narrative experience in two studies using different game genres. In both studies, we compare the narrative evaluation of low- and high-fidelity graphics, but in study 1 (N=78), we kept interaction mechanics intact, and in study 2 (N=124), we removed game interaction in the text prototype. We observed no significant differences in player experience or narrative engagement in either study, indicating that text-based narrative prototypes could be an effective playtesting tool for game studios to integrate into their development cycle early.

Author Keywords
narrative; digital games; transportability; prototyping

ACM Classification Keywords
K.8.0 [Personal Computing]: General - Games.

INTRODUCTION
Game evaluation is an important component for a game’s eventual critical and commercial success, and as such, should be done early in the development lifecycle and often throughout [31]. Some aspects of games are evaluated easily early in development (e.g., basic mechanics and procedures); however, other aspects are more difficult to evaluate outside of the context of a complete prototype (e.g., accessibility, balance) [17]. Because of the benefits of early evaluation, methods have been proposed for play-testing prototypes of games under development, such as physical prototypes [17] or low-fidelity digital prototypes [17]. Furthermore, researchers have been exploring how the fidelity of the prototype affects the player’s experience of the game mechanics and have shown that the evaluation of core mechanics in low-fidelity prototypes of casual games is reliable [19]. Although games user research (GUR) tends to focus on methods that evaluate a game’s mechanics and dynamics [25] [26], there are other factors that contribute to play experience – such as narrative experience – that are less well understood or represented in GUR.

The evaluation of game narratives is an important aspect of playtesting because incoherent, inconsistent, or boring stories can result in poor critical and commercial reception of a game. In a series of blog posts, interviews, and talks [37,38,50], Bioware, a game studio praised for their successful merging of game and narrative, discuss how they employ narrative prototypes to evaluate the story early in a development process in which the writing team is well integrated with the development team. Bioware partially attribute their critical and commercial success of narrative-based games to this early evaluation and the integration of the different teams. In a counter-example, a blog post-mortem of the game Destiny [7], developed by Bungie [44], revealed that the game narrative was rebooted just months before the original release date, leading to delays, and a story that was criticized by reviewers for its vague plot, thin characters, and opaque dialogue. These examples lend support to the idea that narratives should be evaluated early in the development lifecycle of a game.

Questions around the process of early evaluation of game narratives are being asked by games user researchers in both academia and industry. At the 2016 GamesUR Summit [23], there was a panel on the evaluation of narratives that questioned whether or not they could be evaluated early and what effects there might be from evaluating narratives in isolation from the completed game. However, there is little guidance for game designers on the effects of evaluating narratives on prototypes; games user researchers do not currently understand the effects of evaluating an interactive narrative early in the design process – that is, outside of the context of a completed game. For example, will a game narrative be experienced differently if it is disconnected from the immersive and transportative aspects of completed games, such as the high-fidelity graphics? Will a game narrative be experienced differently if disconnected from the control scheme used to interact with the game’s mechanics?
In this paper, we ask how game narrative evaluation is affected when embedded in early game prototypes. Specifically, we systematically investigate how the game narrative evaluation is affected if it is disconnected from the high-fidelity graphics of a completed game (study 1) and the mechanics used to control interaction (study 2).

In a first study, we investigated the player experience differences in a narrative-based game with low-fidelity graphics (LoFi) and high-fidelity graphics (HiFi). All other aspects of the games (i.e., interaction and game control) were kept consistent between the two versions to avoid confounding factors that might cause differences in experience. The game used was similar to an interactive narrative, e.g., *The Stanley Parable* [18] We found no statistically-significant differences on any experiential measures and no interactions between a median split of how transportable a player tends to be and fidelity on any experiential measure.

In a second study, we used an adventure game that included more non-narrative interaction (i.e., fighting, running, and shooting). Furthermore, we did not mirror all of the controls and interaction as we did in study 1, but rather prototyped only a very minimal amount of game interaction to replicate a situation closer to what would be seen in industry when games are under development. We also added measures of narrative engagement and transportation. Again, we found no statistically significant differences on experiential measures or narrative engagement. There were also no significant interactions between player transportability and fidelity on any measure. We did, however, see a minor difference in the transportation of players, which showed that players were slightly more transported when playing the LoFi version rather than the HiFi version, consistent with ideas suggested in the literature for other types of narrative immersion, such as Avatar Identification [13].

Our results showed that there were no significant differences in player experience or narrative engagement as a result of playing a narrative prototype as opposed to the full game. These results hold regardless of whether or not the game interaction itself was prototyped and in two different game genres. There was a small positive effect on experienced transportation in the narrative prototype of study 2. Our results suggest that text prototypes of game narratives can be used early in the development of a game, increasing the likelihood of integrating narrative development with system development, and maximizing a game narrative’s chance for a successful reception.

**RELATED WORK**

We first present player and narrative experience, then discuss iterative development and prototyping in games.

**Player Experience**

Research in studying how players experience games is primarily rooted in psychology. In particular, a common model used as a basis for explaining player experience is needs satisfaction, which is part of the larger theory called Self-Determination Theory (SDT) [42]. SDT explains intrinsic motivation, i.e., doing something for the pure desire to do it, as opposed to being driven by the promise or threat of external rewards [42]. When engaging with an activity, such as a game, an individual’s intrinsic motivation to engage in the activity can be broken down into four dimensions: interest-enjoyment, effort-importance, perceived competence, and the amount of tension-pressure felt during the activity. [36] The building blocks that lead to the experience of intrinsic motivation are based on our basic psychological needs: competence, autonomy, and relatedness. In using needs satisfaction in the specific context of play, Ryan et al. added two additional constructs: intuitive controls and immersion-presence [42].

