

Immersive Learning Technologies in English Language Teaching: A Meta-Analysis

Hamide Kubra Altun¹ and Jeongmin Lee^{2,*}

¹ Ewha Womans University; Master of Arts; altun.kubra2@gmail.com

² Ewha Womans University; Associate Professor; jeongmin@ewha.ac.kr

* Correspondence

<https://doi.org/10.5392/IJoC.2020.16.3.018>

Manuscript Received 22 July 2020; Received 23 September 2020; Accepted 23 September 2020

Abstract: *The aim of this study was to perform a meta-analysis of the learning outcomes of immersive learning technologies in English language teaching (ELT). This study examined 12 articles, yielding a total of 20 effect sizes. The Comprehensive Meta-Analysis (CMA) program was employed for data analysis. The findings revealed that the overall effect size was 0.84, implying a large effect size. Additionally, the mean effect sizes of the dependent variables revealed a large effect size for both the cognitive and affective domains. Furthermore, the study analyzed the impact of moderator variables such as sample scale, technology type, tool type, work type, program type, duration (sessions), the degree of immersion, instructional technique, and augmented reality (AR) type. Among the moderators, the degree of immersion was found to be statistically significant. In conclusion, the study results suggested that immersive learning technologies had a positive impact on learning in ELT.*

Keywords: immersive learning technologies; augmented reality; virtual reality; english language teaching; meta-analysis

1. Introduction

Novel notions regarding instructional paradigm come into prominence due to transition to post-industrial period. Notions such as student-centered instruction, experiential learning, and entertaining instruction receive attention from many educators [1] in various disciplines such as English Language Teaching (ELT). Learners are supposed to gain control over their learning and monitor their progress while learning English. There are a variety of online and offline resources that can facilitate learners to regulate their learning process. Immersive learning technologies can be considered as such resources since they allow learners to choose what, how and when to learn. Moreover, they may give learners the opportunity to learn at their own pace. The aim of this study was to examine Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR) within the scope of immersive learning technologies.

The significance of VR in ELT has been investigated by several studies. VR offers students with means for communicating and interacting with other users or avatars, which contributes to the process of practicing English skills [2, 3]. Furthermore, students can experience not only audio-visual stimuli, but also tactile stimuli [4] as well as kinesthetic activities [3, 5] all of which enhance their engagement in learning process. In addition, when learners are exposed to new contents or pressure from their friends, they might become too anxious to demonstrate their English skills. VR, however, enables learners with a surrounding which is stress-free. Virtual characters can be particularly encouraging for them to use English while interacting as they lessen learners' anxiety [6, 7]. With the embodiment of virtual characters [8], [3] by means of roles allotted to them, learners get to learn English in stress-free environments.

The role of AR in ELT has also been discovered by many studies. An enjoyable atmosphere makes learners, especially young learners, to lose the sense of time while they are playing, which leads to higher learning outcomes [9, 10]. Learners may have difficulty in staying focused on learning contents for a long while, yet AR not only enhances their attention span but also develops their comprehension [11], encouraging the

implementation of AR. AR integration becomes more and more common since it improves memories of learners [12, 13] thinking skills and imagination positively by combining learning with play [14]. Also, it gets easier for learners to make associations between virtual images and real objects, thereby contributing to effective vocabulary learning [10].

MR implementation, on the other hand, is still restricted in ELT in contrast with VR and AR because of technical limitations it poses. As documented by [15], context awareness is still constricted to “machine learning and sensor technologies, hindering the capacity for contextual affinity under certain circumstances” (p. 2177). In their study, WordSense was employed as the application to discover MR, and labelling objects incorrectly or categorizing imprecisely were reported to be the main problems. To illustrate, cats might be categorized as mammals. That is why, learners may become confused as the system can present unrelated embeddings because of the contents linked dynamically. Besides, latency might cause challenges when the network is problematical [15]. Thus, VR and AR were focused under the concept of immersive learning technologies in this study.

