Computers & Education 87 (2015) 24–34

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Exploring the influence of game design on learning and voluntary use in an online vascular anatomy study aid

Andrea Gauthier [a,*], Michael Corrin [b], Jodie Jenkinson [b]

[a] Institute of Medical Science, Faculty of Medicine, University of Toronto, 1 King's College Circle, Toronto, ON, M5S 1A8, Canada
[b] Biomedical Communications, University of Toronto, Terrence Donnelly Health Sciences Complex, 3359 Mississauga Road, Mississauga, ON, L5L 1C6, Canada

ARTICLE INFO

Article history:
Received 27 November 2014
Received in revised form
17 March 2015
Accepted 18 March 2015
Available online 4 April 2015

Keywords:
Serious games
Interactive learning environments
Media in education
Teaching/learning strategies
Computer-mediated communication

ABSTRACT

This research explores the educational impact of an online study aid-game for studying human vascular anatomy ($n = 24$) versus a similar non-game study aid ($n = 22$) and how it relates to medical students' demographic traits and voluntary use over a 35-day period. Hierarchical linear regression models revealed that *study aid success rate* (a metric for assessing performance through the study aids) was a significant predictor of anatomy test improvement with the game ($\beta = 0.41$, $p = 0.05$), but not for the non-game ($\beta = 0.14$, $p = 0.56$). Our analyses suggest that game mechanics encouraged more specific problem-solving strategies than did the control study aid, leading to greater predictability of learning outcomes. There was a non-significant trend among game treatment participants, who were more likely to complete study tasks than those assigned to the control treatment ($p = 0.11$). It would appear that students' studying habits had the greatest influence (though opposite in both tools) on level of engagement in study aid use. However, contrary to expectations, self-reported gaming habits did not impact participation. Overall, these findings support the integration of game design into undergraduate study aids as a means of increasing use of supplementary educational tools and assessing knowledge.

1. Introduction

1.1. Background and theoretical framework

Learning human vascular anatomy is difficult. It requires an understanding of systemic and spatial interrelationships comprised of complex interconnected and branching networks (Hmelo-Silver & Azevedo, 2006)—the aorta alone has over 160 branching vessels. In addition, fine vascular structures are often difficult to preserve in dissection by novice learners, making it a challenge to study these structures in lab. To help students grapple with this material, interactive tools have successfully been used in anatomy instruction to increase engagement, knowledge retention, and systems thinking (Hilbelink, 2009; McCarroll, Pohle-Krauza, & Martin, 2009; Mcintire, 1995; Petersson, Sinkvist, & Wang, 2009). Many of these tools contain self-testing mechanisms; it has been shown that undergraduates' use of self-testing while studying is positively correlated with well-regimented studying habits and academic achievement (Hartwig & Dunlosky, 2012; Stewart, Panus, Hagemeier, Thigpen, & Brooks, 2014). When given a choice, high-performing students are more likely to choose a self-testing method over review-type studying (Kornell & Son, 2009).

Educational, or “serious”, games are designed to support the learning process by integrating cognitive techniques for learning, such as self-testing mechanisms, with play and gaming elements (Gee, 2007; Landers & Callan, 2011; Squire, 2013). Game elements may include a storyline, goals, rules, penalties, achievements, intellectual challenges, scoring systems, and leaderboards (Westwood & Griffiths, 2010). These game design elements have potential to increase a student's willingness to participate in meaningful and intellectual play, thereby enhancing his or her understanding of target content and concepts (Gee, 2007; Squire, 2011).

* Corresponding author. Tel.: +1 905 599 0243.
E-mail addresses: andrea.gauthier@utoronto.ca (A. Gauthier), m.corrin@utoronto.ca (M. Corrin), j.jenkinson@utoronto.ca (J. Jenkinson).

http://dx.doi.org/10.1016/j.compedu.2015.03.017
0360-1315/

Recent research has focused on using serious games and other digital applications both as a means of intervention and of assessment. Emerging frameworks for assessing learning through users' digital interactions are based largely on *educational data mining methods* and *evidence-centred design* (Halverson & Owen, 2014; Mayotte, 2010; Owen, Ramirez, Salmon, & Halverson, 2014; Reese, Tabachnick, & Kosko, 2014; Shute & Ventura, 2013). Educational data mining involves tracking users' click-stream data (actions taken by users while interacting with a digital application). These data can then be summarized and analysed for patterns and trends that, in the case of serious games, are indicative of learning. Evidence-centred design involves the interplay between three models: 1) the student model, that defines the intervention's learning objectives; 2) the evidence model, which describes behaviours that suggest that the student has met those learning objectives; and 3) the task model, that defines activities or problems that are presented to the user in order to elicit the behaviour of the evidence model (Mislevy & Haertel, 2006). When implemented carefully into a game's design, this framework enables researchers to make connections between students' in-game performance and learning.

For example, Halverson and Owen (2014) designed and tested a game, *Progenitor X*, with high school students to examine the link between gameplay performance and learning about stem cell biology. They developed their own framework ADAGE (Assessment Data Aggregator for Game Environments), based largely on evidence-centred design and educational data mining methodologies. The investigators found players' efficiency ratios—described as successful game cycle completions over cycle attempts—were significantly correlated with pre-post test gains. In evidence-centred design terms, they were able to conclude that higher completion efficiency (evidence model) of game puzzle cycles (task model) meant that players have a better grasp of stem cell concepts (student model). Additionally, by tracking players' behaviours through the click-stream data, the investigators were able to show that differential in-game behaviours led to variable game success and knowledge gains. This research contributes evidence linking gameplay interactions directly to learning. However, this type of methodology is still developing and additional studies are needed to add robustness to this field (National Research Council, 2011).

Additionally, it is important to consider the potential determinants of a serious game's impact, such as student type, previous gaming habits, or gender (National Research Council, 2011). As previously mentioned, high-achieving students with good studying habits may be more likely to engage in self-testing practices (Hartwig & Dunlosky, 2012; Stewart et al., 2014) and, if a game employs this technique, then these individuals may be more likely to play it; this effect may be enhanced if the student also displays high video gaming habits outside of school. Gender also has a known influence on gaming preferences, where females typically play less than males and generally gravitate toward more socially oriented games and males toward action-oriented games, though there is considerable variability (Greenberg, Sherry, Lachlan, Lucas, & Holmstrom, 2010). All of these aspects should be taken into account when evaluating the effectiveness and appeal of a serious game.

Many serious game-based research protocols take place in formal settings, such as investigators' labs or classrooms (Barab & Dede, 2007; Barab, Gresalfi, & Ingram-Goble, 2010; Charsky & Ressler, 2011; Kafai, 2009; Kafai, Quintero, & Feldon, 2009; Klopfer, Sheldon, Perry, & Chen, 2012), or in structured, informal environments like museums and after-school programs (Miller, Chang, Wang, Beier, & Klisch, 2011; Squire, 2011; Squire, DeVane, & Durga, 2008). Overall, there is a lack of research that implements serious games in unstructured, informal learning environments where students have absolute choice on where and when to play (National Research Council, 2011).

1.2. Summary of pilot study and goals of the current research

This research uses evidence-centred design and educational data mining techniques to explore the educational impact of learning with an online study aid-game versus a similar non-game study aid and how it relates to students' voluntary use and demographic characteristics. A pilot study by Gauthier and Corrin (2013) found that undergraduate anatomy students were significantly more likely to complete and attempt study aid tasks and engage in use sessions when given a game-study aid than a non-game-study aid over a period of one week. However, learning gains and how these related to gameplay were not considered. Additionally, though the study aids both contained the same learning material, the aesthetic differences between the two treatments possibly confounded the results. By ensuring that both treatments in a randomized trial have equivalent interaction and visual designs, we can tease out the unique contribution that game design makes to learning and engagement. In the present study, we sought to minimize these differences between both digital tools in order to explore how learning is related to study aid interactions.

