The keys to success: Supplemental measures of player expertise in Massively Multiplayer Online Games

ABSTRACT
In this article we describe an investigation of player expertise deployed as part of a mixed-methods longitudinal, multi-site study that examined whether and how players’ offline characteristics are recognizable in their online interactions in Massively Multiplayer Online Games (MMOGs). After detailing our methodology and analytical toolkit, we narrow our focus to a case study that examines three players with previous experience in First-Person Shooter (FPS) games playing Rift (Trion Worlds 2011) (a fantasy-themed MMOG) for the very first time. This case study illuminates how interpretation of data can be inadvertently influenced by the researcher’s choice of technologies and methods employed in their study design. In particular, we demonstrate that initial research assessments of a player’s level of skill may be inaccurate and how the use of multiple data sources acts as a means for triangulating observations and analyses providing a richer – yet more complicated – view of player expertise.

KEYWORDS
Massively Multiplayer Online Games
Rift
keystroke logging
expertise
assessment
study design
INTRODUCTION

The persistent online worlds of Massively Multiplayer Online Games (MMOGs) have been of interest to researchers studying topics ranging from economic decision-making (Castronova 2005; Castronova et al. 2009) to observing how human actors spread communicable (virtual) diseases (Balicer 2007; Boman and Johansson 2007; Kafai et al. 2010). The methods used in these studies vary as widely as their sample sizes, from participant observations of small (yet tightly focused) groups of players (Chen 2012; Eklund and Johansson 2010; Nardi and Harris 2006) to accessing game servers that contain data about an entire game population (Feng et al. 2007; Shen et al. 2011).

‘Usage logs’ have been extremely fruitful for understanding how users interact with the affordances of a digital game. While the specific inputs and sample sizes vary depending on the type of logging being conducted (e.g. keylogging, eye tracking, in-game metrics), this automated capture of data provides a minimally intrusive means of collecting a detailed account of a user’s interaction with game software. The subsequent analysis of log data has led to a better understanding of how norms develop within an in-game community, such as the use of in-game metrics collected in the World of Warcraft (Blizzard Entertainment 2004) Armoury to explore patterns among active players in regards to avatar sex and preferred in-game activities (Yee et al. 2011) and the armour and weapons they equip (Lewis and Wardrip-Fruin 2010). While in-game metrics allow for an expedient collection of data on a game’s population (Williams 2010), other forms of logging, such as eye tracking data, rely on much smaller sample sizes. Despite drawing on smaller numbers, advances in eye tracking software have allowed researchers to compile detailed ‘gaze logs’ to record a user’s visual interaction with the interface of a virtual environment (Istance et al. 2009; Nacke et al. 2011), which can in turn be used to increase accessibility to those who would like to play digital games, but are unable to do so with the traditional inputs of a game controller, keyboard and/or mouse. Logs provide a powerful tool for the study of digital gameplay; however, we argue that without the context provided by other sources of data, logged data – especially in-game metrics – may only provide a partial picture of the events being captured.

In this article we describe how automatic keylogging software was used to supplement other forms of data collection as part of a mixed methods longitudinal, multi-site study that examined whether and how players’ offline characteristics are recognizable in their online interactions within the gameworld of an MMOG. As the academic field of game studies matures, it becomes essential that findings be replicable across multiple sites and studies. In an effort to reduce the ‘reinventing of the wheel’ occurring in empirical studies of MMOGs that Warmelink and Siitonen (2011) caution against, in this article we detail our study design to invite other researchers interested in similar questions about what learning to play a new MMOG might ‘look’ like to repeat the study in a different context. The first half of this article begins with an explicit description of our methods and resulting dataset. Moving on, we present a tightly focused case study that combines gameplay logs with observational and survey data to demonstrate our analytical techniques. A case study of three players (The FPS guys) was selected from our larger dataset as it exemplifies how first impressions of research participants expertise can
be inaccurate. Indeed, this trio represents the most dramatic shift from initial researcher observation to the final assessment of player expertise in the dataset that took into account all of the data sources collected.

As our data collection techniques will likely be of interest to other games researchers, we describe the combination of more traditional qualitative data sources (field notes, participant interviews, observations of play, open ended survey responses) with an automated log that transcribed each mouse movement and key press of a participant’s first time playing Rift (Trion Worlds 2011), a commercially produced fantasy-themed MMOG.¹ By drawing on these very different data sources throughout our analysis, oversights and mischaracterizations of our initial expertise assessments of this particular trio of players are made clear. The methods and results, we argue, form an important case study to illuminate how interpretation of data is effected by the technologies and methods employed, and that the use of multiple data sources can provide a more robust and accurate account when reporting on research findings. While this kind of triangulation of data is common in qualitative research, it has, so far as we can tell, been underutilized in the study of digital games.

**KEYLOGGER SOFTWARE**

‘Keyloggers’ are a type of software used to capture keystrokes, and are often associated with illicit or clandestine surveillance of a computer user. Indeed, there is extensive literature about how best to detect unwanted Trojan software and implications for network security (Howard and Hu 2012; Laubscher et al. 2005; Sreenivas and Anitha 2011). However, the deliberate installation of such software to collect keystroke logs for research purposes can act as an interesting and insightful means of data collection. For example, such software has proven useful to better understand writing and editing processes, as it allows for repeated, exact replays of an emerging text exactly as it is written long after the research session has ended and the participant has left the lab (Leijten and Van Waes 2013; Lindgren and Sullivan 2002).

