Iyer, S. et al. (Eds.) (2022). Proceedings of the 30th International Conference on Computers in Education.
Asia-Pacific Society for Computers in Education

Changing Students' Perceptions of a History Exploration Game Using Different Scripts

Stefan SLATER [a*], Ryan S. BAKER [a], David GAGNON [b], Erik HARPSTEAD [c], Juliana Ma. Alexandra L. ANDRES [a], Luke SWANSON [b]
[a] *University of Pennsylvania, United States of America*
[b] *University of Wisconsin-Madison, United States of America*
* slater.research@gmail.com

Abstract: As digital games and simulations continue to see use in both formal and informal learning contexts, it becomes increasingly important to understand the goals, motivations, and interests of the learners using them. Recently, there has been increased interest in identifying the different engagement profiles of learners playing games, descriptions of groups of learners' tendencies and preferences when playing digital games. These engagement profiles represent a powerful tool for designers looking to create personalized and adaptive learning environments. In this work, we explore two aspects of these engagement profiles. First, we explore the different profiles of engagement in a large sample of elementary and middle school learners playing the history game *Jo Wilder and the Capitol Case* (2019) in a variety of contexts, using the data from a self-report survey. Second, we explore the effect that minor game dialogue changes had on learners' perceptions of different game elements, such as character likeability and humor, as well as personal preferences, such as enjoyment of history. These findings highlight opportunities for personalized and adaptive game design that leverage players' goals and motivations.

Keywords: Games, engagement profiles, typology, history education

1. Introduction

Games are powerful tools for learning and see ever-increasing use in many different educational contexts. Games and other interactive digital environments are used in both physical and digital classroom environments (Obedoza & Sison, 2020), in after-school programs (Pusey, Wong & Rappa, 2020), and in informal learning contexts such as museums (Kumar, Tissenbaum & Berland, 2017). Games have been shown to have positive learning outcomes in a wide variety of educational settings and domains, such as language learning (Culbertson et al., 2016), physics (Sun et al., 2022), and geometry (Ruiperez-Valiente & Kim, 2020). Part of the appeal of using games for learning lies in the way that they allow students to have experiences and do activities that are difficult to do in more traditional forms of media, such as assuming the role of urban planners (Metcalf, Clarke, & Dede, 2009; Foster et al., 2018) or participating in simulations of authentic science practices (Gaydos & Squire, 2010).

Players in educational games have a wide variety of strategies, goals, and motivations for play (Yee, 2006; Bonny & Castaneda, 2022; Kordyaka et al., 2019). These differences between players may influence both engagement with the *context* of the game, such as when and where players like to play games, as well as their engagement with the *content* of the game, such as whether a student playing a science game is interested in science as a subject. Examining and identifying students' game preferences provides a window into the experiences that they are having during gameplay. Recent research examining these engagement profiles suggests considerable variance in how different groups of students play the same game. For example, Ruiperez-Valiente et al. (2020) explored how students engaged with the STEM MMO *The Radix Endeavor*. Using data collected from students' interactions with the game itself, they constructed multiple metrics of players' gameplay engagement. By clustering learners' actions across these metrics, they identified four overall groups of players in the game: integrally engaged learners, who scored high across all recorded metrics; lone achievers, who scored low in social metrics but high in completion metrics and overall playtime; non-engaged learners, who scored low

across all metrics; and social explorers, who scored high in social measures such as chatting, but scored low in quest completion and game progress.

Whether students' states (their responses to the game and the context in which it is being played) or traits (preferences of the player that persist across games and contexts) give rise to these engagement profiles is not currently well-understood, though there is evidence that patterns vary considerably between different groups of learners (Tabanao & Rodrigo, 2016). If engagement profiles were better understood, they could be an opportunity for implementing personalization in games, as the method of personalization should depend on the qualities of the learner during play.

In this paper, we explore how players' perceptions of a game, its characters, and its subject matter, relate to salient and easily modifiable features of the game environment. In doing so, we aim to explore the following research questions:

RQ1: What different types of student players exist in our game?
RQ2: Do changes to game dialogue change groups' perceptions of gameplay elements?