**Narrative Experience**

Researchers have recently begun extending constructs and ideas from media psychology into the realm of digital games [13,22]. For the purposes of this paper, we will distinguish between narrative engagement, experiences which pertain primarily to the game’s narrative [8] and narrative immersion, which depend more strongly on non-narrative game elements to create an immersive experience.

**Narrative Immersion**

Two types of narrative immersion have been used in the context of games: Identification [24] and Transportation [11]. Both types of immersion involve a similar cognitive process in players: a deep immersion into the narrative aspects of the game in which the player loses awareness of the outside world.

A state of immersion through Identification involves the creation of a connection between the player and a character in the game so that the player loses awareness of their outside self, experiencing the game from a different point of view [13]. There are three main ways a player can identify with a character [32]: *similarity identification* - identifying through perceived physical similarities with a character; *embodied presence* - seeing the narrative world through the perspective of a character; and *wishful identification* - when a player sees a character as a desirable version of themselves. Previous work has shown that aspects of player experience, e.g., enjoyment, are predicted by the amount of Identification [4,5] experienced.

Transportation is often described using a traveler metaphor, in which the player is so immersed in the narrative world of a game that they lose sense of their physical surroundings [22]. Transportation is facilitated through the interaction and presentation of the game (i.e., the depth, complexity, and attractiveness of the narrative world) but also through the personality trait *Transportability*, which has been shown to reliably predict transportation [12].

**Narrative Engagement**

The Narrative Engagement Scale was created by drawing items from a wide selection of narrative measures to create four new subscales measuring narrative-specific experienc-
es: narrative understanding, attentional focus, narrative presence, and emotional engagement [8].

**Iterative Development in Interfaces and Games**

We describe the iterative cycle (design, development, and evaluation) of interactive systems and games.

**Prototyping**

In interactive systems and interface design, it is common practice to start with LoFi prototypes and iteratively work up to HiFi versions [1,21]. This process allows designers to quickly explore a wide array of different design directions and possibilities, throwing out those ideas that didn’t work without losing a significant amount of time and effort. Various types of prototyping methods have been used, each with varying levels of fidelity: paper prototypes [14], digital mockups and wireframes [28], and interactive prototypes [16]. There is also research that directly compare prototypes of varying fidelity that have found no difference in the number of usability problems between prototyping fidelity [10,47,48], and that users tended to prefer working with a computer prototype compared to a paper prototype [45].

Game designers have used similar approaches to prototype their ideas because games can be seen as a specific type of interactive system. However, previous work has argued that there is more to test in a game than usability, such as playability, which requires game designers to use slightly different prototyping methods [20]. Fullerton describes two levels of fidelity in prototyping: physical paper prototypes and digital prototypes, as well as various tools to create them [17]. Game designers often prototype by starting with simple LoFi assets, such as capsules for characters and boxes for objects in the environment. Previous work has shown that playtesters find the same amount of usability issues in LoFi and HiFi graphic levels [29].

**Evaluation**

As mentioned, both interface and game designers follow an iterative design strategy. Part of this iterative process is the notion that the product starts at a LoFi stage and increases in fidelity through iteration, but the other essential part of this iterative cycle is the evaluation component that feeds into the next iteration of design.

For interfaces, evaluation may include formal usability inspections [27] heuristic evaluations [27] and cognitive walkthroughs [27] to name a few examples. For games, a different type of user testing can be required because of the increased interaction, which can lead to different kinds of bugs and design problems, often unforeseen by developers. Evaluation strategies have been created for games that have adapted strategies used in interface design – for example heuristic evaluation [40] and usability testing [26], whereas others are more unique to games, such as analytics [15] and playtesting [17]. More advanced playtesting approaches have used modelling techniques to evaluate experience using physiological data [33,34], biometric storyboards [39], and analysis of facial expressions [46]. Additionally, specific tools have been created or adapted from other areas to empirically measure various aspects of players and player experience in games [12,22,32,36,43].

The main idea behind iterative evaluation is for designers to test their current prototype with users to make sure they like it and are able to use it in the intended way. These goals differ for games and interfaces, but the common thread is that users test the prototype and designers use the feedback to make it better. This process is repeated throughout the entire design and development process with quick bursts of development and testing and updating the design to make the product better and easier to use [30].

**INDUSTRY APPROACHES TO GAME NARRATIVES**

Because industry practices are rarely published in academic venues, and development processes can be a closely guarded secret for game studios, there is not a large amount of publicly-available information regarding the actual processes that studios use to prototype, test, and integrate narrative structures into their games. To gain knowledge on how some prominent companies do this, we searched for blog posts written by developers and game designers, post-mortem analyses of games, and conference talks (e.g., Game Developer’s Conference – GDC) to try to piece together the kinds of processes and systems that are used to evaluate narratives.

Bioware, a studio well-known for their use of narrative structures in their games, seems to be the most forthcoming about their narrative design process, though there are only a few public resources that give an indication of what they do. Combing through the statements and interviews indicate three important development practices that demonstrate their commitment to producing excellent game narratives: 1) the close integration of their quality assurance (QA) team with the development team, 2) their use of narrative prototyping to iteratively test and develop game narrative, and 3) the merging of story design and system design.

Across several blog posts, team leads at Bioware have discussed several important QA practices that the studio has embraced over the years to improve their final products, in both narrative and design. In one such post [37], a QA lead talked about how the QA department plays an increasingly important role in the development cycle, moving away from reactionary to proactive involvement in the development process from the beginning. QA leads began attending designer and programmer meetings to be more informed about the game, and the QA team was eventually split into two separate disciplines – Tech, which supported the programmers, and Design, which supported designers and writers. In another post [38], the lead describes the role QA has at Bioware compared to other large studios and reiterates the integration of QA with development.

