The Impact of Language on Students’ Emotional States in Educational Games: A Comparative Study

Duha Awni Assaf
Information and Computing Technology Division,
College of Science and Engineering,
Hamad Bin Khalifa University
Doha- Qatar
doha_awni@hotmail.com

Dena Al-Thani
Information and Computing Technology Division,
College of Science and Engineering,
Hamad Bin Khalifa University
Doha- Qatar
dalthani@hbku.edu.qa

Bilikis Banire
Information and Computing Technology Division,
College of Science and Engineering,
Hamad Bin Khalifa University
Doha- Qatar
banire.bilikis.o@gmail.com

Capturing students’ emotions while playing an educational game is one approach to assess their motivation towards learning. The use of educational games could serve as a motivating factor for players. This study compares two languages (Arabic and English) in an educational game to understand and compare the effect of the two languages on learning motivation via emotions. An experimental study was conducted with 30 Arabic-speaking students (Male n=13, Female n=17) while playing an educational game in both Arabic and English languages, and their emotions were recorded. The result shows that participants express significant negative emotions (anger [p < 0.05], contempt [p < 0.05], and sadness [p < 0.05]) while playing the Arabic version of the game than the English version. This indicates that participants preferred the English version. These findings suggest that emotion might help evaluate language preference in educational games development.

1. INTRODUCTION

Language is involved in every aspect of our lives. It is a tool used to communicate with others, make decisions, and understand given instructions (Antón & Soleto, 2020). However, people tend to use more than one language to communicate worldwide. English is commonly used as a second language for non-native speakers (Eberhard, Gary, & Fennig, 2021). The worldwide organization conducts meetings mostly in English on important and crucial global decisions (Antón & Soleto, 2020). In Qatar and other Arab countries, the Arabic language is a source of great pride and cultural identity. Still, Arab tend to use English more frequently than Arabic to improve the chances of fitting into a globalized market and accessing modern knowledge (Mustafawi & Shaaban, 2019). However, this increased emphasis on learning English could threaten Arab identity and losing interest in any cultural content such as Arabic TV shows, cartoons, books, and educational games.

Educational computer games are becoming a popular essential tool for learning due to their impact on students’ learning outcomes. These games also improve players’ learning motivation and engagement skills (Bontchev & Vassileva, 2016). Moreover, these games elicit emotional reactions in the players, such as fear, anger, surprise, or joy (Squire, 2003).

There are seven basic emotions which include joy, anger, sadness, surprise, fear, contempt, and disgust. These emotions are classified as positive (joy, surprise) or negative (anger, sadness, fear, contempt, and disgust). Similarly, facial action units (AUs) are referred to as low-level expression that further defines emotions. Therefore, emotions play a crucial role in learning, especially to understand why positive emotions such as joy can increase students’ engagement in a game. In contrast, students’ engagement during learning declines due to negative emotions, such as anger and fear induced by the game (Sawyer et al., 2017). Recently, affective computing (AC) has made it possible to develop systems that can recognize, interpret human emotions (Picard, 1999). The techniques of AC have been used extensively in the educational field since 2010 (Yadegaridehkordi et al., 2019) to reduce the frustration of users during an interaction.

There is an increasing interest in the use of digital games for language learning and teaching nowadays(Goumas, Terzopoulos, Tsompanoudi, & Iliopoulou, 2020; Peterson, 2016; Reinders, 2012) Specifically, studies used commercial games to examine how games enhance teaching and language learning. Educational games have been identified as an evidence-based strategy for learning support (Sykes, 2018). These games are usually developed in both native and foreign languages as they help students develop their vocabulary skills in an actively interactive manner as compared to traditional methods. Goumas et al., (2020) created a digital educational game called “Wordsearch”, to
assist users to learn the vocabulary of a foreign language. Such advantages have been observed by Wang (2010), who studied the impact of communicative language games on learning English in elementary schools in Taiwan and found that 150 teachers benefited from these games by making the lessons and language learning more interesting and pleasurable to learn.

Research has shown the benefit of educational games as a tool for learning local and foreign vocabulary, however, little research has been done in exploring the impact of foreign language on a local language. Antón and Soleto (2020) explored the impact of foreign language on decision making by giving two groups of students a set of instructions and rules in their native language and a foreign language. Results showed that the students who were instructed in a foreign language followed the rules better than the students instructed in their native language. To our best of knowledge, no study has explored the effect of the English language in designing an educational game for Arabic native speakers through analyzing the player’s emotions.

