Editorial for EJEL Volume 15 Issue 4  
Special Issue on Games-Based Learning

The papers in this volume tackle a range of contemporary issues that reflect the varied interests of the games-based learning community. Topics include the learning and motivational features of games, characteristics of the players and the context in which games are played, as well as methodological issues relating to game evaluation. The papers also reflect the diverse content areas that digital games now address, such as games to teach computer programming to children, games to teach sign language and games to illustrate the interdependent nature of planning operations in a container terminal.

There is recognition in the literature on games for learning of the need to provide a strong evidence base showing that games are more effective than traditional methods of learning. Kamnardsiri et al report a carefully designed evaluation comparing their game for learning American Sign Language (Kinect) with the traditional face-to-face learning that is typically used in sign language skill training for students with hearing impairments. They found that, while both groups improved performance following training, the game group showed higher scores than the control post training.

An important question for games for learning concerns the kind of thinking that they can support. Games that can support higher level thinking such as problem solving are of particular interest. Rose, Habgood and Jay endorse Papert’s view that computational thinking is just as important for children in the 21st century as the 3 Rs – reading, writing and arithmetic. Rose at al. aimed to extend our understanding of how Visual Programming Tools can help to develop computational thinking in children. They found that children using a ScratchJr-like interface performed more program manipulation or ‘tinkering’ than those using a Lightbot style programming interface and that non-verbal reasoning predicted program manipulation, but only for those in the ScratchJr-like condition.

A number of papers addressed the need to consider individual differences between players in designing games. Recognising that different players may enjoy different kinds of games, Lukosch et al examined how gender and cultural differences impact on player experience and game performance while playing a complex problem solving game that aimed to illustrate the interdependent nature of planning operations in a container terminal. They found that females performed better than males on this task and explained this in terms of the superior abilities of females on planning and attention tasks. They also reported cultural differences in performance.

An important advantage offered by the interactivity of games concerns their capacity to adapt to characteristics of the learner. Gardner’s theory of Multiple Intelligences suggests that people exhibit multiple dimensions of intelligence, such as interpersonal, intrapersonal, bodily-kinesthetic and musical-rhythmic, as well as the more traditionally recognised visual-spatial, linguistic and logical-mathematical intelligences. Sajjadi, Vlieghe and Troyer aimed to map these different intelligences to the fundamental building blocks of games, i. e. game mechanics. They surveyed game experts’ self assessments of their preferred games, categorised these by game genre and game mechanics and linked these to the experts’ self evaluations on multiple intelligences. The authors suggest that these mappings illustrate which game mechanics suit which MI dimensions, and can provide guidelines in designing games for people exhibiting dominance for specific MI dimensions.

Lukosch et al.’s game included a collaborative element where success depended on working together. Collaboration was also examined by Scoular, Care and Awwal who described an Approach to Scoring Student Collaboration in Online Game Environments. Scoring is an important aspect of many games and careful consideration has to be given to how it is implemented. Scoring collaboration between players raises additional issues that do not arise with scoring individual players. Scoular et al. presented a generalised approach to scoring collaboration based on identifying common behaviours across 3 different games.

The immense popularity of entertainment games gave rise to the hope that the features of entertainment games that make them so engaging might also lend themselves to more effective learning. Mozellius, Fagerström and Söderquist examined how motivating factors and intrinsic integration of knowledge in
educational games might be related to players' perceived knowledge acquisition. This paper also addressed the interesting issue of tangential learning, where players find that they enjoy playing a game and are encouraged to follow up on this new interest out-with the gameplay. The key finding was that intrinsic integration of information in games was linked to knowledge acquisition and tangential learning.

The uptake of educational games in the classroom will be influenced by many factors but the Technology Acceptance Model (TAM) emphasises that the perceived ease of use and perceived usefulness of a technology will impact strongly on how well users accept that technology. Sánchez-Mena, Martí-Parreño and Aldás-Manzano used the TAM model to examine teachers’ acceptance of games in the classroom and they confirmed the links predicted by the TAM model. They also found that age moderated the effect of teachers’ perceived ease of use on perceived usefulness, suggesting that older teachers are less accepting of games as a means of learning.

Collectively these papers provide a snapshot of research in this fast developing area, adding to the evidence that games for learning can provide entertaining and enjoyable gameplay with effective methods of learning in many different content areas, both curricular and non-curricular. The papers also show that there are still many issues to address, especially with respect to how learning outcomes can be supported by different kinds of game mechanics.

Guest Editors:

Elizabeth Boyle and Thomas Connolly
University of the West of Scotland, UK
The Effectiveness of the Game-Based Learning System for the Improvement of American Sign Language using Kinect

Teerawat Kamnardsiri, Ler-on Hongsit, Pattaraporn Khuwuthyakorn and Noppon Wongta
College of Arts, Media and Technology, Chiang Mai University, Chiang Mai, Thailand
teerawat.k@cmu.ac.th
ler-on_h@cmu.ac.th
pattaraporn.khuwuth@cmu.ac.th
noppon.w@cmu.ac.th

Abstract: This paper investigated students’ achievement for learning American Sign Language (ASL), using two different methods. There were two groups of samples. The first experimental group (Group A) was the game-based learning for ASL, using Kinect. The second control learning group (Group B) was the traditional face-to-face learning method, generally used in sign language skill training for students with hearing impairments. This study was separated into two phases. In Phase I: 3D trajectory matching measurement algorithm, the Euclidean distance algorithm was employed to present the similarity between teacher and student datasets. Then, Phase II: Effectiveness of the Game-Based Learning system, showed the proposed framework of the game-based learning for sign language. Moreover, knowledge of sign language together with the corresponding actions were captured from three sign language experts using the knowledge engineering method. Then, the proposed game-based system would be analysed to provide students immediate feedbacks and suggestions based on the knowledge transfer from ASL experts. In the experiment, the students (N=31) were divided into two groups. The first group, Group A (N=17), learnt with the game-based learning while the second group, Group B (N=14), learnt with the traditional face to face learning method. The study result showed a significant difference (p<0.05) on the mean score of the post-tests for both Group A and Group B. It also presented that the game-based learning approach provided a better performance of ASL vocabularies than the traditional face-to-face learning approach. Finally, in the section of Discussion and Conclusion, the effectiveness and the future work opportunity of the proposed game-based learning system were discussed for the improvement of sign language actions.

Keywords: Sign Language; American Sign Language; Recognition System; Kinect; Expert System; Game-Based Learning; Knowledge Engineering;

1. Introduction

Nowadays, schools for the deaf in Thailand commonly use the traditional teaching method to provide education for deaf people or students with hearing impairments. These deaf schools use just black and white boards as a teaching tool to transfer knowledge. All communication learning is taught with an emphasis on learning development in Sign Language (SL) using lip-reading, writing, reading, listening, speaking and using SL to communicate with other people (Kamnardsiri et al., 2009).

Sign Language (or Signed Language, SL) is a visual language which is used to communicate with hearing-impaired people. There are two types of sign language: (1) signs and (2) finger spellings (Yang, 2014). As a result, hearing-impaired people can communicate with normal hearing people using sign language. Nonetheless, not everyone uses the same sign language for communication. In order to communicate with foreign people with hearing impairments, a standard sign language is required, for instance, American Sign Language (ASL) is the sign language for communication with hearing-impaired Americans. However, there is a lack of experts to teach American Sign Language a lack of budget to support deaf education for hearing-impaired students. Other learning alternatives are considered as possible ways to help hearing-impaired students learn Sign Languages best.

There have been developments on sensing technology which are employed in numerous kinds of research studies and also in sign language recognition including motion capture technology, data gloves, colour cameras, Microsoft’s Kinect sensor, etc. (Ren et al., 2013). Many studies have been conducted with different diverse goals, for instance, gesture recognition (Schlömer et al, 2008; Xu et al., 2012; Sreekanth et al., 2017), hand gesture recognition (Cheng et al., 2013; Cheng et al., 2016; Maqueda et al., 2016), physical rehabilitation (Chang et al., 2011; Callejas-Cuervo et al, 2016; Spasojević, 2017) and interactive displays (Morrison et al, 2005; Zhang et al., 2012; Nathan et al, 2016; Wang et al, 2016) etc. In addition, computer games, especially
serious and educational games, have provided significant successes for learning complex topics. Most Game-Based Learning (GBL) researchers emphasize on the effectiveness of learning such as motivation, learning achievements, strategies, engagement and behavioural patterns (Liang et al., 2010; Hamari et al., 2016; Hsieh et al., 2016; Tsai et al., 2016). There are numerous studies which are associated with effective Game-Based Learning; for example, Papastergiou investigated the learning effectiveness and motivational appeal of a computer game for the learning of Greek high school Computer Science (CS) curriculum (Papastergiou, 2009). Moreover, Cheng and Su developed a game-based learning system to improve self-efficacy for students’ learning. They mapped the course content into the game to provide scenarios for learning (Cheng & Su, 2012). Furthermore, Lester and his colleagues designed a game-based learning environment named “Crystal Island: Uncharted Discovery” for upper elementary science education (Lester et al., 2014). Additionally, Hsiao and Chen developed a gesture interactive game-based learning (GIGL) approach for preschool children (Hsiao & Chen, 2016). Despite the essential aspect of game-based learning approach, in pedagogy or educational research, there are still some problems, namely, a lack of software and application tools to improve students’ skills, especially in deaf schools, to learn Sign Languages.

In this paper, we investigated the effectiveness of the Game-Based Learning system for the improvement of American Sign Language (ASL) skills in students with hearing impairments using Microsoft’s Kinect sensor. Some Knowledge Management and Knowledge Engineering techniques were employed to capture knowledge from teachers (Schreiber et al., 1994). Moreover, the Euclidean Distance algorithm was used to compare the similarity of ASL 3D trajectory data between students and their teachers, and also to calculate the similarity score of each student for the feedback (Kamnardsiri et al., 2016a).

2. Background and related works

2.1 Sign Language

Sign Language (SL) is one of the languages used to communicate with hearing-impaired people. An Oxford dictionary defines Sign Language as “a system of communication using visual gestures and signs, as used by deaf people” (Oxford, 2017), whereas a Cambridge dictionary defines Sign Language as “a system of hand and body movements representing words, used by and to people who cannot hear or talk” (Cambridge, 2017). Yang classifies Sign Languages into two types of actions: (1) signs and (2) finger spellings. Signs are hand motion sequence movement and hand configurations (Yang, 2014) and finger spellings are codes to represent letters and numbers of the alphabet using standardised finger positions (The Deaf Society, 2015). Sign languages are applied in various countries but they are not a universal language. There are many different Sign Languages in the world; for instance, British Sign Language (BSL), Thai Sign Language (TSL), Japanese Sign Language (or Nihon Shuwa, JSL), Spanish Sign Language (Lengua de signos o señas española, or LSE) and America Sign Language (ASL) (Wfdeaf, 2015).

2.2 Sign Language Recognition

There are many research studies that apply Sign Language recognition to solve problems such as image transformation (affine transformation, rotation, scaling and illumination) and noise and temporal segmentation (Jangyodsuk et al., 2014). Nai and his colleagues proposed a set of fast-computable depth features to classify static hand posture from a single depth image. The proposed features were extracted from pixels on randomly positioned line segments and evaluated with American Sign Language (ASL) finger spelling dataset with two new hand posture datasets. The classification accuracy of the approach was comparable to state-of-the-art methods. Their results also showed that training and testing speed was much faster compared to other methods (Nai et al., 2017). Lim and his colleagues proposed a feature covariance matrix based serial particle filter to isolate sign language recognition system. The fusion of the median and mode filters was used to extract the foreground and perform hand detection, then the region around the tracked hands was extracted to generate the feature covariance matrix of the tracked hand gesture. The results indicated that the hand trajectories as obtained through the proposed serial of hand tracking were closer to the ground truth at an 87.33 percent recognition rate for the ASL (Lim et al., 2016). Yang and his colleagues applied a hierarchical conditional random field (CRF) for recognising hand movement and a Boost Map embedding method to verify the hand shapes in three-dimensional space with the Kinect sensor. The results showed that this method achieved the recognition rate of 90.4 percent from signed sentence data (Yang et al., 2009; Yang & Lee, 2010; Yang, 2014). Jangyodsuk and his colleagues employed dynamic time warping (DTW) for sign trajectory similarity and the histogram of oriented gradient (HoG) features for hand shape representation to develop sign language recognition system. The results revealed that this method could achieve accuracy rate at 82 percent.
in ranking signs in the ten matches (Jangyodsuk et al., 2014). Chai and his colleagues used 3D trajectory matching for sign language recognition system using Microsoft Kinect sensor to establish translation system. The system comprises two modes: (1) Translation Mode and (2) Communication Mode. This system focused on the communication between a general person and a hearing-impaired person that could be done via an avatar (Chai et al., 2013).

2.3 Expert System

Oxford dictionary defines an expert system (ES) as “A piece of software which uses databases of expert knowledge to offer advice or make decisions in such areas as medical diagnosis” (Oxford dictionary, 2017). In the field of computer science, an expert system is an artificial intelligent (AI). The basic idea of ES is to transfer human knowledge to a computer system. Knowledge information is then created inferences by many experts in the specific field to support users’ problem-solving activities (consider, decide and forecast). Expert systems, for example, DENDRAL expert system, was created to analyse mass spectra (Feigenbaum & Buchanan, 1993), while MYCIN was built to diagnose infectious blood diseases and recommend antibiotics (Shortliffe, 1976). DIPMETER was developed for an Advisor Analysis of Data Gathered During Oil Exploration (Hunt, 1985). CARCUSEUS was assembled for blood-borne infectious bacteria (Miller, 1984). Furthermore, R1 (later called XCON, for eXpert CONfigurer) expert system was established for an ordering process (Kraft, 1984). An expert system generally consists of four components which are (1) users, (2) user interface, (3) knowledge base and (4) inference engine. Figure 1 shows the components of a general expert system; first of all, the users are humans who need to solve problems in a specific field using the expert system. Secondly, the user interface is a display that allows users to interact with the system such as entering problems into the system. Then, the system calculates and shows answers on the display based on the knowledge including facts and rules which are related to the problems. Finally, the inference engine provides a set of reasoning methods from experts, sometimes called axiom (Luconi et al., 1986).

![Figure 1: Components of a general expert system, modified from Luconi et al. (1986).](image)

2.4 Knowledge Engineering

Knowledge engineering is the process of knowledge acquisition to generate knowledge-based systems from art into an engineering discipline. Studer and his colleagues defined three of Knowledge engineering frameworks which comprise: (1) CommonKADS, a further development of KADS (Schreiber et al., 1994) as the creation of a collection model that captures particular features of the knowledge-based systems to develop, (2) Model-based and Incremental Knowledge Engineering (MIKE) (Angele, 1996), a method for developing knowledge-based systems that include all steps of designing and implementing via specification from the initial elicitation and (3) PROTEGE-II (Tu et al., 1995) an approach aims to assist the knowledge-based systems development by reusing problem-solving method and ontologies (Studer et al., 1998).

2.5 Game-Based Learning System

Game-Based Learning (GBL) is one type of the games that focus on learning outcomes to combine an effectiveness of learning principles into game environments. The purpose of GBL is to improve self-confidence and problem-solving skills for learners (Liang, et al., 2010). Prensky and Prensky indicated that Game-Based Learning was “about fun and engagement and come together of serious learning as well as interactive entertainment into a newly-emerging and highly-exciting medium” (Prensky & Prensky, 2007). Liang and his colleagues had reviewed and analysed the GBL design process, they suggested eleven preliminary designs which comprises; (1) Constructing an exploratory learning environment, (2) Setting an explicit learning direction, (3) Screening the entire learning process, (4) Applying multimedia, (5) Placing educational challenges, (6) Applying cognitive apprenticeship, (7) Providing appropriate hints and feedback, (8) Ensuring accurate game-play mechanics, (9) Integrating community, (10) Giving learners an appropriate level of control and (11) Designing a user-friendly interface (Liang et al., 2010). Additionally, Garris and his colleagues designed
the Input-Process-Outcome (IPO) game-based learning model to improve a performance of children’s learning as shown in Figure 2 (Garris et al., 2002).

Some studies employed IPO to develop the system for learning, for example, Ghergulescu and Muntean proposed a Motivation Assessment-oriented Input-Process-Outcome (MotIPO) game model to design an educational game (Ghergulescu & Muntean, 2014). Yang and his colleagues, in addition, used the IPO model to improve the English learning of elementary school third graders (Yang et al., 2008).

Figure 2: The Input, Process and Outcome (IPO) game-based Learning model, modified from Garris et al. (2002).

Furthermore, Liang and his colleagues created four extra designs into the game-based learning environment which were (1) to provide opportunities for system thinking, (2) to apply negative game-play mechanics to stimulate motivation, (3) to give learners an alternative experience of role playing and (4) to construct balanced game rules and to incorporate resource exchange mechanisms (Liang et al., 2010). Moreover, Prensky and his colleagues indicated that a game is a form of “organised play” (Prensky, 2001). Kim suggested that “gamification can add an extra level of motivation and incentive to many higher education activities” (Kim, 2015). Heinich and his colleagues stated that a game is “an activity in which participants follow prescribed rules that differ from those of real life while striving to attain a challenging goal” (Heinich, 2002). Deterding and his colleagues addressed that games for education had a huge diversity of objectives for using and the GBL designed serious games which related learning objectives in a game universe with certain cognitive along with visual immersion and gameplay to gamification as “the use of game design elements in non-game contexts” (Deterding et al., 2011). Romero, furthermore, defined that gamification and also serious games pointed at supporting the learning objective through a positive learning as well as gaming experience using the game mechanics and rules such as competitive rules, a scoring system and learning through playing (Romero et al., 2012). Likewise, Romero specified that a methodology for GBL consisted of six-phase methodology containing learning objectives, learning-centered need analysis, game modalities, game mechanics and rules, learning assessment and feedback and gaming and learning experience. This methodology was called HEXA-GBL (Romero, 2015).

Figure 3: A methodology for Game-Based Learning (GBL) design and evaluation, modified from Romero (2015).
2.6 Kinect Characteristic

According to the detail of Kinect, Smisek and his colleagues described that Kinect could be used to recognise both 3D points and images using a depth camera and an RGB colour camera. Kinect is a device for measuring which includes: IR camera and IR projector. IR camera is used to decode the IR projection pattern to triangulate 3D scene. The specification of IR image is 1280 × 1024 for 57 × 45 degrees FOV, 6.1 mm focal length, 5.2 µm pixel size. RGB camera, on the other hand, offers medium quality images with 1280 × 1024 pixel for 63 × 50 degrees FOV, 2.9 mm focal length and 2.8 µm pixel size. Depth image (d) of Kinect is the primary raw output of the depth in the scene that returns inverse depth (d). The depth resolution of Kinect is 0.5m - 15m from a planar target and around 5’ of the image centre (Smisek et al., 2013).

2.7 3D Trajectory Matching

Several researchers have been studying the processing of motion recognition and classification using trajectory matching for many years. The Euclidean distance is one of the metric functions for calculating a distance that conforms to the metric properties: non-negativity, identity, symmetry and triangle inequality. Moreover, the Euclidean distance is more competitive than other approaches, especially when dealing with a larger data size. The Dynamic Time Warping (DTW) algorithm (Rabiner & Juang, 1993) is efficient for one-to-one matching with a low computational cost. However, the Continue Dynamic Time Warping (CDTW) algorithm solved an accuracy problem better than both first algorithms (Cassisi et al., 2012).

3. Methodology

This study was separated into two phases. In Phase I, 3D trajectory matching measurement, a suitable Threshold (Ts) value was obtained for the comparison between training dataset (teachers) and testing dataset (students). Then, in Phase II, the effectiveness of the Game-Based Learning System was calculated to compare between the game-based learning group and the traditional face-to-face learning group.

3.1 Phase I: 3D trajectory matching measurement algorithm

In this phase, we focused on checking the similarity between America Sign Language (ASL) vocabularies of training and testing datasets with the appropriate threshold (Ts) value. The Euclidean distance algorithm was used for similarity checking.

Data Collection

Participants: Three teachers (expert at ASL) from Anusansunthorn School for the Deaf, Chiang Mai, Thailand and ten students from Chiang Mai University were participants in this study.

Procedures: Data consisted of two sections (1) training data captured from three teachers with five ASL vocabularies (bear, butterfly, fish, goat and lion) and (2) testing data captured from ten students using the same set of the ASL vocabularies from the teachers.

Results

Training data: Five ASL vocabularies which were bear, butterfly, fish, goat and lion were collected from each teacher, using only the right hand. In order to check the similarity between training and testing datasets, the captured data from Kinect device were used to collect trajectory path of three teachers with five ASL vocabularies as shown in Figure 5.
Figure 5: Graph of 3D-Trajectory dataset captured from three teachers with five ASL vocabularies: (a) Bear, (b) Butterfly, (c) Fish, (d) Goat and (e) Lion.

Testing data: The Euclidean distance algorithm was employed to measure the similarity between training and testing datasets. Euclidean distance algorithm was given as shown in Equation 1 (Cassisi et al., 2012).

\[ d(T, S) = \sqrt{\sum_{k=1}^{n} (T_k - S_k)^2} \]  

where, \( k \) was time, \( d(T, S) \) was the distance between teachers \( T \) and students \( S \) datasets as given in Equation 2.

\[ d = \sqrt{(Tx - Sx)^2 + (Ty - Sy)^2 + (Tz - Sz)^2} \]  

And also, the similarity between teacher and student datasets is presented in Equation 3.

\[ S(T, S) = \frac{\sum_{k=1}^{L} d(T, S)_k}{L} \]  

where, \( S(T, S) \) was the similarity value between teacher \( T \) and student \( S \), \( k \) was time, \( d(T, S) \) was the distance between teacher \( T \) and student \( S \) and \( L \) was the length of data.

In order to give flexibility to the measurement of ASL similarity between teacher and student datasets, threshold \( Ts \) values were set at 0.05 and 0.10.

Table 1: The similarity results from ten students of five ASL vocabularies (Threshold =0.05).

<table>
<thead>
<tr>
<th>Signs</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>SD</th>
<th>Variance</th>
<th>Confidence Level (95.0%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bear</td>
<td>0.050</td>
<td>0.637</td>
<td>0.411</td>
<td>0.186</td>
<td>0.035</td>
<td>0.341</td>
</tr>
<tr>
<td>Butterfly</td>
<td>0.087</td>
<td>0.562</td>
<td>0.330</td>
<td>0.114</td>
<td>0.013</td>
<td>0.287</td>
</tr>
<tr>
<td>Fish</td>
<td>0.010</td>
<td>0.462</td>
<td>0.303</td>
<td>0.121</td>
<td>0.015</td>
<td>0.257</td>
</tr>
<tr>
<td>Goat</td>
<td>0.025</td>
<td>0.562</td>
<td>0.338</td>
<td>0.138</td>
<td>0.019</td>
<td>0.286</td>
</tr>
<tr>
<td>Lion</td>
<td>0.175</td>
<td>0.525</td>
<td>0.403</td>
<td>0.079</td>
<td>0.006</td>
<td>0.373</td>
</tr>
</tbody>
</table>

Note: The maximum similarity value is 1.000.

Table 2: The similarity results from ten students of five ASL vocabularies (Threshold =0.10).

<table>
<thead>
<tr>
<th>Signs</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>SD</th>
<th>Variance</th>
<th>Confidence Level (95.0%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bear</td>
<td>0.125</td>
<td>0.825</td>
<td>0.588</td>
<td>0.159</td>
<td>0.025</td>
<td>0.528</td>
</tr>
<tr>
<td>Butterfly</td>
<td>0.250</td>
<td>0.725</td>
<td>0.486</td>
<td>0.119</td>
<td>0.014</td>
<td>0.442</td>
</tr>
<tr>
<td>Fish</td>
<td>0.225</td>
<td>0.575</td>
<td>0.425</td>
<td>0.071</td>
<td>0.005</td>
<td>0.398</td>
</tr>
<tr>
<td>Goat</td>
<td>0.025</td>
<td>0.787</td>
<td>0.528</td>
<td>0.175</td>
<td>0.031</td>
<td>0.462</td>
</tr>
<tr>
<td>Lion</td>
<td>0.362</td>
<td>0.625</td>
<td>0.478</td>
<td>0.051</td>
<td>0.003</td>
<td>0.459</td>
</tr>
</tbody>
</table>

Note: The maximum similarity value is 1.0.
Table 3: Average score results of testing five ASL vocabularies from ten students.

<table>
<thead>
<tr>
<th>Student</th>
<th>Average score from three attempts (N=3) with Threshold (Ts = 0.10)</th>
<th>Total (50)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bear (10)</td>
<td>Butterfly (10)</td>
</tr>
<tr>
<td>1</td>
<td>3.33</td>
<td>2.67</td>
</tr>
<tr>
<td>2</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>1.33</td>
<td>2.00</td>
</tr>
<tr>
<td>4</td>
<td>1.33</td>
<td>1.33</td>
</tr>
<tr>
<td>5</td>
<td>2.67</td>
<td>0.00</td>
</tr>
<tr>
<td>6</td>
<td>5.33</td>
<td>1.33</td>
</tr>
<tr>
<td>7</td>
<td>3.33</td>
<td>0.00</td>
</tr>
<tr>
<td>8</td>
<td>5.33</td>
<td>0.00</td>
</tr>
<tr>
<td>9</td>
<td>6.00</td>
<td>0.00</td>
</tr>
<tr>
<td>10</td>
<td>3.33</td>
<td>2.67</td>
</tr>
</tbody>
</table>

Figure 6: Graph of Average score of testing with five ASL vocabularies comprises: Bear, Butterfly, Fish, Goat and Lion from ten students.

The result of this phase consisted of two measurements which were the similarity testing and the score testing. For the similarity testing, we employed the Euclidean distance algorithm to evaluate between teacher and student datasets with the threshold value (Ts=0.05 and Ts=0.10). Additionally, the score testing used IF-THEN rule for calculating the score and assess the student’s test. The results are presented in Table 1, Table2, Table 3 and Figure 6. The results demonstrated that the sign with the most similarity between two datasets was Bear (Mean = 0.59, SD = 0.156, Variance = 0.025 and Confidence Level (95.0%) = 0.056) with Threshold (Ts) = 0.10. For the score testing, moreover, the highest average scores from the student dataset were 11.30 and 10.67 out of the total score of 50.