This research is unique in that it places a serious game in an unstructured, informal learning environment (i.e. wherever the students see fit to use it). It represents a more contextually relevant learning experience as it is used by students who are actively learning the material and allowing these users freedom to interact with the study aids as they would with any supplementary material outside of the classroom.

Through a randomized trial, this study investigates how the presence of game design influences students' learning in an interactive study aid and how personal characteristics mediate voluntary use of the application. Specifically, we hypothesized that:

1. The presence of game design (experimental treatment) would encourage greater use (measured in completed study aid tasks, attempted tasks, and total usage sessions) and breadth of use (unique completed tasks), resulting in higher measured study aid performance (study aid success rate; refer to section 2.2)
2. Personal characteristics, such as gender, studying habits, and gaming habits would influence the voluntary use described above. For example, individuals who play video games on a regular basis would be more likely to engage in study aid tasks in the game group.
3. Higher study aid performance would be predictive of greater learning in both treatments, supporting the incorporation of evidence-centred, node-based study aids for vascular anatomy into students' regular studying practices.

2. Methods

2.1. Description of the study aids

Vascular Invaders is a web-based study aid with integrated game design elements, which is geared towards supporting undergraduate medical students' understanding of human vascular anatomy. The vascular system is represented by a network of nodes, each node

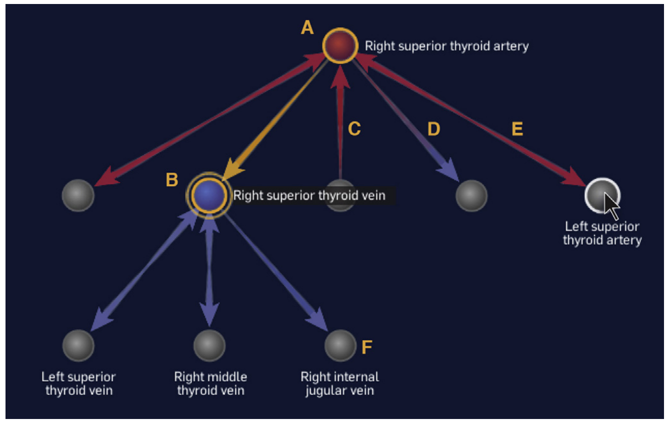

Fig. 1. Detail of vascular "node" network — **A)** Red nodes represent oxygenated blood vessels; **B)** Blue nodes represent deoxygenated blood vessels; **C)** Upward arrows indicate counter-current blood flow; **D)** Downward arrows indicate concurrent blood flow; **E)** Double-headed arrows indicate an anastomosis (blood may flow in both directions); **F)** Grey nodes represent blood vessels that are currently hidden on the 3D anatomical map (not depicted). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

symbolizing one blood vessel (see Fig. 1). The game allows students to self-test their knowledge of nomenclature, vessel supply, and anastomoses by engaging them in vascular-sequencing tasks where players must make their way from one node in the circulatory system to another based on vascular branching patterns. An interactive and anatomically accurate 3D model of vasculature links the nodes and anatomical terminology to structures in 3D space. Most vascular structures are hidden from the user to encourage recall; structures become visible once the player "visits" them or once they purposefully choose to reveal them for help. *Vascular Invaders* (Fig. 2) has a mirrored study aid, the *Vascular Anatomy Study Aid* (Fig. 3), which contains the same learning material, path-finding mechanic, and visual treatment as the game but lacks any game elements, such as rules, penalties, leader board, achievements, points system, storyline, etc. Both versions have the additional challenge of vascular blockages in certain path-finding tasks, forcing the student to consider alternative routes. The academic content of the study aid(s) used for this research was limited to the head and neck segment of the human systemic vascular system; this region was selected because instructors identified it as challenging to learn.

The premise of *Vascular Invaders* is that a rogue nanobot technology, called *Bacterbots*, has infected the University's population. Travelling through the blood stream, the Bacterbots are wreaking havoc on students' immune systems. Players must guide a *Bloodbot* (a good nanobot developed to seek out and destroy these vascular invaders) through the circulatory system to various invasion sites to destroy the Bacterbots. Players have a limited amount of energy with which to reach the invasion site; going against the flow of blood or revealing structures on the anatomical model will use up energy, so they must be strategic about the path that they take and must exercise their knowledge of vascular pathways in order to reach their goal. Lastly, collectable items or power-ups (e.g. a plunger) can be found throughout the circulatory system and can be used strategically for help along the way (plungers unclog blocked blood vessels).

Both study aids were programmed using the Unity Game Engine in C# and exported at 960 by 650 pixels for play in the Unity Web browser plugin. The vasculature and skull models were built using Maxon Cinema 4D and verified for accuracy by an anatomist. To ensure that the study aids had sound interaction and communication design, a round of usability testing was performed with non-medical graduate students; one year prior, these students had completed the same anatomy course with a similar medical student group. Between usability

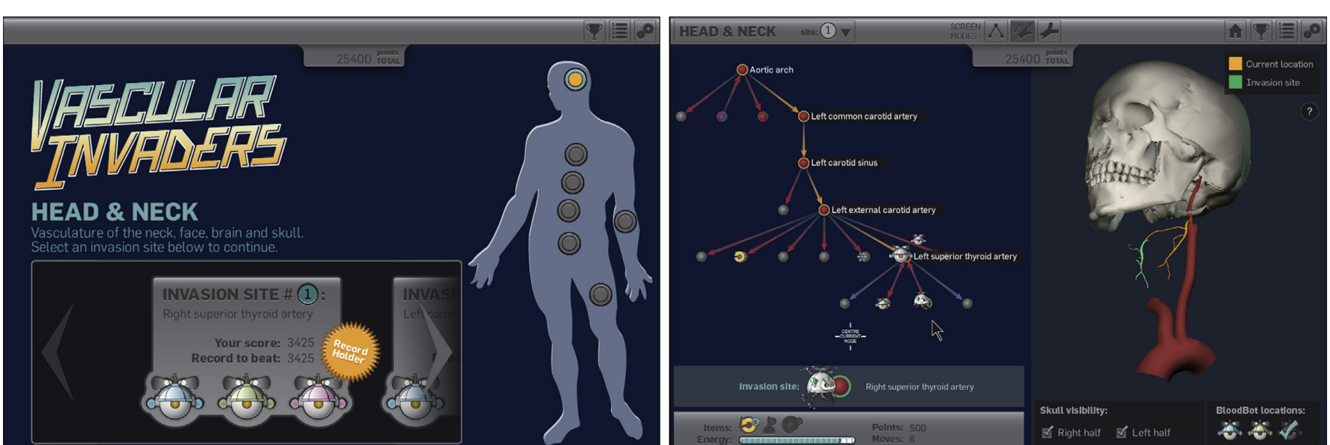

Fig. 2. Screenshots of *Vascular Invaders* — **Left**: Home Page. Users can choose an unlocked path-finding task (Invasion Site), view achievements, view the leaderboard, or edit app settings. **Right**: Path-finding task in progress. Students compare the node network to the anatomical model to find their goal. They may collect and use power-ups, find *Bloodbot* friends, attempt to beat the high score, and must reach their goal without running out of energy.