In this paper, the analysis of emotional states (positive and negative emotions) of school children on the Arabic version (AV) and English version (EV) of a word search game is presented. To guide this analysis, we proposed two hypotheses 1) students show heightened emotional states when playing the EV of the game than AV 2) emotions have significant correlations with engagement.

The rest of this paper is organized as follows: Section 2 describes the method including participant recruitment, procedures, data collection, and analysis. Section 3 discusses the results; Section 4 presents the main findings and the hypothesis; Section 5 narrates the conclusion of the paper with its limitations and future work.

2. METHODS

This section describes the information of the participants, the procedures used in playing the game, and collecting emotional states data.

2.1 Participant Recruitment

An approval from the Institutional Review Board (IRB) of the University (Reference Number: 2021-07-099) was obtained before the study. Participants were recruited from a local School in Doha, Qatar. The inclusion criteria used for selecting participants are students (a) within the age range of 10 and 13 years (b) who have a basic understanding of both English and Arabic (c) who is familiar with using smartphones. These criteria are to ensure that the participants have a similar proficiency level in both languages to support rich data analysis and reliability. A signed informed consent and assent form was obtained from all the participants included in the study. A total of 30 students from grades 5 to 7, between the ages of 10 and 13 years, met the inclusion criteria and participated in the study.

2.2 The Educational Game

“Word of Wonders” (WOW) is a crossword puzzle educational language game designed and developed by a game company called Fugo (Fugo Games)(Figure 1 a and b). This game was designed to improve users’ vocabulary and spelling skills simultaneously. WOW consist of two different languages (Arabic and English). Starting with a collection of related words, the students will then try to find and connect letters to form these words. The game runs on mobile operating systems (Android and IOS). The game can also be played on a standard PC by downloading an emulator such as BlueStacks (Judge, 2013). The screenshots of the WOW game in both AV and EV are depicted in Figure 1. This game is popular with more than 250 million downloads. Unlike other vocabulary games, WOW covers a wide range of languages. it is also suitable for children between 10 to 13 years. In total, the game contains 2800 different levels to solve.

2.3 Procedures

At the start of the experiment, the researcher received the participant and asked them to remove their caps and face masks. Additionally, they were also instructed to refrain from covering their faces with their hands while playing the game to prevent any interference with emotion detection. Next, the participant was seated within the frame of a camera placed on top of the 14-inch monitor (Lenovo ThinkPad X1 Carbon Gen. 3). The participants’ proficiency levels in both English and Arabic were assessed using the Common European Framework of Reference (CEFR) (Council of Europe, 2001). This framework consists of six different levels of language proficiency: beginner, elementary, intermediate, upper-intermediate, advanced and proficiency. The objective of the language proficiency level assessment is to better understand the performance of all the participants. Each participant was randomly assigned to one of two groups where the first group started with EV and
The Impact of Language on Students' Emotional States in Educational Games: A Comparative Study
Duha Awni Assaf ● Dena Al Thani ● Bilikis Banire

then AV. The other group did the opposite to eliminate the effects of selection bias which may affect the study outcome. The participants played both the EV and AV for 10 minutes and were recorded using a Logitech webcam. Also, the emotions of the participants were tracked in real-time during gameplay using iMotions software (iMotions, 2017) as depicted in Figure 2. This software provides instantaneous analysis of single or integrated biometric measures, and different studies have used it for monitoring and evaluating user’s feedback, interaction, and level of attention (Gero & Milovanovic, 2020; Jo et al., 2020; Lei et al., 2017; Sawyer et al., 2017).

Figure 2: Real-Time Face Tracking During the Game using iMotions Software.

2.4 Setting

The study took place in the school’s meeting room isolated from other rooms and free from external noise or distractions. The participants sat in front of a laptop with a 14-inches screen and play the game while a Logitech camera attached at the top of the laptop screen records the participants’ faces (Figure 3).

Figure 3: Experiment Setting

2.5 Study Design

iMotions software is also embedded with Affectiva software (McDuff et al., 2016) which detects engagement levels based on a range value of 0 to 100 which represents not engaged and engaged respectively. The primary objective of this study is to investigate the impact of the Arabic and English languages on students’ emotions while playing the WOW game. Then, we aim to further explore the impact of negative or positive emotions on their engagement.