In the study described herein, participant key presses and mouse movements were recorded for the duration of their time playing the commercial MMOG Rift using Recording User Input (RUI), a free program created by the Applied Cognitive Science Lab at Pennsylvania State University (Kukreja et al. 2006; Morgan et al. 2013) and available for download at http://acs.ist.psu.edu/projects/RUI/ (accessed 29 December 2014). RUI was installed on lab computers and was used to log the keystrokes and mouse movements of participants as they played Rift in the lab-based portion of the study, described below. Like the studies of writing practices, this keylogger data can be used to map a player’s interactions with the MMOG software, for example creating ‘heat maps’ to visualize where they most frequently click on the game’s interface, or to determine whether they tend to use mouse clicks or hot keys (typically the numbers on a keyboard, mapped to shortcuts for offensive attacks) to engage in combat. This, we argue, generates supplementary information to researcher field notes and participant self-assessment to allow for a more nuanced understanding of how players encounter a new game, and the rate at which their proficiency playing increases over time. However, as the case study that follows our description of study will show, remaining overly reliant on one source of data can skew results.

---

¹ Rift (Trion Worlds 2011) is a fantasy themed MMOG set in the fictional world of Telara. Much like other MMOGs drawing on the same genres such as World of Warcraft (Blizzard Entertainment 2004), players can complete quests or defeat monsters to gain experience and increase their avatar’s level and strength. After reaching the level cap players can participate in a variety of activities common in MMOGs such as end game raiding or player vs player battles.
DESCRIPTION OF STUDY

The methods described in this article were used to collect data for an ‘on the ground’ examination of the complex relationships between MMOGs and the localized practices of players in different cities, and in different contexts of play. This study ran from October 2009 to July 2012 and was a longitudinal, multi-site study that asks whether and how players’ offline characteristics are recognizable in their online interactions in MMOGs. A total of 1376 individuals participated in this study, and members of our research team met each participant face-to-face. This gave us ‘ground truth’ on our participants, as we were able to confirm their relative age, sex, and ethnicity, which other studies that rely simply on self-report and/or server scrapes are unable to confirm with the same level of certainty. The sites of data collection varied, ranging from in-school studies of minors, to adults who visited one of two university research labs (Jenson et al. 2013). Data about public MMOG play were also collected at local area network (LAN) parties (events where participants often bring their own computers and/or consoles for the purpose of playing multiplayer games in a co-situated setting) and other gaming-related events (Taylor et al. 2014).

Theoretical framework

The theoretical framework we found most useful for shaping the design of this study is Actor Network Theory (ANT) (cf. McArthur and Jenson’s article in this issue), as developed and articulated in the field of Science and Technology Studies by Bruno Latour (1987, 1992, 2005), John Law (2004) and Michel Callon (1987). ANT reads ‘the social’ – institutions, practices and interactions – as the product of engagement and associations between human and non-human ‘actors’. As Latour insists in *Reassembling the Social* (2005), agency is distributed across actors that co-constitute a network. Non-humans, including artefacts, technologies and other animals, are not simply the ‘backdrop’ to social interactions; in this view, human action is shaped and made possible by non-human agents, to which have been delegated the tasks of prescribing, framing, eliciting and enabling certain forms of action, and prohibiting or discouraging others. The work of researchers, from an ANT perspective, is therefore to ‘re-assemble’ the associations between human and non-human actors in a network so as to better map the relationships between and among them. Working from this framework, then, we were able to map the relationships between the keylogging data, and each of the other data points we had available, and literally ‘reassemble’ an account of players’ experiences playing *Rift*.

Data collection

In this section we describe the types of data collected at two North American university labs. The data collection tools discussed here were used for a subset of the larger study, focusing on research participants who played *Rift* (*N* =154), a fantasy-themed MMOG in the same genre as other popular games such as *World of Warcraft*. At the time of the study, *Rift* was newly released to market and provided an opportunity to observe players of varying skill levels interact with this new game for their very first time. Each participant completed the same interview and survey, created a new *Rift* avatar, and played through (at minimum) the first 45 minutes of gameplay. This standardized approach
to data collection allowed for the unique opportunity to observe a player’s first interaction with an unfamiliar game, and provided a context in which we could test hypotheses such as if participants with experience in similar games (e.g. *World of Warcraft*) were able to acclimatize to *Rift* faster than those who had experience in other genres. What follows is a detailed description our recruitment and data collection techniques, and we hope this accounting will be of use to other researchers interested in conducting similar lab-based research with MMOG players.

**Recruitment:** Participants in the lab-based component of this study were primarily recruited through posters placed around two university campuses. Initial contact was through e-mail, where a scheduling assistant asked a series of questions to determine basic demographic information and to gauge a potential participant’s experience with MMOGs. The study was open to all experience levels. This open approach to recruiting resulted in a diverse sample set, ranging first time MMOG players to those with years of gameplay including experience with MMOGs, other genres of digital games, and/or console games. Participants were encouraged to recruit at least one other friend or acquaintance and attend the session as a group. All participants were compensated for their time.