3.2 Phase II. The Effectiveness of the Game-Based Learning

In this phase, we emphasised on the effectiveness of using the Game-Based system approach to compare with the traditional learning approach. The purpose of this phase was to collect pre-test, post-test data and then evaluate ASL. There were training and testing sections. To do this, ASL vocabularies were collected from teachers for training. Also students were tested with the same set of ASL vocabularies using the Game-Based System.

Figure 7 showed the Framework that we designed for the Intelligent Game-Based System for Learning Sign Language. The system comprised of (1) User Interaction, (2) Graphic User Interface (GUI), (3) Expert System (4)
Trajectory Dataset and (5) Domain Experts. The new game-based system was developed to support both normal hearing people and hearing-impaired people in learning American Sign Language (ASL). The system was able to show the percentage of the correct movement actions as well as that of the incorrect actions and also the action scores. Motor-skills (ASL vocabulary) of users could be improved by themselves.

Game-Based Learning System and Environment: The Game-Based System was developed by the Unity 3D-Game Engine and Kinect Xbox 360 from Microsoft Corporation. The Encode Json format (Peng et al., 2011) was employed to create the data position (X, Y, Z) from both teachers and students. The configuration of the system environment was designed and used in this Phase as shown in Figure 8.

![Framework of the Intelligent Game-Based System for Learning Sign Language](image1.png)

**Figure 7:** Framework of the Intelligent Game-Based System for Learning Sign Language modified from (Kamnardsiri et al., 2016b).

![Configuration of the Game-Based System Environment](image2.png)

**Figure 8:** Configuration of the Game-Based System Environment: Distance between Kinect Xbox 360 and standing point (Kamnardsiri et al., 2016).

**Procedures**

**Participants:** In this Phase, three teachers (ASL experts) from Anusansunthorn School for the Deaf, Chiang Mai, Thailand, along with 31 students from the Animation Department, College of Arts, Media and Technology, Chiang Mai University were participants.

**Data collection:** The experiment was conducted in April 2016. Data comprised of: training data and testing data. Training data were gathered from three teachers with 20 ASL vocabularies: animal, alligator, peacock, zebra, giraffe, cat, tiger, snake, dog, elephant, dolphin, deer, penguin, cow, lion, panda, bear, butterfly, fish.
and goat. Testing data were collected from 31 students with 5 ASL vocabularies including bear, butterfly, fish, goat and lion. Furthermore, the students were separated into two groups: (1) Group A using the game-based learning (N=17) and (2) Group B using the traditional face-to-face learning (N=14). Firstly, all students were given a 10-minute pre-test. Secondly, a 30-minute experiment was conducted by letting students in Group A use the game-based learning to practise ASL vocabularies skills. At the same time, students in Group B were taught by an instructor, using the traditional face-to-face learning. Finally, a 10-minute post-test was given to all students. The procedures of the experiment are shown in Figure 9.

**Figure 9**: The experimental procedures comprise: Group A: Experimental group (17 students) and Group B: Control group (14 students).

**Data analysis**

*Training data:* In this Phase, three sequences of 20 ASL vocabularies were collected from three teachers. Four points (the right hand, the left hand, the right shoulder and the left shoulder) were used to check the similarity between training and testing datasets.

*Testing data:* The Euclidean distance algorithm was used to measure the similarity between training and testing datasets (Cassisi et al., 2012). In order to give flexibility to the measurement of ASL similarity between teacher and student datasets, a threshold (Ts) value was set at 0.10.

**Statistical Measurement**

To measure potential initial differences of five ASL vocabularies between Group A and Group B, the IBM SPSS 17 Statistics software was used for analysis. For the first step, descriptive statistics such as mean, standard deviation (SD) and paired differences of five ASL vocabularies were analysed. The pre-test with Group A and Group B and the post-test with Group A and Group B were evaluated by using Univariate Analysis of Variance (ANOVA). Secondly, inference statistics for the measurement of the different teaching strategies with the 2 x 2 between-groups analysis of covariance (ANCOVA) was conducted to calculate the effectiveness of the students’ skills. The significance level was set at 0.05.

**Results**

*Demographic Characteristics:* Table 4 presents the demographic characteristics of student samples. The age of the majority of the students (54.8 percent) was 21 years old. The table also shows the ratio of the student genders which were male: female (52:48) and also the ratio of student groups which were the Game-based system for learning: the Control group learning (55:45).
Table 4: Demographic profiles of students.

<table>
<thead>
<tr>
<th>Demographic Profiles (N=31)</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>1</td>
<td>3.2</td>
</tr>
<tr>
<td>20</td>
<td>6</td>
<td>19.4</td>
</tr>
<tr>
<td>21</td>
<td>17</td>
<td>54.8</td>
</tr>
<tr>
<td>22</td>
<td>7</td>
<td>22.6</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>16</td>
<td>51.6</td>
</tr>
<tr>
<td>Female</td>
<td>15</td>
<td>48.4</td>
</tr>
<tr>
<td>Group of learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group A (Game-based system for learning)</td>
<td>17</td>
<td>54.8</td>
</tr>
<tr>
<td>Group B (Control group learning)</td>
<td>14</td>
<td>45.2</td>
</tr>
</tbody>
</table>

Descriptive statistics: The mean scores of the learning performance test and ANOVA test which were given to students in both groups (Group A and Group B) as Pre-test and Post-test.

The results included Bear, Lion, Goat, Butterfly and Fish signs. For the Pre-test, the students in Group A (the game-based learning) and the students in Group B (the traditional face-to-face learning) did not provide any significant difference in the mean score. For the Post-test, however, the average results of Group A were higher than those of Group B. According to the ANOVA test results of Group A and Group B, there was a significant difference (p < 0.05) as shown in Figure 10, for example, for the “Lion” sign: the Pre-test of Group A (N=17, Mean=20.38, SD=14.11), and of Group B (N=14, Mean=16.74, SD=10.87). The different results between both groups did not show a significant difference (F=0.626, Sig=0.435, R2=0.021). On the contrary, the Post-test of Group A (N=17, Mean=83.71, SD=4.25), and of Group B (N=14, Mean=78.70, SD=7.99), the results presenting the difference between both groups displayed a significant difference (F=4.992, Sig=0.033<0.05, R2=0.147) as shown in Table 5.

Inference Statistical Measurement: In order to measure the achievement of learning ASL, when using the game-based learning compared to the traditional face-to-face learning, the ANCOVA was used to assess the effectiveness of the student skills via Pre-test and Post-test scores. As shown in the results, the comparison of the mean differences between the Pre-test taken by Group A and Group B demonstrated that the means were not different. This meant that there was no significant difference in the scores for all ASL vocabularies. However, the comparison of the average difference of the Post-test taken by Group A and Group B showed that there was a difference of mean or a significant difference in the scores with the significant level at 5% (P-value < 0.05) as shown in Table 6.

![Figure 10: The comparison of Pre-test and Post-test of five vocabularies between Group A (Experimental group) and Group B (Control group).]
Table 5: Descriptive statistics and ANOVA test for post-test of five ASL vocabularies.

<table>
<thead>
<tr>
<th>ASL Sign</th>
<th>Post-test</th>
<th>Group A (N=17)</th>
<th>Group B (N=14)</th>
<th>F</th>
<th>Sig.</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bear</td>
<td>73.61</td>
<td>6.98</td>
<td>56.15</td>
<td>21.41</td>
<td>0.000</td>
<td>0.42</td>
</tr>
<tr>
<td>Lion</td>
<td>83.71</td>
<td>4.25</td>
<td>78.70</td>
<td>4.992</td>
<td>0.033</td>
<td>0.15</td>
</tr>
<tr>
<td>Goat</td>
<td>83.06</td>
<td>7.27</td>
<td>57.69</td>
<td>18.86</td>
<td>0.000</td>
<td>0.39</td>
</tr>
<tr>
<td>Butterfly</td>
<td>80.11</td>
<td>7.20</td>
<td>58.37</td>
<td>27.07</td>
<td>0.000</td>
<td>0.48</td>
</tr>
<tr>
<td>Fish</td>
<td>81.92</td>
<td>6.25</td>
<td>66.24</td>
<td>25.72</td>
<td>0.000</td>
<td>0.47</td>
</tr>
</tbody>
</table>

Note: Learning methods: Students in Group A used the game-based system learning while students in Group B used the traditional face-to-face learning. Level of significance was set at 0.05.

Table 6: The comparison of the mean differences between Group A (the game-based system learning) and Group B (the traditional face-to-face learning) of Post-test from testing five ASL vocabularies (N=31).

<table>
<thead>
<tr>
<th>ASL Sign</th>
<th>F Mean Difference</th>
<th>Std. Error</th>
<th>Sig. (P-value)</th>
<th>R² Squared</th>
<th>Confidence Interval (95%)</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bear</td>
<td>22.11</td>
<td>1.79+</td>
<td>3.82</td>
<td>0.000</td>
<td>0.44</td>
<td>10.14</td>
</tr>
<tr>
<td>Lion</td>
<td>4.82</td>
<td>5.06+</td>
<td>2.30</td>
<td>0.037</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>Goat</td>
<td>16.98</td>
<td>24.78+</td>
<td>6.01</td>
<td>0.000</td>
<td>0.40</td>
<td>12.46</td>
</tr>
<tr>
<td>Butterfly</td>
<td>26.80</td>
<td>21.99+</td>
<td>4.24</td>
<td>0.000</td>
<td>0.498</td>
<td>13.29</td>
</tr>
<tr>
<td>Fish</td>
<td>28.20</td>
<td>16.35+</td>
<td>3.07</td>
<td>0.000</td>
<td>0.50</td>
<td>10.04</td>
</tr>
</tbody>
</table>

Note: Learning methods: Students in Group A used the game-based system for learning while students in Group B used the traditional face-to-face learning. The sign * indicated significant level at 5% (P-value < 0.05).

4. Conclusions and Discussion

The study presented in this paper explored the effectiveness of the usability of the intelligent game-based system for learning sign language. First of all, for flow experience, it was concluded that: Phase I was about matching captured data from training (teacher) and testing (student) datasets, using the 3D trajectory matching measurement algorithm. This phase employed the Euclidean distance algorithm to measure the similarity between two datasets. In addition, Microsoft’s Kinect Xbox 360 device was used to collect 3D trajectory data with the threshold value ($Ts=0.10$). The data of both teachers and students were encoded to create the data position (X, Y, Z) using JSON format (Peng et al., 2011). The results in Phase I showed that the value suitable for checking the similarity between teacher and student data was $Ts=0.05$.

Secondly, in Phase II, the intelligent game-based system for learning sign language was developed by the Unity 3D-Game Engine with Microsoft’s Kinect device. Afterwards, the effectiveness of the game-based learning system was investigated to compare the learning performance between two experimental groups. Group A (17 students) learnt SL with the game-based system. Group B (14 students) learnt SL with the traditional face-to-face learning. The experimental results of Phase II demonstrated that the performance of Group A students was higher than that of Group B students as shown in Figure 10 and Table 5. There was a significant difference in the scores with (P-value < 0.05) as shown in Table 6.

According to the finding in the study, the benefits and limitations of using the intelligent game-based system to learn sign language are further discussed as follows. Firstly, the benefits of the system are the engagement of students to concentrate on ASL vocabularies and the enjoyment of the learning via the game. When the students attempted to gain victory on each vocabulary by achieving a higher score than the other group, the result was that the students had to use the motor skills to practise their movement. Furthermore, the system could offer a suggestion for the correct movement of each vocabulary. The system, moreover, was user-friendly in a sense that it was easy to use, configure and setup. In terms of the pedagogy, the proposed framework of intelligent game-based learning system was used to increase the performance of students to learn ASL vocabulary skills (actions or movements). Regarding the results from bar graph in Figure 10, it demonstrated that the scores of the post-test of each vocabulary were higher than the score of the pre-test. Hence, it implied that the system offered effectiveness for improving sign language skills of students. In addition, the students had the sense of control and concentration while playing the game, along with a good learning achievement as well.
Nevertheless, there were some limitations for this study: (1) The time limitation for the whole process of the proposed system, only six months, resulted in the small number of vocabularies used in the study. In real life situation, the quantity of vocabularies in the school curriculum is so much greater. The vocabularies used in this study were insufficient to be taught in different levels of the school and the system, therefore, the number of ASL vocabularies for the system could be increased, (2) The cost-effectiveness was directly linked to the increased number of ASL vocabularies. The 3D trajectory data from the teachers (experts) had to be collected, along with the suggestion sentences to correct the mistakes of the students’ movements. Consequently, the cost of production would be higher by direct variation of ASL vocabularies.

Finally, for future works, the larger sample size is recommended due to the small sample size in this study. Also, the suggestions and feedbacks from experts have to be considered and added to the game-based system for sign language learning. In addition, the proposed framework can be applied to the studies in other fields, for instance, in the medical field, a game-based system was used in studies helping brain-injured patients practise their movements (Betker et al., 2007; Lange et al., 2011; Molinari et al., 2016; Putnamet et al., 2016). In the sports field, additionally, game-based systems were applied to improve the performance of athletes (Kamnardsiri et al., 2015a; Kamnardsiri et al., 2015b; Eltoukhy et al., 2016). Hence, there are several opportunities to continue the research and/or to conduct new studies in this area.

Acknowledgements

This study was supported by the Embedded Systems & Mobile Application Laboratory, College of Arts, Media and Technology (CAMT), Chiang Mai University, Thailand. We would like to thank Ms. Apichaya Damrongkiattivej, Animation students of CAMT, Mrs. Preeyawan Pattanakul, Ms. Nattakan Choomjai and Ms. Suttinee Saengsaen from Anusansunthorn School for the Deaf in Chiang Mai, Thailand for their kind support in this study.

References

Angele, J., Fensel, D. and Studer, R., 1996. Domain and task modeling in MIKE. In Domain Knowledge for Interactive System Design (pp. 149-183). Springer US.


An Exploration of the Role of Visual Programming Tools in the Development of Young Children’s Computational Thinking

Simon P. Rose, M. P. Jacob Habgood and Tim Jay
Sheffield Hallam University, Sheffield, UK

Abstract: Programming tools are being used in education to teach computer science to children as young as 5 years old. This research aims to explore young children’s approaches to programming in two tools with contrasting programming interfaces, ScratchJr and Lightbot, and considers the impact of programming approaches on developing computational thinking. A study was conducted using two versions of a Lightbot-style game, either using a ScratchJr-like or Lightbot style programming interface. A test of non-verbal reasoning was used to perform a matched assignment of 40, 6 and 7-year-olds to the two conditions. Each child then played their version of the game for 30 minutes. The results showed that both groups had similar overall performance, but as expected, the children using the ScratchJr-like interface performed more program manipulation or ‘tinkering’. The most interesting finding was that non-verbal reasoning was a predictor of program manipulation, but only for the ScratchJr-like condition. Children approached the ScratchJr-like program differently depending on prior ability. More research is required to establish how children use programming tools and how these approaches influence computational thinking.

Keywords: Visual programming, Education, Computational thinking, K-12, Lightbot, Scratch

1. Introduction

This paper focuses on two tools used to teach programming to young children, ScratchJr (Flannery et al., 2013) and Lightbot (Lightbot Inc., 2016). Existing literature suggests that both these tools encourage computational thinking, yet there are clear theoretical contrasts in the type of programming interfaces that they use. We describe an exploratory study to investigate whether there was a difference in the way that young children use these tools and consider its relationship to developing computational thinking skills.

We live in a digital age where technology plays a key role in almost everything we do, making it increasingly important for us to understand how it works. Today’s children will go on to live a life dominated by computing, both in the home and at work (Barr and Stephenson, 2011). Computing education is receiving increasing attention in classrooms worldwide, with the aim of developing digital, media and information literacies. The need for children to be effective users of computational tools has led to the re-examination of the concept of ‘computational thinking’. Although the term was originally used by Papert (1980), Wing (2006) describes it as the problem-solving processes used by computer scientists. She stated that it should be taught as a basic skill across the school curriculum (2008). Since Wing reintroduced the concept of computational thinking; many researchers have attempted to clarify what it is and how we can teach it (e.g. Grover and Pea, 2013; Yadav, Hong and Stephenson, 2016).

Programming tools are seen as a means of developing computational thinking skills (e.g. Wilson and Moffat, 2010; Brennan and Resnick, 2012; Berland and Wilensky, 2015). This has led to the release of a variety of new tools, such as ScratchJr, Hopscotch and Kodable. Scratch remains the most widely-used of children’s programming tools. It takes inspiration from constructionism and the LOGO programming language (Papert, 1980). Constructionism is a pedagogical theory based on constructivism (Piaget, 1970), which makes specific use of the construction of artefacts as a basis for building knowledge. Papert theorised that by thinking about programming, learners would learn about the process of thinking, and he believed these skills would transfer to other contexts (1980). Scratch provides a constructionist learning environment through block-based programming, where learners combine instruction blocks to form programs (Resnick et al., 2009). Researchers have identified differing approaches when children program in Scratch (Meerbaum-Salant, Armoni and Ben-Ari, 2011).

Several countries have now introduced computer science into national curricula (Heintz, Mannila and Farnqvist, 2016), meaning that children as young as 5 years old are now learning basic programming skills. Whilst there is evidence to suggest that children can learn to program at this age (Bers, 2010; Fessakis, Gouli and Mavroudi, 2013), there is comparatively little empirical research on children’s use of programming tools under the age of 7. This is particularly important due to the cognitive developments that children undergo around this age (Manches and Plowman, 2015). Some researchers are concerned that younger children struggle to understand fundamental computer science concepts like abstraction (e.g. Armoni, 2012).

2. Computational thinking

Seymour Papert first described computational thinking as part of his research into how children develop procedural thinking through computer programming (1980). Wing sparked a renewed interest in the topic (2006), suggesting that “to reading, writing, and arithmetic, we should add computational thinking to every child’s analytical ability” (p. 33). Furthermore, Wing suggested that teaching computational thinking enables children to learn to think in an abstract and algorithmic manner relevant to many disciplines, including mathematics and science. She went on to define computational thinking as “solving problems, designing systems, and understanding human behaviour, by drawing on the concepts fundamental to computer science” (Wing, 2006, p. 33), but ten years later there is still no unanimous agreement on a definition (Garcia-Peñalvo, 2016; Weintrop et al., 2016).

There have been many efforts to clarify what is involved in computational thinking (e.g. Barr and Stephenson, 2011; Grover and Pea, 2013; Kalelioglu, Gulbahar and Kukul, 2016). There is a general agreement that it includes all the concepts that a computer scientist would typically use to solve computational problems (Riley and Hunt, 2014), but the list of concepts is up for debate. Table 1 shows the different concepts used in 7 existing definitions of computational thinking.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstraction</td>
<td>Abstracting and modularising</td>
<td>Abstraction and pattern generalisation</td>
<td>Abstraction</td>
<td>Abstraction</td>
<td>Abstraction</td>
<td>Abstraction</td>
</tr>
<tr>
<td>Algorithms and procedures</td>
<td>Sequences</td>
<td>Algorithmic notions of flow of control</td>
<td>Procedures and algorithms</td>
<td>Algorithms and procedures</td>
<td>Algorithms (including sequencing and flow of control)</td>
<td></td>
</tr>
<tr>
<td>Data collection, analysis and representation</td>
<td>Data</td>
<td>Symbol systems and representations</td>
<td>Data Representation</td>
<td>Data collection, analysis and representation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problem decomposition</td>
<td>Parallelism</td>
<td>Structured problem decomposition, iterative, recursive and parallel thinking</td>
<td>Decomposition</td>
<td>Decomposition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parallelisation</td>
<td>Testing and debugging</td>
<td>Debugging and systematic error detection</td>
<td></td>
<td>Analysis</td>
<td>Automation</td>
<td></td>
</tr>
<tr>
<td>Control structures and loops</td>
<td>Automation</td>
<td>Mathematical reasoning</td>
<td></td>
<td>Automation</td>
<td>Generalisation</td>
<td></td>
</tr>
<tr>
<td>Automation</td>
<td>Simulation</td>
<td>Systematic processing</td>
<td></td>
<td></td>
<td>Modelling and simulations</td>
<td></td>
</tr>
<tr>
<td>Simulation</td>
<td>Events</td>
<td>Efficiency and performance constraints</td>
<td>Systems</td>
<td>Conceptualising</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: The concepts included in existing definitions of computational thinking
For the purposes of this work, we have defined a working definition for computational thinking using the 7 most common concepts included in the definitions above:

- Abstraction and generalisation (removing the detail from a problem and formulating solutions in generic terms)
- Algorithms and procedures (using sequences of steps and rules to solve a problem)
- Data collection, analysis and representation (using and analysing data to help solve a problem)
- Decomposition (breaking a problem down into parts)
- Parallelism (having more than one thing happening at once)
- Debugging, testing and analysis (identifying, removing and fixing errors)
- Control structures (using conditional statements and loops)

This process helped to identify individual concepts and provided a deeper understanding of computational thinking. This is the definition of computational thinking used in the rest of the work and will be used to evaluate two programming tools for their potential to develop computational thinking skills.

3. Programming tools for young children

In the previous section, we defined a working set of computational thinking concepts. This section will analyse two programming tools designed for young children and evaluate their potential to develop computational thinking with respect to this set of concepts.

3.1 ScratchJr

Scratch is a block-based programming tool designed for children aged 8-16. It aims to “support self-directed learning through tinkering and collaboration” (Maloney, Resnick and Rusk, 2010, p. 2) and requires the application of computational thinking concepts (Resnick et al., 2009).

ScratchJr is a version of Scratch redesigned for younger children aged 5-7 (figure 1). It maintains the creative programming elements of Scratch, which allow children to easily create short stories and games. Characters can be added to a scene, and are given behaviours by combining instruction blocks. The interface is entirely symbolic and contains only a third of the original Scratch instruction set because young children can struggle with several levels of decomposition (Flannery et al., 2013). ScratchJr also executes instructions from left to right (the way that the English language is read) instead of the top to bottom approach used in Scratch. It has large buttons for touchscreen use, which apparently compounds difficulties that young children often have with mouse movement. The Cartesian coordinate system used in Scratch has been replaced by a natural coordinate system, and there is a grid that can be overlaid on top of the scene to help children calculate distance. Numerical parameter values have a maximum limit of 25, and children can execute individual instructions simply by pressing on them to help them explore what each instruction does. ScratchJr was developed using several age-appropriate design principles (Flannery et al., 2013). It makes it easy to get started but provides room to use more complex concepts (low floor and high ceiling), it allows many pathways and styles of exploration (wide walls), ideas can be incrementally developed through experimentation (tinkerability), the interface is friendly and playful (conviviality) and it can be used with a wide range of learning outcomes (classroom support).

![Figure 1: A scene from ScratchJr](image)
3.2 Lightbot

Lightbot is an educational puzzle game. The player must arrange a fixed set of block-based instructions in a finite program space that tell a robot what to do (figure 2). The goal is to program the robot to turn all the blue blocks in a level into illuminated yellow blocks. This is done by navigating the robot to a blue block and executing the light command. Players can decompose a level into different sections, which can then be solved one after the other until they have a complete solution. Some of the later levels can only be completed through the correct use of procedures and conditionals. For procedures, the player is given other program spaces below the main program that can be called using special instructions. Conditionals are implemented using a paint tool that colours the robot so that only instructions of that colour are executed. Gouws, Bradshaw and Wentworth (2013) suggest that Lightbot is useful for practising computational thinking. It concentrates on using computational thinking as a problem-solving process, and players are rewarded for producing optimised solutions.

3.2.1 Non-verbal reasoning

Successful Lightbot players can use mental transformations to predict the movement of the robot, recognise patterns from other levels and implement these patterns using known sequences of instructions (Gouws, Bradshaw and Wentworth, 2013). This is comparable to non-verbal reasoning, which is the ability to analyse information and solve problems using visual information. Non-verbal reasoning contains both abstract (or diagrammatic) and spatial reasoning, which includes spatial transformations, recognising visual sequences, and identifying relationships between shapes and patterns. Non-verbal reasoning is not reliant upon or limited by language ability, and research suggests that it can indicate mathematical ability in children (Halberda, Mazzocco and Feigenson, 2008).

Figure 2: A simple level from Lightbot

3.3 Comparison of the tools

These tools were analysed for their support of computational thinking using the definition in the previous section (table 2). From this, it is reasonable to conclude that both tools encourage computational thinking. They both use almost all the common computational thinking concepts identified in section 2. The only difference is that Lightbot doesn’t support parallelism.
Table 2: The computational thinking concepts used in ScratchJr and Lightbot

<table>
<thead>
<tr>
<th></th>
<th>ScratchJr</th>
<th>Lightbot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstraction and generalisation</td>
<td>• Understanding of the grid and character movement</td>
<td>• Understanding of the grid and robot movement</td>
</tr>
<tr>
<td></td>
<td>• Identifying common behaviours using instructions and instruction blocks</td>
<td>• Identifying common solutions to levels</td>
</tr>
<tr>
<td>Algorithms and procedures</td>
<td>• Sequencing instructions to create algorithms</td>
<td>• Sequencing instructions to create algorithms</td>
</tr>
<tr>
<td></td>
<td>• Using procedures to repeat common instructions</td>
<td>• Using procedures to repeat common instructions</td>
</tr>
<tr>
<td>Data collection, analysis and representation</td>
<td>• Counting movement needed using the grid</td>
<td>• Counting movement needed using the level grid</td>
</tr>
<tr>
<td>Decomposition</td>
<td>• Applying behaviours to different characters</td>
<td>• Breaking down and solving levels in parts</td>
</tr>
<tr>
<td></td>
<td>• Having multiple instruction blocks in one character</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Applying behaviours in steps</td>
<td></td>
</tr>
<tr>
<td>Parallelism</td>
<td>• Blocks of instructions are executed in parallel</td>
<td>• The instruction currently being executed is highlighted</td>
</tr>
<tr>
<td></td>
<td>• The instruction currently being executed is highlighted</td>
<td>• Programs can be re-run to check for errors</td>
</tr>
<tr>
<td>Debugging, testing and analysis</td>
<td>• Programs can be re-run to check for errors</td>
<td>• Instructions can be pressed individually to test what they do</td>
</tr>
<tr>
<td></td>
<td>• Instructions can be pressed individually to test what they do</td>
<td></td>
</tr>
<tr>
<td>Control structures</td>
<td>• Using blocks such as repeat and wait to control execution</td>
<td>• Using conditionals in later levels</td>
</tr>
<tr>
<td></td>
<td>• Looping instructions using procedures</td>
<td>• Looping instructions using procedures</td>
</tr>
</tbody>
</table>

Despite their similarities, there is a specific operational difference between the visual programming paradigms employed in ScratchJr and Lightbot. In ScratchJr, a limitless number of blocks can be added to the program space, these blocks are not executed unless they are linked to a trigger block or individually pressed to execute them. Whereas in Lightbot, the play button sequentially executes all the instructions included in the main program. Lightbot also limits how many instructions can be in the program depending on the current level. It is this operational difference which led us to explore how young children used these tools and whether they encouraged a fundamentally different programming approach.