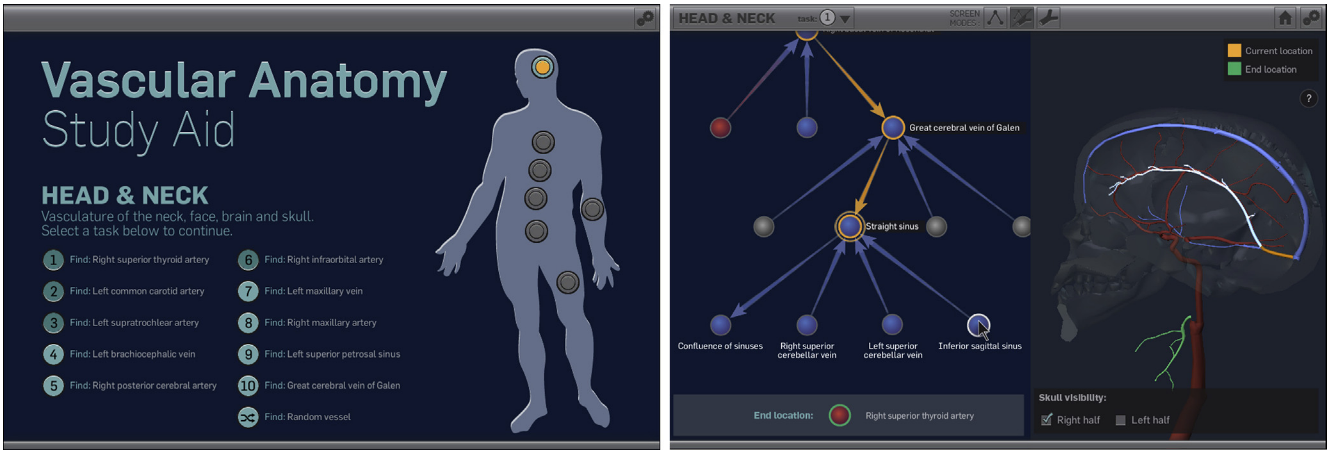

Fig. 3. Screenshots of the *Vascular Anatomy Study Aid* — **Left**: Home Page. Users can choose a path-finding task in any order and edit app settings. **Right**: Path-finding task in progress. Students compare the node network to the anatomical model to find their goal with the same path-finding mechanic as in *Vascular Invaders*, but no additional game design elements are present.

participants, iterative improvements were made to the study aid tools, and the testing-improvement cycle was performed until test users could reliably complete all study aid essential actions, such as task completion and game element access.

2.2. The evidence-centred design framework

As described in Section 2.1, both study aid versions contained the same path-finding tasks and navigation mechanisms and can therefore be assigned the same evidence-centred design framework. This framework is described below.

The student model: The learning objectives include nomenclature and vascular branching patterns of the head and neck, designed around the first-year medical anatomy curriculum at the University of Toronto.

The task model: The study aids present 10 predefined vascular-sequencing challenges (referred to as tasks), each with multiple solutions. Both tools also have a randomized task generator that continually creates new challenges with different starting and ending points. For example, a task might start in the aortic arch and the goal might be to find the right facial artery. The student would then click on a node corresponding to the next artery that leads from the aortic arch to the right facial artery (i.e. the brachiocephalic trunk). Or they may choose to take a longer route through the left side of the face and transfer to the right through arterial anastomoses; many solutions are possible to every task. When designing these study aids, an attempt was made to focus the predefined tasks around blood vessels with high clinical relevance that the students would be tested on during their exam.

The evidence model: Successful completion of a vascular sequencing task would suggest that the student knows the vasculature relevant to that particular task. Successful completion of all tasks would indicate that the student knows the overall vasculature of the head and neck region well since the ten predefined tasks attempted to cover all relevant anatomy. Furthermore, successfully repeating tasks would lead to consolidation of the material and would be indicative of greater learning. All of these variables together are reflective of participants' mastery of the student model.

To represent these data, we developed a metric that we call a *study aid success rate* (SASR), based on the concept of efficiency ratios briefly described in section 1.1 (Halverson & Owen, 2014). The SASR considers users' task-completion efficiency (successfully completed tasks over total attempted tasks) and the amount of available study aid material completed (in this case, out of 11 unique available tasks including the randomized task generator). For example, an individual who completed only one unique task and attempted a total of two different tasks would have an SASR of 9.09%. An individual who completed 10 tasks and attempted 20 tasks, nine of which were unique, would have an SASR of 40.91%. Since participants were allowed to use the study aid to the extent that they wished, it becomes critical to combine efficiency and breadth characteristics as described. In the example above, both of these individuals completed tasks at 50% efficiency, which would not be representative of the overall difference in effort put in by these students. The SASR adjusts for the diversity of material the individual was exposed to, which, we believed, would lead to better predictability of learning.

$$SASR = \frac{completed\ \ tasks}{attempted\ \ tasks} \times \frac{unique\ \ completed\ \ tasks}{unique\ \ available\ \ tasks} \times 100\%$$

As such, the SASR can be implemented as a predictive measure into a multivariate linear regression model to determine an individual's knowledge gains (results summarized in section 3.4). SASR is a helpful metric in that it controls for changes in the number of completed tasks, attempted tasks and unique attempted tasks simultaneously; these are variables that are correlated with each other and fluctuate with respect to one another (e.g. attempted tasks automatically increases as a student completes more tasks). This causes issues of multicollinearity when these variables are included as individual predictors into a regression model, which is a principal reason why the SASR was developed to be included in their place.

2.3. Telemetry data to be collected

Educational data mining techniques were used to support the evidence-centred design model by collecting telemetry—or click-stream—data. Table 2 summarizes the collected statistics for both control and experimental groups. Of note, statistics such as use sessions,

Table 1
Data summary and two-tail analyses of difference (Wilcoxon test at $\alpha = 0.05$) of personal characteristics for the *Vascular Anatomy Study Aid* (VASA, n = 22) and *Vascular Invaders* (VI, n = 24) digital study aids. SD = Standard Deviation.

Personal characteristic	Control group (VASA), n = 22				Experimental group (VI), n = 24				Difference
	Min	Max	Mean	SD	Min	Max	Mean	SD	*p* (W-value)
Age	20	32	23.76	2.32	21	31	23.58	2.41	0.67 (W = 0.67)
Sex	0 (M)	1 (F)	0.68	0.48	0 (M)	1 (F)	0.54	0.51	0.34 (W = 301)
Study habits	1	3	1.64	0.58	1	3	1.67	0.56	0.85 (W = 0.85)
Game-playing habits	0	3	0.86	0.89	0	3	1.21	1.02	0.25 (W = 0.25)

attempted tasks, and completed tasks are used to make judgements about students' amount of use and engagement levels. Average moves per task and revealed 3D structures give us more information about how the user is interacting with the study aid. The number of unique completed tasks provides a sense of the amount of available content covered by the user. Experimental tool-users' interactions with specific game elements (leaderboard views, achievements, power-ups used, etc.) were also collected in order to judge their importance in continued game-use.

2.4. Participants and procedure

The study was implemented with first-year medical anatomy students (n = 46: male = 18, female = 28) at the University of Toronto (UofT). The entire study was administered on a password-protected website accessed by students via their UofT student login credentials. Participants first completed a personal-information questionnaire collecting data on age, gender, self-reported studying habits, and gaming habits, after which the web-based program randomly assigned students to one of two groups (control: n = 22; experimental: n = 24), minimized for gender (Scott et al., 2002). Students next completed a vascular anatomy pre-test, which tested students' knowledge of vascular pathways, connecting vessels, tissue blood supply, and anastomoses. Study participants completed one of two multiple-choice, 15-question tests (A or B). An anatomy expert at the University of Toronto validated the accuracy and equivalency of both versions. Students were then provided with a link to their assigned digital study aid.