**Intake interview:** After describing the project, data collection tools, and answering any questions related to the informed consent documents, participants were asked a short series of questions to roughly gauge their history of MMOG and gameplay more broadly. This interview lasted approximately ten to fifteen minutes, was recorded (and later transcribed), and consisted of a standard set of questions asked of all who took part in the *Rift* study.

**Survey:** All study participants completed a standard survey used across all of our studies (not just the *Rift* study) that included detailed questions about demographics, personality measures, and history of gameplay. This survey was a combination of 91 multiple choice and open-ended questions that on average took 25 minutes to complete.

In addition to standard demographic questions, the survey contained a series of questions to assess experience with digital games broadly, and with MMOGs specifically. Questions about expertise ranged from direct queries (e.g. asking participants to self-assess their level of expertise in particular gameworlds), to determining ludic knowledge, for example by asking if they use game mods and if so, which mods in particular they use. We also asked participants to describe with whom they played (e.g. a sibling, romantic partner, online players who they had not met face to face, etc.) and where they most frequently played MMOGs (e.g. at home, a friend’s house, an Internet cafe, and so on). The survey concluded with a series of questions about personality traits and social values, which were used to add further context to their behaviours and interactions observed as part of their gameplay session.

**Gameplay:** Participants visited the lab in groups of two to four to play *Rift* (a commercial fantasy-themed MMOG) and *Guardian Academy* (SRI International 2011) (a private server game created specifically for this study that emulates other fantasy MMOGs) as a group on our lab computers. For this specific study, all participants played using iMacs running Parallels, a commercially available Windows operating system emulation program. Rather than using the default hardware that ships with the iMacs used in this study, each computer station was outfitted with a full sized keyboard and two-button mouse.
While RUI allows for the logging of mouse cursor movements along the x- and y-axis, the mouse data collected in this particular case study had too much noise and was not easily parsed. Therefore, directional mouse movement data was not used in the analysis described in this article, and is instead limited to a binary metric of whether the mouse was moving or if it was stationary.

2. While RUI allows for the logging of mouse cursor movements along the x- and y-axis, the mouse data collected in this particular case study had too much noise and was not easily parsed. Therefore, directional mouse movement data was not used in the analysis described in this article, and is instead limited to a binary metric of whether the mouse was moving or if it was stationary.

more typically used by PC game players. All of the computers were located in the same room, close enough that participants could communicate using a normal speaking voice. Each session was observed by at least one member of a team of trained research assistants, but wherever possible two researchers were present. Both researchers were responsible for taking field notes, and they would take turns troubleshooting any technical difficulties encountered by the participants.

While participants were given the option to play Rift on their personal account, no one identified as an active Rift player and instead everyone played on one of four lab-owned accounts. The gameplay session began with asking participants to create a new avatar, spending as much (or as little) time as they wanted customizing their class and appearance. When creating a new avatar, Rift requires players to pick between two opposing forces (‘Guardian’ and ‘Defiant’). In-game communication and collaboration is not possible across the two factions. Before beginning avatar creation, we asked participants to negotiate which faction they would all play; video and audio of this negotiation was recorded. Other than requesting everyone in the session to play on the same server and as the same faction to allow in-game text communication, researchers attempted to avoid influencing any player decisions on choosing avatar class, race or sex. To ensure that participants did not feel rushed and were able to look through all the character options in Rift, researchers left the room during avatar customization (see McArthur and Jenson this issue). Participant avatars were not deleted at the end of the session. If they volunteered to return for a follow-up play session, these avatars were made available, allowing participants to ‘pick up where they left off’ and continue advancing through the Rift beginner and intermediate zones. At the end of their time playing Rift (usually 45 minutes, but often longer if they requested further time to complete certain objectives in the game), participants were given a short break and then asked to spend 30-45 minutes playing Guardian Academy. The avatar customization was not nearly as robust in Guardian Academy (players had a choice between male or female avatar, and to be a caster or melee class) and so researchers did not leave the room during avatar creation for the second half of this play session.

Video and audio of play in Rift and Guardian Academy were recorded using ScreenFlow, an inexpensive program that allows simultaneous recording of a webcam, screen capture, and recording of both in-game and external audio inputs. The iMac’s built-in webcam was used to record a participant’s facial expressions, and a microphone headset was used to capture all verbal utterances during the session. The screen capture feature of ScreenFlow was used to record avatar customization and the entire duration of their play in both games (see McArthur and Jenson [this issue] for analysis of this customization). In addition to recording verbal chat, in-game text-based communication was also captured, using the Rift built-in combat and chat log.

Keylogger: Rift is a commercial game, therefore keylogging (to capture in-game movement, as opposed to the chat log mentioned above) had to be done on the local machine, using the aforementioned RUI software program. RUI’s logging was turned on at the start of a gameplay session by a researcher, and remained running in the background for the entire duration of a participant’s play in Rift. While no participants played on a personal Rift account, privacy protocols were in place to ensure no usernames and/or passwords
were captured in our logs. As Guardian Academy was created specifically for this study, in-game chat, key presses and mouse movements were automatically captured and logged by the game server.  

Field notes and initial expertise assessment: Research assistants kept detailed field notes while they observed the play session using a standardized template. In addition to chronological notes, they were asked to specifically comment on any interactions they felt were relevant to assessments of expertise, leadership ability, gender, race and other specific prompts. Finally, all researchers involved in the session were asked to write a reflective note about their specific interventions that may have influenced the behaviours and interactions observed.