2. Methods

2.1 Game and Dataset

Data were collected from the point and click history adventure game *Jo Wilder and the Capitol Case*. In *Jo Wilder*, players control Jo, a pre-teenage girl who is attempting to help her historian grandfather determine the story of two artifacts that will be displayed in a new exhibit at the Capitol. She does so reluctantly, motivated primarily by her dislike of another historian, Wells, who is unkind to her grandfather and who she suspects kidnapped her pet badger. Players are free to move throughout the locations, character conversations and inspectable items freely, but the game's narrative structure moves forward linearly as clues are discovered.

2.2 Survey Development

We developed and piloted a 16 item Likert survey embedded within the game to more closely examine levels of student enjoyment of the game and attitudes towards history. This survey measured four aspects of play, including (1) how immersed are you in the game, adapted from the E-Game Flow Scale (Fu, Su, & Yu, 2009), (2) how much do you like the characters, adapted from the Reysen Likability Scale (Reysen, 2005), (3) how funny do you think the characters are, adapted from the Multidimensional Sense of Humor Scale (Thorson & Powell, 1993), and (4) what are your attitudes towards history, adapted from the Three-Dimensions of Student Attitude Towards Science (TDSAS; Zhang & Campbell, 2011). Students completed four-question item banks, drawing one item from each construct, at four points over the course of gameplay. At the beginning of gameplay students also answered two questions regarding their reading level, as we believed that the text-heavy nature of Jo Wilder could affect the enjoyment and accessibility of the game for strong readers versus weak readers.

2.3 Script Development

For this study, different versions of the game script were developed to make the lead character either more vocally reluctant and "snarky" (critical and sarcastic) or more enthusiastic and positive toward history, and to either introduce jokes and wordplays into the dialog between characters to make it funnier or be more serious, avoiding humor. These scripts were developed by first developing the snarky and funny version, then systematically removing each of those elements. Script changes altered between 20% - 40% of the total dialogue in the game, depending on the condition.

Table 1. *Matrix of different Jo Wilder scripts developed, showing presence of snark and/or humor*

	Snark	Humor
Initial Design: Snark and Humor	X	X
Script 1: No Snark		X
Script 2: No Humor	X	

Script 3: No Humor, No Snark		

2.4 Typology Development

We used Latent Class Analysis (LCA; McCutcheon, 2002; Muthen, 2004) to identify different groups of players based on their responses to the game feedback items. LCA is a method of statistical analysis that identifies one or more latent classes within a broader set of data. These latent classes represent subpopulations of individuals, and membership in a subclass is based on the patterns of variance among an individual's observed variables. Analyses were performed in the statistical software package *MPlus*. Following current recommendations from the literature, we began by fitting a *k*=2 model, hypothesizing that two subclasses existed within our dataset. We assessed the fit of this model, then increased the number of hypothesized subclasses, stopping when model fit no longer substantially improved (Zhou & Bowers, 2020; Nylund-Gibson, Grimm, & Masyn, 2019). We included the game script that students experienced as a covariate in the model, using the 3-step procedure in MPlus (Nylund-Gibson & Choi, 2018) to prevent the covariate from influencing the structure of the indicators on the model.

3. Results

There are several metrics that can be used to assess goodness of fit of an LCA model, and there is no one metric that is identified as the consensus best within the field. Instead, multiple metrics are used and compared in order to determine the correct model to interpret (Jung & Wickrama, 2008). In this work we followed guidelines from Samuelsen & Raczynski (2013) and Nylund-Gibson & Choi (2018) to select the model which we interpret. The fit statistics we considered are listed in Table 2; we chose to interpret the *k*=5 model due to it being the last model for which the Vuong-Lo-Mendell-Rubin statistic was significant (Lo, Mendell, & Rubin, 2001).

Table 2. *Summary of model results for fitted models. The model that we used is in* ***bold****.*

Model	AIC	BIC	VLMR	BLRT	Entropy
K=2	102882	103166	< 0.01	< 0.01	0.847
K=3	100785	101167	< 0.01	< 0.01	0.881
K=4	98940	99421	< 0.01	< 0.01	0.883
***K*=5**	**97602**	**98181**	**0.048**	**< 0.01**	**0.861**
K=6	97166	97166	0.730	< 0.01	0.878

3.1. The k = 5 Model

A plot of the group means by variable for the *k*=5 model is shown in Figure 1. The model consists of five groups: (1) players who disliked most aspects of the game, but especially history (red, N = 184); (2) players who didn't like the characters (Jo in particular, blue, N = 546); (3) players who liked the game but didn't like history (green, N = 489); (4) players who liked both history and the game (pink, N = 665); and (5) players who liked the game, but didn't think it was very funny (brown, N = 533). Respectively, we refer to each of these groups as the "Didn't Like Anything" group (red), "Didn't Like Jo" group (blue), "Didn't Like History" group (green), "Liked It All" group (pink), and "Not Very Funny" group (brown).