Groups proved to be homogeneous in personal traits such as age ($p = 0.67$), sex ($p = 0.34$), studying habits ($p = 0.85$), and game-playing habits ($p = 0.25$) when analysed with two-tail Wilcoxon tests ($\alpha = 0.05$). Results are shown in Table 1. Students also proved to be homogeneous for prior knowledge on the pre-test ($p = 0.33$), evaluated with a two-tail, two-sample t-test ($\alpha = 0.05$). Refer to Table 3.

Both groups had access to their assigned study aid (*Vascular Invaders* or the *Vascular Anatomy Study Aid*) for a period of 35 days. This timeframe was chosen because it was the period from introduction of the anatomy of the head and neck (material covered by study aids) in lecture up until the final exam, a time during which students would naturally be studying the targeted material. Participants were allowed to use their study aid as much or as little as they liked during this time but were required to complete at least one vascular-sequencing task in order to be included in the study. Telemetric data (such as play sessions, tasks attempted, tasks completed, average moves per task, and interactions with various game elements) on their study aid-use were collected digitally (refer to section 2.3).

A vascular anatomy post-test was administered after the 35-day tool-use period. Those exposed to Test A on the pre-test were administered Test B on the post-test and vice versa in order to eliminate priming effects from pre-test to post-test.

2.5. Data analysis

We performed three primary analyses on our telemetric, demographic, and test data in order to prove our hypotheses. Firstly, learning outcomes of treatment groups were compared with two-tailed independent-samples T-tests ($\alpha = 0.05$), and tool-use data were compared with two-tailed non-parametric Wilcoxon tests ($\alpha = 0.05$), to determine whether or not participants exposed to the game learned more and used their study aid to a greater extent than the control group. Secondly, two multiple regression models were formulated to determine how personal characteristics (gender, gaming habits and studying habits) affected use of the study aids (measured in completed tasks) and how

Table 2
Data summary and two-tail analyses of difference (Wilcoxon test at $\alpha = 0.05$) of tool-use telemetry for the *Vascular Anatomy Study Aid* (VASA, n = 22) and *Vascular Invaders* (VI, n = 24) digital study aids, over a period of 35 days. **A**) Basic tool-use data; **B**) Performance metrics; **C**) Game element interaction statistics. SD = Standard Deviation. Trending (*$p < 0.20$) and significant (**$p < 0.05$) differences are indicated.

Tool-use statistic		Control group (VASA), n = 22				Experimental group (VI), n = 24				Difference
		Min	Max	Mean	SD	Min	Max	Mean	SD	*p* (W-value)
A	Use sessions	1	5	2.5	1.34	1	8	2.88	1.87	0.66 (W = 244.5)
	Tasks completed	1	17	6.73	4.86	1	43	11.79	11.14	0.11 (W = 190.5)*
	Tasks attempted	2	30	10.32	4.86	2	59	16	14.69	0.24 (W = 210.5)
	Unique vessels visited	6	100	44.5	29.47	7	207	69.29	57.22	0.19 (W = 204.5)*
	Unique completed	1	11	5.68	3.56	1	11	7	3.81	0.17 (W = 201.5)*
B	Ave. moves per task	4.67	28.22	14.99	6.7	5.6	21.81	11.83	3.7	0.04 (W = 356)**
	Revealed 3D structure	0	121	8.55	26.01	0	13	1.17	2.63	0.50 (W = 292.5)
	Study aid success rate	2.27	85	35.7	26.22	3.03	92.86	48.56	30.13	0.14 (W = 196.5)*
C	Leaderboard views	–	–	–	–	0	23	3	5.95	–
	Achievements earned	–	–	–	–	0	5	2.42	1.47	–
	Friends collected	–	–	–	–	0	70	10.54	17.55	–
	Power-ups used	–	–	–	–	0	38	6.08	8.72	–

Table 3
Data summary and two-tail analyses of difference (t-test at $\alpha = 0.05$) of anatomy test scores for the *Vascular Anatomy Study Aid* (VASA, n = 22) and *Vascular Invaders* (VI, n = 24) groups. SD = Standard Deviation, DoF = Degrees of Freedom. Trending (*$p < 0.20$) differences are indicated.

Statistic	Control group (VASA)				Experimental group (VI)				Difference		
	Min	Max	Mean	SD	Min	Max	Mean	SD	*t*-value	DoF	*p*-value
Pre-test results (out of 15)	0	10	3.22	2.51	0	13	2.46	2.73	0.99	44	0.33
Post-test results (out of 15)	5	12	9.23	2.07	5	14	9.75	2.15	−0.84	43.89	0.41
Pre-post differential	1	10	6	2.85	−2	13	7.29	3.32	−1.42	43.82	0.16*

this differed between treatments. Lastly, an additional two hierarchical linear regression models—one for the control group and one for the experimental group—were implemented using study aid success rate as a predictor for learning, while adjusting for personal characteristics. All analyses were performed using SPSS Statistics version 22.0 (IBM Corporation, 2013).

3. Results

3.1. Summary of major findings

Study aid success rate proved to be a significant predictor of pre-post-test gains for *Vascular Invaders* players, but not for participants using the *Vascular Anatomy Study Aid* (section 3.4). This distinction comes despite the fact that overall tool-use (measured in attempted tasks, completed tasks, tool-use sessions) and study aid performance (SASR) were not statistically different between treatments, contrary to our hypotheses (section 3.2). Further, it was found that studying habits influenced tool-use in unexpected ways (section 3.3). These results are described in more detail below.

3.2. Comparison of study aid use and knowledge gains over the 35-day period

Table 2 summarizes the telemetric tool-use data (section *A*) and performance metrics (section *B*) and Table 3 compares anatomy test results. Students assigned to *Vascular Invaders* used their study aid to a greater extent than those assigned to the control treatment in total completed tasks ($p = 0.11$), total attempted tasks ($p = 0.24$), and unique completed tasks ($p = 0.17$), though these differences did not reach significance. However, the amount of statistical variance seen in completed tasks was significantly different between groups ($p < 0.01$, $F = 0.19$), with individuals completing as many as 43 study aid tasks compared to a maximum of 17 tasks in the control group. Additionally, experimental users completed their tasks in significantly less moves than the control group ($p = 0.04$).

Study aid success rates (SASR) were also compared. The control group recorded a mean SASR of 35.70% and the experimental group of 48.56%, a difference that trends toward significance ($p = 0.14$). Test improvement also trended toward increased performance in the experimental group though not significantly ($p = 0.16$).

It should be noted that metrics on the number of times 3D anatomical structures were revealed were also collected. As described in section 2.1, students could choose to reveal anatomical structures to help them complete a sequencing task; this had an energy cost in the case of the experimental group. Though this function was introduced in the tutorial, only 14 control-tool users (mean 8.55 times) and 18 experimental-tool users (mean 1.17 times) used this function at all, most only using the function once or twice during the entire duration of the study (see Table 2 B). We believe this to be due to a design flaw, making the function difficult to use (discussed further in section 4.4). Further analyses did not reveal any correlation between this statistic and test improvement, demographic traits, or tool-use data and there was no significant difference in use of this function between groups ($p = 0.49$). We will therefore not consider this statistic in our analyses.