After participants left the lab, all researchers observing the session would partake in a debriefing to compare field notes and begin initial analyses including an initial assessment of their answers to the intake interview, as well as participants’ observed skill and proficiency in the two games. Using a checklist adopted from the expertise assessment sliders developed by Taylor et al. (2011) multiple markers were used to determine where a participant fell on the range between novice and expert. A summary of the assessment markers is included in Table 1. Here, we move beyond the use of number of hours spent playing a particular game being the primary marker of expertise a metric (see e.g. Reeves et al.’s [2009] study of Counter-Strike [Valve Corporation 1999] player expertise). Instead, our assessment took into account measures of investment, discourse, game knowledge, and skill. Just like the standardized template for field notes, this standardized expertise assessment schema allows for cross comparison across sessions.

<table>
<thead>
<tr>
<th>Type of expertise</th>
<th>Description</th>
<th>Sample metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Investment</td>
<td>Time and/or resources committed to gameplay</td>
<td>Number of hours spent playing a particular MMOG; Number of in-game friends met offline</td>
</tr>
<tr>
<td>Discourse</td>
<td>Mastery of the social language associated with a gameworld</td>
<td>Use of MMOG-related slang; ability to effectively share information with other players using in-game communication systems</td>
</tr>
<tr>
<td>Game knowledge</td>
<td>Ludic and/or narrative understanding of the internal mechanics of a gameworld</td>
<td>Ludic knowledge: demonstrated knowledge of the different abilities of classes and/or the expectations associated with each class (i.e. tanks, healers, damage)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Narrative knowledge: demonstrated knowledge of game-specific lore and backstory</td>
</tr>
<tr>
<td>Skill</td>
<td>Operational proficiency in a gameworld</td>
<td>Use of hotkeys; Writing customized macros; Multitasking (e.g. concurrently working on several quest objectives)</td>
</tr>
</tbody>
</table>

Table 1: Further details about Taylor et al.’s (2011) expertise evaluation schema. This multimodal assessment was used as part of the researcher debriefing at the end of each session and was used to determine an initial evaluation of participant expertise.
Optional follow-up: At the end of their session, participants were encouraged to schedule another play session in the lab to once again play Rift. With no need to repeat the intake interview, the entire duration of the second session could be spent playing Rift. Many participants did return for follow-up sessions, allowing us to collect hours of audio-visual and keylogger data that spanned in some cases three separate visits to the lab. For novices or those with a little to no MMOG experience, this multi-session data allowed us to begin modelling what learning to play a new game looks like, and how expertise does or does not develop over time.

ANALYSIS

The Rift study described in this article is just one small segment of our larger dataset; members of our research team interacted with 154 participants. Without a doubt, the sheer volume of qualitative data collected presents a challenge. Our strategy has been to break the larger dataset down into manageable chunks to investigate particular hypotheses, such as if players with no prior experience playing fantasy-themed MMOGs choose different avatar classes in Rift than those familiar with this genre of game (Bergstrom et al. 2012) or how participants presented their own play preferences in opposition to the stereotypical archetype of MMOG players (Bergstrom et al. 2014). The entire dataset has also been used to reflect on the theoretical and methodological challenges of studying MMOGs (de Castell et al. 2014). For the remainder of this article, our focus will centre on analyses of participant expertise, which serves as an extension of Taylor et al.’s (2011) critical investigation of MMOG player expertise, which was a model created through an analysis of player data collected during an earlier phase of this largest project. What follows is an overview of the types of analysis we utilized to understand our data, and this information is detailed in order to better contextualize the case study that follows. The case study presented in the subsequent section is shared as a particularly instructive example of how much setting and co-located participants can (unintentionally) influence the inferences made from observational data.

Descriptive statistics: A summary of the survey’s demographic breakdown of lab study participants is included in Table 2. Participants in the lab component of the study skewed heavily towards the young adult (18–24 years old) age bracket, an artefact of recruiting primarily on university campuses. However, this overrepresentation of young adults does not carry across the entire sample. For example, participants in a LAN study skewed heavily towards the adult (over 24 years old) age bracket, as these participants were recruited at an event that markets itself to an older demographic of players (Taylor et al. 2014). This decision to recruit participants from multiple venues was intentional to ensure a diversity of voices represented in the study’s survey data, however, we recognize that despite our recruitment efforts, the demographics of the Rift-specific study unintentionally reflects the stereotype that youth are the primary players of digital games (Bergstrom et al. 2014; Kowert et al. 2012).

Thematic coding: Transcriptions of interviews were thematically coded and used to further contextualize survey responses and observation of a participant’s gameplay. Thematic coding was also utilized to summarize open-ended survey responses. Any reference to gameplay history and/or experiences with MMOGs was noted and we also paid particular attention to the context in
which these games were mentioned, looking for evidence of ludic knowledge about popular MMOG conventions, and expertise in MMOG discourse, for example the degree to which they used slang and/or MMOG-specific terms (e.g. describing their avatar as their ‘toon’, using ‘epic loot’ to describe valuable in-game items, referring to fellow in-game friends as ‘guildies’, etc.) to describe their gameplay history.