4. Programming approaches

Turkle and Papert described two approaches to problem-solving. The first was an analytical top-down approach where solutions to problems are planned. The second was a bottom-up or “bricolage” approach, where solutions are attempted “by arranging and rearranging, by negotiating and renegotiating with a set of well-known materials” (1991, p. 136). In constructivist learning theory, a child builds knowledge through experience. The information they receive through interactions challenges their world view (Piaget, 1970). Constructionism applies this theory to the construction of artefacts (Papert, 1980). It is a pedagogical theory which suggests that learners should be given the opportunity to experiment and explore ideas by tinkering with an artefact. Learners are guided “by the work as it proceeds rather than staying with a pre-established plan” (Papert and Harel, 1991, p. 6), leading to self-directed learning. Scratch is based on these principles (Resnick et al., 2009).

Research has shown that children aged 10-15 can learn computer science using Scratch (Meerbaum-Salant, Armoni and Ben-Ari, 2013; Sáez López, González and Cano, 2016). Despite this, there are some suggestions that Scratch may encourage unusual programming approaches (Meerbaum-Salant, Armoni and Ben-Ari, 2011). A top-down approach is traditionally taught in programming, where software is decomposed into coherent units that can be better maintained. Meerbaum-Salant, Armoni and Ben-Ari observed that 14 and 15-year-olds took the top-down approach to the extreme. They decomposed programs into many small blocks of instructions (sometimes hundreds) that lacked logical coherency. This can make programs particularly difficult to debug in Scratch and ScratchJr due to the way they both execute all instruction blocks in parallel. Children in the study by Meerbaum-Salant, Armoni and Ben-Ari (2011) became frustrated and lost motivation because
they did not understand what was happening in their programs. They also observed that Scratch programs were often developed using a bottom-up approach. In bottom-up programming, components are designed in isolation then linked together to form a complete solution. This can be an appropriate method of software design, but children once again took it to the extreme. When faced with a problem, they would attempt to solve it by "dragging all the blocks that seemed to be appropriate for solving the task, and then combining them into a script" (2011, p. 169). This tinkering behaviour is encouraged in Scratch and ScratchJr by the fact that instructions can be left in the script area without affecting the execution of the program.

We have identified two approaches to programming; top-down and bottom-up. Along with indications that both are used by children in Scratch. In contrast, Lightbot provides a programming interface which doesn’t allow much tinkering. Instructions can be freely added to and deleted from the main program, but when an instruction is visible, it is always part of the program and executed in strict sequence. It was this central difference that provided the basis for this study, exploring the affect that the two programming paradigms had on children’s approaches to programming. Lightbot contains only a subset of the commands available in ScratchJr, so it was decided that the programming tasks used for the study should be based on navigating robots (as in Lightbot), as such tasks could be easily undertaken in both programming environments.

5. Method

5.1 Aims and hypotheses

This was an explorative study examining young children’s approaches to programming using the two different programming paradigms. Although some hypotheses were formed, the study was primarily undertaken to identify questions that could become the focus of future research.

Three hypotheses were initially formed based on the existing literature:

a) A ScratchJr-like programming interface would lead to more “tinkering” than a Lightbot interface.

b) A ScratchJr-like programming interface would lead to improved outcomes on problem-solving tasks.

c) Higher-ability players would benefit more from a ScratchJr-like programming interface.

5.2 Participants

The participants were from a large primary school in a low-income area in northern England. Most pupils at the school are of White British heritage. The school has a well above average proportion of disadvantaged pupils and pupils that require support for special educational need. The study participants included 20 boys and 20 girls between the age of 6 years, 3 months and 7 years, 3 months ($M = 6$ years, 9 months).

5.3 Materials and procedure

A non-verbal reasoning test was created for this study to produce matching pairs, based on the assumption that non-verbal reasoning is required in Lightbot (section 3.2.1). Standardised school worksheets (Primary Leap Limited, 2011) were used as a model for the questions. There were three types of questions; matching shapes (as seen in figure 3), selecting the odd one out from a series of shapes, and selecting the next shape or missing shape in a pattern. The possible answers to some questions were rotated, requiring the participant to perform mental transformations, similar to the rotation process required in Lightbot.

The test took place in the school IT suite in groups of 15. It was 40 questions long, and the participants had 5 minutes to answer as many as they could. Participants were told there was no rush to answer the questions, and that their answers should be carefully thought through. The time-limit and number of questions aimed to produce a greater range of test scores, reducing the possibility of ceiling effects.
Two versions of a Lightbot style game for 6 and 7-year-olds were created for this study. One that used the Lightbot programming interface, and one that used a ScratchJr-like interface (see figure 4). The versions were identical apart from that in the ScratchJr-like version, instructions can be added to the program that will not execute unless linked to the trigger block (see table 3). The game has 15 levels; it begins with simple levels that require only forward and light instructions. The later levels then introduce more complex movements and levels with several lights. The difficulty progression was designed so that it could challenge more able children in the target age group.

Two groups of 20 participants were created using the non-verbal reasoning scores as a matching variable (based on the assumption that non-verbal reasoning ability was required to be successful in the game). Each child then played one version of the game for thirty minutes in a small reading room joined to the children’s classroom. Two laptops were set up facing away from each other so that one child from each group could play the game without being aware that they were using a different version to their classmate. Testing the conditions together meant that any extraneous variables (e.g. time of day) would affect both groups equally. All participants were given a uniform introduction to the game via a tutorial video.

5.4 Measures

A range of measures were used to explore how the participants used each version of the game:

1. The non-verbal reasoning scores for each participant.
2. Program manipulation; additions, moves and deletions of instructions per attempt.
3. The number of attempts needed by a participant to complete a level.
4. The highest level reached by each participant.
5. The time taken by a participant per attempt.
6. The time taken by a participant to complete a level.

As explained in section 5.3, the scores of a non-verbal reasoning test were used as a matching variable to create two even ability groups. It is, therefore, expected that these scores should predict how well a participant performed in the game and this would be demonstrated by a correlation between the participant test scores and the highest level they reached.

Program manipulation was measured by the number of additions, moves and deletions of instructions from the program space between each attempt. An attempt was defined as each time a participant ran their program by pressing the play button. It was hoped that this measure, and the amount of time between each attempt, would provide some indication of how participants are interacting with the game and could be used to infer something about their programming approach (particularly in terms of the amount of ‘tinkering’ taking place).

Overall performance was measured using the highest level reached by each participant. This can be combined with program manipulation to explore the effect of the two conditions on overall performance. Data was also collected on other performance measures, such as how many attempts it took for participants to complete a level and how much time it took them to do so. These measures can be used to show if the two games were as similar as expected.

6. Results

6.1 Comparing the difficulty of the two versions

The programming interface should have been the only difference between the two versions of the game. Independent T-tests were performed on the overall performance measures between groups (see table 4). On average, Lightbot players reached a slightly higher level in 30 minutes than the ScratchJr-like players, but this difference was not significant, \( t(38) = .54, p = .59 \). ScratchJr-like players spent slightly more time (in seconds) on each level than the Lightbot players, again the difference was not significant, \( t(38) = -1.12, p = .27 \). Finally, ScratchJr-like players took fewer attempts on average to complete levels than Lightbot players. This difference was not significant, \( t(38) = .98, p = .33 \).

<table>
<thead>
<tr>
<th>Measure</th>
<th>Lightbot (N = 20)</th>
<th>ScratchJr-like (N = 20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest level reached</td>
<td>10.25 4.25 .95</td>
<td>9.5 4.48 1</td>
</tr>
<tr>
<td>Average time taken to complete a level (seconds)</td>
<td>187 86.1 19.25</td>
<td>224.35 121.33 27.13</td>
</tr>
<tr>
<td>Average attempts needed to complete a level</td>
<td>9.4 6.86 1.53</td>
<td>7.54 5.02 1.12</td>
</tr>
</tbody>
</table>

6.2 Non-verbal reasoning as a predictor of game performance

The non-verbal reasoning test was designed to produce a range of scores, containing 40 questions and using a 5-minute time-limit. The median of the collected scores was 23.5 with a minimum of 15 and a maximum of 35. The middle 50% of scores were between 20.25 and 30.75. The groups were created based on the assumption that the scores would indicate how well participants would perform in the game. This was supported by a strong correlation between the scores and the highest level reached by each participant, \( r(40) = .73, p < .001 \). The game was designed to challenge the more able participants, but not be too difficult for the lower-ability participants. The median highest level that participants reached was 8.5, with a minimum of 4 and a maximum of 15 (the last level). The middle 50% fell between 6 and 15, with 13 participants reaching or completing the last level.

6.3 Comparing program interaction between groups

Participant’s program manipulation was compared to see if there was any difference in how they were interacting with the game. A one-way ANCOVA was used to compare the average instruction additions, moves and deletions per attempt from the program. The non-verbal reasoning scores were used as a covariate because participant ability may have influenced the amount of program manipulation they performed. There was a significant difference between the conditions, \( F(1,37) = 192.19, p < .001 \). Participants in the ScratchJr-
like condition manipulated the program more ($M = 9.06$, $SD = 3.27$, $SE = .73$) than the participants in the Lightbot condition ($M = 4.68$, $SD = 1.9$, $SE = .42$). A similar one-way ANCOVA was used to test for differences in the time (in seconds) taken to formulate an attempt. This also showed a significant difference between conditions, $F(1,37) = 9.58$, $p = .004$. The participants in the ScratchJr-like condition took longer on average ($M = 34.15$, $SD = 11.38$) to construct their programs than the Lightbot group ($M = 24.29$, $SD = 8.87$).

6.4 Using non-verbal reasoning to predict program interaction

We then tested if non-verbal reasoning scores would indicate how much a participant was manipulating their program. A single linear regression was used to predict the average program manipulation per attempt based on the test score. The results of the regression show that overall the scores significantly predicted program manipulation, $\beta = .37$, $t(37) = 3.32$, $p = .002$. Participants with higher non-verbal reasoning scores performed more manipulation. The tests scores also explained a significant proportion of variance in program manipulation, $R^2$ of $.55$, $F(2,37) = 22.46$, $p < .001$. An interaction effect was added to test if there was a difference between groups. This showed that in the ScratchJr-like version, there was a bigger effect of the condition when participants had higher scores of non-verbal reasoning, $\beta = .35$, $t(36) = -2.35$, $p = .025$. This result was supported by a correlation between non-verbal reasoning and program manipulation in the ScratchJr-like condition, $r(20) = .65$, $p = .002$, but not in the Lightbot condition, $r(20) = .22$, $p = .342$ (figure 5).

Figure 5: The correlations between non-verbal reasoning and average program manipulation per attempt for each condition

6.5 Using the highest level reached to predict program interaction

A similar analysis was then conducted using the highest level reached by each participant as an indicator of in-game success, instead of their non-verbal reasoning scores. Participants were divided according to whether they had completed level 8 or not, as this represented a median split. Using the non-verbal reasoning scores as a covariate, a two-way ANCOVA was conducted that examined the effect of the interface-type, and whether a participant completed level 8, on program manipulation. The results showed a significant interaction between the interface-type and program manipulation, $F(1,35) = 45$, $p < .001$, and a significant interaction between the interface-type and whether the participant completed level 8, $F(1,35) = 10.16$, $p = .003$. Completing level 8 alone was not a predictor of program manipulation, $F(1,35) = 1.77$, $p = .192$. The average program manipulation per attempt for both groups is shown in table 5. This is further supported by the graph in figure 6, which shows the average amount of program manipulation per attempt for each level.
Table 5: The average program manipulation per attempt dependant on whether a participant completed level 8.

<table>
<thead>
<tr>
<th>Level Completed</th>
<th>Lightbot (N = 20)</th>
<th>ScratchJr-like (N = 20)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
</tr>
<tr>
<td>&lt; 8</td>
<td>9</td>
<td>4.47</td>
</tr>
<tr>
<td>&gt;= 8</td>
<td>11</td>
<td>4.84</td>
</tr>
</tbody>
</table>

7. Discussion

The results showed that overall performance was similar in both versions of the game. Participants reached a similar level, spent a similar amount of time on each level and took a similar number of attempts to complete each level. This is not surprising given that the programming interface was the only difference between the two versions. We can also say that the game provided a suitable level of challenge, as all participants completed at least the third level of the game and around a third of participants reached or completed the last level.

The non-verbal reasoning test produced a good range of scores (between 15 and 35), suggesting that the difficulty of the questions was appropriate for the age group. The correlation between these scores and the highest level reached shows that non-verbal reasoning was a strong indicator of success in the game, and justifies the use of these scores as a matching variable to create two even ability groups. Furthermore, if both Lightbot and ScratchJr encourage computational thinking as has been suggested (Flannery et al., 2013; Gouws, Bradshaw and Wentworth, 2013) then this correlation could also support the idea that non-verbal reasoning and computational thinking are linked.

There was a pronounced difference in the amount of program manipulation between groups as measured by additions, moves and deletions of instructions. Participants in the ScratchJr-like condition performed 1.9 times more manipulation per attempt than the participants in the Lightbot condition. They also took 1.4 times longer on average to formulate each attempt. These findings are in line with the constructionist principles of Scratch’s design and consistent with the idea that children using the ScratchJr-like interface are being guided “by the work as it proceeds rather than staying with a pre-established plan” (Papert and Harel, 1991, p. 6). The increased tinkering in the ScratchJr-like condition could also suggest a bottom-up, or bricolage, approach to programming.

The role of prior ability in the level of program manipulation was a particularly interesting finding of this study. Non-verbal reasoning test scores were used as an indicator of participant ability and our analysis showed that these were a strong predictor of program manipulation overall. The higher their score of non-verbal reasoning, the more program manipulation a participant performed, but interestingly the effect was only significant for the ScratchJr-like condition. This indicates a contrast between how lower and higher ability players approached tasks in the ScratchJr-like condition. High-ability players performed more tinkering than their low-ability counterparts, suggesting that they were more suited to the free-design approach of ScratchJr-like.
instructions. On the other hand, Lightbot players all manipulated their programs roughly the same amount. This is an interesting finding given the similar overall performance of participants in both groups.

This finding was mirrored by using the highest level reached as an indicator of participant ability instead of non-verbal reasoning scores. Around half the participants progressed beyond level 8, so this point was used to group participants according to their success within the game. The analysis showed that there was a significant difference in manipulation within the ScratchJr-like group, with the more successful participants performing more manipulation on average throughout the game. Whereas the Lightbot participants altered their programs a similar amount per attempt no matter how successful they were within the game. This is further supported by the level-by-level data, which shows an increase in manipulation in the ScratchJr-like condition above level 8 (figure 6).

ScratchJr was designed with a low floor and (appropriately) high ceiling (Flannery et al., 2013). Yet we have found indications that the ability (measured using non-verbal reasoning and game performance) of a child can affect how they interact with a ScratchJr-like programming interface. More-able children tinkered 1.6 times more with their ScratchJr-like programs than less-able children using the ScratchJr interface, while children using the Lightbot interface used the same amount of manipulation throughout the game. Of course, there are many possible explanations for this. Lightbot users may have used less manipulation because programs were faster to create, making a trial and error approach efficient, even as the levels got more difficult. The apparent consistency in strategy could also suggest that the Lightbot interface naturally helped players to decompose levels into smaller sections where instructions could be added incrementally to their programs. Nonetheless, the finding that higher-ability ScratchJr players performed more tinkering is very intriguing and it would be natural to consider whether their approach allowed them to develop a deeper understanding of the game’s abstract concepts and design solutions that more accurately predicted what the robot needed to do. It may suggest that lower ability children need more support when using block-based programming tools like ScratchJr. Their lack of tinkering may be because of underdeveloped working memory and the cognitive load of the task (Sweller, 1988), leading us to question how low the floor should be in low floor and high ceiling design for young children. It also poses several possible questions about how programming tools are used in education; do less-able children get the support they require using these tools to meet learning outcomes? Is this influenced by the teacher’s knowledge of the programming tool? And can cognitive load be reduced by teaching the skills required by these tools individually?

8. Limitations

This study was intended to be exploratory in nature, and clearly, the design has limitations which should be acknowledged. Post-hoc analyses are appropriate to exploratory work, and useful for generating new hypotheses, but limit the validity of the findings. Although a matched design was used, the matching variable would normally have been the same as the dependent variable of the study, rather than a separate measure. Future studies will use a pre-to post-test design based on a common measure of computational thinking to address this, but developing an instrument which reliably measures computational thinking is a non-trivial task (Jenson and Droumeva, 2016).

The software itself had some limitations which could have affected the outcome of the study. It was observed that many participants completed levels using non-optimal solutions. This included over compensating for turns and having to rotate back the other way or having instructions in a direction where the robot could not go (shown by the robot ‘shaking its head’ when the block was executed). Arguably the Lightbot version of the experimental software should have had finite program space per level to contrast with the bottom-up approach employed by ScratchJr. This would have required participants to produce optimal solutions consistent with the original Lightbot game. This could potentially have increased the difference between versions as ScratchJr programming doesn’t have this restriction.

9. Conclusion

Using an exploratory approach, we aimed to establish how children used two different programming tools. We found indications that children aged six and seven interacted differently with the ScratchJr-like programming interface compared to the Lightbot interface. Children using the ScratchJr-like interface performed more program manipulation, which could indicate a more bottom-up programming approach. These findings are in line with the constructionist foundations of Scratch and ScratchJr. A more surprising finding was that more
able children (as measured by non-verbal reasoning skill or game performance), manipulated their programs more. This difference was only found in players using the ScratchJr-like interface, whereas ability had no effect on the program manipulation of Lightbot players.

This exploratory work offers some potential explanations for these findings, but more research is clearly required to establish how young children use programming tools and how they influence their development of problem-solving and computational thinking abilities.

This paper also examined the existing definitions (and expectations) of computational thinking, and we would share the concerns of other authors that these may be too broad (Weintrop et al., 2016). We propose that future research in this area focuses on the individual concepts involved in computational thinking. Investigating whether programming tools can be used to develop concepts such as decomposition, abstraction and algorithmic thinking. This would seem particularly important given concerns that young children may struggle to understand these concepts (Armoni, 2012; Manches and Plowman, 2015) despite the pressure on schools to teach them.

References


Gender and Cultural Differences in Game-Based Learning Experiences

Heide Lukosch¹, Shalini Kurapati¹, Daan Groen² and Alexander Verbraeck¹
¹Delft University of Technology, The Netherlands
²InThere, The Hague, The Netherlands
h.k.lukosch@tudelft.nl
s.kurapati@tudelft.nl
daan@inthere.nl
a.verbraeck@tudelft.nl

Abstract: Games have been successfully used in educational settings for many years. Still, it is not known in detail which factors influence the use and effectiveness of educational games. The game environment, its technology, and other game mechanics are factors directly linked to the game itself. The player’s experience with the subject of the game and/or games in general, his or her motivation and expectations towards the gaming experience influence the outcome of a game-based learning experience. Some of the personal aspects, like age, were already addressed in earlier research. Cultural and gender differences though, were not a main object of study in educational gaming so far. This study started from certain assumptions about differences in game play, related to players’ cultural backgrounds and gender. Literature suggests that gender plays a role when it comes to game performance. This paper introduces outcomes of a study with a so-called Microgame, a brief game used to raise the awareness of interdependent planning operations. It shows that in this game, gender and culture make a difference in relation to the learning experience of the players, measured by game performance.

Keywords: Microgames, learning, gender, culture

1. Introduction

Games are played by the young and old, males and females, and across the whole world. People play violent games, sports games, puzzle games, and action games. Games help players think, force audiences to be active, are social, and engage the body (Shaw, 2010). Since games have become an accepted instrument in educational settings, researchers in the field of game-based learning are constantly in search of factors that affect player motivation and acceptance, as well as challenges of the usage of games for learning (Kirriemuir & McFarlane, 2004). Several studies have identified an influence of social factors such as gender, cultural identity, or ethnicity on the acceptance and performance of digital games used for educational purposes (Raessens & Goldstein, 2011; Yee, 2006). However, empirical studies that confirm these findings are still quite scarce (De Freitas, 2006).

In the gender field, for example, many studies try to investigate the preferences of male and female players for certain game types (Gros, 2007; Klimt & Hartmann, 2006; Greenberg, Sherry, Lachlan, Lucas, & Holmstrom, 2010), or the representation of gender within video games (Dietz, 1998; Cassell, & Jenkins, 2000; Beasley & Collins Standley 2002; Williams, Martins, Consalvo, & Ivory, 2009), while the effect of gender, or the feminine culture, on game performance was poorly researched so far. Most studies show better performance by male players, mainly explained by their more extensive media use, or better computer skills of male players (Imhof, Vollmeyer, & Beierlein, 2007; Kafai, Heeter, Denner, & Sun, 2008). At least, most studies of gender and video games take it for granted that ‘girls’ and ‘boys’ play differently and that finding ways of dealing with that can help make video game culture more accessible to female players (Cassell & Jenkins, 2000), maybe leading to ‘better’ games for learning for both groups.

Culture in general is another interesting, yet underexplored factor in games-based learning. Culture in itself is a concept under debate (Shaw, 2010). Many learning and working contexts nowadays are intercultural contexts (Hofstede & Pedersen, 1999). We follow a conceptualization of culture describing people having certain beliefs and values that can be influential towards the acceptance and comprehension of media like simulation games. Research on the possible effects of culture on performance in games-based learning environments is a poorly developed field. Culture in gaming contexts is often used to describe ‘the’ gamer in the context of the medium, and often shapes a picture of the one or other ‘subculture’ (Shaw, 2010). Only one newspaper article addressed different preferences for games in culture (Schiesel, 2006). Gamers seem to be
differently affected by game worlds. The well-known series Grand Theft Auto for example, with its scenes of glamorized urban American violence, has been tremendously popular in the United States but has largely failed to resonate in Asia and in many parts of Europe (Schiesel, 2006). Hofstede & Pedersen (1999) showed that different cultural backgrounds can hinder players in using games as a universal language, as proposed by Duke (1974).

The main objective of this study is to understand the role of gender and culture on the game-based learning experiences of participants reflected by their game performance. In our study, we used two short single-player and multi-player simulation games, so-called Microgames, called Yard Crane Scheduler (YCS) 1 and YCS3 respectively. For this study, we quantified the learning experience in terms of the game scores of YCS1 and YCS3, and used the results of a post game survey where participants reported on the aspects they learnt from the game. The scores reflect the game performance that acts as a proxy for the learning experience of the participants. We will describe the YCS1 and YCS3 games in the following section.

2. Yard Crane Scheduler games – Microgames to increase awareness for integrated planning operations in individuals and teams

2.1 The YCS1 game

We use the term 'Microgames' for the approach of using brief, situated learning games to address complex problems. These games are grounded on pedagogical considerations (see for more detail Kurapati, Lukosch, Verbraeck, & Brazier, 2015; Lukosch, Groen, Kurapati, Klemke, & Verbraeck, 2015a; Lukosch, Groen, Kurapati, & Verbraeck, 2015b). Due to their shortness, Microgames can only represent a limited part of reality. Still, we could prove that they are capable of being meaningful learning tools (Verbraeck, Kurapati, & Lukosch, 2016; Lukosch, Kurapati, Groen, & Verbraeck, 2016). The design of a Microgame in our approach followed the Triadic Game Design Philosophy (TGD) (Harteveld, 2011), which looks at the aspects Reality, Meaning, and Play. When designing a Microgame in this way, we started from an actual problem in the reality of a complex system as the first aspect of the game. Together with stakeholders, we decided which aspects of the real system should be represented within the game, e.g. what are the biggest challenges in the field, or the most critical elements. For YCS1, it became clear that the interdependent nature of planning operations in a container terminal is a huge challenge for all stakeholders involved. The functions that were transferred into the game are the vessel planning, the yard planning, along with the quay and yard crane handling. In the game, deep-sea vessels arrive to load and unload containers in the terminal. A screenshot illustrates the main screen of the game (see Figure 1).

![Figure 1: Partial screenshot of the Yard Crane Scheduler 1 Game](image)

In the game, the yard is one container high and loading and unloading of ships is not done simultaneously, which is different to reality, and implemented for the sake of playability. The second aspect addressed in the design process is the aspect of meaning, or (learning) purpose of the game. For YCS1, we decided that players of the game should become aware of the multiple functions of planning operations and their integrated nature. Especially the de-briefing phase of the game is used to exemplify and discuss different strategies of integrated planning. Thus, gaining insights in alternative planning strategies is another meaning of the YCS1 game. As the third dimension of the game, we decide about the play aspects of the game during the design process. The play aspect relates to the engaging and fun elements of the game to support the learning process. For YCS1, we decided on a scoring mechanism and a leader board to carefully foster the competition amongst players of YCS1 (not necessarily related to a certain group of players). The objective for the player to win the
game is to make sure the ships are serviced as soon as possible, while making efficient use of terminal resources. Additionally, we followed the approach of designing a game that is easy to play but hard to master. The game’s challenge invites to try over and over again in order to gain a higher score.

2.2 The YCS3 game

YCS3 is a multi-player extension of the YCS1 game. It is similar to the single player YCS1 game in terms of layout and game play but it has some variations regarding roles, especially that of the additional role of the vessel planner. In YCS1, the role of the vessel planner was not visible to the player. In YCS3, the role of the vessel planner is to plan the unloading order of the containers.

In the YCS1 game, the individual player could view all plans and operations of all the roles, whereas in the YCS3, each role has a different access to various planning and operational tasks. The berth planner can only access the quay cranes, while the controller can only access the yard cranes. The yard planner can plan the containers from the ship on the yard, while the vessel planner needs to decide on the order in which the containers need to be unloaded. There is also a sequence of actions that has to be followed by the players. The controller cannot allocate yard cranes before the yard planner finalized the yard plan, and the berth planner cannot unload the ships until the vessel planner created an unloading order. The learning goal of this game that the players understand the need to communicate and collaborate with each other to align their plans with each other. Only by doing so, they are able to reach a high individual and group score. Two screenshots of the YCS3 game are shown in Figure 2.