Telemetric data on the experimental group's interactions with various game elements in *Vascular Invaders* were also collected (Table 2C) to investigate their influence on participation. For each participant, we computed the average interaction per completed task of each game element and adjusted these for the amount of tool completion (e.g. leaderboard views over completed tasks, multiplied by the fraction of tool completion). Similar to the modification of efficiency ratios in the SASR, these adjustments were made in order to compensate for the fact that participants completed available studying material (measured in "unique" tasks) to varying degrees; simply dividing by the number of completed tasks would provide a skewed perspective on the relative persistence of interaction with these elements. Pearson correlation analyses revealed significant positive correlations between completed tasks and these adjusted averages: leaderboard views ($r = 0.47$, $p = 0.02$), *Bloodbot* friends collected (a mini-goal) ($r = 0.86$, $p < 0.01$), and power-ups used ($r = 0.51$, $p = 0.01$). The exception was achievements earned ($r = 0.14$, $p = 0.52$), likely because players could chance upon being awarded achievements in the earlier stages of the game, and therefore this particular statistic is not reflective of intentional interactions.

3.3. Relationship between demographic traits and tool-use

We were interested in how personal characteristics might influence students' use of their assigned tool and ran two multivariate linear regression models to predict use based on demographic traits; results are presented in Table 4. Overall, demographic traits (gender, studying habits, and gaming habits) proved to be trending predictors of tool-use (measured in completed tasks) in the control group ($R^2_{adjust} = 0.16$, $p = 0.11$) and significant predictors in the experimental group ($R^2_{adjust} = 0.27$, $p = 0.03$). In order to avoid issues of multicollinearity with studying habits and gaming habits, age was not included in the model.

Strong and *opposite* effects on tool-use were found for reported studying habits in the two groups. Studying habits were measured on a four-point Likert scale (where 1 = "I cram the night before the exam", and 4 = "I study fastidiously every day"). The models revealed that *Vascular Anatomy Study Aid*-users tended to complete significantly *fewer* tasks if they reported good, regimented studying habits ($b = -3.82$, $\beta = -0.46$, $p = 0.05$), whereas *Vascular Invaders*-players completed significantly *more* tasks if they reported good studying habits

Table 4
Multiple regression model using gender, studying habits and gaming habits to predict **completed tasks** for A) the *Vascular Anatomy Study Aid* (n = 22) and B) *Vascular Invaders* (n = 24) groups. Trending (*p < 0.20) and significant (**p < 0.05) values are indicated.

	b	Std. Error	β	95% Confidence interval for b	Semipartial correlation
A) Control group (Vascular Anatomy Study Aid)					
(Constant)	11.628	4.273	–	(2.65, 20.61)	–
Gender	1.304	2.214	0.128	(−3.35, 5.96)	0.118
Studying Habits	−3.817**	1.806	−0.456**	(−7.61, −0.02)	−0.424**
Gaming Habits	0.527	1.167	0.096	(−1.92, 2.98)	0.091
$R^2 = 0.27$; adjusted $R^2 = 0.16$; ANOVA: F = 2.29; $p = 0.11$*					
B) Experimental group (Vascular Invaders)					
(Constant)	−5.80	7.66	–	(−21.78, 10.18)	–
Gender	−2.24	4.48	−0.10	(−11.58, 7.10)	−0.09
Studying Habits	10.78**	3.69	0.55**	(3.09, 18.47)	0.52**
Gaming Habits	0.70	2.25	0.06	(−3.99, 5.39)	0.06
$R^2 = 0.37$; adjusted $R^2 = 0.27$; ANOVA: F = 3.87; $p = 0.03$**					

(b = −10.78, β = 0.55, $p < 0.01$). In more understandable terms, this data means that for every unit increase on the studying habits Likert scale, control participants will complete on average 3.8 tasks less, and experimental participants will complete on average 10.8 tasks more, than someone with lesser studying habits.

Gender and gaming habits appeared not to have any significant or trending effects in either model.

3.4. Relationship between study aid performance, personal characteristics, and knowledge gains

A significant positive correlation was found between study aid success rate (SASR) and test improvement in the game group ($r = 0.51$, $p = 0.01$) but not in the control group ($r = 0.26$, $p = 0.24$) (Fig. 4). To investigate this relationship further, as well as to test the learning-assessment capabilities of the study aids, two hierarchical linear multivariate linear regression models were constructed (see Table 5). The models were constructed using the study aid success rate (SASR) as the predictor variable and test improvement as the outcome, while adjusting for studying habits, gaming habits, and gender as possible confounders. It tests the hypothesis that higher performance in the study aids is indicative of greater learning. The control model has an adjusted R^2 of 0.04 ($p = 0.33$), meaning that the model accounts for only 4% of the variability learning outcomes; the experimental model has an adjusted R^2 of 0.28 ($p = 0.04$), suggesting that 28% of the variability in learning is accounted for (Table 5 B).

Study aid success rate proved to be a significant predictor of pre-post-test gains for *Vascular Invaders* players (b = 0.05, β = 0.41, $p = 0.05$) but not for participants using the *Vascular Anatomy Study Aid* (b = 0.02, β = 0.14, $p = 0.56$). In translation, for every 10% increase in SASR, control participants can expect, on average, a 0.2-point (1.3%) increase in test performance, whereas experimental participants can expect a 0.5-point (3.3%) increase, after adjusting for other covariates. The accuracy of both models are reflected in Fig. 5, which plots predicted learning outcomes against actual learning outcomes; a stronger relationship with tighter confidence intervals is observed in the experimental group. Semipartial correlations, which indicate the relative "unique" contribution of SASR to each model, confirm beta values; SASR has a semipartial correlation nearly three fold larger in the experimental model ($r_{semi} = 0.37$) than in the control model ($r_{semi} = 0.13$).

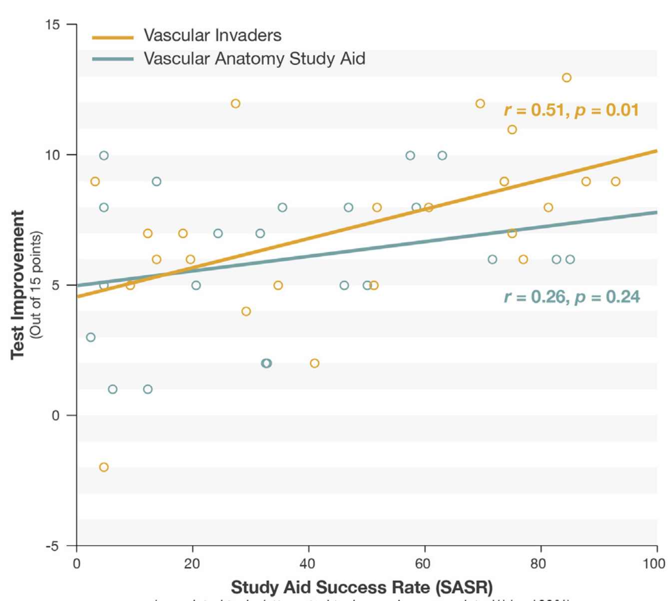

Fig. 4. This graph plots the crude linear relationship between study aid success rate (SASR) and test improvement for participants exposed to *Vascular Anatomy Study Aid* (control tool) or *Vascular Invaders* (experimental tool). Pearson correlation statistics are indicated.