A variety of variables and elements that have an impact on lessons should be examined to integrate immersive learning technologies in ELT in a more effective way and keep its progressive operation. Finding a study indicating what design components affect the use of immersive learning technologies in ELT positively or negatively has been challenging. Hence, a thorough analysis is required to be performed to reveal design components to be followed or avoided. Meta-analysis recommends a direction in terms of what type of research needs to be conducted in related discipline in addition to drawing conclusions regarding the direction among variables [16]. Therefore, the direction of immersive learning technologies in ELT was explored based on the literature, and meta-analysis was conducted to examine whether any changes were observed in the effectiveness of immersive learning technologies in ELT based on moderator variables.

Several reasons can be considered to carry out a meta-analysis on immersive learning technologies by focusing on ELT. First, a comprehensive conclusion is required since the findings reported so far on the impact of immersive learning technologies in ELT have not been consistent. Even though the studies explored the same language skills in identical ways, the same conclusion was not reached, which can result from significant factors including learning environment and learner characteristics. Second, researchers and educators might have difficulties in designing unprecedented lesson plans. Thus, practical guidelines should be provided to them to design lessons by proposing trends based on the current literature and directions for future studies. Third, variables that may manipulate the effect of immersive technologies in ELT need to be presented in order to offer researchers and educators with effective and efficient means to design lessons and activities.

To this end, the main aim of this study was to perform a meta-analysis of the learning outcomes of immersive learning technologies in ELT by examining the effects of moderator variables by calculating and synthesizing the effect size of the individual studies. The research questions of this study were as follows:

- 1) What is the overall effect size of immersive learning technologies on learning effectiveness in ELT?
- 2) What is the mean effect size and the mean effect size of each sub-element on the learning effectiveness by immersive learning technologies in ELT (cognitive domain, affective domain, and interpersonal domain)?
- 3) Are there any differences in the mean effect sizes of immersive learning technologies in ELT according to the moderator variables?

2. Literature Review

AR was defined as a type of VR [17] and the difference between them is as follows: Accordingly, a user is entirely immersed in an artificial surrounding in VR technology withholding the user from seeing the real world. However, in AR, the user can see the real world on which virtual objects are overlapped. Thus, it can be said that AR does not substitute reality but complements it [17].

VR technology embodies a wide range of affordances, which are as in the following: (a) a high level of immersion in the virtual environment and the target language [18] (b) high-fidelity [19], (c) authentic environment [20], (d) ubiquitous learning [2], and (e) immediate feedback [21], (f) promoting motor skills and spatial ability [3], (g) providing social and global meetings in virtual environments for multi-users as well as all kinds of education settings, either formal or informal learning [2], and (h) increasing a sense of presence in the virtual environment [7].

AR technology also offers various affordances such as (a) letting users experience things directly and see things that they cannot see with naked eyes [22] (b) involving several senses in learning process [23],

(c) facilitating information sharing [24], (d) easy access to information [25], and reduced costs regarding virtual laboratories rather than real ones [26].

Meta-analysis studies regarding immersive learning technologies in ELT have not been conducted yet. However, the impact of VR, AR and MR on learning performances in education in general was examined [27]. They included 33 academic journals and dissertations published between 2008 and 2018. They also included experimental control group research design and utilized R program for data analysis. There were 208 effect sizes of 33 studies, and the overall effect size for learning performance was .87 referring to a large effect size. The individual effect size was .74 for VR and .99 for AR. However, MR was not significant in the 95% confidence interval [-.21, .83]. Educational effects, type of curriculum, age of target, experimental size, and type of design were the moderator variables. Educational effect, curriculum type and experimental size were reported to have no statistically significant difference. However, statistically significant difference was found in age of target and design type. Overall, the results showed that immersive learning technologies were effective for students' learning outcomes. Thus, policy and institutional support considering the environment, learning contents, and curriculum as well as individual high-level learning design and individualized applications should be prepared along with the technological development of immersive learning technologies.

A meta-analysis was conducted in K-12 or higher education settings to investigate overall effect size and the influence of instructional design principles in VR which addresses simulation, games, and virtual worlds [28]. Their study analyzed 13 studies for games, 29 for simulations, and 27 studies for virtual worlds (n=69 in total) which were published until November 2011 by utilizing Comprehensive Meta-Analysis (CMA) program for data analysis. They included studies which employed experimental control group research design. The moderator variables of this study were as follows: learning outcome, testing condition, control group treatment, mode of instruction, teacher access availability vs learner-centered environment, individual work vs group work, duration, feedback, research design quality and type of measure. The results indicated that games, simulations, and virtual worlds were effective to develop learning outcome gains. Accordingly, higher learning gains were revealed in games compared to simulations and virtual worlds. It was also inferred that for feedback, elaborate explanation kind becomes rather appropriate in declarative activities, but for procedural activities, knowledge of correct response is more suitable regarding simulation studies. Furthermore, games were concluded to increase students' performance more when played individually rather than in groups. Moreover, an opposite relationship was detected between the number of intervention sessions and learning gains for games. Finally, it was revealed that students' learning gains diminished when they were assessed repetitively.