2.5.1 Study Hypotheses

Experimental studies are commonly based on a predefined hypothesis and not driven by data (Lazar et al., 2017). Hence, the two hypotheses of this study are as follows:

Hypothesis 1
H0: The student’s emotional state when playing the EV of the game is indifferent from that of AV.
H1: The student’s emotional state when playing the EV of the game is significantly different from that of AV.

Hypothesis 2
H0: There are no associations between the seven emotions and engagement.
H1: The emotions have statistically significant correlations with engagement.

2.6 Data Analysis

This section describes the data used for analysis and the statistical tests adopted based on the hypotheses highlighted in this study.

2.6.1 Data Information

The emotional states of 30 participants were recorded during the user study experiment. However, two of them were excluded from the analysis (P22 and P24) because they expressed unexpected emotions before and during the game such as smiling a lot for no specific reason. Our observation shows that these two participants were too anxious to be part of the game as they were smiling and laughing before the game began.

2.6.2 Statistical Testing

The descriptive statistical tests are applied to the quantitative data to understand the nature and distribution of the dataset. A normality test was carried out on the data using the Shapiro-Wilk normality test (Razali & Wah, 2011) and the results showed a significant p-value (p<.001). This indicates that the distribution of the data is not normally distributed. Thus, a nonparametric test was used to make inferences and draw conclusions from the available datasets.

Wilcoxon signed-rank test is a nonparametric statistical test that determines if two measurements from a single group are significantly different from each other based on a defined variable (Fong et al., 2019; Kreitlon et al., 2019). This test was used to answer the first hypothesis of the study which
compares the emotional states of the players in both the AV and EV of the game.

Another nonparametric test, Spearman’s rho correlation coefficient (rS) was used to answer the second hypothesis, the relationship between emotions and engagement. This test is used to measure the strength of association between two variables. It is equivalent to Pearson’s product-moment correlation coefficient but it is performed on the ranks of the data rather than the raw data, where a maximum value of correlation coefficients rS = 1 describes a complete positive association, and rS = -1 describes a complete negative association (Iliou & Anagnostopoulos, 2009). A value of p < 0.05 was adopted to reject the null hypothesis. All the statistical analyses were conducted using JASP software, version 0.15 (Love et al., 2019).

3. RESULTS

The results of the data analysis of the emotional states of children when playing educational games in different languages (Arabic and English). In this section, the descriptive statistics of the users’ demographic data and game language proficiency are discussed. The game levels completed among participants and their emotional states were first analyzed to select the appropriate levels for a fair comparison. Next, the statistical analysis methods used for the emotional states exhibited by the participants were described. This was followed by the descriptive statistics of emotional states for all levels in both languages. Finally, the results from the data acquired from the users’ feedback. Throughout this paper, the abbreviation ‘AV’ will be used to refer to the Arabic version and ‘EV’ will refer to the English Version.

3.1 Descriptive Statistics

The section summarises the participants’ demographic data and proficiency levels in both English and Arabic.

3.1.1 Participant’s Demographic Data

The distribution of participants’ age and grade levels (N=30, Male =13, Female =17) are represented in Figures 4 and 5. According to Al-Manar International School, American curricula were adapted, so grades 5, 6, and 7 are also referred to as school years 6, 7, and 8 in the United Kingdom curriculum.

3.1.2 English and Arabic proficiency level of the participants

The participants’ proficiency levels in English and Arabic were based on the CEFR standard (Common European Framework of Reference for Languages) is summarized in Table 1. The proficiency level assessment for both English and Arabic showed that 86.74% and 92.3% of the participants respectively are above the beginner level. This explains the game-level completion rate in Table 3 where only levels 1 and 2 were completed by all the participants.

3.1.3 Participants’ Game Experience

All the participants were asked about the type of game they have played before and only one of the participants plays educational games every day but not the game used in this study. Fourteen of the participants played entertainment games daily. Furthermore, we observed that participants played educational games (13 out of 30) more than the entertainment games (5 out of 30).

3.2 Game-Levels

The number of completed levels for each version varies among the participants. They played seven different levels in the EV and four levels in the AV. All the 30 participants completed levels 1 and 2 from both versions. Table 2 shows the game-level completion rate for EV and AV. Therefore, to make a fair comparison we selected level 1 and level 2 for the analysis.