Table 5
Two-step hierarchical regression model using study aid success rate (SASR) to predict **test improvement**, while adjusting for gender, studying habits and gaming habits for A) the *Vascular Anatomy Study Aid* (n = 22) and B) *Vascular Invaders* (n = 24) groups. Trending (*$p < 0.20$) and significant (**$p < 0.05$) values are indicated.

	b	Std. Error	β	95% Confidence interval for b	Semipartial correlation
A) Control group (Vascular Anatomy Study Aid)					
Step 1[a]					
(Constant)	5.09	2.62	–	(−0.40, 10.59)	–
Gender	1.84*	1.36	0.31*	(−1.01, 4.69)	0.29*
Studying Habits	−0.72	1.11	−0.15	(−3.05, 1.60)	−0.14
Gaming Habits	0.97*	0.71	0.30*	(−0.53, 2.47)	0.29*
Step 2[b]					
(Constant)	4.02	3.23	–	(−2.79, 10.82)	–
Gender	1.81	1.38	0.30	(−1.10, 4.72)	0.28
Studying Habits	−0.39	1.26	−0.08	(−3.05, 2.26)	−0.07
Gaming Habits	0.96	0.73	0.30	(−0.57, 2.50)	0.28
SASR	0.02	0.03	0.14	(−0.04, 0.07)	0.13
B) Experimental group (Vascular Invaders)					
Step 1[c]					
(Constant)	6.17	2.46	–	(1.04, 11.30)	–
Gender	−2.26*	1.44	−0.35*	(−5.25, 0.74)	−0.30*
Studying Habits	1.08	1.18	0.18	(−1.39, 3.55)	0.18
Gaming Habits	0.45	0.72	0.14	(−1.06, 1.95)	0.12
Step 2[d]					
(Constant)	5.00	2.34	–	(0.10, 9.91)	–
Gender	−1.72	1.35	−0.26	(−4.55, 1.12)	−0.23
Studying Habits	0.21	1.17	0.04	(−2.24, 2.66)	0.03
Gaming Habits	0.56	0.67	0.17	(−0.84, 1.96)	0.15
SASR	0.05**	0.02	0.41**	(0.00, 0.09)	0.37**

[a] $R^2 = 0.21$; adjusted $R^2 = 0.08$; ANOVA: $F = 1.59$; $p = 0.23$.
[b] $R^2 = 0.23$; adjusted $R^2 = 0.04$; ANOVA: $F = 1.23$; $p = 0.33$.
[c] $R^2 = 0.26$; adjusted $R^2 = 0.15$; ANOVA: $F = 2.40$; $p = 0.10$*.
[d] $R^2 = 0.40$; adjusted $R^2 = 0.28$; ANOVA: $F = 3.20$; $p = 0.04$**.

Additionally, we described in section 3.3 that studying habits affected use of the study aids, however they did not appear to have an overall effect on learning outcomes in either group (control: β = −0.08; experimental: β = 0.04), nor did gender or gaming habits (refer to Table 5 A and B).

4. Discussion

The results of this study suggest the following three findings with respect to the integration of game design elements in an e-learning environment:

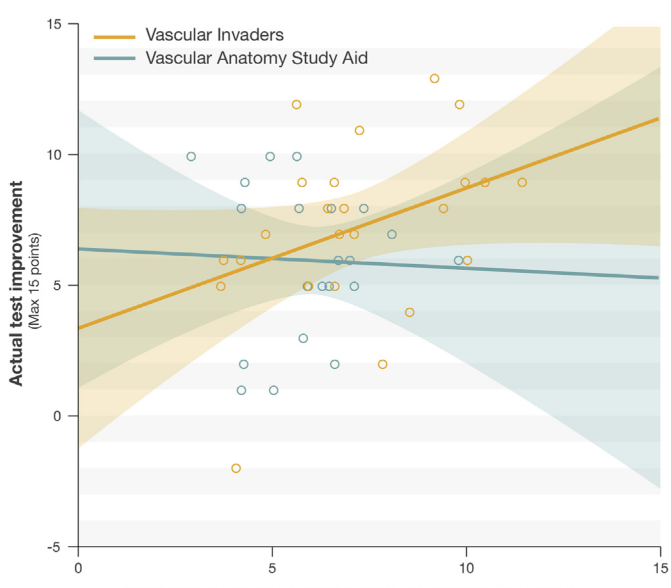

Fig. 5. This graph plots the predicted learning outcomes against actual learning outcomes from participants exposed to *Vascular Anatomy Study Aid* (control tool) or *Vascular Invaders* (experimental tool), after adjusting for gender, studying habits, and gaming habits. Mean 95% confidence intervals are indicated.

1. Game design may engage medical students in increased study aid-use.
2. Game design changes how students perceive the pedagogical value of the study aid.
3. Game design makes learning more predictable by enhancing the assessment capabilities of the evidence-centred design framework.

These findings are discussed below individually.

4.1. Game design may engage medical students in increased study aid-use

We hypothesized that students given access to *Vascular Invaders* would use their study aid significantly more than the control group. This was not the case overall, though the number of completed path-finding tasks and the study aid success rate trended toward significance in the experimental group. This similarity in use may be explained by the fact that the academic content in the two study aid treatments was identical and perhaps the extra motivational elements of game attributes were not required for voluntary use by medical students. It may also be that our sample size was inadequate to detect significance; a priori sample size calculations using G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2009) suggested a sample of 82 participants in order to reach a power of 0.8. Sample size as a limitation is further discussed in section 4.4.

However, while differences in mean study aid use were not significant between treatments, quantitative measures of study aid use, such as task completion, had a significantly greater amount of variance in the experimental group. In the experimental group, we observed some individuals completing as many as 43 tasks; since only ten unique, pre-defined vascular-sequencing tasks were available in the study aids, this suggests that some *Vascular Invaders* players were engaged enough by the gaming environment to repeat the available content up to four times. This is in comparison to a maximum of 17 tasks completed by control group participants.

We hypothesize that this is due to the presence of game design elements in *Vascular Invaders*. Looking at the relationship between game use (measured in completed tasks) and interaction with game elements—such as leaderboard views, Bloodbot friends collected (a mini-goal), and power-ups used—we see a steep, positive correlation. Participants continually accessed these game design elements as they progressed through the game, even though such actions were not at all necessary for successful task completion. This indicates that these interactions were integral to continued use and engagement in the gameplay experience and may be responsible for the marginally higher use seen in the game group.

4.2. Game design changes how students perceive the pedagogical value of the study aid

Analysing how personal characteristics (gender, studying habits, and game-playing habits) are related to tool-use suggests that the presence of game elements changes the way in which students perceive the value of the study aid to their education. Strong and *opposite* effects were found between reported studying habits and tool-use in the two groups. Control participants tended to complete significantly *fewer* tasks if they reported good, regimented studying habits, whereas experimental participants completed significantly *more* tasks if they reported good studying habits. The reasons for this are somewhat unclear. We can suggest that this may be because the game elements in *Vascular Invaders* engaged the "good studiers" long enough for them to discover the potential benefits of the study aid's use; perhaps control participants were not immediately cognisant of such benefits (equal to both study aids) and so "good studiers" sought other studying means without gaining an in-depth appreciation of the *Vascular Anatomy Study Aid*'s value. Alternatively, we discussed in section 1.1 that high-performing students often use self-testing techniques while studying. The point system and rules embedded in *Vascular Invaders* (e.g. scoring higher points for more efficient paths and losing energy for revealing hints) are regulatory self-testing mechanisms that may encourage students with good studying habits to engage with the tool. On the other hand, the *Vascular Anatomy Study Aid* does not include scoring feedback and automated self-testing mechanisms (i.e. the user can reveal hints without penalty). This may have deterred "good studiers" from engaging with this tool.