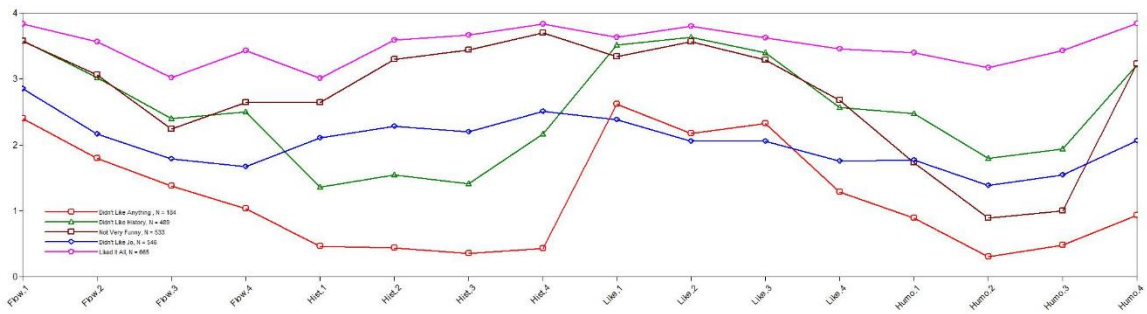

Figure 1. Plot of the sample means for survey items by group.

Once we established the structure of the latent class model, we were able to measure the influence of script type and reading level on students' latent class membership. We followed current recommendations from the literature for the 3-step procedure (Nylund-Gibson & Choi 2018, MPlus Webnote 15 Appendix A) to examine the effect of our covariates, without allowing them to influence the structure of the model itself. A Benjamini & Hochberg post-hoc adjustment (Benjamini & Hochberg, 1995) was used to control for multiple comparisons – however, because this is an exploratory study, we also report differences that were initially statistically significant but did not remain significant after post-hoc adjustment.

3.2. Script Effects on Player Groups

The effects of different scripts on player evaluations are summarized in Table 3. We found no differences between the proportions of players placed into different groups when comparing the No Humor, No Snark script to all other scripts. We also found no differences between the proportions of players placed into different groups when comparing the Initial Design script to all other scripts.

For the Humor No Snark script, the only difference that we found was between the Liked It All and Didn't Like Jo groups of players. For this script, there was a trend suggesting that players were more likely to Like it All, and less likely to Not Like Jo, but this trend did not hold up after a post-hoc correction was applied.

Table 3. *Summary of differences in proportions between scripts.*

-- more likely to be in… --	-- less likely to be in… --	-- when playing… --	Odds ratio
Liked It All	Didn't Like Jo	Humor No Snark	1.519
Didn't Like Jo	Didn't Like History	Snark No Humor	1.660
Didn't Like Jo	Liked It All	Snark No Humor	1.820
Liked It All	Not Very Funny	Snark No Humor	1.506

For the Snark No Humor script, we found three differences between the proportions of groups of players in this script, compared to the other scripts. We found that players were more likely to be in the Didn't Like Jo group, and less likely to be in either the Didn't Like History group or the Liked It All group. We also found that there was a trend for players to be more likely to be in the Liked It All group than the Not Very Funny group compared to other scripts, though again this trend was not quite significant when a post-hoc correction was applied.

3.3. Reading Level Effects on Player Groups

We found strong differential effects of reading level on players' class, which are summarized in Table 4. As reading level increased, players were less likely to be in the "Didn't Like Anything" group, and more likely to be in any other group. Players with a higher reading level were also more likely to be in the "Liked It All" group.