An earlier study identified what level of education, learning environment, and field of education moderate learning outcomes of the students in AR systems [29]. Their study analyzed 64 academic journals which applied the pre-test - post-test control design, the post-test only with control design, and single-group pre-test – post-test design and were published from 2010 to 2018. The findings revealed that AR has a medium effect size on students' learning gains ($d = .68$, $p < .001$). The moderator variables were control treatment, learning environment, level of education, and field of education. AR applications were compared in terms of multimedia sources, conventional lectures, and conventional pedagogical instruments. As a result, the learning gains are reported to be higher in case of the treatment with AR resources. As for the results obtained from moderator variables, informal settings were noted to be more effective to conduct the intervention. AR was the most advantageous for students from bachelor or equivalent levels in terms of the level of education. Besides, a higher effect was provided by AR systems for fields of Engineering, Arts and Humanities.

3. Methods

3.1 Data Analysis

3.1.1 Selection Criteria

This was a meta-analysis study to identify if immersive learning technologies are effective in ELT. To this end, scientific papers were searched from 2010 to 2019 to be examined. To detect the relevant articles, this study carried out a systematic search from the following databases: Web of Science, EBSCOhost, ERIC, and Taylor & Francis as these databases include the highest amount of studies regarding education [30]. The following keywords were searched: *Augmented Reality*, *Virtual Reality* linked with *English*, *English learning*, *EFL*, *ESL*, *language learning*, and *foreign language teaching*. The selection of the articles was completed based on PICOS criteria by [31] as given in the following. First, the relevant studies should contain the terms of VR

and AR as well as their definition and characteristics in their contents. Second, the participants of the relevant studies should be pre-school, primary school, secondary school, high school, university students in ELT, and other EFL/ESL students with diverse backgrounds. Third, the studies under question need to have the control condition (pre-test – post-test or control group – experimental group). Finally, the primary studies should consist of experimental, quasi-experimental, or pre-experimental design. Thus, qualitative studies were excluded.

3.1.2 Selection Process

Based on the selection criteria, the mentioned keywords were searched to find the studies employing VR or AR in ELT, which resulted in 1294 studies. After removing the duplicates, 1224 studies remained to be screened. The extracted studies were scanned based on their titles and abstracts, which resulted in eliminating 328 of those 1224 studies since they were not relevant to VR or AR. Another 825 of the remaining studies were not related to ELT. For eligibility, the extracted 71 studies were reviewed, which resulted in eliminating 3 studies as they were not written in English. Additionally, other 3 studies were removed because they could not be accessed for free. Other six studies were excluded since they provided only descriptive information about immersive learning technologies and were not scientific research studies. Furthermore, 22 studies which did not include any learning gains as an outcome variable were removed. Of the remaining 37 studies, 22 studies were eliminated since they lacked the control condition (pre-test – post-test or control group – experimental group). In addition, 3 studies that did not give adequate information to calculate the effect size (standard deviations, mean scores, and sample sizes) were omitted. At the end of this process, 12 studies were selected for meta-analysis. The detailed process of the studies' selection process is given in Figure 1.