In 2007 at The Australian Game Developers Conference, a lead writer for Bioware, Mike Laidlaw, gave a presentation describing the early part of the production process for a
Laidlaw further describes the writing process for their game *Mass Effect* [2] from early-stage prototyping to late-stage integration. In the early-stage prototyping, the writing team designed the dialogue system and tested how players would interact with it, the tone and shape of the game was formed, and the ESRB rating was decided so the writers knew what type of content to create. In the next phase, the game story was written by the writing team. The written content was then iteratively tested using a LoFi narrative prototype, to test for technical issues with the story content. The end result of the narrative prototyping stage was a playable level with full dialogue in text format. This allowed for major rewrites, because the content was text, which was cheap to rewrite as needed. After this initial prototyping phase, the text was passed to QA, who looked for plot holes, bugs, and inconsistencies. From there, the text moved to voiceovers, cinematics, and integration into the game.

**Integrated Development of Story and System**

A talk at GDC 2015 [50] discussed Bioware’s process of melding story with game systems in the production of *Dragon Age: Inquisition* [3]. The design team split in half to explore different design techniques. The first team created narrative content and the second team designed standalone systems. What they found was that the narrative team’s content scaled poorly, and the system team’s content had a disconnected impact. The main conclusion was that story and technical systems developed separately cause problems and dissonances in the final product. The solution was to develop story and systems together, in tandem.

Bioware’s *Destiny* [7] is an example of a game that was received poorly by players due to the separation of narrative and system teams. The details of what led to the game’s lukewarm reception after the large pre-release hype are fuzzy, but the general idea is documented in a post-mortem of the game [44]. After three years of development, and only months before the game was initially supposed to be released, the lead writer put together the ‘supercut’ for the game – a video that included the game’s cinematics and major story points, and showed it to the studio’s leadership. The senior staff at Bungie were not satisfied with the story and scrapped it, deciding to start from scratch. The result was a hacked together story, criticized by reviewers for its vague plot, thin characters, and opaque dialogue. Furthermore, there were system elements that could not be scrapped as too much development had gone into them and it was too close to release, so they did not hang together well with the final version and seemed to be disconnected.

Laidlaw noted that the writers are part of the design process from day one, as a part of the design team. He also describes three phases of development the team undergoes: prototyping – exploring ideas, setting, and story work; pre-production – systems, pipelines, and story arcs are defined; and production – “a flood of content”.

Fans blamed *Destiny’s* [7] poor reception on its lackluster story, which can likely be attributed to the separation of the development of story and game systems, the same problem Bioware found with *Dragon Age: Inquisition* [3]. That the leadership had not seen the story and cinematics prior to their completion, combined with a mindset that they could develop and implement a full story within the restrictions of existing game systems as a final touch, rather than a fundamental part of the game, resulted in a poorly constructed and inconsistent narrative experience.

The important part of this is that game studios with successful narrative games, such as Bioware, are successful because of how the story is integrated into their process. They develop the story early on, in parallel with the other game systems. The story and game need to fit well together without creating inconsistencies between them and in order to achieve that, iterative testing of these two systems together early and throughout development is important. Many studios, like Bungie, developed these two branches separately, which led to miscommunication between departments and an inconsistent, dissonant narrative experience, even though the game mechanics were innovative and well designed.

**CONSTRUCTING A NARRATIVE PROTOTYPE**

In this paper, we conducted two studies in which we created a LoFi narrative prototype based off of a finished HiFi game. In this section we will document the process we used to construct our narrative prototypes.

The first step was to write the overall story, conveying through text the information that would normally be understood from the visuals. The goal was to try to get the same information across as would be in the final game.

The second step was to simulate player actions and game mechanics using a simple text interface. Player actions refer to simple things like “Walk over to the computer”. Game mechanics refer to interactions such as combat, in which the player has a wider range of options and the interaction is more complex. For these types of game mechanics, we presented options to the player, but in a very simplified version, (e.g., "Attack").

The final step is to implement the narrative. We used a simple tool [41] that we customized to convert a game script (scene descriptions and character dialogue) into an interactive system. Our tool allowed us to display the text in a game-like environment – we displayed the character portraits alongside their dialogue in study 2, but not in study 1 – and use mouse and keyboard controls to move through the prototype. However, a different medium-fidelity tool (e.g., PowerPoint) could just as easily have been used to achieve the same functionality.

**STUDY 1**

Using Bioware’s narrative prototypes as inspiration, we wanted to systematically measure how effective such a technique is for narrative games, and the feasibility of using it as an iterative development tool. That is, we wanted to determine how the player experience of the narrative could be evaluated early in the development lifecycle. In particu-
lar, we wanted to know whether a text-based version of a game would result in reduced player experience, when compared to the same game with high-fidelity graphics.

To investigate this, we used the above-mentioned process to create a text-based version of an existing research game and compared the LoFi text version with the HiFi graphical version. We performed a between-subjects experiment and collected measures of experience, in-game behavioural motivation, and demographic surveys to assess player traits.

**Figure 1: Screenshots for LoFi (left) and HiFi(right) versions**

**Game Design**

We created *Space Adventure*, a narrative game involving an accident on a space ship; participants play the game from the perspective of the ship’s Captain and explore different rooms through a surveillance system to discover the events that occurred on board. The game was implemented in C# using the Unity Game Engine.

*Space Adventure* is played in a full 3D game environment similar to immersive narrative-based commercial games (Figure 1). Players started by designing their customized avatar: they were given the choice to select sex, hairstyle, hair colour, shirt, shirt colour, pants, and pants colour. The player moved around the inside of a spaceship, experiencing the game’s narrative and optionally discovering the backstory that led to the events in the game.