Table 4. *Summary of odds ratios for all reading level effects.*

-- more likely to be in… --	-- less likely to be in… --	-- when a… --	Odds ratio
Liked It All	Didn't Like Any	Strong Reader	2.362
Didn't Like History	Didn't Like Any	Strong Reader	1.456
Didn't Like Jo	Didn't Like Any	Strong Reader	1.377
Not Very Funny	Didn't Like Any	Strong Reader	1.625
Liked It All	Didn't Like History	Strong Reader	1.622
Liked It All	Didn't Like Jo	Strong Reader	1.715
Liked It All	Not Very Funny	Strong Reader	1.454

4. Discussion

In this work, we found evidence for five different groups of players in the game Jo Wilder, based on their self-reported responses to a four-construct, 16-item survey bank. Two groups of players either "Liked It All" (n = 665) or "Didn't Like Any Of It" (n = 184), providing consistently high (or low)

ratings of the game, its characters, and enjoyment of history, respectively. In between these two groups we found "Didn't Like History" (n = 489), "Didn't Like Jo" (the game's main character; n = 546), and "Not Very Funny" (n = 533) groups of players, who generally provided average responses to the survey items, except for specific categories that they did not rate themselves as enjoying. Overall, however, students appeared to enjoy the game experience, with the largest overall group being "Liked It All" (28% of all players) and the smallest group being "Didn't Like Any" (7% of all players).

At present, there is open debate as to whether game typologies represent player *states* – the way that they feel about the activity in the present moment – or player *traits* – relatively fixed attitudes or predispositions towards these types of activities in general. In this work, we have found evidence for both of these hypotheses. Changing the presence of humor and snark in the game scripts did not change the proportions of players in the "Didn't Like Anything" group, suggesting that their attitudes towards Jo Wilder could be based on pre-existing traits related to educational games, or to history. These players may dislike history as a subject, or not enjoy narrative-focused games.

However, in the "Not Very Funny", "Didn't Like History", and "Didn't Like Jo" scripts, we found evidence that players' attitudes (and latent class membership) may be more malleable, and therefore more driven by learners' current states. By removing either humor or snark from the base game script, we found that students' evaluations of Jo as a character changed, as well as their evaluations of history learning and of the game. We believe that this may be because the 'sense of humor' in Jo Wilder comes at the expense of Jo's attitudes towards education. In the base game's script, Jo is not a model student and doesn't particularly care about school, and the humor and snark in the game scripts draw on this character trait. In the Initial Design script, where both humor and snark are present, Jo's character traits could have reinforced players' existing negative attitudes towards history learning. When either of these character traits were dropped from the script, players were more open to positive evaluations of their own history learning. However, these changes did come with a cost – our findings suggest that a "snarky" Jo may have come across to some players as being a mean character, and worsened player attitudes towards her as a character.

For reading level, players with higher self-reported reading level enjoyed the game more as an overall trend, observed across all groups. This finding underscores the importance of reading comprehension and understanding of game content for the employment of games in classroom environments. When designing curricula and pedagogy that utilizes games and digital environments, care needs to be taken to ensure that students will be able to read and understand the content of the game itself, and that its instructions and narratives are worded at an appropriate level for, and are culturally responsive to, the students with whom the game will be used.

Towards the development and refinement of adaptive learning systems, our findings suggest that some players' attitudes towards both game enjoyment and content topics can be substantially influenced by changing the dialogue and tone of the game world. In Jo Wilder, modifying game dialogue was a relatively easy change to make in the game's architecture. Interventions such as this one may then hold promise for improving student attitudes towards desired content topics, or promoting engagement in learning tasks based on players' characteristics. Salient qualities of a game, such as characters' appearances, mannerisms, or senses of humor, could be modified to suit the particular individual playing the game, towards optimizing their engagement with the overall learning activity. These potential changes represent powerful tools for improving players' capacity to engage with games as learning tools, and strengthen learning outcomes for diverse groups of learners.

Acknowledgements

We would like to acknowledge NSF #DRL-1907437 for funding this project, and the Field Day Lab team for the use of their data architecture and guidance in data processing.

References

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289-300.

Bonny, J. W., & Castaneda, L. M. (2022). To Triumph or to Socialize? The Role of Gaming Motivations in Multiplayer Online Battle Arena Gameplay Preferences. *Simulation & Gaming*, *53*(2), 157-174.

Culbertson, G., Andersen, E., White, W., Zhang, D., & Jung, M. (2016). Crystallize: An immersive, collaborative game for second language learning. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 636-647.