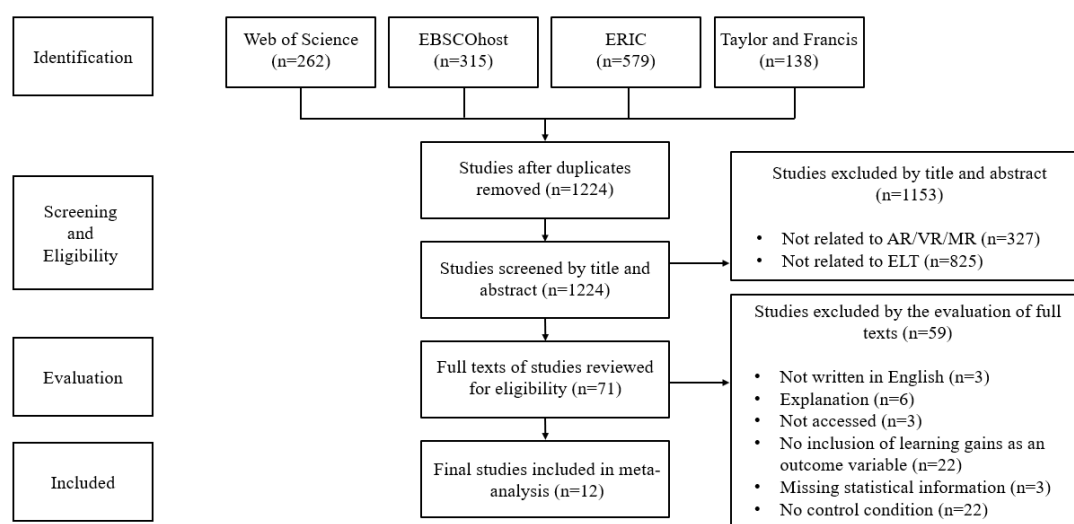


Figure 1. PRISMA Flowchart

3.2 Procedure

3.2.1 Coding of Data

A coding list was adapted from the work of [32] to conduct data coding. The coding list was confirmed by a Ph.D. student majoring in Educational Technology to ensure consistency and reliability. To establish inter-rater reliability, 6 of 12 articles were randomly selected and independently coded by two researchers one of whom was a PhD student majoring in ELT. Microsoft Excel was used to analyze two code sets to find Cohen's Kappa coefficient value. It was found to be 0.78, which indicates agreement between researchers based on [33]. Besides, the rest of the articles were discussed by two researchers and a compromise was reached.

The relevant articles were classified into research characteristics, research contents, research experiments, and calculation of effect size. Basic information regarding authors, publication year, and publication type were coded. Following items were coded to calculate the effect size: the number of the samples, mean and standard deviation values, dependent variables, and whether there is an effect or not. To investigate the effect size based on moderator variables, sample scale, technology type, tool type, work type, program type, duration by session, degree of immersion, instructional technique, and AR type were coded. Degree of immersion was divided into

immersive or non-immersive, and this distinction for VR was made based on the classification by [34]. Accordingly, studies using immersive systems which surround learners completely such as Head-Mounted Display (HMD), Cave Automatic Virtual Environments (CAVE), or Large Screen Projection (LSP) were classified as immersive VR. Others, however, offering learners a conventional projection from a computer through which they interact with the virtual content from the outside were classified as non-immersive VR. When it comes to AR, contents including dynamic videos or animations were categorized as immersive, whereas the ones consisting of static 2D/3D images as non-immersive. Hence, this classification for AR was made based on the contents displayed in AR and the devices for VR.

As for instructional technique, it was categorized into observation or game, which was adapted from [35]. Accordingly, the studies were grouped as observation if students interacted with the contents passively during their exploration via AR. The ones, however, comprising of game elements were sorted as games. There were other instructional techniques such as inquiry (a more active learner compared to observation), and role-play (learners act as the fictional characters that they represent). Yet, the studies scanned in this meta-analysis did not apply these instructional techniques.

Finally, the classification of AR type was made depending on [36]. Accordingly, there are two types as location-based AR systems and image-based AR systems which is divided into marker-based and marker-less AR. Marker-based AR works with markers which are defined as labels that AR system recognizes via camera to display, for example, a 3D image as the same location with the marker. Regarding marker-less AR, real objects are registered as markers. When it comes to location-based AR applications, a precise location information is required for the location-based systems of AR, so the Global Positioning System (GPS) is often utilized. That is why, AR type was divided into marker-based AR and marker-less AR. Markedly, the articles scanned in this meta-analysis did not implement location-based AR.