3.2.1 Time Spent on Each Game

On average, participants took the longest time on Arabic level 2 (mean = 5.6, SD = 1.61) followed by
English level 2 (mean = 2.6, SD = 1.31), Arabic level 1 (mean = 1.78, SD = 0.41), and English level 1 mean = 1.64, SD = 0.48) see Table 3.

Table 1: English and Arabic proficiency level

<table>
<thead>
<tr>
<th>Proficiency level</th>
<th>English</th>
<th>Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1. Beginner (Can understand basic interaction in a simple way)</td>
<td>4(13.3%)</td>
<td>2(6.7%)</td>
</tr>
<tr>
<td>A2. Elementary (Can use expressions fluently)</td>
<td>10(33.3%)</td>
<td>10(33%)</td>
</tr>
<tr>
<td>B1. Intermediate (Can produce simple connected text on familiar topics)</td>
<td>7(23.3%)</td>
<td>8(26.7%)</td>
</tr>
<tr>
<td>B2. Upper-intermediate (Can interact with a degree of fluency and spontanity)</td>
<td>5(16.7%)</td>
<td>3(10%)</td>
</tr>
<tr>
<td>C1. Advanced (Can understand a wide range of texts and converse in present coherently)</td>
<td>3(10%)</td>
<td>4(13.3%)</td>
</tr>
<tr>
<td>C2. Proficiency (Can summarize information from different spoken and written sources)</td>
<td>1(3.3%)</td>
<td>2(6.7%)</td>
</tr>
</tbody>
</table>

Table 2: Game-Level completion rate for EV and AV

<table>
<thead>
<tr>
<th>Levels</th>
<th>Arabic (n = 30)</th>
<th>English (n = 30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Level 2</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Level 3</td>
<td>21</td>
<td>30</td>
</tr>
<tr>
<td>Level 4</td>
<td>1</td>
<td>29</td>
</tr>
<tr>
<td>Level 5</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>Level 6</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Level 7</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 3: Average time spent on each level

<table>
<thead>
<tr>
<th>Levels</th>
<th>M(minutes/seconds)</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic level 1</td>
<td>1.78 / 106</td>
<td>0.41</td>
</tr>
<tr>
<td>Arabic level 2</td>
<td>5.6 / 336</td>
<td>1.61</td>
</tr>
<tr>
<td>English level 1</td>
<td>1.64 / 98</td>
<td>0.48</td>
</tr>
<tr>
<td>English level 2</td>
<td>2.6 / 156</td>
<td>1.31</td>
</tr>
<tr>
<td>Levels 1 &amp; 2 (English and Arabic)</td>
<td>2.91 / 174</td>
<td>1.93</td>
</tr>
</tbody>
</table>

3.2.1 Emotional States in Game Level 1

Table 4 shows the overall arithmetic means and standard deviations for each emotion state based for both EV and AV. On average, students expressed more anger (mean = 1.09, SD = 2.53), disgust (mean = 1.00, SD = 2.05), joy (mean = 1.45, SD = 4.01), fear (mean = 1.49, SD = 3.87), contempt (mean = 1.76, SD = 3.97) in the AV than in the EV. In contrast, students expressed a higher sadness (mean = 1.7, SD = 4.55) in the EV, and for surprise there is no clear difference between both versions.

Table 4: Descriptive statistics for level 1

<table>
<thead>
<tr>
<th>Emotional states</th>
<th>Arabic (n = 28)</th>
<th>English (n = 28)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>0.48</td>
<td>1.20</td>
</tr>
<tr>
<td>Sadness</td>
<td>2.46</td>
<td>8.70</td>
</tr>
<tr>
<td>Disgust</td>
<td>0.73</td>
<td>1.06</td>
</tr>
<tr>
<td>Joy</td>
<td>2.08</td>
<td>5.36</td>
</tr>
<tr>
<td>Surprise</td>
<td>1.74</td>
<td>3.16</td>
</tr>
<tr>
<td>Fear</td>
<td>1.15</td>
<td>3.04</td>
</tr>
<tr>
<td>Contempt</td>
<td>1.22</td>
<td>3.46</td>
</tr>
</tbody>
</table>

3.2.2 Emotional States in Game Level 2

Table 5 shows the overall arithmetic means and standard deviations for each expression based on both EV and AV. On average, students expressed more anger (mean = 1.09, SD = 2.53), disgust (mean = 1.00, SD = 2.05), joy (mean = 1.45, SD = 4.01), fear (mean = 1.49, SD = 3.87), contempt (mean = 1.76, SD = 3.97) in the AV than in the EV. In contrast, students expressed a higher sadness (mean = 1.7, SD = 4.55) in the EV, and for surprise there is no clear difference between both versions.