Foster, A., Shah, M., Barany, A., Petrovich Jr., M.E., Cellitti, J., Duka, M., Swiecki, Z., Siebart-Evenstone, A., Kinley, H., Quigley, P., & Williamson Shaffer, D. (2018). Virtual learning environments for promoting self transformation: iterative design and implementation of Philadelphia land science. In *International Conference on Immersive Learning*, 3-22.

Fu, F. L., Su, R. C., & Yu, S. C. (2009). EGameFlow: A scale to measure learners' enjoyment of e-learning games. *Computers & Education*, *52*(1), 101-112.

Gaydos, M., & Squire, K. (2010). Citizen science: Designing a game for the 21st century. In *Interdisciplinary Models and Tools for Serious Sames: Emerging Concepts and Future Directions*, 289-305.

Jung, T., & Wickrama, K. A. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, *2*(1), 302-317.

Kordyaka, B., Jahn, K., Müller, M., & Niehaves, B. (2019). The Comparative Self: Understanding the Motivation to Play and the Subsequent Video Game Use. In *DiGRA Conference*.

Kumar, V., Tissenbaum, M., & Berland, M. (2017). What are visitors up to? Helping museum facilitators know what visitors are doing. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 558-559.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*(3), 767-778.

McCutcheon, A. L. (2002). Basic concepts and procedures in single-and multiple-group latent class analysis. *Applied Latent Class Analysis*, 56-88.

Metcalf, S. J., Clarke, J., & Dede, C. (2009). Virtual worlds for education: River City and EcoMUVE. In *MiT6 International Conference*, 1-6.

Muthén, B. (2004). Latent variable analysis. *The Sage Handbook of Quantitative Methodology for the Social Sciences*, *345*(368), 106-109.

Nylund-Gibson, K., Grimm, R. P., & Masyn, K. E. (2019). Prediction from latent classes: A demonstration of different approaches to include distal outcomes in mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(6), 967-985.

Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*, *4*(4), 440.

Obedoza, J.B., & Sison, R. (2020). Design of a Game-Based Intelligent Learning Environment to Remediate Fraction Addition/Subtraction Misconceptions. *Proceedings of the 28th International Conference on Computers in Education*.

Pusey, M., Wong, K.W., & Rappa, N. (2020). The effect of a more knowledgeable other on resilience while playing single-player puzzle video games. *Proceedings of the 28th International Conference on Computers in Education.*

Reysen, S. (2005). Construction of a new scale: The Reysen likability scale. *Social Behavior and Personality: An International Journal*, *33*(2), 201-208.

Ruipérez-Valiente, J. A., & Kim, Y. J. (2020). Effects of solo vs. collaborative play in a digital learning game on geometry: Results from a K12 experiment. *Computers & Education*, *159*.

Ruiperez-Valiente, J. A., Gaydos, M., Rosenheck, L., Kim, Y. J., & Klopfer, E. (2020). Patterns of engagement in an educational massively multiplayer online game: A multidimensional view. *IEEE Transactions on Learning Technologies*, *13*(4), 648-661.

Samuelsen, K., & Raczynski, K. (2013). Latent Class/Profile Analysis. *Applied Quantitative Analysis in Education and the Social Sciences*, 304.

Sun, C., Shute, V.J., Stewart, A.E.B., Beck-White, Q., Reinhardt, C.R., Zhou, G., Duran, N., & D'Mello, S. (2022). The relationship between collaborative problem solving behaviors and solution outcomes in a game-based learning environment. *Computers in Human Behavior*, *128*, 107-120

Tabanao, E., Rodrigo, M. M. T. (2016). A comparison of the experience of confusion among Filipino and American learners while using an educational game for physics. In *Proceedings of the International Conference on Computers in Education* (pp. 17-20).

Thorson, J. A., & Powell, F. C. (1993). Sense of humor and dimensions of personality. *Journal of Clinical Psychology*, *49*(6), 799-809.

Yee, N. (2006). Motivations for play in online games. *CyberPsychology & Behavior*, *9*(6), 772-775.

Zhang, D., & Campbell, T. (2011). The psychometric evaluation of a three-dimension elementary science attitude survey. *Journal of Science Teacher Education*, *22*(7), 595-612.

Zhou, X., & Bowers, A. (2020). A Typology of Parental Involvement in Student Experience. *The High School Journal*, *103*(2), 99-132.