3.2.2 Data Analysis

The individual effect size and the overall effect size of the studies were calculated using the CMA program to calculate the effect sizes. A two-step procedure was employed based on the study by [37]. At the first step, effect study of each study was calculated, and every effect size was converted into a general metric at the second step. The classification of the effect level while calculating the effect size is as follows: $-0.15 \leq \text{Cohen } d < 0.15$ insignificant effect, $0.15 \leq \text{Cohen } d < 0.40$ small effect, $0.40 \leq \text{Cohen } d < 0.75$ medium effect, $0.75 \leq \text{Cohen } d < 1.10$ large effect, $1.10 \leq \text{Cohen } d < 1.45$ very large effect, $1.45 \leq \text{Cohen } d$ huge effect [38].

3.2.2.1 Selection of Statistical Model

Fixed effects model should be selected if the samples of the meta-analysis studies are of the same size, whereas random effects model should be selected if the samples are of the different size [32, 39]. When it comes to the selection of the statistical model that is employed for the analysis, p and Q values are considered. To this end, the size of the significance value according to (p) 0.05 or the size of Q value according to df value in the chi-square table is examined. In case of $p > 0.05$ or $Q < df$, it might be concluded that the studies included in the meta-analysis are similar and have a homogenous structure indicating the use of fixed effects model as the statistical model [40]. In case of $p < 0.05$ or $Q > df$, it can be concluded that the studies forming meta-analysis are not similar and have a heterogenous structure suggesting random effects model for the statistical model. In this study, random effects model was employed according to the results.

3.2.2.2 Reliability and Validity of the Study

In this study, Funnel plot, Rosenthal's Fail-Safe N and Orwin's Fail-Safe N were employed to present the reliability and validity of this study as well as publication bias [41]. The funnel plot of the effect size of the studies regarding ELT learning outcomes is given in Figure 2.

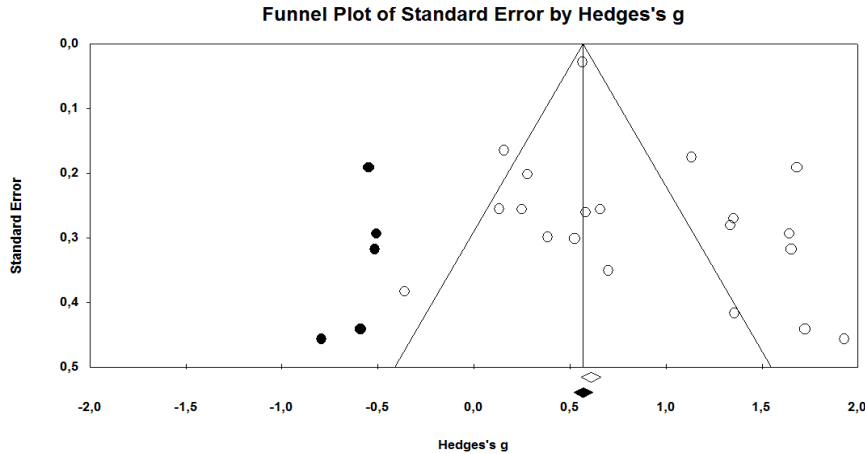


Figure 2. Funnel Plot

The studies that get dispersed symmetrically and inside the funnel do not cause publication bias. However, the ones dispersed unsymmetrically and outside of the funnel lead to publication bias. Figure 2 shows that the studies scanned for this meta-analysis study disperse symmetrically to a large extent. In addition, there are also studies dispersing outside the funnel. The distribution which is close to symmetry indicated that publication bias was low. When Begg-Mazumdar and Egger tests regarding bias indicators of the funnel plot are considered, the values in question were set as Begg-Mazumdar Kendall's tau = 0.25, $p=0.12$ and Egger: bias = 1.21 (95% CI = -0.20 to 2.61), $p=0.09$. In this case, it is expected to show an insignificant difference. Besides, p value is expected to be higher than 0.05. P values obtained are higher than 0.05. According to the overall findings, publication bias is revealed to be at a very low level. The potential causes of publication bias in this study might be the fact that primary studies were excessive and multiple findings from the same study were employed. As a result, publication bias is very low according to the funnel plot.

In addition to funnel plot, Rosenthal's fail-safe number was calculated and Orwin's fail-safe number analysis was conducted in order to investigate publication bias. The data obtained from Rosenthal's fail-safe number is shown in Table 1.