**Narrative/Procedure**

In the game, the player awakens after the ship’s engines had crashed. One of the ship’s soldiers instructs the player about the controls for moving around the environment, before guiding the player to investigate the cause of the failure.

First, players are introduced to the *Surveillance Room* where they can access various cameras around the ship to view footage from different rooms of the ship, shown to the player through a text-based description. They then teleport to the *Hacking Room*, which is necessary in the game to unlock the cameras one at a time for viewing in the *Surveillance Room*. The game primarily consists of players teleporting back and forth between the *Hacking Room* and the *Surveillance Room* throughout the narrative, unlocking a camera in the *Hacking Room*, then viewing a piece of the story by looking through the camera in the *Surveillance Room*, then teleporting back to hack the next camera. This is done eight times throughout the narrative, each time revealing a different piece of the story. In addition, while in the *Hacking Room*, the player can unlock more of the game’s backstory by optionally hacking a console each time they return to the room. Pressing the right arrow key reveals more information about the backstory one word at a time, which is revealed while the player-character is waiting to gain access to the next camera in the surveillance room. Once a console is hacked, the player teleports back to the surveillance room and the cycle continues.

Using cameras as a narrative technique allowed us to guide participants through a narrative from a local point of action, reducing the need for navigation through the virtual world. After hacking the last console, participants could either escape the ship right away by pressing a button, or alternately run down the hallway to escape the ship.

**Controls, Interface, and Versions**

Participants controlled the avatar using ‘W’ to move forward, ‘A’ to rotate the avatar to the left, and ‘D’ to rotate the avatar to the right. A 3rd-person camera was positioned in a fixed location directly behind the avatar.

We included a heads-up-display (HUD) on which players could hack and teleport by clicking on the *hacking and teleport* buttons, using the left mouse button when the options were available (see Figure 1). A mini-map was provided in the bottom right corner with highlighting of the current room so the player could see where the fictional camera was looking (see Figure 1).

We used two versions of the game: the HiFi version as described, and a text-based LoFi version (See Figure 1).

**Participants and Deployment Platform**

We recruited participants by telling them that they would be evaluating a game under development. We had 81 participants (48% female) with an average age of 22 (SD = 15.9) participate in our study through Amazon Mechanical Turk (MTurk), a platform that acts as a broker between parties offering a range of Human Intelligence Tasks (HITs) and paid workers. Participants received $6 USD as compensation and the study took on average 40 minutes (15 minutes of gameplay). Although it has been shown that MTurk is a reliable research tool [35], we measured the time spent per questionnaire to evaluate task performance and ensured that participants were attentive despite the online setting. As a compliance check, we removed participants who spent less than 1.5 seconds per item on two or more questionnaires, as these participants likely did not carefully read and consider the questions. Three participants were detected and removed using this filter, leaving a total of 78 participants to be used in the final analysis (N<sub>low-fidelity</sub>=40, N<sub>high-fidelity</sub>=38).