Event-based coding: The multiple sources of audio-visual data recorded through ScreenFlow were synced and rendered into a single file. This file was then loaded into an event logging software package, Noldus The Observer XT. Research assistants were trained to ensure intercoder reliability and provided with a standardized coding schema to capture multiple markers of in-game proficiency such as:

- Successful or unsuccessful interaction with a Non-Player Character (NPC)
- Starting a quest, time to completion of quest objectives and time until they returned to the NPC they must visit to turn in the quest
- Number of avatar deaths

Researchers also annotated the video to indicate when a participant was engaged in avatar customization, instances of text-based chat with other participants and/or other players in the gameworld, and the duration of combat against hostile NPCs. These annotations were used in conjunction with the keylogger visualizations described below.

<table>
<thead>
<tr>
<th>Value</th>
<th>(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>No answer</td>
<td>3</td>
</tr>
<tr>
<td>Youth (&lt;18 years)</td>
<td>11</td>
</tr>
<tr>
<td>Young Adult (18–24 years)</td>
<td>56</td>
</tr>
<tr>
<td>Adult (&gt;24 years)</td>
<td>30</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>31</td>
</tr>
<tr>
<td>Male</td>
<td>69</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>37</td>
</tr>
<tr>
<td>Caucasian</td>
<td>41</td>
</tr>
<tr>
<td>Indian</td>
<td>12</td>
</tr>
<tr>
<td>Other</td>
<td>10</td>
</tr>
<tr>
<td>Education</td>
<td></td>
</tr>
<tr>
<td>No answer</td>
<td>1</td>
</tr>
<tr>
<td>Primary/Elementary</td>
<td>1</td>
</tr>
<tr>
<td>Secondary/High School</td>
<td>40</td>
</tr>
<tr>
<td>Trade school, technical college, or community college</td>
<td>11</td>
</tr>
<tr>
<td>Undergraduate</td>
<td>33</td>
</tr>
<tr>
<td>Graduate school or professional school</td>
<td>13</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
</tr>
<tr>
<td>Native Language</td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>53</td>
</tr>
</tbody>
</table>

Table 2: Summary statistics for demographics of lab-based participants.
Keylogger visualizations: RUI outputs data into a text file in a format that is not easily parsed by the human eye and requires some manipulation before any meaningful analysis can occur. The raw log-file is comprised of the time-ordered sequence of mouse/key presses. All of the log files were cleaned and parsed using a Python script, and the MatPlotLib plotting library was used to visualize the data. To begin, histograms were made to easily determine the distribution of keys pressed during a play session. Here, researchers looked for an emphasis on WADS (keys typically used for avatar movement) and/or ‘hot key’ usage (number keys mapped with shortcuts to activate spells and physical attacks). Using Taylor et al.’s (2011) assessment criteria, frequent use of one or both of these key combinations are read as an indicator of skilled play.

A useful visualization involved a categorical binning of the key-presses, plotted as a function of the relative time since the beginning of data capture (see Figure 1). With the keylogger data plotted over time, the results of the event-based coding could be overlaid to segment the visualization into different activities (see Figure 2). These coloured overlays allow for easy demarcation between various activities, such as avatar customization and creation (seen here in red), in-game chatting (green), and combat against hostile monsters (blue).
A CASE (STUDY) OF MISTAKEN IDENTITY

In the previous sections, we detailed the multiple modes of data collection and the types of analyses conducted as part of this study. To demonstrate the (dis)advantages of such a robust and varied data collection protocol, in this section we narrow our focus to a case study of three participants who played co-situated in the lab in two sessions. We describe the results from each of the data sources provided above, and then detail specific conclusions about player MMOG expertise derived from each source. This case study is meant to serve as an example of the ways in which inferences from observational data can be inadvertently skewed by initial observations. The goal of this section is to illustrate the analytical strength that arises from approaching studies of games using multiple data sources, describing how we worked through conflicting information to result in our final conclusions.

Introducing the case

The case study described here details a group of three male friends who visited the lab on two separate occasions. The name given to this case study, ‘the FPS guys’, comes from this group of player’s self-described extensive experience playing FPS games. A fourth friend responding to our posters on a
university campus recruited the trio. The friend who made initial contact with the scheduling assistant ultimately did not participate in either play session. The other three visited the lab for two extended sessions, resulting in over three hours of Rift and 45 minutes of Guardian Academy play footage.

All three participants were East Asian males in their mid 20s and students at a local technical college. None had played Rift prior to visiting the university lab. A summary of survey responses about their gameplay history is included in Table 3. All three were avid FPS players, primarily playing Counter-Strike with their mutual friends and coworkers. In their self-assessments of general gameplay competency (‘how good are you at games?’) asked as part of their intake interviews, two indicated that they were ‘good’, and the third described himself as ‘seven out of ten’. According to observational data this trio seems to present as an example of novice MMOG play, however, this case study also provides unexpected revelations to the study of expertise. To reiterate our earlier statement, this case study was selected to illustrate how initial observational data led the researchers to believe that this was a trio of rank novices, however, upon adding other sources of data to this analysis, a much more complex view of player expertise begins to emerge.