Figure 2: Screenshots of YCS3

2.3 Participants

The case studies were conducted in a quasi-experimental controlled setting, with classes of bachelor and graduate students from universities in the Netherlands, the United States and Germany. Although a total of 172 students participated in the actual study, we could only use only the data from 169 students for the gender study and 164 students for the cultural differences study, due to incomplete surveys or student refusal to provide data to the study. Among the students considered for the gender study, 95 students were male, and 74 students were female. Among the 164 students considered for the culture study, the nationalities of the students included Dutch (42), Chinese (39), American (37), German (16), Indian (9), Taiwanese (5), Vietnamese (3), Colombian (2), Pakistani (2), Greek (2), South Korean (2), Costa Rican (1), Ecuadorian (1), Finnish (1), French (1), Syrian (1). 41 additional students participated in a pre-test of the experimental sessions. Their results were excluded from the reported data, as they played a pre-final version of the game to support the development of the game itself. 26 students from a Dutch University participated in the YCS3 study. Of the 26 students, 11 were female and 15 were male. The mean age of the participants was 23.3 with a standard deviation of 1.7. The majority of the participants were Dutch (20), followed by three Chinese students, and one each from Belgium, Mexico and Greece. We received the approval of the Institutional Review Board of the University of Maryland and the Human Research Ethics Committee of Delft University of Technology to conduct this research study with students.

2.4 Methods and Materials

Participation in the experiments was voluntary. The participants were encouraged to perform well by the possibility of winning a small prize through a weighted lottery method based on their task performance score. Each experimental session took about 2hr 15 minutes to execute. Participants were asked to complete a pre-
survey that collected their demographic information. After that, the participants were presented with a briefing lecture on container terminal planning operations, and they received a tutorial and practice session for the YCS game that was used to assess their planning task performance (Kurapati, Lukosch, Groen, & Verbraeck, 2014). Two exercises with varying levels of difficulty (mission 1 and mission 2) were provided to the participants after the tutorial session. The scores of mission 1 were not counted for the data analysis. The participants then had to complete the more difficult mission 2 twice. The highest score of mission 2 was considered for evaluating the planning task performance or the game performance of each participant. YCS1 was played individually; YCS3 was played in teams of four players. After playing YCS, the facilitator conducted a de-briefing session to gather the players’ insights and strategies related to YCS and to discuss the challenges in container terminal planning operations.

For the experiments, the researchers provided the needed hardware and software to the students. All experiments were held within a classroom setting. One teacher was involved in this study and at least two more researchers for observation and technical support participated in each of the sessions.

3. Results

3.1 Gender and game performance in YCS1

We chose to use the highest score per participant for mission 2 in the YCS1 game as an indicator of learning experience to make sure that players had a certain, comparable level of proficiency in playing the game before we counted their scores for our study. The overall maximum score of YCS1 in mission 2 was 12819 and the minimum score was -364.

In mission 2 of YCS1, the mean high score of female participants was 6502.56 and the mean high score for male participants was 4322.26. In order to check the significance of this difference, we employed the Mann-Whitney U test, a non-parametric statistical test, to compare the YCS1 game performances between male and female participants. We chose the Mann-Whitney test because we found that the YCS1 score was not normally distributed among male and female participants in our study.

Figure 3: Comparison of YCS1 game performance based on gender

Figure 3 illustrates the results of the Mann-Whitney test. The y-axis denoted by YCS_HIGH_1_2 represents the distribution of scores of the female and male participants, while the x-axis represents the frequency. We can observe that the mean rank of female participants (99.64) is higher than the mean rank of male participants (66.2). This difference is also statistically significant as the Mann-Whitney standardized test statistic is -4.4 with a p value of 0. Therefore we can confirm that female participants outperformed their male counterparts in the YCS1 game performance in mission 2.

3.2 Culture and game performance in YCS1

Figure 4: Comparison of YCS game performance based on nationality
To compare the cultural differences in the game performance, we classified the participants based on nationality. The 4 major groups turned out to be Dutch (42), Chinese (39), American (37), German (16) and Other (30). We employed the Kruskal Wallis test, which is a non-parametric statistical test to compare different groups. We can observe from figure 4, which represents the YCS1 mission 2 high score on the y-axis and the nationality on the x-axis that the Dutch, and German participants outperform their counterparts. This observation is further fortified by examining the pair-wise comparisons of each of the 4 groups represented in Table 1. We can observe that the pair-wise comparisons of “Other”-“Dutch”, and “American”-“Dutch” are statistically significant. This also holds true to “Other”-“German” and “American”-“German”. The other pairwise comparisons did not yield statistically significant differences. Therefore we can state that Dutch and German participants performed significantly better than their American counterparts. Since the “Other” group doesn’t represent a particular culture, we don’t consider this group in our results.

Table 1: Pair-wise comparisons of the game performance of different nationalities

<table>
<thead>
<tr>
<th>Pair-wise comparisons</th>
<th>Std. Test Statistic</th>
<th>Significance level (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Other”-“American”</td>
<td>.026</td>
<td>1.000</td>
</tr>
<tr>
<td>“Other”-“Chinese”</td>
<td>1.023</td>
<td>1.000</td>
</tr>
<tr>
<td>“Other”-“German”</td>
<td>2.871</td>
<td>0.041*</td>
</tr>
<tr>
<td>“Other”-“Dutch”</td>
<td>4.223</td>
<td>0.000**</td>
</tr>
<tr>
<td>“American”-“Chinese”</td>
<td>-1.054</td>
<td>1.000</td>
</tr>
<tr>
<td>“American”-“German”</td>
<td>-2.949</td>
<td>0.032*</td>
</tr>
<tr>
<td>“American”-“Dutch”</td>
<td>4.448</td>
<td>0.000**</td>
</tr>
<tr>
<td>“Chinese”-“German”</td>
<td>-2.157</td>
<td>0.310</td>
</tr>
<tr>
<td>“Chinese”-“Dutch”</td>
<td>3.422</td>
<td>0.000</td>
</tr>
<tr>
<td>“German”-“Dutch”</td>
<td>.411</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*p<0.05, **p<0.01

3.3 Gender and Culture

We conducted an analysis of the relationship between gender and culture in our study population. The results of the descriptive cross tabulation data representing the count and percentage of male and female students of each of the nationalities are presented in Table 2. The valid data points amounted to 164 because the culture study had only 164 valid responses. The table shows a relatively even distribution of male and female players in all groups, expect of the Dutch and the German, where more female than male players can be found.

Table 2: Cross tabulation data: Gender vs. Nationality

<table>
<thead>
<tr>
<th>Gender</th>
<th>Female</th>
<th>Count</th>
<th>American</th>
<th>Chinese</th>
<th>Dutch</th>
<th>German</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>18</td>
<td>19</td>
<td>28</td>
<td>11</td>
<td>17</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>48.6%</td>
<td>48.7%</td>
<td>66.7%</td>
<td>68.8%</td>
<td>56.7%</td>
<td>56.7%</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>Count</td>
<td>19</td>
<td>20</td>
<td>14</td>
<td>5</td>
<td>13</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Percentage</td>
<td>51.4%</td>
<td>51.3%</td>
<td>33.3%</td>
<td>31.3%</td>
<td>43.3%</td>
<td>43.3%</td>
<td></td>
</tr>
</tbody>
</table>

Both gender and nationality are categorical variables. Therefore we performed the Pearson’s chi-square test also known as the chi-square test for independence between these variables. The results of the chi-square test are presented in Table 3.

Table 3: Chi-Square Test: Gender vs. Nationality

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>df</th>
<th>Asymptotic Significance (2-sided)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Chi-Square</td>
<td>4.635</td>
<td>4</td>
<td>.327</td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>4.691</td>
<td>4</td>
<td>.321</td>
</tr>
<tr>
<td>Linear-by-Linear Association</td>
<td>.149</td>
<td>1</td>
<td>.699</td>
</tr>
</tbody>
</table>

From Table 3, we can notice that the Pearson chi-square value $\chi^2(1) = 4.635$, with $p = .327$. The high p value shows that there is no statistically significant association between gender and nationality. In addition to the chi-square test, we also performed the Phi and Cramer’s V tests to test the strength of association between gender and nationality shown in Table 4. We can see that the strength of association between the variables is very weak with very high p values of .327.
Table 4: Tests for the strength of association between gender and nationality

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Approximate Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal by Nominal</td>
<td>.168</td>
<td>.327</td>
</tr>
<tr>
<td>Cramer’s V</td>
<td>.168</td>
<td>.327</td>
</tr>
<tr>
<td>N of Valid Cases</td>
<td>164</td>
<td></td>
</tr>
</tbody>
</table>

From the above tests we can conclude that gender and nationality are sufficiently independent as their association is statistically insignificant and weak.

3.4 Culture, gender and game performance in a team in YCS3

We conducted the non-parametric Mann-Whitney U test to test if the YCS3 scores were different for male and female participants within one team playing the game. The null hypothesis of the Mann-Whitney U test is that there is no difference in YCS3 scores across gender. We compared the scores of three difficulty missions of YCS3. The results of the Mann-Whitney U test for mission 1 were U = 68, z = -.753, p = .474, for mission 2 they were U = 50.5, z = -1.661, p = .097 and for mission 3 they were U = 60, z = -1.168 and p = .259. None of these results is significant. Therefore we can retain the null hypothesis and state that the distribution of the individual scores of YCS3 was the same across both the male and female participants. The majority of the participants, 20 out of 26, were Dutch. The sample population was quite small (26) to draw any significant conclusions based on the quantitative data analysis. Therefore we chose to use video analysis as a qualitative measure to analyse the teams in terms of their gender and cultural diversity. We transcribed the video based on the information flow, communication and interaction patterns among team members and related them to their overall team performance.

We compared two teams—Team A with 2 male, 2 female participants all from the Netherlands and Team B with three Dutch male participant and one Chinese female participant. Team A showed a uniform information sharing pattern among all members and they collaborated efficiently to align their plans. In Team B, we noticed that a Dutch male player who played the role of the yard planner constantly advised the Chinese female participant who played the role of a controller. She followed the advice without much questioning.

From the qualitative analysis we observed that team A with a balanced gender composition and a uniform nationality performed better than team B with three Dutch males and one Chinese female. Nam, Lyons, Hwang & Kim (2009) support the notion that homogeneous teams perform better than heterogeneous teams, because their common background helps them to show agreement, give opinions, and demonstrate both positive and negative emotions with minimal inhibitions. In our study the behaviour of the Chinese female can be attributed to the fact that the likelihood of Asian team members criticising the decisions of other team members assuming leadership roles is very small (Militello, Ormerod, Schraagen, & Lipshitz, 2009).

3.5 Summary of the results

Our results indicate that gender and cultural differences can be an important factor affecting game performance. We discovered from our results that female participants outperformed male participants in a difficult level of the YCS1 game, which was measured as an indicator for learning experience of planning tasks in container terminals. Similarly, cultural differences had an important influence. Our results showed that Dutch and German participants outperformed their American counterparts in a difficult level of the YCS game. The qualitative analysis of the multi-payer YCS3 game showed that homogenous teams performed better, but we could not find significant relations between gender and culture aspects and the learning results. Overall, we could show that in our study population, gender and culture were two independent aspects. With this, our study showed the effects of gender and culture on game performance. In the following section, we will relate our results to theory.

4. Discussion and Conclusion

4.1 Gender differences in game performance

Already for a long time, researchers are interested in gender differences in cognitive abilities. This interest stems from the wish to explore ‘appropriate roles of men and women’ from the early 19th century (Hyde & McKinley, 1997). As these differences could also have implications for preferences and different uses of games by men and women, game designers and researchers are looking into gender differences as well. In gaming
literature, male culture is usually related to better game performance (Imhof et al., 2007; Kafai et al., 2008), while our study showed that female players outperformed the male population on average for a planning task in a technical environment. Studies show that in general, male individuals perform better in virtual environments and games and use them more often. For example, differences in spatial abilities in VR are shown in the studies of Astur, Ortiz, & Sutherland (1990), Sandstrom, Kauffman, & Huettel (1998), Cutmore, Hine, Maberly, Langford, & Hawgood (2000), and Castelli, Corazzini, & Geminiani (2008). In these studies, male test persons performed better in spatial orientation than female participants. One reason for this is found in the level of anxiety that test persons show when learning to navigate through a virtual environment, with a higher level of anxiety within the female population (Bryant, 1982; Lawton, 1994; Lawton, 1996). Results from another study indicated that gender differences in game play could also emerge from a different ability in the use of interfaces (Waller, 2000). A collection of game-related studies showed that, while the number of female players of digital games has increased over the past decades, the content of games still seems to address male players (Kafai et al., 2008). Gender differences in socialization towards the use of computers were also shown by (Whitley, 1996) and could be one reason for actual differences found in the studies mentioned above. In summary, the focus on computer use and simulation game performance does not provide us with an explanation for the gender differences we found in our study.

As the YCS game is about planning performance, and gender studies on game-play differences cannot serve as an explanation for our findings, we looked into research on planning performance itself to find some elucidation of the phenomenon found in our study. Naglieri & Rojahn (2001) found better performance of girls in planning and attention tasks scales as result of a test based on the Planning, Attention, Simultaneous, Successive (PASS) cognitive-processing theory. Cognitive research seems to yield consistent findings from studies on gender differences related to specific patterns of intellectual abilities. Women in these studies perform better on verbal and memory tasks, while men do better on spatial cognition and spatial learning (see for an overview Boghi, Rasetti, Avidano, Manzone, Orsi, D'agata, Caroppo, Bergui, Rocca, Pulvirenti, Bradac, Bogetto, Mutani, & Mortara, 2006). The same study showed that men and women use different strategies in a planning task, which was represented by their brain activity. It cannot be said whether this is a biological or a social phenomenon, with both genders learning different approaches towards complex planning tasks throughout their life (Boghi et al., 2006). From these studies, we can draw the conclusion that different cognitive approaches towards planning tasks could be an explanation for the differences found in our study. Still, there is plenty of variance between the performance of men and women that might rather be related to individual differences than to gender differences (Unterrainer, Ruff, Rahm, Kaller, Spreer, Schwarzwalid, & Halsband, 2005).

In summary, we can say that planning ability could be one explanation for the difference in scores between males and females we found in our study. This is at least more probable than the influence of the test instrument being a computer game. Nonetheless, being able to provide a clear statement on the gender differences in general would mean to conduct further analyses thoroughly grounded on related theories, to include both neurological and sociocultural aspects that could partially explain the differences found (Naglieri & Rojahn, 2001). We are aware of the fact that being a woman or a man does not necessarily mean that the one or other performs better or worse based only on her or his biological sex, but that social and individual differences play a very important or even more important role as well (Unterrainer et al., 2005).

### 4.2 Cultural differences in game performance

In our study, we made use of the model of Hofstede that contains six dimensions of national cultures (Hofstede, 1984; 2011). Hofstede conceptualizes culture as the collective programming of the mind that distinguishes the members of one group or category of people from others (Hofstede, 2011). Following this concept, culture is always a collective phenomenon, which can be connected to different collectives, or nationalities. Very broadly speaking, culture is the way one is brought up, making it a collective phenomenon (Hofstede & Pedersen, 1999). Culture includes certain beliefs and values, and rules of behaviour. Hofstede identified six dimensions, along which cultures can be classified (Hofstede, 1980; 2011): large vs. small power distance, individualism vs. collectivism, achievement-orientation vs. cooperation-orientation, uncertainty avoidance vs. uncertainty tolerance, masculinity vs. femininity, and long-term vs. short-term orientation. Especially the last two are heavily related to planning performance, as planning is a symbolic activity, which may or may not have an impact on what happens in the future (Hofstede, 1984).
Along these dimensions, German culture can be characterized as being a culture of a rather small power distance, being more collective, uncertainty tolerant, and rather achievement-oriented. The Netherlands show a similar rather smaller power distance, being more individualistic, with a slightly higher level of uncertainty tolerance, but being mainly cooperation-oriented. The culture of the USA is characterized with a comparable power distance, being highly individualistic, having less uncertainty tolerance than Germany and The Netherlands, while being achievement-oriented (Hofstede, 1984). Therefore the main cultural difference between the Western-European (German, Dutch) and the American participants is that of uncertainty avoidance. Cultures with low tolerance to uncertainty are more rigid in their behaviour, while cultures with high tolerance are more flexible (Hofstede, 1984). With respect to the YCS game, to achieve a good score, a player needs to adopt a flexible planning strategy, where he/she needs to pay attention to the arriving ships, and has to constantly reallocate containers and resources. This could be one possible explanation of the effect of cultural difference on the YCS game performance, which showed that Western-Europe (German, Dutch) students outperformed the players with a US-American cultural background.

While some studies explored the role of cultural differences in team and group performance and behaviour (Earley, 1994; Cox, Lobei, & McLeod, 1991), we were not successful in finding studies that explained individual performance and behaviour related to culture. After all, culture is a collective phenomenon (Hofstede, 1984). However, we do acknowledge that culture is a complex artefact not limited to nationality alone. The Merriam-Webster dictionary defines culture as “beliefs, arts, customs of a particular group, place or time”, “a particular society that has its own beliefs, ways of life, art, etc.”, “a way of thinking, behaving or working that exists in a place, group or organization”. We haven’t considered culture in such a complex and wide range, so it can be attributed as one of the limitations of the study. We hope to study culture in game-based learning contexts in a more holistic manner in future studies. Additionally our sample population consists of students from different universities with different study programs and we haven’t considered the effect of such diversity in our study. It is also a possible topic for future work. We will now provide our overall conclusions of this study in the following section.

4.3 Overall conclusions

In conclusion, we would like to highlight that educators, instructional designers and policy makers need to consider factors like gender as well as several social, demographic and cultural aspects regarding the design and use of technology-enhanced learning systems such as games to nurture independent, thoughtful, resourceful and responsible students (Chen, Mashhadi, Ang, & Harkrider, 1999). In a study conducted by Militello et al. (2008), a Chinese interviewee was quoted saying "As a team member, you must accept and support the leader’s decision even if you do not agree with it and even if it leads to a bad outcome" [p.149]. This observation is also consistent with the hierarchical nature of Asian cultures that rate high on the power-distance scale of Hofstede and Pedersen (1999). Although diversity in teams has been valued as a positive performance enabler in a team due to the rich new perspectives it brings into the team, it should be noted that this positive effect has been observed only in the long term (Nam et al., 2009). For ad-hoc teams, it could be that a highly diverse backgrounds of team members hinders effective collaboration, as team members first have to develop a common language to understand each other. As Duke (1974) proposes, games can serve as a means to support this development. For educators and trainers this means that they should be aware of the challenges of heterogeneous teams to learn and work towards effective team task performance in socio-technical work organizations, potentially with the help of games. We believe that our study made first steps towards identifying such factors, by empirically studying the role of gender and culture in game based learning.

Acknowledgment

We thank the Dutch Institute for Advanced Logistics DINALOG for funding the research presented in this article. We also thank the student participants from Delft, The Netherlands, and College Park (MD), USA, who participated in this study.

References


Exploring the Relation between the Theory of Multiple Intelligences and Games For the Purpose of Player-Centred Game Design

Pejman Sajjadi, Joachim Vlieghe and Olga De Troyer  
Vrije Universiteit Brussel, Department of Computer Science, WISE, Brussels, Belgium  
Ssajadi@vub.ac.be  
Joachim.Vlieghe@vub.ac.be  
Olga.DeTroyer@vub.ac.be

Abstract: A large body of research work demonstrates the importance and effectiveness of adapting a learning game to its players. This process is driven by understanding the differences between individuals in terms of abilities and preferences. One of the rather interesting but least explored approaches for understanding individual differences among learners is Gardner’s theory of Multiple Intelligences (MI). Gardner suggests that people exhibit multiple dimensions of intelligence or abilities. In the literature, it is suggested that people with different types of intellectual strengths (intelligences) often exhibit clear preferences toward specific modalities and types of interaction and content in relation to learning. This raises the question whether this knowledge could be transferred and employed in adapting learning games to players, more in particular for the purpose of improving the game and/or learning experience, as well as the learning outcome of the players. Although various claims regarding the existence of a relationship between MI and games have been made, none of them are substantiated with empirical evidence.

This paper presents the results of an empirical study that has led to evidence-based mappings between the different dimensions of intelligences proposed in MI and the fundamental building blocks of games, i.e. game mechanics. These mappings indicate which game mechanics suit which MI dimensions, and can therefore act as design guidelines when designing games targeting people exhibiting dominance for specific MI dimensions. A tool that visualizes these mappings and facilitates their use in the design of such player-centred (learning) games is also presented.

Keywords: Multiple intelligences; Game preferences; Game mechanics; Evidence-based; Game design; Learning games

1. Introduction

Adapting learning games to their players has proven to be an effective way for improving the game experience and learning outcome of the players (e.g. Grappiolo et al. 2011; Kickmeier-Rust & Albert 2010; van Oostendorp et al. 2013; Muir & Conati 2012; Parnandi & Ahmed 2014). Many different aspects of players have already been considered as the driving force in such an adaptation process: e.g. knowledge level, attention, engagement, task skill, and learning style. The so-called intelligences of the players based on the theory of Multiple Intelligences (MI) (Gardner 2011) however, have been largely neglected. Gardner has stated that human intelligence is multi-dimensional. He defined eight distinct dimensions of intelligences, where each dimension focuses on specific abilities to solve problems or to create products. Furthermore, he stressed that everyone possesses every intelligence dimension but to different degrees. Chan (2005) suggested that people with different intelligences or intellectual strengths often exhibit clear preferences toward specific modalities and types of interaction in relation to learning and self-expression. This raises the question whether these MI dimensions could be useful to the process of adapting a game to (some of) the characteristics of players, and if so how could this be done. In particular, we are interested in using the MI dimensions in player-centred game design for the purpose of improving the game and learning experience of the players. The term player-centred game design is used to refer to the design process of a game targeting a group of players with specific characteristics.

Possible relationships between the theory of MI and game constructs, as well as its potential for adaptation of games have been discussed by different scholars (McCue, 2005; Becker, 2007; Jing et al., 2012; Chuang & Sheng-Hsiung, 2012; Starks, 2014; Lepe-salazar, 2015). Although this seems to be a promising direction to follow, empirical evidence for the existence of such relationships is virtually non-existent. Such evidence is not only crucial for determining whether this theory can be potentially used in the adaptation of (learning) games, but also for supporting game designers and developers in making informed decision about the game constructs to incorporate in their design when taking players’ intelligences into consideration.

This paper reports on a research aimed at establishing empirical grounds for incorporating the theory of MI in player-centred game design. As such, we first investigated the existence of aforementioned relationships by means of a survey study, measuring the participants’ intelligences and their preferences for specific games. Secondly, to understand the underlying causes for the observed correlations, we analysed the considered games based on their fundamental building blocks (i.e. game mechanics). Based on this analysis, mappings between the different MI dimensions and game mechanics were established. The rest of this paper is structured as follows: we first briefly explain the theory of MI (section 2) and explore related work (section 3). We then explain the survey (section 4) and next the derivation of the mappings between MI dimensions and game mechanics (section 5). Furthermore, we discuss the implications of our findings (section 6), and lastly we elaborate on limitations and future direction of this research (section 7).

2. The theory of multiple intelligences (MI)

The theory of MI draws a framework for defining individual differences between people in terms of their abilities and preferences. According to Gardner (2011), an intelligence is “the ability to solve problems, or to create products, that are valued within one or more cultural settings” (ibid. page xxviii). Eight distinct intelligences or so-called MI dimensions were proposed. Each dimension represents a different way of thinking, problem solving and learning. They are defined as follows (The Components of MI, 2016):

- Visual-spatial intelligence represents the ability to conceptualize and manipulate large-scale spatial arrays (like a pilot does), or more local forms of spaces (like an architect).
- Bodily-kinesthetic intelligence is the ability to use one’s whole body, or parts of the body, to solve problems or create products (like a dancer).
- Musical-rhythmic intelligence implies having sensitivity to rhythm, pitch, meter, tone, melody and timbre (like a musical conductor). This may entail the ability to sing, play musical instruments, and/or compose music.
- Linguistic intelligence suggests sensitivity to the meaning, order, sound, rhythms, inflections, and meter of words (like a poet).
- Logical-mathematical intelligence is the capacity to conceptualize the logical relations among actions or symbols (like a mathematician).
- Interpersonal intelligence represents the ability to interact effectively with others and being sensitive to others’ moods, feelings, temperaments and motivations (like a negotiator).
- Intrapersonal intelligence implies being sensitive to one’s own feelings, goals, and anxieties, and the capacity to plan and act in the light of one’s own traits. Intrapersonal intelligence is not particular to specific careers; rather, it is a goal for every individual in a complex modern society, where one has to make consequential decisions for oneself.
- Naturalistic intelligence is the ability to make consequential distinctions in the world of nature as, for example, between one plant and another, or one cloud formation and another (like a taxonomist).

Note that everyone possesses every MI dimension but to different degrees (Gardner, 2011). All dimensions work together in an orchestrated way and they can influence each other. For instance, Castejon et al (2010) analysed and compared the different theoretical models of the structure of those intelligences using Confirmatory Factor Analysis (CFA) and showed that the different dimensions of MI are not independent.

We recognize that controversies exist concerning the theory of MI. Opponents have argued that there is a lack of strong empirical evidence to support the claim that eight distinct intelligence dimensions exist (see Brody, 2006; Waterhouse, 2006a; 2006b). Proponents have argued that “a theory is not necessarily valuable because it is supported by the results of empirical test, rather its value depends on the contribution it makes to understanding and to practice in the field” (Chen, 2004, page 22). Furthermore, studies using instruments developed to measure the values for the different dimensions show that people indeed differ in terms of their so-called intelligences (e.g. Akbari & Hosseini 2008; Castejon et al. 2010; Marefat 2007; Naeini & Pandian 2010). Therefore, we believe that it is worthwhile to investigate whether MI dimensions can be useful for adaptation of learning games.