**Ethical Compliance**

Ethical approval was obtained from the University of Saskatchewan behavioural ethics board; participants were asked to give informed consent at the beginning and debriefed on the experiment at the end. To comply with ethical guidelines, the HIT was only made available to workers in the USA who were older than 18. Also, only workers with an approval rate above 90% were offered the HIT as a means of quality control. After the experiment, we asked participants to respond to simple questions about the story to ensure that the experiment was understood.
Table 1: Study 1 results – Means and SD for Low-Fidelity and High-Fidelity conditions, as well as F, p, and β values from a Multivariate Analysis of Variance (MANOVA) for Condition, Transportability, and the interaction.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Low-Fidelity Mean (SD)</th>
<th>High-Fidelity Mean (SD)</th>
<th>Low-Trans Mean (SD)</th>
<th>High-Fidelity Mean (SD)</th>
<th>F</th>
<th>p</th>
<th>Transportability Mean (SD)</th>
<th>F</th>
<th>p</th>
<th>Condition * Transportability Mean (SD)</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enjoyment</td>
<td>4.88 (3.35)</td>
<td>5.10 (3.35)</td>
<td>4.00 (1.65)</td>
<td>5.28 (1.08)</td>
<td>2.82</td>
<td>0.10</td>
<td>0.60</td>
<td>19.26</td>
<td>0.00</td>
<td>0.01</td>
<td>1.59</td>
<td>0.21</td>
</tr>
<tr>
<td>Effort</td>
<td>5.44 (1.77)</td>
<td>5.53 (1.27)</td>
<td>4.99 (1.33)</td>
<td>5.97 (0.93)</td>
<td>0.23</td>
<td>0.64</td>
<td>0.92</td>
<td>13.36</td>
<td>0.00</td>
<td>0.04</td>
<td>0.04</td>
<td>0.84</td>
</tr>
<tr>
<td>Tension</td>
<td>3.14 (1.52)</td>
<td>2.81 (1.21)</td>
<td>2.79 (1.33)</td>
<td>3.29 (1.42)</td>
<td>0.95</td>
<td>0.33</td>
<td>0.83</td>
<td>3.97</td>
<td>0.05</td>
<td>0.47</td>
<td>0.95</td>
<td>0.33</td>
</tr>
<tr>
<td>Competence</td>
<td>5.09 (1.19)</td>
<td>4.90 (1.41)</td>
<td>4.67 (1.38)</td>
<td>5.37 (1.29)</td>
<td>0.31</td>
<td>0.57</td>
<td>0.91</td>
<td>3.72</td>
<td>0.01</td>
<td>0.25</td>
<td>0.50</td>
<td>0.48</td>
</tr>
<tr>
<td>Autonomy</td>
<td>3.93 (1.44)</td>
<td>3.40 (1.46)</td>
<td>3.28 (1.47)</td>
<td>4.20 (1.33)</td>
<td>2.55</td>
<td>0.12</td>
<td>0.63</td>
<td>11.87</td>
<td>0.00</td>
<td>0.06</td>
<td>2.01</td>
<td>0.16</td>
</tr>
<tr>
<td>Relatedness</td>
<td>4.07 (1.40)</td>
<td>3.93 (1.34)</td>
<td>3.79 (1.28)</td>
<td>4.27 (1.40)</td>
<td>0.14</td>
<td>0.71</td>
<td>0.93</td>
<td>3.04</td>
<td>0.09</td>
<td>0.57</td>
<td>0.51</td>
<td>0.48</td>
</tr>
<tr>
<td>Presence</td>
<td>4.48 (1.39)</td>
<td>3.99 (1.66)</td>
<td>3.50 (1.36)</td>
<td>5.05 (1.23)</td>
<td>2.08</td>
<td>0.15</td>
<td>0.68</td>
<td>28.85</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.96</td>
</tr>
<tr>
<td>Intuitive Controls</td>
<td>5.74 (1.09)</td>
<td>5.61 (0.94)</td>
<td>5.32 (1.03)</td>
<td>6.00 (0.88)</td>
<td>0.21</td>
<td>0.64</td>
<td>0.92</td>
<td>8.33</td>
<td>0.01</td>
<td>0.16</td>
<td>0.10</td>
<td>0.75</td>
</tr>
<tr>
<td>Similarity ID</td>
<td>3.31 (1.56)</td>
<td>3.73 (1.56)</td>
<td>2.85 (1.30)</td>
<td>4.21 (1.52)</td>
<td>2.43</td>
<td>0.12</td>
<td>0.65</td>
<td>19.61</td>
<td>0.00</td>
<td>0.01</td>
<td>0.21</td>
<td>0.65</td>
</tr>
<tr>
<td>Embodied Presence</td>
<td>4.02 (1.67)</td>
<td>3.47 (1.77)</td>
<td>2.85 (1.42)</td>
<td>4.69 (1.52)</td>
<td>1.88</td>
<td>0.14</td>
<td>0.72</td>
<td>30.33</td>
<td>0.00</td>
<td>0.00</td>
<td>0.03</td>
<td>0.85</td>
</tr>
<tr>
<td>Wishful ID</td>
<td>3.07 (1.50)</td>
<td>3.12 (1.40)</td>
<td>2.58 (1.24)</td>
<td>3.64 (1.45)</td>
<td>0.09</td>
<td>0.76</td>
<td>0.94</td>
<td>32.18</td>
<td>0.00</td>
<td>0.06</td>
<td>0.57</td>
<td>0.45</td>
</tr>
<tr>
<td>Motivation</td>
<td>0.95 (0.13)</td>
<td>0.94 (0.18)</td>
<td>0.91 (2.21)</td>
<td>0.97 (0.16)</td>
<td>0.10</td>
<td>0.75</td>
<td>0.94</td>
<td>2.48</td>
<td>0.12</td>
<td>0.64</td>
<td>0.25</td>
<td>0.62</td>
</tr>
</tbody>
</table>

**Measures**

In this section we present the measures used for study 1. Unless otherwise specified, all items were presented to participants using a 7-point Likert scale.

**Needs Satisfaction** (i.e., competence, autonomy, relatedness, presence, intuitive controls) was measured using the Player Experience of Needs Satisfaction Scale (PENS). Participants rated their agreement to a series of statements, e.g., “I feel competent at the game”. (Cronbach’s alpha = .75, .85, .65, .94, .712).

**Enjoyment, Effort, and Tension** were measured using the Intrinsic Motivation Inventory (IMI), e.g., “I enjoyed the game very much.” (Cronbach’s alpha = .94, .85, .85).

**Identification** was measured using the Player Identification Scale (PIS). Participants reflected on their experience and answered several questions on how they identified with the character they played. The PIS consists of three subscales: similarity identification, embodied presence, and wishful identification. (Cronbach’s alpha = .95, .96, .91).

**Transportability** is a personality trait that describes an individual’s susceptibility to become transported into a narrative world. The transportability scale [12] was presented prior to playing the game to measure this trait. Participants rated their tendency to become transported on a series of items, e.g., “I have vivid images of the events in the story” (Cronbach’s alpha = .93). High- and low- transportability were calculated using a median split.

**Motivation** was measured using the in-game behavioural measurements from the hacking terminals. At each terminal, the player could reveal the backstory, but could leave the task at any time with no negative consequences to the game. We calculated the motivation score by combining the number of words revealed in each terminal and normalizing the score to an overall average percentage. The purpose of this measure was to get an idea of how much effort participants put into the task, since as crowd workers they could primarily be motivated by the money rather than the task.

**Results**

The purpose of this study was to investigate differences in player experience of narrative-based games with high or low graphic fidelity. Based on previous research [19] that investigated the experiential effects of HiFi and LoFi casual games, we expected there to be no main effects by condition on any measure. Also, because the game was primarily an interactive narrative, we expected there to be significant differences on all dependent measures between participants who rated high and low in transportability. Based on this, we came up with the following hypotheses:

**H1:** There will be no differences in any game experience measures due to graphical fidelity.

**H2:** There will be significant differences in experiential measures between high and low transportable participants.

We conducted an a priori power analysis using [6] to ensure that we recruited enough participants to find medium to large effect sizes, if they existed. We performed the test expecting an effect size of 0.33, a significance level of 0.05, a β of 0.8, and two different groups in our between-subjects experiment design. We decided that medium-large effect sizes were sufficient, because we only need to show that any differences between dependent measures are quite small, rather than statistically identical. The power analysis suggested a sample size of 37 participants per group, which we rounded up to 40. For all of our analyses, we use p < .05 as a threshold for accepting statistical differences, and β > .80 as a threshold for accepting statistical similarities.