Gender did not have a significant impact on the amount of study-aid use contrary to much literature surrounding gender-based video game preferences (Brown, Hall, Holtzer, Brown, & Brown, 1997; Greenberg et al., 2010; Hartmann & Klimmt, 2006; Hoffman & Nadelson, 2009; Ogletree & Drake, 2007). Similarly, game-playing habits did not appear to affect the amount of tool-use in either group, contrary to our initial hypothesis. This result is somewhat surprising since the inclusion of game elements was intended to appeal to and engage regular game-players. It should be investigated whether or not students' preferences for commercial, entertainment-based video games actually transfers to educational games. Nonetheless, it is reassuring to note that students, regardless of video gaming habits, may be interested in using games to enhance their academic experience.

4.3. Game design makes learning more predictable by enhancing the assessment capabilities of the evidence-centred design framework

Our evidence-centred design framework allows us to compute study aid success rate (evidence model), from data collected during study aid task completion (task model), to predict learning outcomes (student model). Since both study aid treatments contained the same framework and academic content, we predicted that students' SASR would reflect learning equally in both *Vascular Invaders* and the *Vascular Anatomy Study Aid*. However, this was not the case; SASR was a significant predictor of learning outcomes in the experimental group but not in the control group (Table 5; Fig. 5). This may be explained by analysing metrics that describe how the participants interacted with the study aids. For instance, participants in the control group completed their tasks in significantly more moves than did participants in the experimental group. This is likely because the two study aid treatments encourage different kinds of interaction with their rules, or lack thereof. In *Vascular Invaders*, the player is constrained by an energy meter; moving against the blood flow or backtracking in one's path reduces energy, so the player must strategically plan their route to successfully arrive at their goal, usually resulting in shorter paths. Furthermore, the player is rewarded with greater point values for finding the end-point efficiently (i.e. in fewer moves). The *Vascular Anatomy Study Aid* lacks such constraints and incentives, so the user may move around the system freely, often resulting in more circuitous paths without an emphasis on completing the task. It is possible that the user may also find the end through random node selection since there are no negative consequences for such actions.

Since *Vascular Invaders* places greater emphasis and restrictions around efficient task completion, we can better predict how students will interact with the digital environment and, ultimately, how they are thinking about the material. This renders *Vascular Invaders* a better knowledge-assessment tool—and, possibly, learning tool—in comparison to the control treatment, even though the quality of the interventions is comparable as evidenced by similar use-statistics (i.e. tasks attempted and completed).

Test improvement also trended toward increased performance in the experimental group but, since we do not see a significant difference in test improvement between treatments, we cannot determine that strategic problem solving seen in *Vascular Invaders* is more beneficial for learning vascular anatomy than the more exploratory nature of the *Vascular Anatomy Study Aid*.

Medical students are recognized efficient, highly motivated learners (Amin, Tani, Eng, Samarasekara, & Huak, 2009; Dunnington, Markwell, & Dutta, 2001; Hojat & Zuckerman, 2008). This group was undoubtedly studying this material using several means—lecture notes, atlases, dissections, and prosections (Zurada et al., 2011)—in preparation for their exam. Since we aimed for high external validity by implementing the study aids over a long term and in the students' natural environment, we cannot include the influence of other studying methods in our analyses. However, we can suggest that, since the experimental model accounted for a much larger portion of the variance in learning, the majority of the learning recorded in the control group came from other sources (textbook review, dissection, etc.).

4.4. Limitations

This study presented three main limitations: 1) internal validity surrounding the relationship between study aid use and test scores; 2) design flaw rendering the revealing of 3D anatomical structures obsolete; and 3) sample size.

As discussed in section 4.1, the experiment was designed with high external validity but lower internal validity. We strived to keep the study authentic to a real-world scenario, offering students the study aids to use, as desired, over a 35-day period while learning about head and neck anatomy. While this is a great way to see trends and differences in voluntary use between treatments, it becomes difficult to understand what other factors (e.g. studying via other resources) may have lead to knowledge improvement. These other factors likely make up the variance that remains unaccounted for in our regression models (section 3.4). Since we use our regression statistics to claim that game interactions account for a certain proportion of learning, we do not mislead our readers into believing that the significant increases in learning are due solely to gameplay.

An information design feature of both the *Vascular Anatomy Study Aid* and *Vascular Invaders* was that 3D anatomical structures remained hidden until needed, allowing the student to exercise recall skills before seeking visual help. Revealing structures required the user to right-click on blood vessel nodes. Since a large proportion of participants did not use this function, we surmise that this was perhaps a difficult action to perform, especially for individuals on a laptop who did not have access to a mouse. This may in turn have rendered the studying tasks too difficult for some users and may have discouraged them from continued use of the study aids. Nonetheless, since we do not see differences in use of this function between groups, we can assume that both groups suffered equally from this design flaw and the validity of our results remain in tact.

Lastly, a third potential limitation concerns the small sample size used in this study. This may be explained by the timing of the study. Students were intentionally given access to the study aid for a 35-day period prior to their exam. While many students (75) used the aid throughout the term they failed to complete the post-test, scheduled in the 48 h preceding the exam (thus excluding them from inclusion in the study).

Despite the limitations of the present study, it suggests some interesting implications for the integration of gaming features in study-aids, and these findings are supported by the literature.

5. Conclusion

This study presented a game, *Vascular Invaders*, and a non-game-study aid, the *Vascular Anatomy Study Aid*, to medical anatomy students in order to explore how game design affects learning and voluntary study-aid use in an unstructured, informal context. Ultimately, the findings of this study suggest that adding game design to a vascular anatomy study aid may motivate increased voluntary use by medical students, though only trending differences were observed. The study also suggests that an individual's studying habits are likely to mediate use in opposite ways in the presence and absence of game design.

Perhaps the most pertinent finding was an unexpected one: study aid performance was a significant predictor of learning in the game group but not in the control group. The rules of the game treatment encouraged more specific, desired actions from the player, allowing for greater predictability and assessment of learning outcomes in comparison to the non-game. Future educational games should be created with this in mind; game mechanics should enhance evidence-centred design by eliciting actions from the player that require him or her to reflect on the target concepts so that patterns of interactions can be related directly to the players' knowledge. This represents a unique contribution of game design to a digital application.

In the case of this study, significant differences in test performance were not observed between experimental groups and it may be worth investigating whether or not the more exploratory environment of the non-game version impacted learning in a different way, possibly overlooked here. Future studies should consider expanding the study aids to include modules on the vascular anatomy of the entire body, as well as offering a wider array of carefully scaffolded tasks in each module.

Acknowledgements

This research is supported in part by grants SSHRC #SIG-13/14 from the Social Sciences and Humanities Research Council and the University of Toronto's Instructional Technology Innovation Fund (ITIF). The authors would also like to acknowledge Dr. Mike Wiley, University of Toronto, Division of Anatomy, for his collaboration in reviewing anatomical models, contributing to the development of pre-post-testing instruments, and providing access to first year medical students.

References

Amin, Z., Tani, M., Eng, K. H., Samarasekara, D. D., & Huak, C. Y. (2009). Motivation, study habits, and expectations of medical students in Singapore. *Medical Teacher, 31*(12), e560–e569. http://dx.doi.org/10.3109/01421590903193554.

Barab, S., & Dede, C. (2007). Games and immersive participatory simulations for science education: an emerging type of curricula. *Journal of Science Education and Technology, 16*(1), 1–3. http://dx.doi.org/10.1007/S10956-007-9043-9.

Barab, S., Gresalfi, M., & Ingram-Goble, A. (2010). Transformational play: using games to position person, content, and context. *Educational Researcher, 39*(7), 525–536. http://dx.doi.org/10.3102/0013189X10386593.

Brown, R. M., Hall, L. R., Holtzer, R., Brown, S. L., & Brown, N. L. (1997). *Gender and Video Game Performance, 1*(36), 793–812.