**Evidence of novice MMOG play**

The preliminary assessments of these three participants indicated that the researchers firmly believed that they were all rank novice MMOG players. This conclusion was drawn from the following evidence:

*Recruitment:* In the initial contact with the scheduling assistant, the group was described by the fourth participant (who, as stated above, did not attend either session) as experienced MMOG players. However, their intake interviews presented a different story. Their session began with a request to play EVE Online (CCP Games 2003) – another MMOG – notorious for its difficulty and punishing learning curve (Paul 2011, 2012). They were, however, not current EVE players. In fact, they had never set (virtual) foot inside this particular game world. When researchers asked why they wanted to play EVE, the trio stated they were confident they were prepared, as they had spent some time the previous evening watching YouTube videos about the game. Ultimately the trio agreed to play Rift, but this interaction set the tone for the remainder of their first play session.

*Intake interviews:* Further evidence of their novice MMOG play was drawn from participants’ intake interviews. For example, when asked if he played MMOGs, Participant B responded yes, he played Counter-Strike (a game usually considered a FPS, not an MMOG). Field notes from the research assistant who conducted the interview indicate that Participant B was quite
confident that he considered Counter-Strike an MMOG and so she proceeded to ask a series of follow up questions about his play in this particular game. This probing resulted in further information about Participant B’s Counter-Strike play, determining that he primarily plays offline on an Xbox, co-situated with a small group of friends.

Researcher notes indicate that similar confusion arose once again while Participant B was completing his survey as he was having difficulty answering some of the questions about his other MMOG of choice – online poker – a game that just like Counter-Strike is not typically considered a MMOG. This confusion surrounding what constitutes a ‘Massively Multiplayer Online Game’ was taken as evidence that this participant was unfamiliar with the genre, as Counter-Strike played on a shared television screen via an Xbox or online poker are generally not considered examples of MMOGs. These two interactions resulted in Participant B’s scoring very low on the research assessment of his knowledge about MMOG-specific discourse (Ensslin 2011).

Observations of gameplay: Through researcher field notes we see a very clear description of three players struggling to play Rift even after two extended sessions of playing this MMOG. Despite playing Rift for over an hour on their first visit to the lab and not much time passing before their second visit, researchers had to give Participant B and Participant C an extensive refresher about the basic mechanics of the game (e.g. movement keys, right clicking to interact with NPCs, etc.).

After about fifteen minutes of play in their second visit to the lab, a researcher came to Participant B’s aid once again to answer his question about why he was repeatedly dying at the hands of a low level hostile NPC. According to researcher field notes, she watched Participant B engage combat with the NPC and observed that after hitting it once with his weapon, he repeatedly pressed the z key. This key was not bound to any abilities or attacks, so his avatar stood there motionless while the NPC slowly chipped away at his health pool. She writes that she suggested he try again, this time using the number keys (to which his offensive spells were currently bound) and his number of avatar deaths seems to drop off significantly.

Interactions such as this, or the refresher described above, were taken into consideration when assigning these participants a relatively low score in the MMOG skill assessment metric. Based on the observation of gameplay over two sessions, the researchers were quite confident that all three players were novices, and it was likely that this was their very first time encountering a MMOG. However, as described above, researcher observations and field notes make up two segments of the data collected as part of this research. In the following section we describe the addition of other sources of data to this expertise assessment, describing how we worked through conflicting data sources to arise at our final assessments. Following this, we describe the impact this study had on our understanding of what learning to play a MMOG entails, and speculate what impact this might have on the larger study of MMOG expertise.

Further analysis: How does one climb a virtual tree?

The initial assessment and field notes pegged this trio as firmly belonging as part of the novice category. However, when additional sources of data were consulted to provide further evidence of their novice status, rather than the proof of their (lack of) expertise becoming clearer, a muddier picture began to develop:
Survey: Due to the structure of the survey software utilized as part of this study, results had to be downloaded in batches and were not usually made available to researchers until after the session had ended. In this case, the survey responses for the ‘FPS guys’ were not collected until after the conclusion of their second session. Assessment of their gameplay experience was done through information gathered during their intake interview and any information disclosed as part of their play sessions. This means researchers did not know of the information in Table 3. Therefore, initial assessments did not include Participant A’s previous experience playing World of Warcraft or Second Life (Linden Lab 2003), as he did not disclose this in his intake interview. Researcher field notes indicate that they were observing the session with the assumption that none of the three had any prior MMOG experience and rated their expertise accordingly.

Event-based Coding: When research assistants were asked to code each session individually, it became easier to bracket off each player’s actions from their co-situated friends. In this case study, the revisiting of participant game footage for the purpose of event-based coding actually afforded the opportunity to revisit observational field notes. For example, when his play footage was viewed individually, Participant A did not seem to have nearly as much difficulty with the game’s interfaces as researcher field notes indicated his fellow participants were having. Indeed, Participant A can be seen to successfully engage in combat against hostile NPCs and complete quests. Through the use of the consistent coding schema for all participants who played Rift, Participant A’s in-game activities were annotated and these annotations are, in fact, not consistent with the annotations of other first time MMOG players who visited the lab. Instead, this annotated coding is more in line with what would be expected from an advanced novice to a low intermediate MMOG player.