**Differences in Experiential Measures**

We conducted a multivariate analysis of variance (MANOVA) in SPSS 24 using all of the dependent measures and using both condition and a median split of transportability as between-subject factors. We used the application g*Power [6] to calculate the post-hoc power for each of the analyses performed and derived the β values. We present the means and standard deviations for both factors, as well as the F, p, and β values for both factors and the interactions in Table 1.

**Effects of Fidelity.** There were no significant differences between conditions for any experiential measure, and the β values were sufficiently high enough to determine that there were no differences between competence, intuitive controls, relatedness, effort, tension, wishful identification, and be-
havioural motivation. Autonomy, presence, enjoyment, similarity identification, and embodied presence did not have a high enough post-hoc statistical power to say there was no difference, but their non-significant differences were still very small (about 0.5 points on a seven-point Likert scale). These results support H1, with the exception of the measures that didn’t have enough statistical power to determine that there was no difference.

Effects of Transportability. There were significant differences between players with high and low transportability for all dependent measures, except for relatedness (p=.09) and motivation (p=.12). With the exception of these two non-significant differences, our results support H2.

Additionally, there were no significant interactions between fidelity condition and transportability, showing that the lack of differences between the LoFi and HiFi versions did not depend on the transportability of the players.

Study 1 Discussion
Our results suggest that for a short narrative game, player experience does not change very much between a LoFi prototype and a HiFi version with the same story and interactions. The narrative immersion measures of identification were also not statistically different.

Because the goals of this study were to compare play experience between different levels of graphic fidelity, our measurements were primarily focused on measures of play experience, such as need satisfaction and game enjoyment. We decided to include avatar identification as a measure of narrative immersion, but we did not specifically evaluate the players’ experience of the narrative itself.

We also decided in this study to keep the controls consistent between the two versions – the game interaction was made as similar as possible by minimizing the necessary navigation and exploration in the HiFi condition and making the interaction in the hacking mini-game identical between conditions. As a result, the HiFi game was essentially an interactive narrative rather than a complete polished game. While this means the results are not generalizable to games in general, the game we created was fairly representative of a particular genre of narrative-based games, which are often focused on following a story and being present in a 3D world (e.g., walking simulators, such as The Stanley Parable [18]). Furthermore, we know that the lack of game interaction or control did not confound our results.

While the results from study 1 indicate that graphic fidelity may not be as important to certain kinds of games, the short play time is not enough to make a general claim. What we can say, however, is that in a short amount of time playing this type of game, graphic fidelity had very little to do with the overall play experience and narrative immersion. This indicates that this type of LoFi text interface could be useful for game developers when designing and testing game narratives early in development, before graphics and mechanics are fully implemented.

In our second study, we addressed the limitations from study 1 to help generalize to commercial narrative games.

STUDY 2
In study 1 we showed that measures of experience remain similar when comparing LoFi and HiFi graphics. We made several changes in study 2. First, we included measures of narrative engagement and removed PENS [43] to reduce the study complexity and shift the focus to narrative experience. We kept the IMI as a means of replicating the results from study 1 and understanding effects on experience. Second, we added simple, controlled game mechanics to generalize our results beyond an interactive narrative. Third, we made the LoFi version with simplified game controls to test if the fidelity of game interaction affects experience.

Game Design
We created a simple, 2.5D fantasy game with 5 main levels: A prologue, three levels, each with a single unique game mechanic, and an epilogue. The prologue showed a Princess running through a forest, chased by the shadow of a dragon, accompanied by text setting up the story. The first level used sword-based combat, the second level was an infinite runner, and level 3 used a shooting mechanic. The epilogue was a conversation to complete the narrative arc. The game was built initially to study sexism in games [5] however, a neutral version was used in the present study.

Narrative/Procedure
The underlying narrative of the game followed a damsel-in-distress trope. The player took on the role of a knight character on a quest to rescue a Princess captured by an evil wizard. In the prologue, the Princess’s kidnapping is shown, and the premise of the story is described to the player through text scrolling across the screen. In level 1, the Knight is investigating the forest where the Princess was last seen when he runs into a woman trapped in a cage in the middle of a bandit camp. The knight fights the bandits and rescues the prisoner from her prison. The woman, named Victoria, thanks him, but runs off into the forest, prompting the knight to follow her to learn what happened to the Princess. In level 2, the Knight runs along a forest path, trying to catch up to Victoria to find out where the Princess was taken. Catching up to her, she promises to lead him to the wizard’s castle if he rescues her village from a bandit raid. In level 3, the knight sits outside the village walls with a crossbow and must shoot bandits, while avoiding the villagers, as they run out through the main gate one at a time. Once the Knight has killed enough bandits, Victoria leads him to the wizard’s castle, as promised. In the epilogue, the Knight and Victoria arrive at the castle only to find it abandoned, with Victoria claiming that “The Princess is another castle”, before deciding to permanently join the Knight in his quest to find the Princess.

Low-Fidelity Version
The LoFi version of the game follows the same story with identical dialogue as the HiFi version (Figure 2). As place-
holders for gameplay mechanics, this version used simple text-based interactions to simulate the actions the player undertook in the HiFi version. For example, the knight fighting off the bandits was replaced with player text options such as “Perform sword slash” and the infinite runner level had options such as “Dodge to the left” or “Jump over the fallen tree” to provide players with high-level actions to replace the actions taken in the HiFi condition (see Video Figure). Unlike in study 1, the game controls in the LoFi version were not intended to be comparable to the HiFi version and were more representative of what might be possible to prototype very early in the development of a game, when the details of the mechanics are still being developed.