3. Related work

The related work considering the theory of MI and games can be divided into two groups. The first group is about the potential relationship between the theory of MI and games. The most prominent works in this group
are the ones of Becker (2007) and Starks (2014); they have suggested that relationships between MI dimensions and certain characteristics of games exist. We briefly explain these suggestions. Becker (2007) argues that there is a link between the written and spoken elements and instructions in games and the development of the linguistic intelligence. According to Becker, “this is one reason why children often experience success in learning to read through games like Pokémon” (Page 371). Similarly, she maps musical intelligence to a game’s soundtrack and auditory feedback, referring to games such as Karaoke Revolution; logical-mathematical intelligence to in-game strategizing, arithmetic, management style and puzzle games such as Pikmin; visual-spatial intelligence to the graphical environment, visual elements of games and how they are perceived through the screen; bodily-kinesthetic intelligence to games that promote physical movement as well as the different physical states a player experiences while playing a game such as Dance Dance Revolution; intrapersonal intelligence to games that involve ethical dilemmas and moral decision making such as Black & White; interpersonal intelligence to multiplayer collaboration, communication and competition; and naturalistic intelligence to realistic portrayal of natural environments in games such as Zoo Tycoon.

Starks (2014) provides similar arguments, stating that in-game graphics engage a person’s visual intelligence, while the way a player moves in the game environment engages their spatial intelligence. She also states that social relationships inside and surrounding a game refers to the use of the interpersonal intelligence, like in MMO games; that empathy provoking situations inside a game, such as in Darfur is Dying, engage a person’s intrapersonal intelligence; that music and sounds engage a player’s musical intelligence; that narrative and language used inside the game engage the linguistic intelligence; that components like arithmetic, calculations and geometry, as well as pattern detection and logical deduction activate logical-mathematical intelligence; that in-game actions requiring actual physical movement engage bodily-kinesthetic intelligence; and that realistic representations and simulations of natural environments engage a player’s naturalistic intelligence. It is important to note that the observations of Becker (2007) and Starks (2014) are solely based on their theoretical analyses, and they do not provide empirical evidence for their claims.

The second group of studies focus on investigating the development of players’ intelligences through games. Unlike the first group, they provide some empirical evidence for their claims. In (Crescenzi-Lanna & Grané-Oró, 2016) the importance of developing the MI of children at an early age is stressed. The study analyses 100 educational apps (including games) for children under the age of eight. The results indicate that the majority of the apps focus on the visual-spatial and logical-mathematical dimensions. The results also show that other dimensions such as kinaesthetic, interpersonal, intrapersonal or musical are neglected, even though they are developmentally essential for children at that age. Jing et al (2012), provide an overview of several educational games that can aid in the development of a player’s logical-mathematical intelligence. Similarly, Chuang and Sheng-Hsiung (2012) claim that games can be used as a tool to enhance players’ intelligences and learning outcomes. Li et al (2013) have investigated the effect of Role Playing Games on intrapersonal intelligence.

To the best of our knowledge, there is no academic work that provides a mapping between the theory of MI and game mechanics. However, research work that relates other types of personal differences with game mechanics could offer an interesting frame of reference. In this context, we can mention the work of Jason VandenBerghe, called “Engines of Play”. VandenBerghe (2012, 2013) focused on personality traits rather than intelligences, and defined a mapping between the Big Five (Goldberg 1990) and game mechanics. The purpose of this mapping is to know which game mechanics to incorporate in the design of games to boost the players’ motivation. In the domain of learning games, the Learning mechanic-Game mechanic model (Arnab et al. 2014; Lim et al. 2015) links game mechanics to different pedagogical and learning theories. The links were extracted from serious games following different pedagogical or learning theories.

4. Empirically-validated relationships between MI dimensions and preferences for games

To investigate whether MI dimensions can be used for adapting learning games to players, we started by performing a comprehensive survey study to investigate the existence of relationships between MI dimensions and preferences for games. This survey study is described in section 4.1. In search for an explanation for the results, we investigated whether the type or genre of the games could explain the results. This is described in section 4.2.
4.1 Survey

This section elaborates on the methodology used for the survey study (sections 4.1.1, 4.1.2 and 4.1.3), and reports on its findings (section 4.1.4)

4.1.1 Data collection

An online survey (http://goo.gl/5v6wOR), comprised of three sections was created and launched on July 17th of 2015. After 8 days, 308 participants had completed the survey and the response rate started to drop. We were confident that this number of participants would suffice for drawing conclusive statements, and therefore we proceeded with the collected data.

The first section of the survey contained seven questions that inquired about the demographic information of the participants (gender, age range, and level of education), as well as their game-related background information (frequency of gaming, experience with different game platforms or devices, preferred game genres, and hands-on experience with game design or development). This information would enable us to determine the heterogeneity of the sample, as well as investigating the effect of personal and contextual factors such as age, education and prior experiences.

The second section of the survey was designed to assess the strength of the MI dimensions of the participants. This was performed using the Multiple Intelligences Profiling Questionnaire (MIPQ) (Tirri & Nokelainen, 2008; Tirri & Nokelainen, 2011). The participants were asked to rate 31 statements on a scale of 1 to 5 to measure the eight MI dimensions. Each dimension was measured using four questions, except for the naturalist intelligence dimension that was measured by three questions.

Table 1: 47 game titles selected for the study. (S represents game series)

<table>
<thead>
<tr>
<th>Intelligence dimension</th>
<th>Selected games</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual-spatial</td>
<td>World of Warcraft, Minecraft, Dirt, Portal, Angry Birds, Tetris</td>
</tr>
<tr>
<td>Bodily-kinesthetic</td>
<td>Xbox Fitness, Street Fighter, Boom Blox, Kinect Sports, Wii Sports Resort, Just Dance, Dance Central Spotlight, Dance Dance Revolution, Fantasia: Music Evolved</td>
</tr>
<tr>
<td>Musical</td>
<td>Guitar Hero, Audiosurf, Rock Band, SingStar, Bit. Trip Runner, Just Dance, Dance Central Spotlight, Dance Dance Revolution, Fantasia: Music Evolved</td>
</tr>
<tr>
<td>Linguistics</td>
<td>The Typing of The Dead, Wordament, Scribblenauts, Wordfeud, Ace Attorney</td>
</tr>
<tr>
<td>Logical-mathematic</td>
<td>The Room, 2048, Braid, Where’s My Water?, L.A. Noire, Heavy Rain</td>
</tr>
<tr>
<td>Interpersonal</td>
<td>The Sims, DayZ, Life is Strange, Second Life, Farmville, Word of Warcraft, The Walking Dead, Heavy Rain</td>
</tr>
<tr>
<td>Intrapersonal</td>
<td>Fable, Black &amp; White, Infamous, Mass Effect, Fallout, Heavy Rain, The Walking Dead</td>
</tr>
<tr>
<td>Naturalist</td>
<td>Endless Ocean, Spore, Plan It Green, Flower, Afrika</td>
</tr>
</tbody>
</table>

The third section of the survey contained a list of 47 game titles (see Table 1) that the participants were asked to rate (1 to 5 stars) to reflect their enjoyment of and preference towards the game (i.e. 1 star represents lowest and 5 stars highest enjoyment of and preferences towards a game). The participants were explicitly instructed to only rate the games that they had previously played. The list of games was compiled such that for each MI dimension five game titles were provided (i.e. 40 titles in total). The list of games was compiled in collaboration with a team of avid gamers and academic game experts, and based on the suggestions found in the literature on the mapping between games and MI dimensions (Becker 2007; Starks 2014). Seven games that could be related to more than one MI dimension were also added because of their unique design and popularity. We decided to limit the size of the list to reduce the time required to complete the survey in order to maximize participation.

4.1.2 Sampling and population

As we targeted an international population of frequent gamers, we spread calls for participation through social media targeting online communities of avid gamers, game designers, developers and researchers (Facebook, LinkedIn, Google+, Reddit, and Twitter). We also reached out by email to academic communities focusing on game research (DIGRA, IGDA, DIGRA Australia, IFIP and CHI-WEB). In the aforementioned period, 465 people responded, of which 308 participants (97 females and 211 males) completed the survey. An overview of the age range distribution is presented in Figure 1.
Based on the self-reported frequency of gaming activity, 82.79% of our sample can be considered avid gamers (playing every day or 3-6 times a week). Based on the results of MIPQ, the distributions of the participants’ dominant MI dimensions were obtained (see Figure 2). This distribution shows that in our sample, there were more players with dominant logical-mathematical and intrapersonal intelligence, while the rest of the MI dimensions were somewhat heterogeneously distributed, allowing us to draw conclusions about all MI dimensions. A MI dimension was considered “dominant” if the sum of the scores on the individual questions for the dimension was above 15 (out of 20) or 12 (out of 15) in the case of naturalistic intelligence.

Instead of directly hypothesizing that specific correlations exist between MI dimensions and preferences for certain games, we first opted to unveil any existing patterns in our dataset. If patterns would emerge, then we could hypothesize that correlations exist within our dataset. We therefore first analysed the data by means of factor analysis (FA). Using the extraction method “Principal Component Analysis” and the “Promax” rotation method, the FA provided us with sufficient evidence that patterns do exist in our data. We then opted to test for every possible significant correlation between MI dimensions and game preferences by means of bivariate correlation analyses using the Spearman’s rho on two levels. On the first level, the correlations between the participants’ MI levels and preferences for all game titles were tested. This aided us in identifying significant correlations, as well as the direction of the correlations. On the second level, in order to understand the cause and nature of the correlations, correlations between preferences for game genres and MI dimensions were tested.

FA between each MI dimension and the 47 games was repeated 8 times to identify patterns (once for each MI dimension). Not all game titles were found to be part of these patterns, however those that do (20 in total) can be grouped into factors, here called Intelligence-game factors, shown in Table 2. More precisely, each column...
Pejman Sajjadi, Joachim Vlieghe and Olga De Troyer

Table 2: Summary of the 8 pattern matrices of FA tests. KMO for linguistics .767, logical-mathematical .765, visual-spatial .766, bodily-kinesthetic .763, musical .766, interpersonal .765, intrapersonal .767, naturalistic .765; Sig. levels for KMO and Bartlett's tests .000 (represents a game series)

<table>
<thead>
<tr>
<th>MI dimension</th>
<th>Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>intelligence-game factor 1 (linguistics)</td>
</tr>
<tr>
<td></td>
<td>.409</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Games</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Boom block</td>
<td>.357</td>
<td>.432</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wii sports resort</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dirt*</td>
<td>.338</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angry Birds</td>
<td>.544</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tetris</td>
<td></td>
<td>-.460</td>
<td>-.329</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Street Fighter*</td>
<td>.344</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Just Dance*</td>
<td>.461</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dance Central Spotlight*</td>
<td>.395</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dance Dance Revolution*</td>
<td>.469</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guitar Hero*</td>
<td>.530</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rock Band</td>
<td>.599</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wordfeud</td>
<td>.402</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SingStar?</td>
<td>.432</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2048</td>
<td>.308</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Where’s My Water?</td>
<td>.574</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heavy Rain</td>
<td>.543</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Sims*</td>
<td>-.398</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endless Ocean</td>
<td>-.371</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spore</td>
<td>.358</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flower</td>
<td>.531</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Next, bivariate correlation analyses between game preferences and MI dimensions’ levels (or strength), showed significant correlations for each of the eight MI dimensions and several game preferences. Note that for this analysis, each MI dimension was represented by a single value (see section 4.1.2).
To gain a deeper understanding of these correlations, the bivariate correlation analysis was repeated focusing on the individual questions of the MIPQ, rather than on the single value for each MI dimension. The results highlighted which questions were correlated to specific game preferences. Note that it is possible for a game preference to show correlation with an MI dimension presented by a single value, as well as with the questions for that dimension. However, these two correlations could be of different degrees. To ease the application of our results, we have combined the results of both levels into a single table that shows positive (‘+’) or negative
(‘-‘) correlations and their significance levels (indicated by * for $P < 0.05$ and ** for $P < 0.01$). The results of this multi-level approach are summarized in Table 3. Cautious readers will note that the results of this test (i.e. the specific correlations) sometimes deviate from the results of the FA. This slight discrepancy is caused by the difference in assessment techniques used by both analytic methods. Nonetheless, we have attained empirical evidence for the existence of a relationship between MI dimensions and games.

The results of our study provide empirical evidence that partly confirms the theoretical suggestions made by Becker (2007) and Starks (2014). Although we have managed to provide empirical evidence for the existence of significant correlations between MI dimensions and games, our results are in some cases not neatly in line with the theoretical mappings suggested by Becker and Starks. For instance, Becker states that there is a link between logical-mathematical intelligence and in-game strategizing, arithmetic, management style and puzzle games. However, the logically-mathematically intelligent people in our population also exhibit a significant preference for games that require extensive physical movement such as Fantasia: Music Evolved and Xbox Fitness (see Table 3). This is a rather interesting finding, which raises the question: What makes certain games be preferred by players who exhibit dominance for a certain MI dimension? Our first hunch was that this could be the genre of the games.

4.2 Follow-up analysis and results

To investigate whether the obtained correlations could indeed be explained in terms of preferences for particular game genres, the genres of the game titles that showed correlations were extracted using the official website of Pan European Game Information (PEGI) at [http://www.pegi.info/en/index/](http://www.pegi.info/en/index/) (see Table 3), commonly used in the gaming industry. The results suggest that the correlation between MI dimensions and game preferences cannot be explained independently in terms of unique preferences for one or multiple game genres. To confirm our claim, we performed a bivariate correlation analysis using the participants’ explicit game genre preferences obtained from the first section of the survey (see Table 4).

Table 4: Bivariate correlation analysis between MI and explicit game genre preferences. + (positive), - (negative), the values represent the strength of the correlation at $P < 0.01$ ** or $P < 0.05$*

<table>
<thead>
<tr>
<th>Game genre</th>
<th>Linguistics</th>
<th>Logical-Mathematical</th>
<th>Visual-Spatial</th>
<th>Bodily-Kinaesthetic</th>
<th>Musical</th>
<th>Interpersonal</th>
<th>Intrapersonal</th>
<th>Naturalistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action/adventure</td>
<td>-.095*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adventure</td>
<td></td>
<td></td>
<td>.115*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMO</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.112*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Platform/platformer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.145**</td>
</tr>
<tr>
<td>Puzzle</td>
<td>.146**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPG</td>
<td>-.119*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Racer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhythm/dance</td>
<td>.198**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shoot `em up</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.135**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sims</td>
<td>-.118*</td>
<td>-.100*</td>
<td>-.105*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sports</td>
<td>.114*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strategy</td>
<td>+141**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+150**</td>
</tr>
</tbody>
</table>

The result of this analysis shows that, for instance, the “action/adventure” genre is correlated with the logical-mathematical and musical dimensions, whereas MI dimensions such as visual-spatial are not correlated with any genre. Moreover, certain genres seem to correlate with the majority of MI dimensions. These overlaps and internal conflicts indicate that no unique combinations of game genres could independently explain the relationships between MI dimensions and game preferences. An explanation could be that the categorization into game genres is only based on one aspect of the game and that preference is probably based on different aspects (Sajjadi, Vlieghe, et al. 2016b). Therefore, to identify the underlying cause for the obtained correlations, we decided to investigate whether we would be able to establish mappings between MI dimensions and the games’ fundamental building blocks (i.e. game mechanics).
5. Evidence-based mappings between MI dimensions and game mechanics

In order to draw relationships between MI dimensions and game mechanics, the games that correlated to the different MI dimensions were analyzed based on their mechanics. Many scholars have defined the term game mechanic (e.g. Sicart 2008; Lundgren & Bjork 2003; Rouse III 2010). Sicart provided one of the most comprehensive definitions: “methods invoked by agents, designed for interaction with the game state” (ibid 2008) (Paragraph 25).

To identify which game mechanics were utilized by each of the games, a comprehensive repository of game mechanics was needed. Numerous existing repositories of game mechanics were found (e.g. Arnab et al., 2014; “Game Mechanic Mixer; Gamification Wiki; SCVNGR’s Secret Game Mechanics Playdeck; Social game mechanics; Hamari & Järvinen, 2011; Järvinen, 2008; Lim et al., 2015; Louchart & Lim, 2011; Oberdörfer & Latoschik, 2013). Many of these repositories are quite elaborate and show significant overlaps. Despite the overlaps, the different repositories do show discrepancies in terms of the number of mechanics and their labels and definitions. Thus, for the purpose of this research, we have compiled a comprehensive list of game mechanics based on existing repositories and complemented it with game mechanics that we have identified ourselves during the analysis of the games used in our study. This process led to the creation of a repository of 236 distinct game mechanics, which can be found online1.

Furthermore, we decided to also consider the role a game mechanic plays inside a game. For this, we used the categorization of Fabricatore (2007), i.e. core and satellite mechanics. Core mechanics are “the set of activities that the player will undertake more frequently during the game experience, and which are indispensable to win the game” (Page 12), and the satellite mechanics are “special kinds of mechanics, aimed at enhancing already existing activities” (Page 13) (ibid 2007). Next, the games that were either negatively or positively correlated to one or more MI dimensions were analysed to discover which game mechanics they were using and in which role.

5.1 Methodology

For each MI dimension, two clusters were created: a positive cluster containing the games having a positive correlation with the dimension, and a negative cluster containing the games having a negative correlation with that dimension. Subsequently, each of the games were analysed with respect to the 236 game mechanics to determine whether the mechanic was used as either core (c) or satellite (s), or not used. An example of such analysis with respect to four mechanics and some of the games that were correlated to the logical-mathematical intelligence dimension is given in Table 5. The positive cluster is indicated in grey. This process was repeated 8 times, once for each MI dimension.

Table 5: Example analysis of game mechanics for the Logical-mathematical intelligence dimension (fragment)

<table>
<thead>
<tr>
<th>Mechanics</th>
<th>Portal</th>
<th>2048</th>
<th>Braid</th>
<th>Fable</th>
<th>Fallout</th>
<th>Xbox Fitness</th>
<th>Wordament</th>
<th>Heavy Rain</th>
<th>The Sims</th>
<th>World Of Warcraft</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovery</td>
<td>c</td>
<td>s</td>
<td>c</td>
<td>c</td>
<td></td>
<td></td>
<td>c</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Epic meaning</td>
<td></td>
<td>s</td>
<td>s</td>
<td>c</td>
<td></td>
<td></td>
<td>c</td>
<td>s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infinite gameplay</td>
<td>c</td>
<td>s</td>
<td>s</td>
<td>c</td>
<td></td>
<td></td>
<td>c</td>
<td>s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motion</td>
<td></td>
<td></td>
<td>s</td>
<td>c</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We established the following rule to determine whether a mechanic is related to an MI dimension: the mechanic should be utilized by at least half of the games correlated to that dimension in either the negative or the positive cluster. This rule considers the fact that if more than half of the games correlated to a MI dimension in each cluster are not using a particular game mechanic, there is not sufficient evidence that this

---

1 Available at: https://goo.gl/hGYRdM
game mechanic plays an important role in the preference (or lack thereof) for these games. However, if the majority of the games in either of the two clusters utilize the game mechanic, then it is reasonable to conclude that the game mechanic has an influence on the game preference (or lack thereof) of the players. For example, the results for the game mechanics from Table 5 can be seen in Table 6. The “Discovery” game mechanic for the logical-mathematical dimension is used in 4 out of 6 games in the positive cluster, and in 2 out of 4 games in the negative cluster. This means that at least half of the games in each cluster utilize the game mechanic “Discovery”, and therefore the mechanic is related to the dimension. For example, the mechanic “Motion” (i.e. the players’ bodily stances (postures, gestures, etc.) produce input to the game system or benefit in dealing with its challenges) turns out not to be related to the logical-mathematical dimension: only 2 out of 6 game utilize this mechanic in the positive cluster, and no game utilizes it in the negative cluster.

Table 6: Example of the decision protocol for a relationship between the “discovery” mechanic and the logical-mathematical dimension of MI

<table>
<thead>
<tr>
<th>Mechanics</th>
<th>Portal</th>
<th>Portal</th>
<th>Braid</th>
<th>Fable</th>
<th>Fallout</th>
<th>Xbox Fitness</th>
<th>Total weight</th>
<th>Wordament</th>
<th>Heavy Rain</th>
<th>The Sims</th>
<th>World Of Warcraft</th>
<th>Total weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovery</td>
<td>c (+2)</td>
<td>s (+1)</td>
<td>c (+2)</td>
<td>c (+2)</td>
<td></td>
<td></td>
<td>7</td>
<td>c (+2)</td>
<td>c (+2)</td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Epic meaning</td>
<td>s (+1)</td>
<td>s (+1)</td>
<td>c (+2)</td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>c (+2)</td>
<td>s (+1)</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Infinite gameplay</td>
<td>c (+2)</td>
<td></td>
<td>s (+1)</td>
<td></td>
<td>c (+2)</td>
<td></td>
<td>3</td>
<td>c (+2)</td>
<td>c (+2)</td>
<td>s (+1)</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Motion</td>
<td>s (+1)</td>
<td></td>
<td>c (+2)</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

If there is a relationship between a game mechanic and an MI dimension according to the previous rule, we distinguish between three types of relationships: “positive”, “dubious” (uncertain), and “negative” (see Table 7). To decide on the nature of the relationship we assigned a weight to the game mechanic for the positive and for the negative cluster and compared these two weights (see Table 6). The weight is calculated as follows: each time the game mechanic is used as a core mechanic in the cluster +2 is added and when it is used as a satellite mechanic +1 is added.

If the total weight in the positive cluster is larger than the one for the negative cluster by at least 2, that game mechanic is declared to have a “positive” relation with that dimension. If the total weight for the negative cluster is larger than the one for the positive cluster by at least 2, the game mechanic is declared to have a “negative” relation with that dimension. In the other case (i.e. equal weights or the difference is at most one) the relationship is called “dubious”, meaning that we do not have evidence that is mechanic is either strongly positively or strongly negatively related to that dimension.

Table 7: Results of the decision from Table 6

<table>
<thead>
<tr>
<th>Mechanics</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovery</td>
<td>Positive</td>
</tr>
<tr>
<td>Epic meaning</td>
<td>Dubious</td>
</tr>
<tr>
<td>Infinite gameplay</td>
<td>Negative</td>
</tr>
</tbody>
</table>

For example, the “Discovery” game mechanic is related to the logical-mathematical dimension. This game mechanic plays three times the role of “core” and once the role of “satellite” in the positive cluster, giving it a total weight of 7 in this cluster; it plays two times the role of “core” in the negative cluster, giving it a total weight of 4 in this cluster. Therefore, this game mechanic is declared to have a positive relation with the
logical-mathematical dimension. Using this protocol, the game mechanic “Infinite gameplay” has a negative relation with logical-mathematical dimension, and the game mechanic “Epic meaning” has a dubious relation with this dimension. The decisions for the mechanics given in Table 6 are given in Table 7.

5.2 Results

Our analysis showed that within our sample, the 8 MI dimensions have relationships with 116 different game mechanics. These mappings can be used to help game designers to select game mechanics when creating games for a specific audience (i.e. player-centred game design). A positive relationship indicates that players with the particular intelligence will generally respond positively to the game mechanic. As a consequence, we can recommend the use of this game mechanic to enhance the game experience of players with this particular intelligence. Game mechanics with a negative relationship to a particular MI dimension evoke mostly negative responses and therefore it is recommended not to use them when targeting players with this particular intelligence. Dubious relationships point towards a fairly equal mix of positive and negative responses. Therefore, game mechanics with a dubious relationship can be used for players with this particular intelligence but require extra caution. As an example, the “quick feedback” mechanic has a positive relationship with most of the MI dimensions, indicating that it can be recommended for all dimensions. On the other hand, the “infinite gameplay” mechanic has a negative relationship with logical-mathematical, linguistics and intrapersonal dimensions, indicating that it is recommended not to incorporate this mechanic in games designed for players with high levels for those intelligences. The “helping” mechanic has a dubious relationship with two dimensions. This means that its incorporation in games designed for those dimensions is neither encouraged nor discouraged. If more than one dimension is targeted, the choice for using a dubious mechanic can be based on the type of relationship that this mechanic has with the other dimension(s). For example, the game mechanic “reaction time” has a positive relationship with the logical-mathematical, and a dubious relationship with the bodily-kinesthetic dimension. If a game is to be designed targeting both these dimensions, this mechanic could be used (Sajjadi, Vlieghe, et al. 2016a).

5.3 Recommendation tool

The mappings between 8 MI dimensions and 116 game mechanics is a large amount of data. In order to facilitate its use for game designers and researchers, we developed a kind of recommender tool. The tool visualizes the mappings between MI dimensions and game mechanics. The user can filter on one or more MI dimensions. The “concept network” visualization technique is used for this and implemented using D3js (https://github.com/d3/d3). This visualization is shown in Figure 3. A MI dimension is represented by a node in the centre, where the different game mechanics are the surrounding nodes and are linked to the dimension when there is a relationship. The user can hover over the MI dimensions to highlight (in blue) the relationship. In order to see the type of relationship between a dimension and the correlated game mechanics, the user can click on a dimension node. This generates a different visualization that provides the definition of that dimension and the type of the relationships (positive, negative or dubious) using 3 different colours (green for positive, yellow for dubious, and red for negative) (see Figure 4). The user also has the ability to click on a game mechanic node. This generates a visualization that provides the definition of that mechanic and the relationships it has with the different MI dimensions.

Since the number of possible relations between each MI dimension and game mechanics can be very large, a classification of game mechanics (10 classes) (Figure 5) has been introduced, which can be used as a filtering mechanism. Each class is represented by a different colour. For example, an orange dot is used for mechanics of the class “motivation” (see Figure 4). This feature makes it easier to recognize game mechanics of a particular type. In addition, if a game designer is looking for mechanics related to motivation, he can use the filter to only receive mechanics of this class.

Lastly, a game designer can select the game mechanics he wants to include in his design. From this selection, a report can be generated providing an overview of the nature of the relationships with the selected dimensions. An example report is given in Figure 6. This report helps the game designer to detect possible conflicts that could arise as a result of targeting more than one MI dimension. For example in Figure 6 we see mechanics that have a positive relation with one of the selected dimensions, but a negative one with the other selected dimension. Note that it is up to the designer to decide whether to replace the mechanics by more suitable ones, or to keep them when there are good reasons for this, e.g. because they are essential for the game or goal under consideration.
6. Discussion and conclusions

In this paper we addressed the question whether the theory of MI could be useful for the process of player-centred game design, and if so how to do this. We tackled this question by first investigating the existence of empirical relationships between the different dimensions of MI and preferences for games. This investigation was performed by means of a survey study. Using factor analysis (FA) on the data obtained from the survey, we managed to find patterns between MI dimensions and game preferences. This result provided us with enough incentive to hypothesize the existence of correlations and test for their significance levels. Based on the results of bivariate correlation analyses, we established empirical evidence for the existence of correlations. Furthermore, our results indicated that the theoretical mappings suggested in the literature (Becker 2007; Starks 2014) could be refined.