Participant B struggled to find his bearings within the unfamiliar affordances of a MMOG and his frequent requests for help were documented throughout researcher field notes. Event logging indicates multiple deaths and no successful completion of quests. And yet, when watching his first and second play session unfold on the recording, we observed Participant B attempting to scale walls and climb trees. Recall above that Participant B had to call the researcher over to troubleshoot why he kept dying to the low level mobs that populate the Guardian starting zone. What is missing from her initial field notes, but observable when watching the video documentation, are the events leading up to his asking for assistance. The audio recording reveals that he is talking to himself, trying to find a vantage point where he can attack the hostile NPCs but they cannot to reach him. He reasoned that his constant deaths were because he wasn’t able to fire enough shots on his ranged weapons before the mob was within melee range, and therefore the solution would be to move out of harm’s way. This is not typical behaviour in a MMOG, but when this is viewed in conjunction with his experience playing other games, perhaps Participant B was attempting to play Rift as if it were a FPS. This sort of testing of the in-game environment to find elements present in other game genres is not something that would otherwise be captured in our coding schema, as the schema had been designed to assess based on MMOG-specific criteria.

It would seem then that Participant B’s initial assessment of being a novice was at least in part a result of researcher expectations of what MMOG-experience would look like. A similar oversight is found upon re-watching
Participant C’s footage for event-based coding. Participant C quickly stopped engaging in combat in favour of exploring the in-game environment. This was read as ‘wandering’ in the field notes, carrying with it an assumption that Participant C was having difficulty finding his way back to the beginner’s zone of Rift. However, when revisited as part of the event-based coding, this ‘wandering’ seems to be more akin to exploring an unfamiliar gameworld. Much like the example described above where Participant B repeatedly tried to find a sniper-like vantage point, Participant C’s exploration would not be fully accounted for through event-based coding, nor would it be acknowledged through a study of only his keylogger data (described below).

Keylogger data: According to the RUI visualizations, Participant A’s logged data looks more like advanced novice to intermediate play. Participant A made frequent use of the WADS keys to move, and some use of hotkeys to engage in combat. This was an unexpected result, as based on the researcher observations, Participant A was originally thought to be a rank novice MMOG player like his friends. However, this can be accounted for when further context is added by combining the visualized log with the information obtained by revisiting his gameplay footage for the event-based coding, and his survey responses where he revealed prior MMOG (World of Warcraft) and virtual world (Second Life) play. When viewed in isolation from his co-situated participants, Participant A’s footage indicates that he didn’t actually seem to struggle all that much with Rift. Given the wider context, it is likely that Participant A’s initial expertise was underestimated by virtue of his proximity to Participants B and C.

The other interesting keylogger result comes from Participant C. In this case, his keyboard usage drops off completely quite early in the play session. According to this log alone, this may seem like a disengagement with the game and may be read as non-activity. However, as described above in the event-based coding of Participant C, re-watching the footage indicates that what from the researcher’s vantage point looked like disengagement with the game and null activity in RUI, Participant C was still engaged in Rift. By combining his keylogger visualizations with the new observational notes arising from the event-based coding, we see evidence that Participant C was continually exploring the gameworld, using mouse clicks to move his avatar forward. It is only through the combination of multiple data sources that we were able to piece together the full account of Participant C’s time in Rift.

Summary: When the initial play session with the ‘FPS guys’ was scheduled, we expected a group of experienced MMOG players to arrive at the research lab. This case study was selected because it highlights how researchers had to reconsider their assessment of these players’ expertise at each stage of the research process – what was originally read as ‘MMOG novice’ had to be continually re-evaluated each time a new data source had to be accounted for as part of the analysis process, and this was the singular case where we turned out to have an inaccurate initial assessment. By detailing this particular trio’s visits to our lab and the subsequent analysis of their resulting data, we argue, shows the importance of reflection as part of the research process, but also the ways in which a single data source can lead researchers astray when attempting to analyse a user’s expertise in a MMOG. When viewed in isolation, Participant C’s absence of keystrokes may have been discarded as research tool malfunction and/or null activity, but when placed in conversation with footage of his gameplay, we see a player attempting to explore the in-game environment of a MMOG. Similarly, when researcher assessment
of novice play based on observations is in conflict with keylogger data that demonstrates behaviours typical of players with prior MMOG experience, an overlooked survey response where the participant mentions prior MMOG play becomes the missing link. By describing, in detail, our methodological and analytical tools, our goal here is to provide a framework for further mixed methods investigations of the sociotechnical assemblages of digital games. Furthermore, by describing our missteps as we worked through the analysis of multiple sources of data, we demonstrate the importance of reflection as argued by de Castell et al. (2014), especially the oversights that can arise if researchers do not leave room for surprise in their study design.

**RETHINKING THE STUDY OF MMOG EXPERTISE**

In this article we have discussed the difficulty that arises when attempting to assess player expertise by a single metric alone. Other studies of MMOG players that either specifically or tangentially deal with expertise assessment are frequently authored by researchers approaching gameplay from an educational context and are firmly grounded in cognitive theory (Ang et al. 2007; Chen 2012; Steinkuehler 2004, 2008), but other approaches include an assessment of the number of hours played as indicative of level of mastery of a specific gameworld (Huffaker et al. 2009). In part arising from the findings of this case study analysis, our subsequent work investigating player expertise is informed by critiques of such approaches, namely through the work of Jonas Linderoth (2009, 2011, 2012). Linderoth calls for a shift away from cognitive theory, arguing that ecological psychology is a more appropriate theoretical lens for understanding where and how learning could happen within a game.