Table 1. Rosenthal's Fail-Safe Number Analysis

| | |
|--|-------|
| Z-value for scanned articles | 17.96 |
| P-value for scanned articles | <.01 |
| Alpha | .05 |
| Tail | 2 |
| Z for alpha | 1.96 |
| Number of scanned articles | 20 |
| Number of absent articles which could bring p-value to > alpha | 1659 |

Table 1 shows that fail-safe number obtained from this meta-analysis study based on Rosenthal method is 1659. In order for p-value for observed studies to be $p>0.05$, namely 1659 studies with an effect size of zero are required to eliminate the significance of the meta-analysis result [42]. This result revealed that there should be at least 1659 studies in the literature in contrast to the findings obtained from this study for the findings of the 12 studies included in the meta-analysis to be deemed as invalid. Furthermore, Orwin's method was also employed to specify publication bias and similar findings were reached. The findings are presented in Table 2.

Table 2. Orwin's Fail-Safe Number Analysis

| | |
|--|-----|
| Standard difference in means in reviewed articles | .61 |
| Criterion for a 'trivial' standard difference in means | .10 |
| Mean standard difference in means in absent articles | .00 |
| Number of absent articles required to bring standard difference in means under 0.1 | 103 |

Table 2 presents that the mean effect size obtained from the findings of this meta-analysis is .61 according to Orwin's method. The mean effect size of .61 found needs to be decreased to .10, that is, 103 papers having

an effect size value of zero should be conducted to consider the overall effect size values as insignificant. Thus, it is indicated that there is no publication bias detected in this study. As a result, it can be concluded that this meta-analysis is reliable considering the fail-safe numbers obtained from both methods. Moreover, the data of 1771 participants composing of 1047 in the experimental groups and 724 in the control groups were examined in the total of the studies included [42]. Accordingly, the number of the samples of this study is large. The fact that the number of the studies, and therefore samples is large can be considered as another factor that increases the reliability of the analysis.

3.2.3 Checklist for Reporting Meta-Analysis

PRISMA Statement aims to make sure the way of reporting systematic reviews is both clear and transparent [43]. A 27-item checklist is provided through PRISMA Statement increase the reliability of the reviews. Hence, these checklist items were followed to offer transparency in this study.

3.2.4 Instrument for Data Analysis

Excel and CMA programs were employed to analyze the effect size of learning of VR and AR technologies in ELT classrooms. Meta-analysis experts from Biostat in the United States developed CMA which is a well-known program for meta-analysis resolving the challenge of utilizing syntax during statistical analysis through SPSS, SAS, STATA, and so forth.

4. Results

4.1 Overall Effect Size of Immersive Learning Technologies in ELT

The articles scanned showed different effect sizes, which is necessary for conducting the study statistically. Heterogeneity tests were applied to figure out whether the effect sizes are appropriate for normal distribution. According to fixed and random effects model, the overall effect size values of the articles are summarized in Table 3.

Table 3. Effect Size Values based on Fixed and Random Effects Model

| Model | Effect size and 95% confidence interval | | | | | Test of null (2-Tail) | | Heterogeneity | | |
|--------|---|-----|----------|-------------|-------------|-----------------------|------|---------------|-------|----------------|
| | Hedges's g | SE | Variance | Lower limit | Upper limit | Z | P | Q | df(Q) | I ² |
| Fixed | .61 | .02 | .00 | .55 | .65 | 7.17 | 0.00 | 121.23 | 19 | 84.32 |
| Random | .84 | .11 | .01 | .61 | 1.07 | | | | | |

As indicated in Table 3, the effect size of immersive learning technologies on learning outcomes in ELT was calculated as .61 based on fixed effects model and .84 based on random effects model. In contrast, Q value that is calculated by applying homogeneity test revealed the distribution of the effect size regarding learning outcomes has a heterogenous structure (Q=121.23; df=19; $X^2(.95) = 30.14$). Considering the results, random effects model was implemented. Thus, it was aimed to eliminate the illusions led by the heterogeneity of the sample.

The effectiveness of immersive learning technologies in ELT and traditional ELT practices on learning outcomes was compared by employing random effects model. The results showed that the overall effect size was .84 indicating a large effect size depending on Cohen's classification. Since the overall effect size was proved to be positive, it can be inferred that immersive learning technologies in ELT yields more positive results than teaching practices without these immersive technologies. The fact that I² values are greater than 75% indicates that the distribution of the effect size on the learning outcomes of the studies is extremely heterogeneous [16]. Figure 3 presents the effect size values and weighs of the articles that examined learning outcomes in ELT.