To understand how the theory of MI could be used in the process of player-centred game design, recognizing the cause of the observed correlations was imperative. Therefore, we started by considering the genres of the games as possible cause. A comparison between the analysis of the implicit and explicit preferences for game genres revealed that the genres of games on their own are too limiting to explain the common characteristics shared among the game titles that correlate to certain MI dimensions. Therefore, we decided to investigate
the fundamental building blocks of games, also known as game mechanics, as a possible explanation for the
correlations. Following a protocol, relationships between game mechanics and the different MI dimensions
were drawn. These relationships, i.e. mappings, can be positive, negative or dubious. These mappings can be
of great value for player-centred game design, as they indicate what game mechanics are appropriate or not
when designing games that target players with specific dominant MI dimensions.

Based on the results presented in this paper, we can conclude (1) that gamers’ MI dimensions and their game
preferences have multiple strong relationships; (2) that the relationships between MI dimensions and game
preferences cannot be explained only by game genres; and (3) evidence-based mappings between game
mechanics and the different MI dimensions exist, which can facilitate the use of the MI theory for player-
centred game design. Note that these mappings could also be used in other contexts than player-centred
game design, e.g. to dynamically adapt learning games to the player or to stimulate the development of
certain intelligences.

7. Limitations and future work

Although our study was designed with the utmost care, inevitably, there are some limitations. Firstly, we
recognize that our selection of game titles represents a snapshot of the current landscape of popular video
games, and that any selection unavoidably influences the outcome of the study. However, to minimize any
effect, we carefully selected a broad range of games. Naturally the list should be updated for future studies.
Furthermore, although the focus and intended application of our findings is in the context of games for
learning, a compilation of commercially available entertainment games have been used in the survey. This
could be seen as a limitation, but the use of learning games instead of popular commercial games would have
certainly reduced participation, as these games are not as widely known and played. For future research it
would be interesting to perform a similar survey (probably with a smaller sample size) to investigate whether
the findings obtained from the commercial games also apply for learning games.

Secondly, we acknowledge the risk of bias associated with self-evaluation. We are convinced, however, that
this was the best approach given the subject of our study, the sample size and the resources. Additionally, the
instrument used for measuring the MI dimension levels of the participants was an already existing and
validated instrument that has also been used in other research works.

Thirdly, we admit the subjectivity associated with the analysis concerning the game mechanics used in a game
and their role, and the proposed protocol for establishing the mappings. In future work, the sensitivity of the
results due to this subjectivity should be investigated, e.g. by asking different people to perform the same
analysis and by trying out different protocols for establishing the mapping.

The results presented in this paper enable game designers to take the dominant MI dimensions of their players
into account when designing (learning) games. Whether doing so would create the desired effect, i.e.
positively influencing the game experience and learning outcome of the intended audiences, is a question that
needs to be addressed in future research. Already two experiments were conducted in this context and

References

Becker, K., 2007. Pedagogy in commercial video games. Games and simulations in online learning: Research and
development frameworks, Hershey, PA: Information Science Publishing.
rebel psychologist faces his critics, pp.73–94.
factor or multiple intelligences? Intelligence, 38(5), pp.481–496. Available at:
Chan, D.W., 2005. Perceived Multiple Intelligences and Learning Preferences Among Chinese Gifted Students in Hong Kong.


 Erick Schonfeld, SCVNGR’s Secret Game Mechanics Playdeck. Available at: <http://goo.gl/e5G1qs> [Accessed March 21, 2016b].


Lim, T. et al., 2015. The LM-GM framework for Serious Games Analysis.


Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3910127/>.


An Approach to Scoring Collaboration in Online Game Environments

Claire Scoular, Esther Care and Nafisa Awwal
University of Melbourne, Australia
c.scoular@unimelb.edu.au
e.care@unimelb.edu.au
n.awwal@unimelb.edu.au

Abstract: With technological advances, it is now possible to use games to capture information-rich behaviours that reveal processes by which players interact and solve problems. Recent problem-based games have been designed to assess and record detailed interactions between the problem solver and the game environment, and thereby capture salient solution processes in an unobtrusive way (Zoanetti, 2010; Bennett et al., 2003; Shute & Wang, 2009). The Assessment and Teaching of 21st Century Skills project used an innovative approach to capture these processes through responses (activities and communication patterns) within collaborative games (Griffin & Care, 2015). Game player response patterns are identified as behaviours that are indicative of skills and are captured in real-time game play within time-structured log files. The analysis of these log files allows for inferences to be drawn in regard to the efficiency and quality of player performance. A concern with this approach is that game development for this purpose is complex, time consuming and expensive, with unique scoring applied to each game. This paper presents another approach that identifies, across games, common behaviours. A systematic scoring system for assessing player behaviours in games could provide access to useful data not only from new games but from existing games. This paper presents such an approach using collaborative games situated in a problem-solving context.

Keywords: Collaboration, problem solving, online assessment, log stream data, measurement, e-learning

1. Introduction

It is unsurprising that references to 21st century skills have started many debates over what and how today’s students should learn in order to become productive citizens (Kim & Shute, 2015). Lists of 21st century skills typically include problem solving, critical thinking, decision making and collaboration (Binkley et al., 2012). While these skills are not new, their relevance is greater than ever in this information age, where dissection of information, rather than just consumption of information, is critical. The increasingly rapid spread of sophisticated technologies into every facet of society is causing significant shifts, including how educational systems should be structured to prepare students effectively for life and work in the 21st century. Knowledge workers, now confronted by complex problems, need to be able to think systemically, creatively and critically, and at the same time be able to apply extensive problem solving and complex communication (McClarty et al., 2012) skills. Continuing to provide the same types of education to students as the world continues to change will not serve them well. The skills and knowledge exposed by traditional education are no longer seen as adequate preparation for success in life. Moreover, traditional assessments often fail to assess complex skills or measure what students can do with the knowledge and skills acquired through education (Shute & Kim, 2014).

2. Embedding assessments in games

How and what we learn is rapidly changing and researchers argue that games possess unique affordances to address complex problem-solving skill development that our current educational system is failing to provide (OECD, 2013). Assessment generally refers to the process of gathering information about competencies and attributes (Kim & Shute, 2015). Games can be powerful vehicles to assess competencies and dispositions, such as persistence, flexibility, creativity, self-efficacy, critical thinking, systems thinking, openness, problem-solving and teamwork, which can positively impact student academic achievement and other aspects of life (Delacruz, 2011; Zein & Diab, 2015; Care et al., 2015; Wang et al., 2015; Deniz et al., 2014; Kim & Shute, 2015). Technologies, as well as educational and psychological measurement approaches, have advanced in recent times (Wang, Shute, & Moore, 2015). Technologies allow us to embed assessments into games by extracting multifaceted evidence from a learner to infer competency levels across a network of skills, addressing what the person knows and can demonstrate (Shute & Kim, 2014).
One area of promise is a movement toward the use of well-designed games to provide meaningful student assessment environments that require the application of various complex competencies. Traditional assessments, such as paper and pencil multiple choice tests, do not necessarily capture the dynamic and complex nature of 21st century skills. Many researchers are exploring the possibility that games are well suited to improve instruction and differentiate learning while also providing more effective and less intrusive measurement than traditional assessments can offer (McClarty et al., 2012). Gee (2007) argues that if games are designed appropriately, they will stimulate higher-order thinking and collaborative activities among participants. Games are likely to address 21st century skill development through active interactions, collaborative experiences and complex problem solving (Schifter, 2013). Thus, embedding an approach to scoring within games provides a way to monitor participants’ level of competency and provide useful feedback (Shute & Kim, 2014). Assessment is a process of drawing reasonable inferences about what a person knows by evaluating what they do, say, make or write in a given situation (Griffin & Care, 2015). Games can be designed to increase in complexity and cognitive demand as play progresses, such that gameplay can be used for assessment purposes (Delacruz, 2011; Gee, 2007).

The use of games and simulations in assessment and learning is expected to increase over the next few years (McClarty et al., 2012). Additionally, assessments in this format have the advantage of providing teachers with tools that help them tailor instruction to the needs of students at different achievement levels (Mayrath, Clarke-Midura, Robinson, & Schraw, 2012). Test takers’ skills in these virtual environments can be automatically assessed by tracking their behaviours. Moreover, analysis of log files can highlight the strengths and weaknesses of game design (Ifenthaler et al., 2014; O’Neil, Chuang, & Chung, 2004).

The use of games in education so far reports mixed results, despite theoretical endorsement of its benefits. Design, development and implementation of games is expensive, time-consuming, resource-demanding and challenging (Awwal, Alom, & Care, 2015), and the goals of games do not necessarily align with learning or assessment goals: ostentatious game features can distract players from their task. A scientific attitude with regard to the design of educational games requires validated measures of learning outcomes and associated assessment methods in order to determine which design elements work best, when, and why (Ifenthaler et al., 2014). Scholars from various disciplines have found that well-designed games require problem solving skills, persistence, and creativity (Gee, 2007; Prensky, 2004; Shaffer, Squire, Halverson, & Gee, 2005); these are competencies critical to success in the 21st century, but to date are not explicitly supported by education systems. With appropriate design, student learning of skills through games can be promoted (Moreno-Ger, Burgos, Martínez-Ortiz, Sierra, & Fernández-Manjón, 2008).

3. Measuring collaboration

Theorists have for decades emphasised the importance of social interaction during learning (e.g. Vygotsky, 1978; Dillenbourg, 1999). From a learning perspective, the benefits of working collaboratively lie in the process rather than the end goal. If an individual is faced with a problem, they may have ways and means to solve it alone. However, if a student collaborates to solve a problem they enter into an iterative process in which they engage with others, understand perspectives, and regulate social interaction. A collaborative approach provides different learning opportunities than working individually. However, it is reasonable to presume that it is in the individual’s interests to learn how to work alone as well as to collaborate.

Game theory has been based in two long-standing categories of people working together: competitive and cooperative. Marschak and Radner (1972) recognised that a team is an organisation in which players can possess different information and skills, and individuals need to work together to maximise a joint outcome. Recent research places competition, cooperation and collaboration on a spectrum with competition and collaboration at opposite ends, determined by whether players have a shared goal (Zagal, Rick, & Hsi, 2006). In competitive games, participants typically work against one another, whereas in cooperative games they work together. The key difference in collaborative games is that there is a necessity for players to work together, with each player contributing uniquely to the game. Collaborative games can be designed to reduce any competitiveness players may initially possess. This can be addressed by providing shared goals and ensuring dependency between players through division of information and resources. Such strategies may constitute effective learning mechanisms. Cooperation, collaboration, and competition are combined in real life, where workers at times understand that they need to work together in order to achieve their goals, notwithstanding that they may also be competitors.
Critical to the integration of collaboration into current education systems is development of tools and strategies for assessing the skill. There is recognition that collaborative assessment requires human-human interactions, complex problems that cannot be solved by a single individual, and diverse sets of knowledge, skills and experience (Hesse et al., 2015). Assessing individuals' collaboration while also ensuring a reliable, valid, and fair assessment of their skills represents a major advance in educational measurement (von Davier & Halpin, 2013). Measuring less explored skills such as collaboration therefore requires methods of assessment design and psychometric analysis that may not be contained within traditional methods of assessment. In order to assess 21st century skills such as collaboration, inferences about student ability are required, and technology is providing ways to capture the relevant data from which these inferences may be drawn. Log stream data are a familiar feature of computer use, but have only recently been generated at scale to provide new insights and perspectives in education. Log stream data capture student activities and retain a time-stamped record as students work through an online game. Students can thus be scored in real time in an automated approach that reduces the time and effort required for teachers to observe student performance in the classroom.

4. Method

4.1 Design

The online games described harness log stream technology transforming them into assessments of collaboration. Behaviours such as action and chat communication are captured in real-time in a time-series log file, allowing for reasonable interpretations of such behaviours to be made in relation to a player's collaborative ability. The games presented act as a neat illustration of the similarity of the skills applied across collaborative games and therefore of the opportunity to capture similar behaviours. The method outlined was designed for a pilot study to test an approach to scoring involving common behaviours across games. The analysis presented in this paper involved three collaborative games developed by the ATC21S project: Laughing Clowns, Balance Beam, and Game of 20.

4.2 Participants

Each game was administered to 410 students (186 males and 224 females) across six schools in Victoria, Australia. The principal in each school was responsible for selecting which classes and, therefore, which students were involved in the trial. The principal selected classes primarily on the basis of availability for timetabled visits. The students were 12 to 15 years old (X = 14 years).

4.3 Materials

The games presented in this study were designed for two players to work together in an online environment. The players are provided a login and required to work through the assessment together. Their communication is facilitated by an embedded chat box. Players can collaborate on the game irrespective of physical location, provided they login simultaneously. Players cannot progress through the assessment independent of one another, and often cannot solve the problem without input from their partner. From the players' perspective, these assessments present as games. The log stream data generated allows for identification and scoring of player behaviours during game play.

In the Laughing Clowns game, illustrated in Figure 1, the players are presented with a clown machine each and 12 balls to be shared between them. The goal for the players is to determine whether the two clown machines work in the same way. In order to do this, the two players need to share information that is available to each of them and discuss the rules, as well as negotiate how many balls they should each use for problem resolution. This game requires players to find patterns, share resources, form rules and reach conclusions (Care et al., 2015).
The Balance Beam game, illustrated in Figure 2, is presented so that each player can interact with only one side of the beam and is unable to see their partner’s side. Both players are then provided with a limited number of weights. To make the process more collaborative, one player begins with all the weights, hence requiring sharing across players to progress through the task. Initially, weights must be passed from one player to the other; then each must place their weights in the correct position on the balance beam in order to achieve balance. Players need to apply and test rules in order to balance the beam, possibly leading to multiple correct solutions.

The Game of 20, illustrated in Figure 3, presents a mathematics based problem, in which the object is to work out the ideal numbers to reach a sum of 20 before the competitor, which in this game is the computer agent, reaches that sum. The two players must collaborate in order to win against the computer. Players need to take turns at selecting a number, with their aggregated number the one ultimately played. Instructions provided to the players include “Where will your team place the counter to be sure of winning the game?” and “Remember that the number of counter advances will be a combination of your number and your partner’s number”. These instructions make clear both the competitive and collaborative nature of the game – that together the players need to win against a third party – the computer.
4. Procedure

This paper outlines four behaviours identified as common across players and across each of the games. Each player completed all three games with the same partner within a classroom environment, taking approximately one hour. Log stream data was used to identify actions and communications undertaken and interactions between players. These observable behaviours are identified from the log stream data by employing search algorithms. The algorithms search the stream for the behaviours and score each behaviour as either present or absent for each player in each game. All activities and interactions that are possible within a game environment, if recorded systematically (records of player–game interactions), can measure salient solution processes in an unobtrusive way (Bennett et al., 2003; Zoanetti, 2010). These recorded detailed interactions between the problem solver and the game environment can be linked to each player’s level of proficiency and used to evaluate the process and efficiency with which problem solvers complete games (Pellegrino, Chudowsky, & Glaser, 2001; Williamson, Mislevy, & Bejar, 2006).

5. Results

The pilot study identified four common behaviours across the three games (Table 1). The percentages in the columns on the right refer to the frequency of each behaviour in each game, measured across all participating players.

<table>
<thead>
<tr>
<th>Number</th>
<th>Behaviours</th>
<th>Scoring</th>
<th>Clowns (%)</th>
<th>Balance (%)</th>
<th>Game of 20 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sends partner resource</td>
<td>0 (absent); 1 (present)</td>
<td>54</td>
<td>44</td>
<td>57</td>
</tr>
<tr>
<td>2</td>
<td>Presence of chat immediately before selecting final answer</td>
<td>0 (absent); 1 (present)</td>
<td>34</td>
<td>23</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td>Trials all resources</td>
<td>0 (absent); 1 (present)</td>
<td>34</td>
<td>55</td>
<td>62</td>
</tr>
<tr>
<td>4</td>
<td>Asks partner a question</td>
<td>0 (absent); 1 (present)</td>
<td>75</td>
<td>71</td>
<td>74</td>
</tr>
</tbody>
</table>

Behaviour number one identifies the percentage of players sending their partner a resource. This is an important component of collaboration, since to work together players need to identify and share each other’s resources. Further, the asymmetrical nature of the games makes this sharing critical for solving the problem. Figure 4 provides an example of the log stream data from a pair of players in the Balance Beam game. This excerpt shows the data used to identify behaviour number one. The identification of the ‘pass resource’ data event, as indicated by the green arrow, represents the passing of a resource from one player to another. In this example, it is the passing of the 300g mass from player A to player B. To record the presence of this behaviour, the search algorithm for behaviour number one searched the log stream data for the presence of a ‘pass resource’ data event. If present, a value of 1 was recorded for that player against that behaviour. If absent, a 0 value was recorded. Instances of sending resources were identified in all three games.
Figure 4: Example of log stream data from the Balance Beam game showing the ‘pass resource’ data event.

Table 1 presents the percentage of players for whom the presence of each behaviour was recorded in each game. Recording a present behaviour, resulting in a ‘1’ being recorded for that player in that game on that specific behaviour, is much like being scored correctly on a traditional test: for the purpose of assessment, behaviours are like test items. The frequencies of percentages correct (or, in this case, presence of a behaviour) can be taken as a proxy measure of the perceived difficulty of that behaviour (Adams et al., 2015). If a large proportion of players demonstrate a particular behaviour, this suggests that behaviour was easier to complete than those behaviours demonstrated by only a small proportion of the players. In the case of behaviour number one, the frequency appears reasonably consistently across the three games, at a mid-range (between 44-57%).

Behaviour number two identifies the presence or absence of chat immediately before the final answer in the game is selected. Discussion of suggestions or answers to the problem before entering them suggests good collaboration. From a game design perspective, there needs to be an opportunity to select a final answer to make this behaviour viable. As can be seen in Table 1, there is a relatively low frequency of players demonstrating this behaviour and this is consistent across games (between 23-34%).

Behaviour number three identifies whether players engage with every feature of the game. In problem solving, a systematic approach requires trialling, and in the collaborative context, this is made more complex by virtue of different resources being available to each player. In games of this kind, players benefit from pooling and trialling all joint resources. In the games demonstrated, some resources are redundant or of little use to the current activity, and players need to trial resources to assist in identifying the useful ones. Behaviour number three increases in frequency across the games suggesting that more players engaged with all of the features in the last game, Game of 20 (62%), than they did in the earlier game, Laughing Clowns (32%). It is possible players become more familiar with the interactive elements of the design in later games, or they become more confident in engaging with the game features.

Behaviour number four identifies whether a player asks their partner a question, indicated by the presence of key question words such as ‘what’, ‘why’ or ‘how’ in the chat box. Seeking more information from others is critical to building a shared understanding of the problem and will provide a foundation from which further collaboration can be built. The frequencies for this behaviour were consistently high across the three games (between 71-75%) suggesting that most players recognised the benefits of potential contributions from their partner.

6. Discussion

The approach taken in this paper highlights the ability to measure collaboration through games. In addition, it demonstrates that common behaviours can be identified across games to make inferences about player collaborative ability. Although the three games presented in this study demonstrated different problems and different requirements for collaboration, behaviours common to all three could still be identified. The identification of common behaviours makes scoring much more efficient, eliminating the laboriousness of scoring each game separately. Not only does the log stream data provide insight into what players are learning within the games, it provides an approach for assessing that learning.

Advances in technology have expanded the availability and opportunities for assessment game design. Computer-based performance assessment allows information to be captured, in relation to the processes and responses of the player (Bennett, Persky, Weiss, & Jenkins, 2007), that paper-based assessments cannot provide. This development also clearly possesses administrative advantages, such as automated scoring.
greater reach, instant feedback and a reduction in subjective scoring. The approach presented here allows for games to be transformed into assessments without the player experience or interface changing. This presents an engaging situation for students in the classroom and an opportunity to generate data for teachers to evaluate their students’ collaboration skills. Future work will focus on identifying additional common behaviours in games from which the same inferences might reasonably be drawn.

7. Conclusion

The approach outlined represents an innovation in assessment, moving from a model in which abilities are identified by direct responses to questions with correct and incorrect answers to a behavioural inference model. In this model, inferences about capabilities are made on the basis of individuals’ actions and communications in a game environment. The challenge for educators who want to employ games to support learning is to make valid inferences about what the students know and can do without disrupting their engagement with the game (Shute & Kim, 2014; Wang et al., 2015). The use of capture of actions through log streams provides this facility without students needing to consciously respond to assessment prompts.

References


www.ejel.org

ISSN 1479-4403


Motivating Factors and Tangential Learning for Knowledge Acquisition in Educational Games

Peter Mozelius\textsuperscript{1}, Andreas Fagerström\textsuperscript{2} and Max Söderquist\textsuperscript{2}
\textsuperscript{1}Department of Computer and Systems Sciences, Mid Sweden University, Sweden
\textsuperscript{2}Department of Computer and Systems Sciences, Stockholm University, Sweden
mozelius@dsv.su.se

Abstract: Game-based learning has been a strong emerging trend in the 21st century, but several research studies on game-based learning reports that the educational potential of games has not been fully realised. Many educational games do not combine learning outcomes with entertaining gameplay. At the same time as there is a tendency to digitise and personalise education by the use of digital games, there still exists a lack of knowledge about efficient educational game design. To identify design factors that are important for players’ learning motivation this study has analysed three popular entertainment games that were selected for their educational values.

The aim of the study is to explore, analyse and discuss, if and how motivating factors and intrinsic integration of knowledge in educational games might be related to players’ perceived knowledge acquisition. Test players with experience of the selected digital games were recruited from online gaming forums where a questionnaire also was used to collect data. Lepper’s and Malone’s set of heuristics for intrinsic motivation in interactive learning environments were used in a combination with Habgood’s and Ainsworth’s theory of intrinsic integration to examine the relationship between these factors in the educational games. Beside the direct acquisition of knowledge from gaming there was also an analysis of the concept of tangential learning.

Results from a \textit{t}-test showed that tangential learning was significantly more important for two of the tested games. Correlation analysis revealed several relationships between factors, where intrinsic integration was pointed out as particularly interesting for knowledge acquisition and tangential learning. Results showed weak or no relationships for Lepper and Malone factors, but with some tendencies for control, imagination and competition.

Keywords: Educational games, Intrinsic integration, Tangential learning, Game-based learning, Learning motivation

1. Introduction and Aim

There is a long tradition of using games in educational contexts and in the 1980s Game-based learning (GBL) began as a research topic (Habgood & Ainsworth, 2011). Educational games differ from traditional education in the sense that a game explicitly requires that the learner/gamer is active (Gee, 2003; Portnow, 2008). This can promote active learning instead of the traditional view of students as mere passive knowledge consumers. GBL has a potential to provide active learning instead of students as passive knowledge consumers (Gee, 2003).

However, despite the hype of GBL in the 21\textsuperscript{st} century, educational games do not seem to have reached their full potential to engage learners (Gunter, Kenny & Vick, 2008; Brusse, Neijens & Smit, 2010; Sigurdardottir, 2012).

One possible explanation to the failure might be that educational games do not have a gameplay that motivates gamers to play and learn. Papert (1998) has described the worst GBL artefacts as Shavian reversals, when games are inheriting the worst properties from both its parents in the creation of a boring e-book instead of a learning stimulating game (Wiklund, & Mozelius, 2013). Researchers’ advice is to design games more as platforms for autonomous learning than e-books and to integrate content in gameplay and game mechanics (Habgood & Ainsworth, 2011).

Another branch of GBL is to instead use Commercial Off-The-Shelf (COTS) games for learning purposes (Charsky, 2008; Wiklund, & Mozelius, 2013). According to Van Eck (2006) COTS games are suitable for educational contexts since they are affordable and well-designed by professional game developers. On the other hand COTS games need a thorough analysis before they can be integrated in curricula and course syllabi since part of the content can have a fictive nature (Van Eck, 2006).

In a study by Charsky (2010) it is reported that both the academia and the industry have a demand for serious and educational games that can provide something more than just entertainment. An identified problem is the...
lack of a thorough design framework (Charsky, 2010). Recommendations are that such a framework should be grounded in research on a combination of COTS and educational games, inspired by earlier work on motivational model for game construction by Lepper and Malone (1987) and Gee (2003). The aim of the study is to explore and discuss if and how motivating factors, intrinsic integration of knowledge and tangential learning in educational games might be related to students' perceived acquisition of knowledge.

1.1 Problem

The problem that this study have tried to address is that many educational games have failed with the intended combination to be entertaining and educational at the same time. A problem that has been identified to have its roots in the fact that the games do not meet the players goals and that they, in a classroom setting, are not efficient enough to replace traditional teaching methods and neither entertaining enough to attract learners (Kerawalla & Crook, 2005). In the increasingly digitised society today, it seems like a waste of resources to not use the identified motivating effect and learning potential of digital games (Gros, 2007; Prensky & Prensky, 2007; Mozelius, 2014).

Despite the strong interest among researchers as well as among consumers and learners it seems that the educational games that are designed and developed today do not grab the target audience's attention. One factor that partly can explain the absence of success might be that many educational games do not reach the same standards of graphics and game mechanics as players of COTS games are used to (Kerawalla & Crook, 2005). A theory on how educational games can be designed to stimulate motivation can be found in Lepper & Malone's (1986) taxonomy of intrinsic motivations for learning and ideas on intrinsic integration have been described by Habgood & Ainsworth (2011). However, despite several research studies on the design of educational games (Gee, 2003; Dickey, 2007, Brusse et al, 2010; Tang & Hannaghan, 2014), there still seems to be a gap between theories and ideas in research, and the actual implementation and construction of educational games.

1.2 Aim of the study

This study has explored the relationship between learners' perceived acquisition of knowledge in educational games, and the concepts of intrinsic integration and tangential learning. The study was based on three games that all are described in detail under 'Selected games' in the method chapter. Furthermore, the study aims to analyse how theories and design ideas from research may be applied in actual game construction. This is also a type of study that earlier have been asked for by game researchers (Belanich et al., 2004). Hopefully, the outcomes of the study should be of use for researchers and game designers interested in motivation factors, intrinsic integration and tangential learning.