Participants, Deployment, and Ethics
We used identical deployment methods as with study 1, except we chose to recruit 125 participants (46.8% female), one of which was removed using the outlier filter from study 1 (N_{low-fidelity}=65, N_{high-fidelity}=59). The study took an average of 19 minutes to complete (10 minutes of gameplay), and participants were compensated $5 USD.

Measures
Player Experience was measured as in study 1 using the Enjoyment, Effort, Tension, and Perceived Competence subscales from the Intrinsic Motivation Inventory. Because our goal with this study was to further investigate narrative experience, we chose to leave out PENS [43].

Transportability and Identification were measured using the same scales as in study 1 [22,32] (Cronbach’s alpha = .94, .94, .96, .91 for ID_{sim}, ID_{emb}, ID_{wish} and transportability).

In addition, we added the following measures in study 2.

Transportation was measured using the Transportation Scale [22] which measures an individual’s transportation into a narrative world during their experience with a particular narrative (e.g., “I wanted to learn how the narrative ended”). Transportation is indicative of the quality of a narrative that is experienced (Cronbach’s alpha = .83).

Narrative Engagement was measured using the Narrative Engagement Scale (NES, [8]), which measures four aspects of narrative engagement: narrative understanding - reverse coded items to measure the understanding of the narrative, e.g., “I had a hard time recognizing the thread of the story”, attentional focus - reverse coded items asking participants about how well they were able to mentally focus on the story, e.g., “I had a hard time keeping my mind on the game", narrative presence - how strongly they felt present in the narrative, e.g., “At times during the game, the story world was closer to me than the real world”, and emotional engagement - how emotionally connected they felt to the events of the story, e.g., “The story affected me emotionally”.

To maintain consistency across the four subscales, results from narrative understanding and attentional focus are presented reversed, so for all four a higher value refers to stronger narrative engagement (Cronbach’s alpha = .79, .89, .75, .82).

In the results and discussion below, we distinguish between player experience, narrative engagement, and narrative immersion as discussed in the related work section.

Results
We conducted an identical analysis as was done with study 1. We present the means and standard deviations for both independent factors, as well as the F, p, and β values from the MANOVA by condition, a median split of transportability, and the interaction between the two in Table 2.

Differences in Experiential Measures

Play Experience. In terms of play experience, there was enough statistical power to assert that there were no signifi-

![Figure 2: LoFi (top-left), HiFi Level 1 (top-right), Level 2 (bottom-left), and Level 3 (bottom-right) Screenshots](image-url)
cant differences between conditions for enjoyment, effort, tension, and competence.

**Narrative Engagement and Immersion.** There was enough statistical power to assert that there were no significant differences between narrative understanding, attentional focus, and embodied presence. Narrative presence and emotional engagement were also not statistically different and had a relatively high power but didn’t meet the threshold of 0.8 (ß=0.71 and ß=0.72, respectively) to accept the null hypothesis.

There was, however, a significant difference between the LoFi and HiFi versions in terms of transportation (p=.04, η²_p=.034), similarity identification (p=.05, η²_p=.03), and wishful identification (p=.02, η²_p=.04), although the actual difference between means were small (less than 0.5 on a seven-point Likert scale). An independent-samples t-test showed no significant differences in the transportability between the groups (t_{23}=0.18, p=.859), this shows the differences cannot be explained by random assignment to the game versions. The results showed that experienced transportation was slightly higher in the LoFi version than the HiFi version.

**Effects of Transportability.** Comparing high and low transportability revealed similar results as in study 1 – there were significant differences between highly-transportable and low-transportable players for all dependent measures, with only experienced tension failing to reach significance. Additionally, there were no significant interactions between condition and transportability, showing that the lack of differences between the LoFi and HiFi versions did not depend on the transportability of the players.

**Study 2 Discussion**
The main interpretation of results is very similar to study 1. For all of the measures of play experience there were no significant differences between conditions, indicating that LoFi narrative prototypes can be used to assess the game narrative in early stage development. Furthermore, there were no significant differences in experienced narrative engagement, showing that the quality of the narrative could be assessed outside of the context of the full game experience (i.e., HiFi graphics and game interactions).

Interestingly, the two measures of immersion, transportation and identification, were two of the exceptions to this pattern, though the differences were very small in terms of the big picture of overall experience. This could be an indicator that for story immersion, while the story itself is important, there are other factors beyond the story that go into it, such as the presentation of the narrative world and the interaction with and presentation of different characters.

**DISCUSSION**
In this section, we will discuss the implications of our results, industry use cases and how game studios could integrate narrative prototypes into their existing design processes. We will also discuss the limitations of our studies and future opportunities.

**Implications of Results**
While we did see similarities between many measures of player experience and narrative immersion, these similarities were often not significantly similar, and in some cases were significantly different.

From a practical point of view, the most useful result from a narrative prototype is the narrative engagement. In Study 2, we showed that level of fidelity had no effect on narrative understanding and attentional focus, and narrative presence and emotional engagement were very similar between the conditions. This shows that the addition of visual graphics and game mechanics had very little effect on how well participants understood the narrative, the level of focus they placed on the story, their presence in the narrative, and the emotional engagement they felt from the story. These four measures suggest that early testing of a story in a text-based format would be useful for game designers to test for these particular narrative properties early in the story’s conception, since they appear to be fairly independent from the game mechanics and visuals.

**Industry Use Cases**
There are two different situations in which the ideas underlying our proposed narrative prototypes could be used and applied to game studios. The most obvious, and the main motivation for this paper, is a tool to be integrated into the iterative design process that most game studios now use.