Charsky, D., & Ressler, W. (2011). Games are made for fun": lessons on the effects of concept maps in the classroom use of computer games. *Computers & Education, 56*(3), 604–615. http://dx.doi.org/10.1016/j.compedu.2010.10.001.

Dunnington, G., Markwell, S., & Dutta, S. (2001). An evaluation of study habits of third-year medical students in a surgical clerkship. *American Journal of Surgery, 181*, 268–271.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2009). *G*power*.

Gauthier, A., & Corrin, M. (2013). Exploring how the incorporation of video game design elements into an online thoracic vasculature study aid affects use patterns of undergraduate anatomy students. *Journal of Biocommunications, 39*(2), 50–56.

Gee, J. P. (2007). *What video games have to teach us about learning and literacy* (2nd ed.). New York, New York, USA: Palgrave MacMillan.

Greenberg, B. S., Sherry, J., Lachlan, K., Lucas, K., & Holmstrom, A. (2010). Orientations to video games among gender and age groups. *Simulation & Gaming, 41*(2), 238–259. Retrieved from http://myaccess.library.utoronto.ca/login?url=http://search.proquest.com/docview/742870627?accountid=14771 LA - English.

Halverson, R., & Owen, E. (2014). Game-Based Assessment: an integrated model for capturing evidence of learning in play [Special Issue: Game-Based Learning] *International Journal of Learning Technology, 9*(2), 111–138.

Hartmann, T., & Klimmt, C. (2006). Gender and computer games: exploring females' dislikes. *Journal of Computer-Mediated Communication, 11*(4), 910–931. http://dx.doi.org/10.1111/j.1083-6101.2006.00301.x.

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review, 19*(1), 126–134. http://dx.doi.org/10.3758/s13423-011-0181-y.

Hilbelink, A. J. (2009). A measure of the effectiveness of incorporating 3D human anatomy into an online undergraduate laboratory. *British Journal of Educational Technology, 40*(4), 664–672. http://dx.doi.org/10.1111/j.1467-8535.2008.00886.x.

Hmelo-Silver, C. E., & Azevedo, R. (2006). Understanding complex systems: some core challenges. *Journal of the Learning Sciences, 15*(1), 53–61. http://dx.doi.org/10.1207/s15327809jls1501_7.

Hoffman, B., & Nadelson, L. (2009). Motivational engagement and video gaming: a mixed methods study. *Educational Technology Research and Development, 58*(3), 245–270. http://dx.doi.org/10.1007/s11423-009-9134-9.

Hojat, M., & Zuckerman, M. (2008). Personality and specialty interest in medical students. *Medical Teacher, 30*(4), 400–406. http://dx.doi.org/10.1080/01421590802043835.

IBM Corporation. (2013). *SPSS statistics*. IBM Corporation.

Kafai, Y. B. (2009). World of whyville: an introduction to tween virtual life. *Games and Culture, 5*(1), 3–22. http://dx.doi.org/10.1177/1555412009351264.

Kafai, Y. B., Quintero, M., & Feldon, D. (2009). Investigating the why'' in whypox: casual and systematic explorations of a virtual epidemic. *Games and Culture, 5*(1), 116–135. http://dx.doi.org/10.1177/1555412009351265.

Klopfer, E., Sheldon, J., Perry, J., & Chen, V. H.-H. (2012). Ubiquitous games for learning (UbiqGames): weatherlings, a worked example. *Journal of Computer Assisted Learning, 28*(5), 465–476. http://dx.doi.org/10.1111/j.1365-2729.2011.00456.x.

Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory (Hove, England), 17*(5), 493–501. http://dx.doi.org/10.1080/09658210902832915.

Landers, R. N., & Callan, R. C. (2011). *Serious games and edutainment applications*. In M. Ma, A. Oikonomou, & L. C. Jain (Eds.) (pp. 399–423). http://dx.doi.org/10.1007/978-1-4471-2161-9.

Mayotte, S. (2010). Online assessment of problem solving skills. *Computers in Human Behavior, 26*(6), 1253–1258. http://dx.doi.org/10.1016/j.chb.2010.03.033.

McCarroll, M. L., Pohle-Krauza, R. J., & Martin, J. L. (2009). Active learning in the classroom: a muscle identification game in a kinesiology course. *Advances in Physiology Education, 33*(4), 319–322. http://dx.doi.org/10.1152/advan.00013.2009.

Mcintire, C. L. (1995). Corpus morph us anatomy. *The American Biology Teacher, 57*(8), 538–543.

Miller, L. M., Chang, C.-I., Wang, S., Beier, M. E., & Klisch, Y. (2011). Learning and motivational impacts of a multimedia science game. *Computers & Education, 57*(1), 1425–1433. http://dx.doi.org/10.1016/j.compedu.2011.01.016.

Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice, (Winter)*, 6–20.

National Research Council. (2011). In M. A. Honey, & M. Hilton (Eds.), *Learning science through computer games and simulations* (pp. 1–157). Washington, DC: National Academies Press.

Ogletree, S. M., & Drake, R. (2007). College students' video game participation and perceptions: gender differences and implications. *Sex Roles, 56*(7–8), 537–542. http://dx.doi.org/10.1007/s11199-007-9193-5.

Owen, V. E., Ramirez, D., Salmon, A., & Halverson, R. (2014). Capturing learner trajectories in educational games through ADAGE (assessment data aggregator for game environments. In *A click-stream data framework for assessment of learning in play. In American Educational Research Association Annual Meeting (pp. 1–7). Philadelphia, PA*.

Petersson, H., Sinkvist, D., & Wang, C. (2009). *Web-based interactive 3D visualization as a tool for improved anatomy learning, 68(April)* (pp. 61–68).

Reese, D. D., Tabachnick, B. G., & Kosko, R. E. (2014). Video game learning dynamics: actionable measures of multidimensional learning trajectories. *British Journal of Educational Technology*. http://dx.doi.org/10.1111/bjet.12128. n/a–n/a.

Scott, N. W., Hons, M. A., Sc, M., Mcpherson, G. C., Hons, B. S., Ramsay, C. R., et al. (2002). *The method of minimization for allocation to clinical trials: A review* (Vol. 23, pp. 662–674).

Shute, V. J., & Ventura, M. (2013). *Measuring and supporting learning in games: Stealth assessment*. Cambridge, MA: The MIT Press.

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

Squire, K. (2013). Video game – based learning: an emerging paradigm for instruction. *Performance Improvement Quarterly, 21*(2), 7–36. http://dx.doi.org/10.1002/piq.

Squire, K., DeVane, B., & Durga, S. (2008). Designing centers of expertise for academic learning through video games. *Theory into Practice, 47*(3), 240–251. http://dx.doi.org/10.1080/00405840802153973.

Stewart, D., Panus, P., Hagemeier, N., Thigpen, J., & Brooks, L. (2014). Pharmacy student self-testing as a predictor of examination performance. *American Journal of Pharmaceutical Education, 78*(2), 32. http://dx.doi.org/10.5688/ajpe78232.

Westwood, D., & Griffiths, M. D. (2010). *The Role of Structural Characteristics in Video-Game, 13*(5).

Zurada, A., Gielecki, J. S., Osman, N., Tubbs, R. S., Loukas, M., Zurada-Zielińska, A., et al. (2011). The study techniques of Asian, American, and European medical students during gross anatomy and neuroanatomy courses in Poland. *Surgical and Radiologic Anatomy: SRA, 33*(2), 161–169. http://dx.doi.org/10.1007/s00276-010-0721-6.