Table 5: Descriptive statistics for level 2

<table>
<thead>
<tr>
<th>Emotional state</th>
<th>Arabic (n = 28)</th>
<th>English (n = 28)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>1.09</td>
<td>2.53</td>
</tr>
<tr>
<td>Sadness</td>
<td>1.34</td>
<td>3.19</td>
</tr>
<tr>
<td>Disgust</td>
<td>1.00</td>
<td>2.05</td>
</tr>
<tr>
<td>Joy</td>
<td>1.45</td>
<td>4.01</td>
</tr>
<tr>
<td>Surprise</td>
<td>2.04</td>
<td>3.19</td>
</tr>
<tr>
<td>Fear</td>
<td>1.49</td>
<td>3.87</td>
</tr>
<tr>
<td>Contempt</td>
<td>1.76</td>
<td>3.97</td>
</tr>
</tbody>
</table>

3.2.3 Emotional States in Game Levels 1 and 2

Table 6 shows the overall arithmetic means and standard deviations for each expression when using the EV and AV. On average, students expressed more anger (mean = 0.99, SD = 2.21), disgust (mean = 0.95, SD = 1.69), fear (mean = 1.43, SD = 3.62), and contempt (mean = 1.64, SD = 3.51) in the AV than the EV. In contrast, students expressed a higher sadness (mean = 1.58, SD = 3.63), joy (mean = 1.78, SD = 4.24), and surprise (mean = 2.01, SD = 2.53) in the EV.

Table 6: Descriptive statistics for levels 1 and 2

<table>
<thead>
<tr>
<th>Emotional state</th>
<th>Arabic (n = 28)</th>
<th>English (n = 28)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>0.99</td>
<td>2.21</td>
</tr>
<tr>
<td>Sadness</td>
<td>1.52</td>
<td>3.89</td>
</tr>
<tr>
<td>Disgust</td>
<td>0.95</td>
<td>3.89</td>
</tr>
<tr>
<td>Joy</td>
<td>1.58</td>
<td>3.77</td>
</tr>
<tr>
<td>Surprise</td>
<td>1.99</td>
<td>3.16</td>
</tr>
<tr>
<td>Fear</td>
<td>1.43</td>
<td>3.62</td>
</tr>
<tr>
<td>Contempt</td>
<td>1.64</td>
<td>3.51</td>
</tr>
</tbody>
</table>
3.3 Effect of language-based games on emotions

To evaluate the effect of language on emotion, Wilcoxon signed-rank test was used to test the first hypothesis.

Hypothesis 1

Null hypothesis (H0): The student's emotional state in EV is different from that of AV of the game. Alternative hypothesis (HA): The student's emotional state when playing the EV is significantly different from their emotional state when playing the AV of the game. In this case, we have a two-sided test problem:

\[ H_0 : \mu = 0 \quad \text{vs} \quad H_A : \mu \neq 0 \]

The results in Table 7 show that there is a significant negative emotional state difference between EV and AV in game level 2: anger (w = 302, p = 0.02; < 0.05) and contempt (w = 311, p = 0.01; < 0.05) as well as in and game level “1 and 2”: anger (w = 292, p = 0.04; < 0.05) and (w = 313, p = 0.01; < 0.05). Thus, the null hypothesis is rejected for game level 1 and level “1 and 2”. However, there was no significant difference found between emotional state in level 1 of AV and EV (Table 7). Hence, we can accept the null hypotheses for level 1.

3.4 Correlation of Engagement with Emotions

In this section, the nature of the correlation between engagement with emotions is identified based on EV and AV. The overall game completed by all the participants: levels “1 and 2” was considered since level 1 represents only 2 minutes out of the 20 minutes duration and there was no significant difference observed between EV and AV. A Spearman's rank-order correlation (rS or ρ) was run to test the second hypothesis.