1.3 Research question

How might motivating factors and intrinsic integration of knowledge in educational games have an impact on learner's perceived acquisition of knowledge?

2. Extended background

Game-based learning has great potential to give many people a chance to learn new things on their own terms. Today this potential is not fully realised, mainly because many games used in education is of poor quality, both in terms of entertainment and education. It is an endeavour to digitise and personalise education by using computer games but at the same time it is problematic to not have any guidelines for educational game design. By examining popular entertainment games this study seeks to identify and understand which factors are important for players' learning motivation.

2.1 Extrinsic and intrinsic motivation

Motivation can be divided into two main parts, extrinsic and intrinsic motivation (Cameron et al., 2005). Extrinsically motivated activities are activities where a person can achieve rewards or try to avoid punishment. These sticks and carrots are not directly related to the activities and examples of carrots in educational contexts are grades, certificates and diplomas (Mozelius, 2014). In digital games extrinsic motivation can be implemented by achievements and badges (Filsecker & Hickey, 2014).

Intrinsic motivation can be compared with gaming for gaming’s sake or learning for the sake of learning (Mozelius, 2014). Intrinsically motivated learners tend to be more aware of inconsistencies, complexities, and
unexpected possibilities (Kapp, 2012). A model that breaks down intrinsic motivation into two levels with seven components is Frank Lepper’s and Thomas Malone’s taxonomy of intrinsic motivation (Malone & Lepper, 1987). The levels and the components in the taxonomy of intrinsic motivation are:

The personal level
- **Challenge** in terms of goals, uncertain outcomes and performance feedback
- **Curiosity** in terms of sensory and cognitive inquisitiveness
- **Control** in terms of contingency, choice and power
- **Fantasy** with emotional and cognitive aspects interwoven

The interpersonal level
- **Cooperation** in terms of players working together to achieve goals
- **Competition** where players compete against each other to achieve goals
- **Recognition** in terms of making achievements available for others

(Malone & Lepper, 1987)

Dickey (2007) used the Lepper and Malone taxonomy in a study to analyse intrinsic motivation aspects of the online game *World of Warcraft*.

### 2.2 Intrinsic integration

An important factor for motivation in the design of educational games could be what Habgood & Ainsworth (2011) call **intrinsic integration**. Their claim is that the pedagogic or didactic quality of learning games is depending on how well learning content is integrated and interwoven in the gameplay. The problem might be that many educational games today separate the joyful gaming from breaks with mandatory ‘learning exercises’ to open up the next game level (Habgood & Ainsworth, 2011). An example of this is when the gamer should solve mathematical exercises by shooting at tiles to enter the right solution. Shooting at tiles is a game mechanic without any relevant alignment to the educational theme of solving mathematical problems.

Habgood & Ainsworth are not the only researchers advocating intrinsic integration, there exist several other studies conducted in the last decades that have emphasised the importance of integrating learning content with game mechanics and narration (Kafai, 1995; Ke, 2008; Klopfer et al, 2009; Ke, 2016). Furthermore, it has been pointed out that the extent to which content engagement is intrinsic to the main gameplay will influence gamers’ learning (Richards et al, 2013). However, there is no guarantee that the intrinsic design of ‘learning-play integration’ in the core gameplay necessarily develops any knowledge (Ke, 2016).

According to Jacob Habgood (2005) two basic guidelines for the implementation of intrinsic integration in games are to:

1. Include the learning content in the parts of the game that are the most fun to play without interrupting the flow experience.
2. Integrate the learning content in the game world and in the player’s interaction with it, to provide an external representation of the learning material that should be explored through the game mechanics.

### 2.3 Tangential learning

Educational games may not necessarily have to teach topics directly. In a video by Floyd and Portnow (2008) and in an article by Portnow and Floyd (2008), the concept of tangential learning is suggested as an alternative or complementary way for games to stimulate learning. The basic idea of the term is that a game should introduce a theme, a technique or a concept to inspire learners to further self-studies. Instead of direct teaching and learning a game should engage and stimulate learning by putting abstract knowledge in an attractive and engaging context.

Squire, DeVane & Durga (2008) explored the existence and potential of tangential learning in a study where lower secondary school students played Civilization III during a year. Civilization is a history game with a relatively high degree of realism where the player can follow a civilization from its beginning to present time...
(Squire, 2005). Participants could be described as low and average performing students with low commitment to traditional History classes. By playing Civilization they started to seek information outside school and outside the actual gaming. During the year there was also an improvement in the formal subject grades (Squire, DeVane & Durga, 2008).

Several studies indicate that games have enormous potential to encourage tangential learning (Portnow & Floyd, 2008; Porcino et al, 2014; Rath, 2015), but there are also several identified obstacles. As highlighted by Robert Rath (2015) some problems to address are the lack of learning structures, online misinformation and players who are poor at source criticism. His method to handle these issues is an ‘explanatory criticism’ that should provide vetted information in a didactic structure that gives the learner a ground to start his or her own learning (Rath, 2015).

3. Methods for data collection and analysis

To understand what makes some educational games more successful in terms of transferring knowledge to the player than others, a questionnaire was distributed via online fora to the players of a number of educational entertainment games. The games were chosen because they contained a significant amount of information relevant for educational contexts, and that they have players active online in Internet fora.

The purpose of the questionnaire was to collect data about how players perceive the game and measure how much the players learned from playing the game. Data was collected and used to measure how well the motivating factors described by Lepper and Malone (1987), and Habgood and Ainsworth (2011) correspond to the user’s perceived increase of knowledge.

3.1 Method of analysis

In order to determine whether the results obtained from the survey are statistically significant two methods are used, a T-test to determine differences between the results and a Pearson correlation analysis to determine the correlation between pairs of values.

The mean of the results are first analysed game by game. According to Denscombe (2010), the use of a median value be more suitable when working with ordinal data, where average values are more affected by outliers in the results. The problem with using the median value is that it allows little mathematical analysis.

There is a small risk that some extreme values affect the results when working with mean values, although in this case the number of respondents were so high that even very extreme values would have little impact. In order to better determine whether the mean values are actually saying something, the standard deviation of the results needs to be obtained. Mean values and standard deviations of the results were compared to see if some variables are significant.

By comparing values from the different games it can be determined if one is more successful at mediating knowledge than another. The independent variables that affect motivation and integrated knowledge can then be examined to see if any other variable is particularly strong in the game(s) that seems to be better at mediating knowledge. It will also be examined whether any of the factors that are prominent in the game has a weaker presence in the other games, and would therefore be able to explain a reduced perceived amount of learning for the respondents.

3.2 Selected games

Kerbal Space Program

Kerbal Space Program is a game developed by the Studio Squad, where the goal is to build spacecrafts, send them into orbit and visit moons and other planets. The game features a realistic physics engine and can simulate the physical laws that exist in our solar system (Hall, 2014). The company Teacher Gaming has made a modified version of Kerbal Space Program under the name KerbalEdu, a light version of the original game developed specifically for learning in classroom settings (Teacher Gaming, 2015).

NASA recognised the game’s ability to spread interest in space research, and in 2013 went into a collaboration with Squad to develop an addition to the game where asteroids can be captured and placed in orbit, a project that NASA plans to perform in 2025 (The Guardian, 2015). Kerbal Space Program won the 2014 Game
Developers Choice Awards, appointed by the players as the year’s games in the category Audience Awards (Game Developers Choice Award, 2015).

Crusader Kings
Crusader Kings II, is a grand strategy game developed by Paradox Interactive. It allows the player to experience and control large parts of the Middle Ages in Europe, the game has been described as one of the more historically accurate in the genre (The Public Medievalists, 2014). The game has a large player base and has influenced players to seek further knowledge about the events that occur in the game.

Civilization
Civilization is a series of grand strategy games developed by Firaxis Games, where the player gets to control a civilization from its beginnings to modern times. The games possess a high degree of realism, which makes it possible to acquire knowledge about historical events (Battersby 2010). Furthermore, Civilization games have a so-called Civipedia containing information about the civilizations that the player can control. An idea of including Civiipedias is that players should be able to gain real world knowledge and thus make them more involved in the game.

3.3 The survey
Slightly different surveys were sent out depending on the game, where the game’s main theme is represented by either one of three subjects: history, astrophysics or politics. The word game in brackets will be replaced with the name of the game being examined. The questions are designed by the study authors to measure the variables studied: knowledge acquisition, tangential learning, intrinsic integration and the Lepper and Malone factors. Each issue also has an associated, optional input field for comments, where additional comments can be collected to more clearly describe the respondent’s opinions.

- The first four questions are census questions regarding age, gender and education
- The first three Likert-scale questions deal with prior knowledge
  - The question regarding prior knowledge of the subject is to verify that past knowledge or previous interest is not affecting acquired knowledge since previous knowledge is an underlying variable that can affect the outcome.

I knew a lot about (history/astrophysics/politics) before I started playing (game).

- The question regarding gained knowledge measures how much the respondent experience that they have learned from playing the game. Perceived acquired knowledge is one of the dependent variables and is believed to be affected by the motivational factors.

I have learned a lot about (history/astrophysics/politics) from playing (game).

- The question regarding having sought knowledge on their own outside the game is to measure the occurrence of tangential learning. Tangential learning is the second dependent variable, and is believed to be affected by the motivational factors.

I've felt motivated to seek out additional information about (history/astrophysics/politics), either outside or in the game, when it was not necessary for completing goals in the game.

- The fourth Likert-scale question examines how well the respondents feel that knowledge is integrated into the gameplay. This question is based on Habgood and Ainsworths (2011) study which showed that a high degree of knowledge integration facilitates learning.

I feel that information about (history/astrophysics/politics) is very well integrated within the game mechanics of (game).
• The last seven questions of the questionnaire explores the respondent’s perception of the game in regards to Lepper’s and Malone’s (1987) heuristics: challenge, curiosity, control, imagination and cooperation, competition and recognition.

I would describe (game) as challenging in regards to finishing goals and progressing.
The amount (or lack of) information available to me in game was enough to keep me curious.
When playing (game) I feel in control of choosing my own strategies and controlling the mechanics of the game.
I have felt immersed in (game), in that I can either relate to the characters or feel like I’m a part of the game.
I rely on cooperation with other people, either in or outside the game, to understand the game better and finish goals.
I compete with other players of (game), in measuring our in game accomplishments.
Receiving recognition from other players of (game) is important for my motivation to continue playing.

4. Results

The questionnaire had a total of 635 respondents (N = 632), where players of Kerbal Space Software accounted for 442 of the responses, Crusader Kings accounted for 145 of responses and Civilization constituted 45 of the answers. The average age of the respondents overall was 21 years, and evenly distributed between games with the average age of Kerbal Space Program players being 21 years, Crusader Kings 22 years and Civilization 21 years.

Table 1: Respondents’ average age

<table>
<thead>
<tr>
<th>Game</th>
<th>Respondents</th>
<th>Respondents %</th>
<th>Average age</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kerbal Space Program</td>
<td>442</td>
<td>70%</td>
<td>20.94</td>
<td>7.17</td>
</tr>
<tr>
<td>Crusader Kings</td>
<td>145</td>
<td>23%</td>
<td>22.47</td>
<td>7.55</td>
</tr>
<tr>
<td>Civilization</td>
<td>45</td>
<td>7%</td>
<td>21.45</td>
<td>8.85</td>
</tr>
<tr>
<td>Total</td>
<td>632</td>
<td>100%</td>
<td>21.33</td>
<td>7.37</td>
</tr>
</tbody>
</table>

The respondents were mostly male, a small number of females and even fewer people that would not disclose their gender or identified as intergender participated.

Table 2: Gender distribution

<table>
<thead>
<tr>
<th>Game</th>
<th>Male</th>
<th>Female</th>
<th>Intergender</th>
<th>Unspecified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kerbal Space Program</td>
<td>97%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>Crusader Kings</td>
<td>93%</td>
<td>4%</td>
<td>0%</td>
<td>3%</td>
</tr>
<tr>
<td>Civilization</td>
<td>91%</td>
<td>7%</td>
<td>0%</td>
<td>2%</td>
</tr>
<tr>
<td>Total</td>
<td>96%</td>
<td>2%</td>
<td>0.5%</td>
<td>1.5%</td>
</tr>
</tbody>
</table>

Table 3 illustrates the respondents’ level of education, which is close to equal for all games with a majority of respondents without any experience of higher education.

Table 3: Respondent’s education level

<table>
<thead>
<tr>
<th>Game</th>
<th>Secondary school</th>
<th>Upper Secondary school</th>
<th>Tertiary education</th>
<th>University Degree</th>
<th>Postgraduate studies</th>
<th>Postgraduate Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kerbal Space Program</td>
<td>31%</td>
<td>17%</td>
<td>29%</td>
<td>14%</td>
<td>4%</td>
<td>6%</td>
</tr>
<tr>
<td>Crusader Kings</td>
<td>19%</td>
<td>14%</td>
<td>28%</td>
<td>21%</td>
<td>9%</td>
<td>8%</td>
</tr>
<tr>
<td>Civilization</td>
<td>31%</td>
<td>22%</td>
<td>27%</td>
<td>16%</td>
<td>5%</td>
<td>6%</td>
</tr>
<tr>
<td>Total</td>
<td>26%</td>
<td>17%</td>
<td>27%</td>
<td>16%</td>
<td>5%</td>
<td>6%</td>
</tr>
</tbody>
</table>

Table 4 shows the educational orientation of the respondents who have post-secondary education. The respondents are mostly technically oriented in their education, with some significant numbers in social sciences and humanities as well.
Table 4: Respondent’s educational orientation

<table>
<thead>
<tr>
<th></th>
<th>Natural Sciences</th>
<th>Technical</th>
<th>Medicine</th>
<th>Social Sciences</th>
<th>Humanities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kerbal S P</td>
<td>20%</td>
<td>62%</td>
<td>3%</td>
<td>7%</td>
<td>7%</td>
</tr>
<tr>
<td>Crusader K</td>
<td>9%</td>
<td>44%</td>
<td>4%</td>
<td>23%</td>
<td>20%</td>
</tr>
<tr>
<td>Civilization</td>
<td>8%</td>
<td>42%</td>
<td>0%</td>
<td>23%</td>
<td>27%</td>
</tr>
<tr>
<td>total</td>
<td>17%</td>
<td>56%</td>
<td>3%</td>
<td>12%</td>
<td>12%</td>
</tr>
</tbody>
</table>

Respondents perceived knowledge before playing differed between the games, where players of Kerbal Space Program showed a mean value of 2.5, while players of Crusader Kings and Civilization produced a mean value around 4.

4.1 Knowledge acquisition and Tangential learning

Mean knowledge acquisition shows that respondents feel they have acquired new knowledge from playing games, where Kerbal Space Program show a mean value of 5.29, Crusader Kings got a mean of 4.75 and Civilization 4.24. Regarding tangential learning players of Kerbal Space Program answered about the same as for the acquisition of knowledge, but players of Crusader Kings and Civilization valued tangential learning higher than acquisition of knowledge directly from the games with mean values of 5.31 and 4.87.

Table 5 shows the mean value and standard deviation of the respondent’s prior knowledge, knowledge acquisition, and the degree of tangential learning.

Table 5: Respondent’s prior knowledge, knowledge acquisition and tangential learning

<table>
<thead>
<tr>
<th>Prior Knowledge</th>
<th>std.dev.</th>
<th>Knowledge acquisition</th>
<th>std.dev.</th>
<th>Tangential learning</th>
<th>std.dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kerbal S P</td>
<td>2.53</td>
<td>1.38</td>
<td>5.29</td>
<td>1.01</td>
<td>5.25</td>
</tr>
<tr>
<td>Crusader K</td>
<td>4.02</td>
<td>1.35</td>
<td>4.75</td>
<td>1.2</td>
<td>5.31</td>
</tr>
<tr>
<td>Civilization</td>
<td>4.16</td>
<td>1.53</td>
<td>4.24</td>
<td>1.32</td>
<td>4.87</td>
</tr>
<tr>
<td>total</td>
<td>2.99</td>
<td>1.53</td>
<td>5.09</td>
<td>1.13</td>
<td>5.24</td>
</tr>
</tbody>
</table>

4.2 Summary of answers

Table 6 shows the mean value and standard deviation of the respondents’ answers to the questions regarding prior knowledge, increase of knowledge, tangential learning, integration of knowledge in the game, challenge, curiosity, control, fantasy, cooperation, competition and recognition.

Table 6: Mean values for motivating factors, prior knowledge and knowledge acquisition.

<table>
<thead>
<tr>
<th></th>
<th>Mean value</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior knowledge</td>
<td>2.99</td>
<td>1.53</td>
</tr>
<tr>
<td>Knowledge acquisition</td>
<td>5.09</td>
<td>1.18</td>
</tr>
<tr>
<td>Tangential learning</td>
<td>5.24</td>
<td>1.11</td>
</tr>
<tr>
<td>Intrinsic integration</td>
<td>4.57</td>
<td>1.13</td>
</tr>
<tr>
<td>Challenge</td>
<td>4.93</td>
<td>1.09</td>
</tr>
<tr>
<td>Curiosity</td>
<td>4.88</td>
<td>1.11</td>
</tr>
<tr>
<td>Control</td>
<td>5.20</td>
<td>0.92</td>
</tr>
</tbody>
</table>
5. Analysis

A t-test was performed to investigate whether there was any difference between how much respondents learned from the games directly and how much they learned through tangential learning via other sources. For Kerbal Space Program there was no significant difference with a p-value of .4255. For Crusader Kings tangential learning was significantly greater than acquisition of knowledge from within the game, where p < 0.0001. For Civilization tangential learning was also significantly greater, where p = 0.0067. For all the games combined tangential learning was significantly higher than direct knowledge acquisition, where p = 0.0038.

Table 7: Mean values for direct knowledge acquisition and tangential learning, and significance from the t-test of the difference between the two.

<table>
<thead>
<tr>
<th></th>
<th>Knowledge acquisition</th>
<th>Tangential Learning</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kerbal</td>
<td>5.29</td>
<td>5.25</td>
<td>0.4255</td>
</tr>
<tr>
<td>Crusader</td>
<td>4.75</td>
<td>5.32</td>
<td>&lt;0.0001**</td>
</tr>
<tr>
<td>Civilization</td>
<td>4.24</td>
<td>4.87</td>
<td>0.0067**</td>
</tr>
<tr>
<td>total</td>
<td>5.09</td>
<td>5.24</td>
<td>0.0038**</td>
</tr>
</tbody>
</table>

*Significance at 0.05 (two-tailed)
** Significance at 0.01 (two-tailed)

5.1 Correlation analysis

To investigate the relationship between the factors examined in the questionnaire a Pearson correlation analysis was performed for each combination of two values. Due to the high number of respondents there was a risk that even weak correlations yielded significant results. Therefore the analysis only treats correlations of around 0.3 or above and -0.3 and below as significant.

Prior knowledge showed low or non-significant correlations with other factors. A weak positive correlation was found between prior knowledge and direct knowledge acquisition for Civilization (r = 0.35), which could indicate that more knowledge leads to improved knowledge acquisition. However, this tendency was not found for the other games, which despite low r-values or insignificant results rather indicated a negative relationship where more prior knowledge would lead to less knowledge acquisition.

A weak positive correlation was found between knowledge acquisition and tangential learning (r = 0.38). This correlation implies a beneficial effect between learning from a game and seeking additional knowledge from outside the game. While it does not show which of the two affects the other, it does suggest that tangential learning should not be dismissed when discussing the educational potential of a game. A medium strong correlation was found between knowledge acquisition and intrinsic integration (r = 0.43) with the strongest relationship being for Civilization (0.58), which supports Habgood and Ainsworths (2011) theory that knowledge must be integrated into the game's mechanics to effectively be taught. For Lepper and Malone's factors for intrinsic motivation a few weak positive correlations were found. A weak correlation between knowledge acquisition and control (r = 0.26) was found. In Kerbal Space Program and Crusader Kings a weak correlation was found between knowledge acquisition and imagination (r = 0.27, r = 0.28), and for Crusader Kings there was also a weak correlation between knowledge acquisition and competition (r = 0.29). This indicates that the player's sense of immersion and control is important for learning. Competition may be important for some games depending on how competitive they are. Crusader Kings showed such a relationship.
while Kerbal Space Program, which lacks built-in elements of competition between players, does not show such a correlation.

**Table 8:** Correlations between knowledge acquisition and the studied factors.

<table>
<thead>
<tr>
<th>Corr w/ Knowledge Acquisition</th>
<th>Total</th>
<th>Kerbal</th>
<th>Crusader</th>
<th>Civilization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior knowledge</td>
<td>-0.16879**</td>
<td>-0.0805</td>
<td>-0.10005</td>
<td>0.34623*</td>
</tr>
<tr>
<td>Tangential</td>
<td>0.37878**</td>
<td>0.43151**</td>
<td>0.25255**</td>
<td>0.38391**</td>
</tr>
<tr>
<td>Integration</td>
<td>0.42654**</td>
<td>0.37694**</td>
<td>0.37625**</td>
<td>0.5827**</td>
</tr>
<tr>
<td>Challenge</td>
<td>0.21766**</td>
<td>0.16042**</td>
<td>0.16638*</td>
<td>0.2659</td>
</tr>
<tr>
<td>Curiosity</td>
<td>0.1716**</td>
<td>0.15175**</td>
<td>0.22092**</td>
<td>0.22076</td>
</tr>
<tr>
<td>Control</td>
<td>0.25911**</td>
<td>0.23944**</td>
<td>0.20213*</td>
<td>0.18868</td>
</tr>
<tr>
<td>Fantasy</td>
<td>0.19503**</td>
<td>0.26718**</td>
<td>0.27706**</td>
<td>0.1191</td>
</tr>
<tr>
<td>Cooperation</td>
<td>0.13566**</td>
<td>0.1551**</td>
<td>0.14496</td>
<td>0.10325</td>
</tr>
<tr>
<td>Competition</td>
<td>0.15301**</td>
<td>0.13481**</td>
<td>0.28585**</td>
<td>0.02116</td>
</tr>
<tr>
<td>Recognition</td>
<td>0.16802**</td>
<td>0.10809**</td>
<td>0.19423*</td>
<td>0.23018</td>
</tr>
</tbody>
</table>

*Significance at 0.05.
**Significance at 0.01.

Tangential learning showed overall fewer associations with the different factors than the direct acquisition of knowledge. Intrinsic integration showed a weaker positive correlation with tangential learning (r = 0.24) than for direct acquisition of knowledge (r = 0.38). A weak to medium strong correlation was found between tangential learning and curiosity for Crusader Kings (r = 0.33) and Civilization (r = 0.52), suggesting that the more curious a player, the more they are motivated to seek out additional knowledge on their own. Civilization also had a weak positive correlation between tangential learning and cooperation (r = 0.33).

**Table 9:** Correlations between tangential learning and the studied factors.

<table>
<thead>
<tr>
<th>Corr w/ tangential</th>
<th>Total</th>
<th>Kerbal</th>
<th>Crusader</th>
<th>Civilization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior knowledge</td>
<td>0.01258</td>
<td>-0.02421</td>
<td>0.09647</td>
<td>0.28494</td>
</tr>
<tr>
<td>Integration</td>
<td>0.24164**</td>
<td>0.26185**</td>
<td>0.22488*</td>
<td>0.13163</td>
</tr>
<tr>
<td>Challenge</td>
<td>0.07872*</td>
<td>0.07071</td>
<td>0.0387</td>
<td>0.2802</td>
</tr>
<tr>
<td>Curiosity</td>
<td>0.022148**</td>
<td>0.15087**</td>
<td>0.33337**</td>
<td>0.51835*</td>
</tr>
<tr>
<td>Control</td>
<td>0.18932**</td>
<td>0.21448**</td>
<td>0.17708*</td>
<td>0.09867</td>
</tr>
<tr>
<td>Fantasy</td>
<td>0.17133**</td>
<td>0.20801**</td>
<td>0.08413</td>
<td>0.04967</td>
</tr>
<tr>
<td>Cooperation</td>
<td>0.21318**</td>
<td>0.23648**</td>
<td>0.11794</td>
<td>0.33167*</td>
</tr>
<tr>
<td>Competition</td>
<td>0.13803*</td>
<td>0.14592*</td>
<td>0.11835</td>
<td>0.27608</td>
</tr>
<tr>
<td>Recognition</td>
<td>0.06709</td>
<td>0.06218</td>
<td>0.03613</td>
<td>0.20279*</td>
</tr>
</tbody>
</table>

*Significance at 0.05.
**Significance at 0.01.
6. Conclusion

Results showed that tangential learning accounts for a substantial part of knowledge learned for the observed games. The concept of inspiring players to effectively seek out knowledge and teach themselves is an exciting notion, and it may be a piece of the puzzle that’s been missing in educational game development. If tangential learning is proven as a viable means of education it may shift the role of educational games from being merely instructional to be more motivating and inspirational.

The most obvious, and expected finding, is that intrinsic integration of learning content seems to be a crucial design factor. This can also, like in the study by Habgood and Ainsworth (2011), be aligned to educational games potential to stimulate tangential learning. In two of the three tested games respondents perceived that they learnt more by tangential learning than by direct learning and a design idea worth further exploration is to construct games that includes a model or a module for further self-studies. An interesting example of this are the built-in encyclopaedias in the Civilization games where gamers can dig deeper for knowledge about historical events in the gameplay.

In games where knowledge has been successfully integrated, the game mechanics can serve as platforms for trying out and experimenting with ideas that have been acquired outside of the game. Imagine learning about avionics and how wing designs affect speed and manoeuvrability, or reading historical descriptions of the British longbow and how it changed the way battles were fought, and imagine having a tool to try out and experiment with those concepts, seeing how they play out and how changes made by the player effects the outcome.

The relationship shown between intrinsic integration and tangential learning further strengthens this. More research is needed testing the extent of tangential learning in games that either only motivates the player to learn more and games that require the player to gain knowledge to complete the game. An idea for future research is to implement and test this more specific in learning games on computer science where authors have the required domain knowledge. This should follow the recommendation from Portnow & Floyd (2008), to enable learning but to keep entertainment as the main goal of the games. Domain knowledge is important to not violate the concept of that players transcending from the game to the related content, should be introduced with quality content to ensure a sufficient knowledge acquirement (Portnow & Floyd, 2008).

7. Future work

What would be an interesting continuation is to test the games in a real educational setting and to compare students more formal mandatory gaming with voluntary gaming. In such a study students’ learning outcomes should be measured by pre- and post-testing.