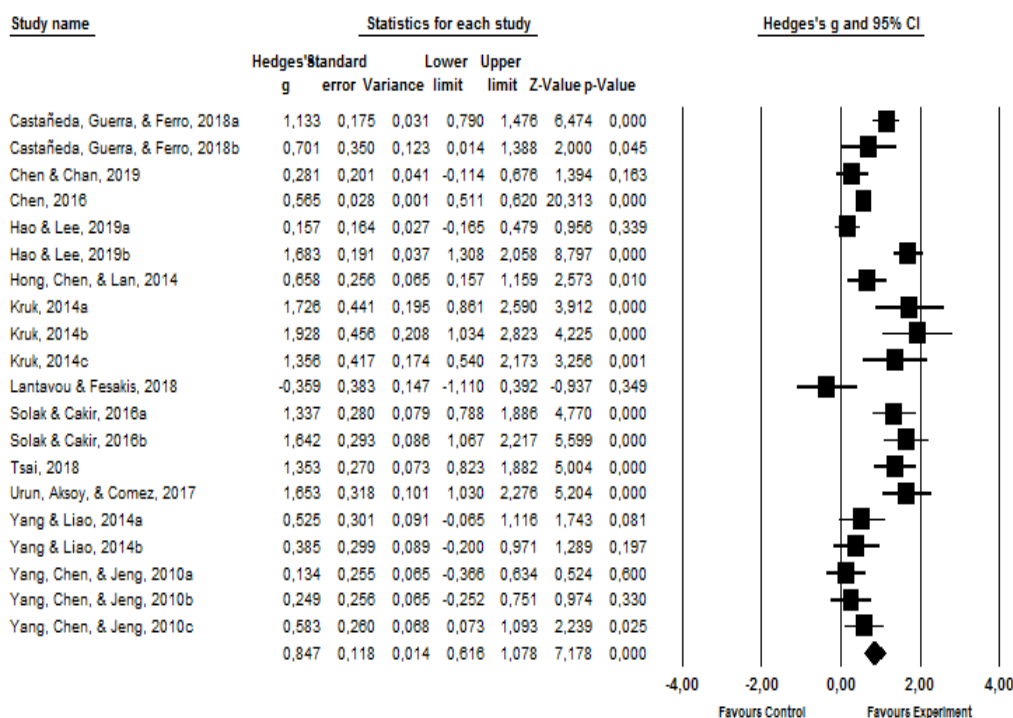


Figure 3. Forest Plot

4.2 Mean Effect Size of Immersive Learning Technologies in ELT on Dependent Variables

Dependent variables were categorized as cognitive, affective and interpersonal domains. Table 4 presents the effect size of cognitive .85 and affective domain .98 with a large effect size, and interpersonal domain .39 revealing a small effect size. Consequently, a statistically significant difference was not explored among the dependent variables ($Q_b=2.08, p=.35$).

Table 4. Dependent Variables

| Moderators | k | Hedges's g | SE | %95 CI | $Q_{between}$ | p |
|---------------|----|------------|-----|------------|---------------|-----|
| Domains | 20 | | | | 2.08 | .35 |
| Cognitive | 17 | .85 | .12 | .61, 1.10 | | |
| Affective | 2 | .98 | .71 | -.42, 2.39 | | |
| Interpersonal | 1 | .39 | .30 | -.20, .98 | | |

4.3 Mean Effect Sizes of Immersive Learning Technologies in ELT according to Moderator Variables

This study examined a total of nine moderators which are sample scale, technology type, tool type, work type, program type, duration (session), degree of immersion, instructional technique, and AR type. First, sample scale was grouped as small scale and large scale. The effect size of the studies for both large sample scale (.82) and small sample size (.87) revealed a large effect size. Hence, a statistically significant difference between the studies with regards to the sample scale was not explored ($Q_b=.03, p=.85$).

Next, technology type was classified as VR or AR. The studies employing VR (.96) and AR (.81) showed a large effect size. Consequently, no statistically significant difference was revealed among the studies according to technology type ($Q_b=.02, p=.86$).