Hypothesis 2

Null hypothesis (H0): There is no association between the seven emotions and engagement. Alternative hypothesis (HA): There is an association between the seven emotions and engagement. In this case, we have a two-sided test problem:

\[ H_0 : \rho = 0 \quad \text{vs} \quad H_A : \rho \neq 0 \]

Table 7: Paired Samples T-Test for levels 1, 2, and 1 & 2

<table>
<thead>
<tr>
<th>Measure 1</th>
<th>Measure 2</th>
<th>Level 1 W (p)</th>
<th>Level 2 W (p)</th>
<th>Level 1 and 2 W (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger Arabic</td>
<td>Anger English</td>
<td>222 (0.678)</td>
<td>302(0.023*)</td>
<td>292 (0.043*)</td>
</tr>
<tr>
<td>Disgust Arabic</td>
<td>Disgust English</td>
<td>210 (0.884)</td>
<td>231 (0.537)</td>
<td>261 (0.194)</td>
</tr>
<tr>
<td>Surprise Arabic</td>
<td>Surprise English</td>
<td>203 (1.000)</td>
<td>213 (0.831)</td>
<td>227 (0.598)</td>
</tr>
<tr>
<td>Contempt Arabic</td>
<td>Contempt English</td>
<td>186 (0.711)</td>
<td>311 (0.013*)</td>
<td>313 (0.011*)</td>
</tr>
<tr>
<td>Sadness Arabic</td>
<td>Sadness English</td>
<td>237 (0.451)</td>
<td>277 (0.095)</td>
<td>240 (0.412)</td>
</tr>
<tr>
<td>Joy Arabic</td>
<td>Joy English</td>
<td>187 (0.728)</td>
<td>244 (0.362)</td>
<td>210 (0.884)</td>
</tr>
<tr>
<td>Fear Arabic</td>
<td>Fear English</td>
<td>264 (0.171)</td>
<td>261 (0.194)</td>
<td>248 (0.316)</td>
</tr>
</tbody>
</table>

Note. Wilcoxon signed-rank test. *p < 0.05

Table 8: Spearman’s Correlations of engagement with emotions in AV and EV

<table>
<thead>
<tr>
<th>Variable</th>
<th>Engagement AV Spearman’s Rho (P)</th>
<th>Variable</th>
<th>Engagement EV Spearman’s Rho (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger Arabic</td>
<td>0.425 (0.025)</td>
<td>Anger English</td>
<td>0.533** (0.004)</td>
</tr>
<tr>
<td>Sadness Arabic</td>
<td>0.180 (0.359)</td>
<td>Sadness English</td>
<td>0.062 (0.754)</td>
</tr>
<tr>
<td>Disgust Arabic</td>
<td>0.394(0.039)</td>
<td>Disgust English</td>
<td>0.288 (0.136)</td>
</tr>
<tr>
<td>Joy Arabic</td>
<td>0.177(0.385)</td>
<td>Joy English</td>
<td>0.248 (0.202)</td>
</tr>
<tr>
<td>Surprise Arabic</td>
<td>0.471* (0.012)</td>
<td>Surprise English</td>
<td>0.389* (0.042)</td>
</tr>
<tr>
<td>Fear Arabic</td>
<td>0.340 (0.077)</td>
<td>Fear English</td>
<td>0.554** (0.003)</td>
</tr>
<tr>
<td>Contempt Arabic</td>
<td>0.511** (0.006)</td>
<td>Contempt English</td>
<td>0.381* (0.046)</td>
</tr>
</tbody>
</table>

Note. *p < 0.05, **p < 0.01
It was important to check how engagement correlates with emotions in the two game versions: EV and AV. At AV's level “1 and 2”, Table 8 shows a statistically significant positive relationship between engagement and the emotions of anger (rs (26) = .425, p < .05), disgust (rs (26) = .394, p < .05), surprise (rs (26) = .471, p < .05), and contempt (rs (26) = .511, p < .01). Similarly, a statistically significant positive relationship was found in the EV between engagement with anger (rs (26) = .533, p < .01), surprise (rs (26) = .389, p < .05), fear (rs (26) = .554, p < .01), and contempt (rs (26) = .381, p < .05, see Table 8). Therefore, we can reject the null hypothesis and accept the alternative hypothesis.

3.5 User’s Feedback

All the participants were asked to give their feedback on their preference for the EV and AV of the game after they completed the experiment. The users' feedback consisted of three questions about their overall experience, which language version they found more enjoyable, and if they had tried Arabic educational games before. The purpose of the feedback was to have a better insight into the statistical analysis.

According to the feedback, none of the participants found the EV hard. All the participants stated that they found the EV easier and more exciting than the AV. This result is supported by their scores; only one of them attempted level 4 while 29 finished level 4 in the EV. Surprisingly 24 out of the 30 participants added that they had never played Arabic games. As a result, the students tend to enjoy playing the EV more for many reasons including a smaller number of rows and columns in the different levels of the game, and hence fewer words to find. Some of the excerpts from the participants’ feedback are as follows.