A second application for narrative prototyping is to cheaply evaluate game ideas before committing to a project. While more study is needed to be able to confidently say that our technique would apply across games of different types and genres, our results show that, at least for a certain type of narrative game, we can get similar levels of narrative experience from a simple, text-based narrative compared to a complete experience. This would allow game studios to create fast and cheap prototypes for several story-based games in a short amount of time and quickly test them with users to get a feel for which game, or games, would be the best to invest further resources into.

**How to Integrate Prototypes into the Design Process**
Here, we discuss how a narrative prototype could be embedded in the iterative design process of games, based on the current use of LoFi prototypes, and using our understanding of industry practices. In our approach, we adhere to the following three guidelines for effectively making use of a narrative prototype: close integration with the current state of the game, iterative testing and feedback, and increasing fidelity over time.

A lesson Bioware learned from working on *Dragon Age: Inquisition* [3], and the lesson Bungie learned with *Destiny* [7], is the need for story to be developed in tandem with the game itself. If the narrative team doesn’t consider how players will interact with the game, and if the game design-
ers don’t consider the game narrative while developing their content, they may introduce dissonances between the narrative and the game. The goal, then, is to match the narrative prototype with the game’s current level of development. This is similar to the goal of interactive system prototypes in which the fidelity of the sketch or prototype should match the maturity of the idea [9].

Early on in the development cycle, the narrative prototype would be similar to the text-based prototypes used in our studies. The goal would be to test and evaluate the story, looking for narrative dissonances, plot holes, and travel through branching and fold-back story structures. As the story is tested and refined, the game would be developed in parallel, constructing LoFi versions of the systems and interactions. As the game gets fleshed out, the narrative prototype should slowly add in the assets and mechanics as they become available (for example, how we added the static character image in study two). This stage can also be used to test multiple options with users (e.g., different avatars or character personalities). As a game level gets constructed, the descriptions in the narrative prototype could be replaced by the actual level, placing the prototype inside the level, with the tester now focusing more on the dialogue, and ensuring the setting makes sense with that version of the story. As mechanics get developed, e.g., a sword attack, these can also be added to the prototype, replacing the early-stage text-based interactions. The goal with iteratively developing the story in this way is to always be aware of how the story fits with the game and vice versa. By always testing the latest version of the story, game developers have a better chance of finding dissonances between the two, as with Destiny [7], as early as possible, instead of waiting until the game is nearing completion.

Limitations and Future Opportunities
In both studies, the two conditions differed in graphical fidelity and interaction. While we tried to minimize differences in terms of game interaction for study 1, there were still slight differences due to the closely-coupled nature of graphics and interaction. We argue that this prioritizes external validity, as both conditions represent a snapshot in the development timeline of a game. In study 2, we iterated on this closely-coupled aspect, further differentiating the interaction between conditions. The design choice of representing the mechanics as simple text options is representative of how such a prototype would be created early in a game’s development, with no access to graphics or systems.

A second limitation of our work is that both games we used in the studies were already developed and the narrative prototypes were developed after the games, rather than before as would happen in a real-world use case.

A third limitation is that our games – although enjoyed by participants according to the means – were representative of what a small, indie studio would create as a demo and not what a triple-A studio would generate. Because the goal is for the fidelity of the narrative prototype to match the maturity of the game’s development, we don’t see this as a potential source of error; however, higher-fidelity narrative prototypes may yield different play experiences than their game-based counterparts of commercial quality.

Finally, participants were only exposed to the game for a short time (ten to fifteen minutes). Our results show no significant difference in play experience between the LoFi and HiFi versions, but we speculate that with a significantly longer play session, differences may emerge, due to a lack of visual stimulation in the LoFi versions over a long period of time. Short sessions are ideal for playtesting, however, which fits the use case of our prototyping method.

Future Work
We would like to expand on the ideas in this paper and more deeply consider narrative prototyping. In both of our studies, we used games we had created. While this gave us an advantage in having control over every aspect of their design, the experiences were short and were not comparable to commercial game in terms of content and quality. The next step would be to replicate study 2 with an off-the-shelf narrative game, such as The Stanley Parable [18]. This would increase the generalizability of our results, showing that narrative prototypes can be effective on commercial games with complex content and interactions, built by professional writers. Second, we would like to document the processes we describe in the paper by iteratively developing a game using narrative prototyping, to test the feasibility of the method in practice. Finally, the narrative prototypes we used are focused on early-stage narrative evaluation and, to some extent, character dialogue. We would like to expand on this idea and look at creating different types of narrative prototypes, focusing on specific aspects of story development, such as world building, designing character personalities, and evaluating procedurally-generated narratives.

CONCLUSIONS
Iterative development and evaluation is important for all types of interactive systems, including interfaces and games. While iterative methods have become common practice for most game designers, the integration of game narratives early in development is still new and unfamiliar to many developers. We proposed using text-based narrative prototypes as a means to iteratively test and develop game narratives, allowing writers to work closely and in parallel with the designers of the game systems. Through two studies, we compared high and low levels of graphic fidelity and player interaction with game mechanics and show that all experiential measures are not affected by playing the narrative prototype. Our findings provide evidence that narrative prototyping could feasibly be used by game studios to quickly and cheaply evaluate game ideas before committing development resources to them and to more closely integrate their narrative design into existing iterative development processes.

ACKNOWLEDGEMENTS
We thank NSERC and the UofS for funding.
REFERENCES


25. Robin Hunicke, Marc LeBlanc, and Robert Zubek.


and Ergonomics Society Annual Meeting, 399–403.


50. 2015. GDC: See how BioWare melded story with systems to make Dragon Age.