According to ecological psychology, and building on Linderoth’s preliminary accounts, learning happens by becoming attuned to the world around us and cultivating our perception in order to identify tools that allow for the accomplishment of particular goals (2011: 5). He argues that a successful performance in a game might actually be the result of identifying the tools provided by the gameworld, and that often those performances are so heavily scaffolded that failure is nearly impossible. Linderoth (2012) cautions that any skill or expertise is highly specific to a particular gameworld, and not necessarily as easily transferrable to other digital environments. Here we reflect on Participant B’s play, especially his repeated attempts to find a wall or tree that he could scale and remain out of reach of the NPCs he was attempting to attack. This sort of mechanic is not typically found in a MMOG, but recall that Participant B’s gameplay history was primarily in FPS style games such as *Counter-Strike*. We argue that Participant B was attempting to apply his knowledge of in-game tools and affordances – heavily steeped in FPS conventions – to the unfamiliar gameworld of *Rift*. However, with little to no shared conventions between these two genres, Participant B’s FPS experience impeded his ability to adapt to the in-game environment of a MMOG. While Linderoth’s theorizing is still in its early stages, case studies such as this add empirical weight to emerging critiques of the dominant paradigm framing discussions about expertise and learning within digital games.

Our goal here was not to present a set of findings that are generalizable to a larger population (e.g. how FPS players encounter a MMOG for the first time). Instead, this case study is a reminder that research on games and
gamers is not always as straightforward as what we might think, nor is it then necessarily generalizable to other players and/or other games. For instance, this case study is not meant to, nor should it have to, speak to all players of any game. Instead, we hope it has served as an example of the necessity of triangulation in research methods, as well as ensuring that we have data collection methods that make triangulation possible. The mixed methods approach we have carefully articulated here is meant to serve as one example of the importance of using multiple data sources in studies, as well as pursuing lines of enquiry that offer alternative perspectives, readings and interpretations of data that might not be so clean and clear as it is presented. This lab-based study contributes to literature that is attempting to understand whether and how players learn to play and the ‘FPS guys’ certainly challenged the ways in which we were understanding gamer ‘expertise’.

ACKNOWLEDGEMENTS

This work was funded by the Air Force Research Laboratory. The authors would also like to thank the journal’s editorial team and anonymous reviewers for their helpful comments about this manuscript. We would also like to thank the anonymous reviewers from the 2014 International Communication Association who provided feedback on an earlier draft of this manuscript.

REFERENCES


—— (1992), ‘Where are the missing masses? The sociology of a few mundane artifacts’, in W. E. Bijker and J. Law (eds), *Shaping Technology/Building


Valve Corporation (1999), Counter-Strike, Kirkland, WA: Valve Corporation.

SUGGESTED CITATION


CONTRIBUTOR DETAILS

Kelly Bergstrom is a Ph.D. candidate in York University’s Language, Culture and Teaching programme. Her research falls into two overlapping areas of interest: (non)participation in online communities, and the ways gender stereotypes are replicated in and through digital culture.
Jennifer Jenson is Professor of Pedagogy and Technology in the Faculty of Education at York University. Her research and publications include work on gender and technologies, gender and digital gameplay, players and identities in Massively Multiplayer Online Games (MMOGs). She has also published research on technology and education, and on technology policies and policy practices in K-12 education in Canada.

Richard Hydomako received his Ph.D. in physics from the University of Calgary. His expertise includes data analysis and visualization.

Suzanne de Castell is Professor and Dean of the Faculty of Education at University of Ontario Institute of Technology. She specializes in qualitative research methods and her work has focused on literacy, new media, and educational technologies, on learning and attention in game-based and non-formal learning environments, and on multimodal analysis of communicative interaction.

Kelly Bergstrom, Jennifer Jenson, Richard Hydomako and Suzanne de Castell have asserted their right under the Copyright, Designs and Patents Act, 1988, to be identified as the authors of this work in the format that was submitted to Intellect Ltd.
Ubiquity

ISSN: 2045-6271 | Online ISSN: 2045-628X
2 issues per volume | Volume 2, 2012

Aims and Scope
‘Ubiquity’, the ability to be everywhere at the same time, a potential historically attributed to the occult, is now a common feature of the average mobile phone. The journal anticipates the consequences for design and research in a culture where everyone and everything is connected, and will offer a context for visual artists, designers, scientists and writers to consider how ubiquity is transforming our relationship with the world.

Call for Papers
The journal invites contributions on subjects such as:
• The evolution of media forms as they seep off the page and screen and into the environment
• The emergence of the ‘Internet of Things’ and the advent of spimes, blogjects and ambient intelligence
• The consumption of networked digital devices and the transgressions and trauma implicit in the cultural shift that follows
• The behaviours and technologies that cultivate a networked culture
• New inter-disciplinary vocabularies for understanding a social, environmental and technical sense of ‘place’

Editors
Mike Phillips
University of Plymouth
M.Phillips@plymouth.ac.uk

Chris Speed
Edinburgh College of Art
c.speed@eca.ac.uk

Editorial Assistant
Jane Macdonald
Edinburgh College of Art
jane.macdonald@eca.ac.uk