“*The Arabic game was harder because the number of rows and columns was much more than the English version.*” (P2)

“*English is easier because the words were shorter so I could find them quickly.*” (P21)

“*English is easier because it has fewer words.*” (P25)

4.1 Effect of game language on emotions

The analysis of participants' performance shows that they spent longer time during level 2 in the AV (5.6 minutes on average) than in the EV, this result indicates that AV level 2 is the most difficult. A possible explanation for this might be that each level has a different number of rows, columns and words to find. For instance, in level 2 Arabic, the participants must find 7 words across 9 rows and 8 columns, which is more than the words in the EV (e.g. 5 words across 6 rows and 5 columns). This difference may be explained by the fact that Arabic contains 28 letters while English has only 26 letters. Together, these results provide important insights into why the students could not reach the same level as in the EV. These findings could help the game developer to take into consideration an important factor regarding languages used for creating educational games.

The results of level 1 show no significant difference between Arabic and English probably because of the shorter completion time (2 minutes) as compared to levels 2 (5.6 minutes) and levels 1 and 2 (7.6 minutes). Also, players show more significant negative emotions: anger, contempt, and sadness, than they did when playing the EV. This suggestion also conforms with the feedback given by the participants.

Negative emotions can highly affect the game outcomes. For example, if the students expressed more anger, disgust, fear, or contempt toward the game then this may lead to a loss of interest and poor learning performance (Meyer & Turner, 2006). Wiklund et al., (2015) found that the negative emotions toward educational games could be indicative of a poor educational game. What emerged from the results reported here is that the AV educational game represented a higher degree of negative emotions than EV. A possible explanation for a negative emotion found in the AV might be due to the differences in the difficulties between the game stages instead of the language proficiency.

4.2 Correlation of Engagement with Emotions

According to the result from the correlation matrix, the four emotions that correlate with engagement in both AV and EV were identified as anger, fear, contempt, and surprise. The results of this study indicate that the students may exhibit both negative and positive emotions when they are engaged which corresponds to the findings in similar studies (Tekinbas & Zimmerman, 2003; Yannakakis & Paiva, 2014). What is surprising is that a strong correlation between negative emotions such as anger, fear, and contempt with engagement was found. A possible explanation for this might be that the participant felt like they were being watched and recorded, as the recording camera was mounted on top of the monitor and it was visible to the participants. Wiklund et al. (2015) mentioned that
the effect of the observer may impact how participants express their emotional state in both directions (negative or positive). Using a survey to get the feedback from participants on what degree they felt that they were observed could be a way to control this factor. These results warrant further investigation to determine exactly how conspicuous video recording may affect the participant’s emotions during study observation.

### 4.3 Limitation

The challenges with emotion recognition are face obstruction, where the face is covered by an object, hand, etc. In our study, the participants touch their lower faces with one or two hands often. These gestures may indicate important information about the participant’s emotions. For instance, students place a hand on their face while thinking or put both their hands on their heads because they were tired or bored. All of these gestures make the processing of emotions impossible (Sawyer et al., 2017). The solution for this limitation is either to keep reminding the participant to remove their hands or use an additional method to recognize body language (Behoora & Tucker, 2015).

There are some confounding factors that effects students’ performance emotions and engagement such as the communication language used at home; for instance, they might communicate in the English language most of the time; hence it will reflect on their performance when it comes to Arabic language games. Lastly, the most important limitation lies in the fact that the game design does not consider the differences between Arabic and English languages for the same stage. According to our findings, stage 2 of the AV has a higher difficulty than stage 2 of the EV due to the number of rows, columns, and words to find. Therefore, it might affect players’ emotions and engagement, as more negative emotions would be expressed towards particularly difficult stages, leading to a discontinuity in playing a particular language version.

### 4.4 Recommendations

There are several recommendations to researchers and designers or game developers who are interested in this research are as follows:

1. The emotion recognition technology is relatively simple and inexpensive. Therefore, it may be a reasonable choice for evaluating educational games with different languages to understand the user experience via emotions throughout the game.

2. Designers need to consider the differences between languages when it comes to vocabulary learning. Game designers utilize various techniques to raise user engagement and motivation such as in serious games. If the designers do not take into consideration the differences between these languages and their vocabularies, then the game itself might seem difficult for a specific language; while it is not the case, it is because the design of the game makes it looks difficult.