The concept of tangential learning is still relatively new and warrants more research in exploring the concept and its relationship to direct knowledge acquisition in educational games. It would be interesting exploring if different types of knowledge and skills favour one method over the other. A deeper understanding of the relationship between intrinsic integration and tangential learning would also be interesting, mainly, if games that require knowledge from the player to become better at the game affects tangential learning, or if the power of tangential learning merely stems from the players growing interest in the subject.

This study will also be included as a part of the ResearchGate project ‘Learning to Program by Playing Learning Games’ where authors will try to apply the concepts of intrinsic integration and tangential learning in the field of programming education. For ResearchGate members the project can be accessed via: https://www.researchgate.net/project/Learning-to-Program-by-Playing-Learning-Games

To implement the concepts in game-based learning environments takes time and requires resources. One of the authors are currently part of a large scale project application where these ideas could be part of a three year collaboration between five universities with experience and expertise in the field of game-based learning for programming education.
References


The Effect of Age on Teachers’ Intention to Use Educational Video Games: A TAM Approach

Antonio Sánchez-Mena1,2, José Martí-Parreño1 and Joaquín Aldás-Manzano3
1Universidad Europea de Valencia, Valencia, Spain
2Universidad Europea de Canarias, Tenerife, Spain
3Universitat de València, Valencia, Spain
antonio.sanchezmena@universidadeuropea.es
jose.marti@universidadeuropea.es
joaquin.aldas@uv.es

Abstract: Educational video games (EVGs) are gaining momentum as a means of increasing students’ motivation in their learning process. Nevertheless, teachers might face several barriers that dissuade them from using educational video games in their courses. This study analyses factors affecting teachers’ behavioural intention to use educational video games in their courses using a Technology Acceptance Model (TAM) approach. The research model is tested via structural equation modelling (SEM) on a sample of 312 teachers in Higher Education institutions. Results suggest that perceived usefulness influences in a direct and positive way teachers’ intention to use educational video games. Results also suggest that perceived ease of use indirectly influences intentions through perceived usefulness. Age was found to moderate the effect of teachers’ perceived ease of use on perceived usefulness of EVGs. Regarding managerial implications, our findings highlight the importance of addressing specific Teacher Training Programmes focusing on teachers’ age and perceived usefulness of EVGs in order to encourage teachers to adopting this educational innovation in their courses. Limitations of the study and future research lines are also addressed.

Keywords: Educational Video Games; TAM (Technology Acceptance Model); Higher Education; Behavioural intention; Age

1. Introduction

Educational video games (EVGs) represent a great opportunity to motivate and to engage students in their learning process and are gaining momentum among teachers and educational researchers. In fact, game-based learning research has increased fivefold over the last five years (Martí-Parreño, Méndez-Ibáñez, and Alonso-Arroyo, 2016). Recent academic research on the topic has approached the subject from different angles including learning outcomes when using EVGs (e.g. Boyle et al. 2016), instructional effectiveness (e.g. Sitzmann, 2011), or citation analysis (e.g. Harman, Koohang & Paliszkiewicz, 2014) to name a few.

Academic literature suggests a wide variety of subjects in which a game-based learning approach to education has been successfully applied. These subjects include citizenship education (Lim and Ong, 2012), nanotechnology (Blonder and Sakhnini, 2012), energy education (Yang, Chien and Liu, 2012), health education (Sung, Hwang and Yen, 2015), veterinary education (De Bie and Lipman, 2012), Newtonian physics (Shute, Ventura and Kim, 2013) and language teaching (Reinders and Wattana, 2014) among others. Extant academic literature also suggests that game-based learning can be used to develop high order cognitive abilities such as problem solving skills (Oblinger, 2004; Klopfer and Yoon, 2005) and the so-called 21st Century skills such as teamwork, communication skills, and social/cultural skills (Romero, Usart, and Ott, 2015).

Although a game-based approach to education can be traced back to the sixties (Piaget, 1962), the benefits of using EVGs have been pointed out more recently (Prensky, 2001). These benefits include a higher learning motivation for digital natives who make an intense use of technology and digital interactivity and for whom traditional learning methodologies do not appeal or motivate anymore (Prensky, 2001). Video games also favour a trial-and-error process that makes mistakes recoverable (Hanus and Fox, 2015) and gives students the freedom to fail without fear when learning (Lee and Hamer, 2011). Moreover, video games provide immediate and frequent feedback (Kapp, 2012) and allow teachers to tailor difficulty progression that facilitates scaffolded instruction based on each individual student’s needs (Hanus and Fox, 2015). Students also benefit from the visual display of their learning progress for example through badges (Kapp, 2012) and can be motivated through competition for example through leaderboards (Camilleri, Busuttil, and Montebello, 2011).
Despite all these potential benefits of EVGs, a game-based approach to education has also been criticized because game-based learning competition can foster adverse effects in social interaction amongst students in the classroom (Brom et al., 2014). Gaming can also potentially cause anxiety and embarrassment among students (Henderson, 2005) especially when leaderboards are used and students ranking in lower positions feel they are performing badly if compared to other students in the classroom. It has also been stated that game-based learning often concerns repetition of cyclic content that provokes persistent re-engagement which tends to address lower level learning goals rather than higher level goals (Ma et al., 2007).

EVGs also represent a challenge for teachers who are not familiar with this type of educational innovation. Previous research found that teachers can face new technological developments in education as a threat and technological innovations can be a cause of much anxiety (Goodwyn, Adams, and Clarke, 1997). Factors such as stress or anxiety towards technological innovations along with uncertainty in the expected outcomes of technological innovations can act as barriers to teachers adopting technological innovations such as EVGs. However, it has been stated that teachers are the true agents of change in schools (Teo, 2008) and the use of EVGs depends largely on the acceptance by classroom teachers (Bourgonjon et al., 2013; Niederhauser & Stoddart 2001).

Thus, the main goal of this research is to explore teachers’ intention to use EVGs using the technological approach proposed by the Technology Acceptance Model or TAM (Davis, 1985). TAM (Davis, 1985) was developed to predict an individual’s likelihood of accepting a technological innovation. Because EVGs can be considered a technological educational innovation using TAM is a suitable framework to exploring teachers’ intention to use EVGs. In fact, a meta-analysis conducted by Sumak, Hericko, and Pusnik (2011) pointed out that TAM was the most popular theory among models exploring e-learning acceptance. Because age might affect attitude towards technological innovations the moderating role of age is also analysed.

This paper is structured as follows. First, a literature review of game-based learning and educational video games is presented. Second, we present TAM and set the hypotheses. Third, the method is explained. Then we discuss the results. Finally, conclusions, limitations of this study and future research lines are addressed.

2. Game-Based Learning and EVGs

A review of the literature clearly suggests that different constructs have been used to refer to game-based learning. These constructs include gamification and serious games. Gamification has been defined as “the use of game design elements in non-game contexts” (Deterding et. al. 2011, p. 9), and seems to broaden previous constructs such as serious games and educational games in the sense that a game (or video game) is not needed. Gamification focuses on game design and game mechanics, that is, contrasts “gamification” “against other related concepts such as serious games via the two dimensions of playing/gaming and parts/whole” (Deterding et. al. 2011, p. 5). Rather than using a (video)game in the classroom the teacher “makes the class itself a game” (Hanus & Fox 2015). Van Eck (2006) summarizes the use of video games in education in three main strategies: a) the use of commercial off-the-shelf videogames (COTS) that take advantage of the existence of contents in these games that can be used for educational purposes, b) the use of serious games – a type of video games developed with non-recreational purposes where learning is the primary goal–, and c) to make students build their own games allowing the development of problem-solving abilities, programming skills and game design skills. Examples of the use of commercial off-the-shelf videogames include the use of SimCity to strengthen leadership decision-making (Lin and Lin, 2014) and RollerCoaster Tycoon 3 which has been used to support student learning of systems thinking (Shah and Foster, 2014). Serious games have been defined as “video games (VGs) intended to serve a useful purpose” (Girard, Ecalle, and Magnant, 2012, p. 207) where the “useful purpose” is learning. Examples of the use of serious games include mass-market serious games like Where in the World is Carmen Sandiego? which was developed to teach geography (Sitzmann, 2011). Other examples of serious games are ETIOBE Mates which was developed to improve children’s nutritional knowledge (Baños et al., 2013) and ECOPET which was developed to educate learners to use home-energy conservatively (Yang, Chien, and Liu, 2012). One example of making students build their own video games as part of their learning process is the case provided by Yang and Chang (2013) where students designed digital games based on biology course content to increase retention of both course content and critical thinking skills.
3. Hypotheses development and research model

The Technology Acceptance Model or TAM (Davis, 1985) was developed to predict an individual's likelihood of accepting a technological innovation. More specifically, TAM was developed to predict computer-based systems acceptance such as email services. Since its development TAM has been widely applied to different technological innovation including mobile commerce (Wu, and Wang, 2005), internet banking (Lai and Li, 2005), and the adoption of mobile internet (Hong, Thong, and Tam, 2006). TAM has also been applied to different educational contexts such as online education (Ngai, Poon, and Chan, 2007) and mobile learning (Liu, Li, and Carlsson, 2010). One of the main goals of TAM was to identify the major motivational variables that mediate between system characteristics and the actual use of the system. Davis (1985) identified two major variables influencing attitude towards a given technological innovation: a) perceived usefulness and b) perceived ease of use. Both perceived usefulness and perceived ease of use are directly influenced by design features. Finally, attitude influenced the adoption of the system, i.e. whether it was used.

3.1 Perceived usefulness

Perceived usefulness is defined as ‘the degree to which an individual believes that using a particular system would enhance his or her job performance’ (Davis, 1985, p. 26). Hence, this variable measures the utilitarian dimension of adopting the technology on the basis that users expect that the use of the technology will facilitate them in accomplishing the tasks to be developed. Specifically, perceived usefulness refers to effectiveness at work, productivity understood as time saving and the relative importance of the system for the individual's work (Davis, Bagozzi, and Warshaw, 1989). Within our conceptual framework, perceived usefulness of educational video games can be conceptualized as the degree to which teachers believe that using educational video games would enhance their job, that is, help them to improve students’ learning (for example, students will learn faster). It is important to point out that we adopt a student-centric approach to perceived usefulness, this is, instead of asking teachers if they believe that educational video games enables them “to enhance their teaching” we ask them if they believe educational video games enhance students’ learning. The expectation-value model asserts that a relationship between perceived usefulness and attitude exists (Fishbein and Ajzen, 1975). Thus, the following hypothesis is posited:

H1: teachers’ perceived usefulness of educational video games has a positive influence on attitude towards educational video games.

3.2 Perceived ease of use

Perceived ease of use is defined as ‘the degree to which an individual believes that using a particular system would be free of physical and mental effort’ (Davis, 1985, p. 26). Hence, this variable can be used to measure both teachers' physical costs (e.g. time devoted to preparing gamified classes) and mental costs (e.g. switching from traditional teaching methodologies to new teaching methodologies). Perceived ease of use can affect attitude in two ways: perceived self-efficacy and instrumentality (Davis, Bagozzi, and Warshaw, 1989). Perceived self-efficacy refers to the expectancy of performing well when interacting with the system while instrumentality refers to the expectancy that performing well will lead to rewards. The easier the interaction within the system (for example EVGs) the higher an individual’s sensation of efficiency and control (Bandura, 1982). If teachers experience difficulties working with educational video games they will show negative attitudes towards educational video game. Thus, the following hypothesis is posited:

H2: teachers’ perceived ease of use of educational video games has a positive influence on attitude towards educational video games.

3.3 Perceived ease of use and Perceived usefulness

Perceived ease of use influences perceived usefulness because a system’s simplicity can improve results and a technology is perceived as being more useful if it is easier to use (Davis, 1985). Later studies focused on e-learning services show evidence that perceived ease of use has influence on perceived usefulness (Okazaki and Renda Dos Santos, 2012). Therefore, if teachers perceive educational video games easy to use they will perceive educational video games useful. Hence, the following hypothesis is posited:

H3: teachers’ perceived ease of use of educational video games has a positive influence on their perceived usefulness of educational video games.
3.4 Attitude and behavioural intention

Attitude is an individual’s positive or negative evaluation of a given object or behaviour (Ajzen, 1991) and includes feelings or affective responses. Attitude also refers to an individual’s general willingness to engage in a given behaviour. This attitude is the result of individuals’ beliefs concerning the behaviour, the results of that behaviour and the importance attached to such beliefs. Social psychology literature clearly suggests that attitude has two components: affective and cognitive (Bagozzi and Burnkrant, 1979). The affective component refers to what extent a person likes the object of his thoughts (McGuire, 1985) and measures the degree of emotional attraction to the object. The cognitive component refers to an individual’s specific beliefs about the object (Bagozzi and Burnkrant, 1985) and consists of a value-based assessment, judgment, reception or perception of the object (Chaiken and Stangor, 1987). Behavioural intention is defined as ‘an individual’s subjective probability that he or she will perform a specified behaviour’ (Fishbein and Ajzen 1975, p. 288) and is a better predictor of actual behaviour than attitude when an intention has been formed (Warshaw and Davis, 1985). Literature review clearly shows a direct and positive relationship between a person’s attitude towards an object or behaviour and that person’s behaviour (Brown and Stayman, 1992) so the following hypothesis is posited:

H4: teachers’ positive attitude towards educational video games will have a positive influence on teachers’ behavioural intention to use educational video games.

3.5 The moderating role of age

Older teachers are more experienced than newer teachers are and previous research found that prior experience influences teachers’ actions (Pajares 1992). Even for student teachers prior experience has been found to inform their beliefs about practice (Calderhead and Robson 1991).

Academic literature also suggests that age is a factor that might moderate teachers’ behavioural intentions, including teachers’ use of technology. In fact, Goodwyn et al. (1997) found that older teachers perceive Information and Communication Technologies (ICTs) as a threat and cause of anxiety. Hamari and Nousiaien (2015) found that age affected teachers’ perceived value of EVGs. Therefore, the following hypotheses are posited:

H5: age will moderate the effect of teachers’ perceived usefulness on attitude towards educational video games.

H6: age will moderate the effect of teachers’ perceived ease of use on perceived usefulness of educational video games.

H7: age will moderate the effect of teachers’ perceived ease of use on attitude towards educational video games.

Figure 1 depicts a graphical representation of the research model.
4. Method

An exploratory transversal study involving teachers serving in Higher Education institutions was used in this research. Data was gathered through an online questionnaire. Snowball sampling (Goodman, 1961) was used for selection of respondents in this study. Although snowball sampling is unlikely to obtain a representative sample because there is no real control of the snowball effect (Hall and Hall 1996), this form of sampling is often used in online questionnaires to target hard-to-reach population subgroups (Sadler, Lee, Lim, and Fullerton, 2010).

4.1 Sample

A total of 312 teachers serving in Higher Education institutions completed the online questionnaire. The age of the participants ranges between 26 and 65 years, with an average of 42.8 years and 52.4% are males.

4.2 Survey instrument

All items used to develop the questionnaire were adapted from existing scales previously validated in the academic literature: ten items were adapted from Davis (1985) to measure perceived ease of use (e.g. “It is easy for me to use educational video games in my classes”); ten items were adapted from Davis (1985) to measure perceived usefulness (e.g. “Overall, I find educational video games useful to learn”); three items were adapted from Taylor and Todd (1995) to measure attitude (e.g. “My attitude towards educational video games is positive”); and three items were adapted from Shimp and Kavas (1984) to measure behavioural intention (e.g. “My general intention to use educational video games in my classes in the future is high”). All questionnaire items were measured using a 5-point Likert-type scale where (1) = strongly disagree, and (5) = strongly agree.

4.2.1 Measurement model validation

As shown in tables 1 and 2, the measurement model shows evidence of reliability and convergent validity as composite reliability indicators and Cronbach’s alpha are higher than .90 and AVEs are also higher than .50. All the standardized loadings are significant and higher than the generally accepted minimum cut off score of .70 (Nunnally and Bernstein, 1994; Fornell and Larcker, 1981). Regarding discriminant validity correlations among factors are always lower than their correspondent AVE square root and ratio HTMT is lower than .50.
**Table 1:** Reliability and convergent validity

<table>
<thead>
<tr>
<th>Factor</th>
<th>Item</th>
<th>Loading</th>
<th>t-value</th>
<th>CA</th>
<th>CR</th>
<th>AVE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Attitude</strong></td>
<td>ATT1</td>
<td>0.94</td>
<td>62.84 **</td>
<td>0.95</td>
<td>0.97</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>ATT2</td>
<td>0.98</td>
<td>201.75 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ATT3</td>
<td>0.95</td>
<td>61.14 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Perceived ease of use</strong></td>
<td>EOU1</td>
<td>0.76</td>
<td>23.02 **</td>
<td>0.92</td>
<td>0.93</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>EOU2</td>
<td>0.80</td>
<td>25.90 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EOU3</td>
<td>0.64</td>
<td>11.73 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EOU4</td>
<td>0.82</td>
<td>29.72 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EOU5</td>
<td>0.68</td>
<td>15.69 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EOU6</td>
<td>0.82</td>
<td>26.92 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EOU7</td>
<td>0.76</td>
<td>17.79 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EOU8</td>
<td>0.74</td>
<td>14.35 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EOU9</td>
<td>0.77</td>
<td>17.44 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EOU10</td>
<td>0.81</td>
<td>22.00 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Intention to use</strong></td>
<td>INT1</td>
<td>0.91</td>
<td>76.72 **</td>
<td>0.87</td>
<td>0.92</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>INT2</td>
<td>0.92</td>
<td>62.97 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>INT3</td>
<td>0.83</td>
<td>17.94 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Perceived usefulness</strong></td>
<td>PU1</td>
<td>0.89</td>
<td>52.22 **</td>
<td>0.96</td>
<td>0.96</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>PU2</td>
<td>0.81</td>
<td>26.45 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PU3</td>
<td>0.83</td>
<td>32.58 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PU4</td>
<td>0.81</td>
<td>29.16 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PU5</td>
<td>0.77</td>
<td>16.22 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PU6</td>
<td>0.90</td>
<td>50.46 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PU7</td>
<td>0.79</td>
<td>27.44 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PU8</td>
<td>0.91</td>
<td>64.20 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PU9</td>
<td>0.88</td>
<td>45.73 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PU10</td>
<td>0.85</td>
<td>31.59 **</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

** ** p<0.01

Note: CA=Cronbach’s alpha; AVE=Average Variance Extracted; CR=Composite Reliability

**Table 2:** Discriminant validity

<table>
<thead>
<tr>
<th>Factor</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>F1. Attitude</strong></td>
<td>0.96</td>
<td>0.40</td>
<td>0.88</td>
<td>0.75</td>
</tr>
<tr>
<td><strong>F2. Perceived ease of use</strong></td>
<td>0.38</td>
<td>0.76</td>
<td>0.34</td>
<td>0.46</td>
</tr>
<tr>
<td><strong>F3. Intention to use</strong></td>
<td>0.81</td>
<td>0.31</td>
<td>0.89</td>
<td>0.70</td>
</tr>
<tr>
<td><strong>F4. Perceived usefulness</strong></td>
<td>0.72</td>
<td>0.45</td>
<td>0.64</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Note. Diagonal: square root of AVE; Lower triangle: factor correlation; Upper triangle: ratio HTMT
Figure 2 depicts a graphical representation of the measurement model with the factor loadings.

Figure 2: Graphical representation of the measurement model with the factor loadings

5. Results

The structural model was estimated using PLS-SEM. R² of all the dependent latent variables are higher than 10% and Stone-Geisser Q² are positive confirming the predictive relevance of the model. The effect of perceived usefulness (H1, β=0.68; p<0.01) on attitude is significant. Although ease of use improves perceived usefulness (H3, β=0.43; p<0.01), it has not a significant direct effect on attitude (H2, β=0.06; p>0.05), showing that ease of use itself cannot foster EVGs use without the mediation of perceived usefulness. A significant attitude significantly causes intention to use EVGs favourably perceived (H4, β=0.81; p<0.01). Age was found to moderate teachers’ perceived ease of use on perceived usefulness of EVGs (H6, β=0.07; p<0.05). Nevertheless no moderating effect was found neither for perceived usefulness and attitude (H5, β=0.03; p>0.05) nor for ease of use on attitude (H7, β=0.00; p>0.05). Table 3 shows the hypotheses testing results.

### Table 3: Hypotheses testing

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>Std. Beta</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1 Perceived usefulness -&gt; Attitude</td>
<td>0.68 **</td>
<td>9.28</td>
</tr>
<tr>
<td>H2 Perceived ease of use -&gt; Attitude</td>
<td>0.06</td>
<td>0.85</td>
</tr>
<tr>
<td>H3 Perceived ease of Use -&gt; Perceived Usefulness</td>
<td>0.43 **</td>
<td>6.30</td>
</tr>
<tr>
<td>H4 Attitude -&gt; Intention to use</td>
<td>0.81 **</td>
<td>23.43</td>
</tr>
<tr>
<td>H5 Age x (Perceived usefulness-&gt;Attitude)</td>
<td>0.03</td>
<td>0.55</td>
</tr>
<tr>
<td>H6 Age x (Perceived ease of Use -&gt; Perceived Usefulness)</td>
<td>0.07 *</td>
<td>2.04</td>
</tr>
<tr>
<td>H7 Age x (Perceived ease of Use -&gt; Attitude)</td>
<td>0.00</td>
<td>0.06</td>
</tr>
</tbody>
</table>

**p<0.01 *p<0.05

R² (Intention to use)=0.653; R²(Attitude)=0.524; R² (Perceived usefulness)=0.219
Q² (Intention to use)=0.485; Q²(Attitude)=0.441; Q² (Perceived usefulness)=0.148

6. Discussion

Results suggest that perceived usefulness is a main antecedent of attitude towards EVGs, that is, the higher the teachers’ perception of usefulness of EVGs the better the teachers’ attitude towards EVGs. This is consistent with previous findings that have used TAM to test attitude and adoption of technological developments in education such as a Skype-Based E-Learning System (Martí-Parreño et al, 2013). This finding is also consistent with previous TAM studies analysing secondary school teachers’ acceptance of EVGs (Bourgonjon et al., 2013). Although a direct effect of ease of use on attitude was not found, ease of use indirectly influences attitude through perceived usefulness. This has an explanation following this rationale:
teachers might not have a positive attitude towards EVGs just because they perceive EVGs as easy to use but teachers must perceive that EVGs are useful to enhance their teaching activity. Attitude towards EVGs directly and positively influences teachers’ intention to use EVGs in their courses, that is, teachers having a positive attitude towards EVGs show behavioural intention to use EVGs. This finding is also consistent with extant literature that found the relationship between attitude and behavioural intention (Warshaw and Davis, 1985).

Regarding the moderating effect of age, results suggest that the effect of teachers’ perceived ease of use on teachers’ perceived usefulness of EVGs is moderated by this factor. Broadly speaking, how useful EVGs are perceived is a direct effect of how ease of use teachers perceived EVGs differs among older and younger teachers. This is an important finding that must be taken into account when developing Teacher Training Programmes as it suggests a generational gap that might affect teachers’ adoption of EVGs. Moreover, considering that perceived usefulness has been found as the main factor contributing to teachers’ attitudes towards EVGs this moderating effect of age deserves more attention to delve into factors that contribute age differences on the effect of teachers’ perceived ease of use of EVGs on teachers’ perceived usefulness of EVGs. However, no moderating effect of age was found either for perceived usefulness and teachers’ attitude or for perceived ease of use and teachers’ attitude. That is, age is independent of how teachers’ perceived usefulness and perceived ease of use of EVGs affects their attitude towards EVGs. Teacher Training Programmes should focus then on different approaches to older and younger teachers because perceptions in both groups of teachers vary regarding how their perceived ease of use of EVGs affects their perceived usefulness of EVGs and consequently their attitude and finally their behavioural intention to use EVGs in their courses.

7. Conclusions, limitations and future research

One main conclusion of this study is that teachers’ perceived usefulness is a key factor to predict teachers’ intention to use EVGs. This finding suggests that managers of Higher Education institutions who want teachers to use EVGs in their courses should encourage them to enrol in specific Teacher Training Programmes that focuses in highlighting how useful EVGs can be in their teaching activity and how EVGs can benefit students’ learning. Because ease of use does not affect teachers’ attitude towards EVGs (only indirectly through perceived usefulness), Teacher Training Programmes should not encourage teachers to use EVGS just because EVGs are easy to use but to use EVGs because they can be a useful resource to apply in the classroom. For example, EVGs can be used as a means of motivating students to learn in a more entertaining way and also to increase students’ engagement with the learning activities. Teacher Training Programmes should also consider the age of teachers and especially how age moderates the effects of perceived ease of use on perceived usefulness. For example, the effect of perceived ease of use on perceived usefulness in younger teachers might rely on their greater familiarity with video games compared to older teachers. Future research should delve into factors that better explain this moderating role of teachers’ age. Because both perceived usefulness and perceived ease of use are directly influenced by design features, future research should also explore EVGs design factors (e.g. difficulty level, course and subject suitability, learning goals…) that might affect teachers’ perceived usefulness and perceived ease of use of EVGs. For example, teachers’ perceptions on how the content of EVGs can fit the subject they teach or how EVGs’ features can increase students’ learning could provide a better understanding of key drivers of teachers’ perceived usefulness of EVGs.

One main limitation of this study is the use of a convenience sample that does not allow the representation of the target population. Future research should use a probabilistic sample that better represents the target population. Because culture affects human behaviour, future research should be developed within a cross-cultural framework in order to gain a better knowledge of cultural differences when adopting EVGs. Gender can also affect attitudes and behavioural intentions to use EVGs. More research is needed in order to better understand the role of gender in teachers’ intention to EVGs. The educational level (primary school, secondary school…) can also affect teachers’ decisions about whether to implement EVGs in their course or not. Future research should explore differences in teachers’ intention to use EVGs across educational levels but also across institution characteristics, such as public or private schools. Finally, as this study adopts a teacher-centric approach future research should use a student-centric approach in order to better understanding students’ intention to use EVGs.
Acknowledgments

This work was supported by Laureate International Universities through the David A. Wilson Award for Excellence in Teaching and Learning [LIU-WIL2015].

References


Okazaki, S., & Renda Dos Santos, L. (2012). Understanding E-Learning Adoption in Brazil: Major Determinants and Gender Effects. The International Review of research in Open and Distance Learning, 15(4), 91–106.


www.ejel.org