3. These findings shed light on the observer effects and the degree that well affects the participants to express their emotional state in either direction (Negative or Positive). The participants might feel like they were being watched by the researcher and recorded as the recording camera was mounted on top of the monitor and visible to the participants. We suggested that more research on observer effects is needed to understand precisely how the observer may affect the participants’ emotions in an observational study. This recommendation was also reported by Wiklund et al. (2015).

4. A similar study should be carried out using a multimodal measurement technique, such as a combination of emotions and a physiological recognition technique (e.g., EEG, EDA, RA, HR, ECG) to establish a greater degree of accuracy on users’ assessment. Researchers adopted multimodal because they accurately measure the users’ emotions. Hence, it would help researchers understand the learner’s motivation more effectively (Landowska & Miler, 2016). Additionally, the multimodal method resolved the constraints of using only one model; Wu et al. (2014) suggested that different integrated techniques could recognize the emotional state more accurately.

5. Designers of multi-language games should consider the different languages as an essential component in the game design process. We also discourage direct translation. Boroditsky (2009) showed that the information can be lost or changed due to language differences. Therefore, translating English games to Arabic may affect users’ performance in both languages.

### 5. CONCLUSIONS AND FUTURE WORK

#### 5.1 Conclusions

The Arabic language is a source of great pride. In particular, the cultural identity of all Arab countries, and local Qatari among them, is closely tied to the Arabic language. Still, people tend to use English more frequently than Arabic, primarily in the educational field as evident in this study. The increased emphasis on preference for English over Arabic could threaten learners' Arab identity, making them lose interest in any Arabic content such as Arabic TV shows, cartoons, books, and educational games. This study also shows the potential of emotions for analyzing the engagement and language preference of students. In addition, the recommendation proposed in this study can help game developers in preserving the Arab identity and culture through strategic designs of Arabic games that match the English version.
5.2 Future Work

Despite the significant contribution made by this study, some recommendations are proposed for future work. Other ways of improving the study’s results are by attaining more insight on the target population characteristics, increasing the amount of data acquired, and extending our work using specific methodology such as machine learning. The highlights for the future work are as follows.

5.2.1 Extending the Studies

1. Extending the sample size: In this study, the sample size was relatively small (N = 30) and having a larger sample would increase the accuracy of the results and include a greater variety of participants.

2. Extending the target population: This study focused on secondary students from private schools. However, the same study can be conducted with students of other age groups and different school backgrounds such as public school.

3. Using machine learning model for predicting the level of difficulties based on their emotions and AUs: Identifying the level of difficulty while playing a different language version of the same educational game (e.g., Arabic and English) is essential for understanding the players’ emotional state at each stage. We found that Stage 2 of the Arabic version has a higher difficulty stage compared to Stage 2 of the English version. Hence, it could affect players’ emotions and engagement, as more negative emotions would be expressed towards particularly difficult stages, leading to discontinuity in playing a particular language version. A process similar to (Verma et al., 2020) is used in our next step to capture the variation of negative emotions and AUs by having a pop-up message appear on the screen during the game, whenever the threshold value of a negative emotion exceeds 40 out of the maximum value of 100. The message will ask the players to scale the difficulty level by selecting one of three available options: easy – 1; normal – 2; difficult – 3. Based on that, we can utilize a machine learning model (for example, logistic regression), as each pop-up message may have one of three classes (easy, normal, or difficult). Therefore, we can label them into two main classes—not difficult (easy and normal) and difficult—to be able to apply logistic regression and examine if the tracked emotions and expression data predict the difficulty level of any stage of the game.

4. Investigate on whether the more negative emotional response to the Arabic educational game is due to the different difficulty levels between AV and EV. Also, if the primary learning language of the students is English, which influences their preference to EV than AV.

6. ACKNOWLEDGMENT

We would like to extend our appreciation to all the participants who took part in the study and their teachers. Their dedication and commitment have had an enormous impact on the quality of work produced in this study.

REFERENCES


The Impact of Language on Students’ Emotional States in Educational Games: A Comparative Study
Duha Awni Assaf ● Dena Al Thani ● Biklisis Banire

The Impact of Language on Students’ Emotional States in Educational Games: A Comparative Study
Duha Awni Assaf ● Dena Al Thani ● Biklisis Banire