When it comes to tool type, it was categorized as computers, smartphones, and tablets. The studies employing computers (.90) demonstrated a large effect size. The ones implementing smartphones (.51) and tablets (.70) presented a medium effect size. As a result, a statistically significant difference was not found among the studies regarding tool type ($Q_b=.34, p=.84$).

Work type was classified as studies which employed group work or individual work. The effect size of the articles holding group works (.69) indicated a medium effect size, whereas the ones forming individual works (.90) showed a large effect size. In consequence, no statistically significant difference was detected among the studies in terms of work type ($Q_b=.42, p=.52$).

As for program type, it was divided into studies which designed their own program and the ones which employed an existing program. The effect size of the studies designing their own program (.71) had a medium effect size, while the ones implementing existing program (1.04) displayed a large effect size. Accordingly, no statistically significant difference was discovered among the studies in regard to program type ($Q_b=1.48, p=.22$).

Regarding duration (session), it was classified as 3 and less sessions versus 4 and more sessions. The effect size of the studies holding 3 and less sessions (.71) had a medium effect size, yet the ones holding 4 and more sessions (1.07) indicated a large effect size. Thus, no statistically significant difference was revealed among the articles in terms of duration regarding sessions ($Q_b=1.78, p=.18$).

Degree of immersion was grouped as immersive and non-immersive. The effect size of the studies employing immersive contents or devices (1.28) illustrated a large effect size, and the ones using non-immersive contents or devices (.63) showed a medium effect size. Accordingly, a statistically significant difference was found among the studies regarding degree of immersion ($Q_b=3.95, p=.04$).

As for instructional technique, it was divided into game and observation. The effect size of the articles deploying game (.66) had a medium effect size, whereas the ones applying observation (1.12) revealed a large effect size. Therefore, no statistically significant difference was identified among the articles in terms of instructional technique ($Q_b=2.97, p=.08$).

Finally, AR type was categorized into marker-based AR and marker-less AR. The effect size of the papers utilizing both marker-based AR (.86) and marker-less AR (.77) showed a large effect size. Hence, a statistically significant difference was not seen between the groups by AR type ($Q_b=.04, p=.83$). The summary of the mean effect sizes depending on moderator variables was presented in Table 5.

Table 5. Mean Effect Sizes Based on Moderator Variables

| Moderators | Subcategories | k | Hedges's g | SE | %95 CI | $Q_{between}$ | p |
|-------------------------|---------------|----|------------|-----|-------------|---------------|-----|
| Sample Scale | Large | 7 | .82 | .18 | .45, 1.19 | .03 | .85 |
| | Small | 13 | .87 | .18 | .51, 1.23 | | |
| Technology Type | AR | 11 | .81 | .20 | .42, 1.21 | .02 | .86 |
| | VR | 9 | .96 | .16 | .53, 1.19 | | |
| Tool Type | Computer | 15 | .90 | .12 | .65, 1.15 | .34 | .84 |
| | Smartphone | 2 | .51 | .85 | -1.15, 2.19 | | |
| | Tablet | 3 | .70 | .48 | -.24, 1.66 | | |
| Work Type | Individual | 14 | .90 | .13 | .63, 1.17 | .41 | .52 |
| | Group | 6 | .69 | .29 | .11, 1.26 | | |
| Program Type | Developed | 12 | .71 | .20 | .32, 1.11 | 1.48 | .22 |
| | Used existing | 8 | 1.04 | .18 | .68, 1.40 | | |
| Duration by Session | 3 and less | 12 | .71 | .15 | .40, 1.02 | 1.78 | .18 |
| | 4 and more | 8 | 1.07 | .21 | .64, 1.50 | | |
| Immersion | Immersive | 6 | 1.28 | .31 | .68, 1.89 | 3.95 | .04 |
| | Nonimmersive | 14 | .63 | .11 | .41, .85 | | |
| Instructional Technique | Game | 11 | .66 | .14 | .37, .96 | 2.97 | .08 |
| | Observation | 9 | 1.12 | .12 | .69, 1.54 | | |
| AR Type | Marker-based | 5 | .86 | .34 | .18, 1.54 | .04 | .83 |
| | Marker-less | 6 | .77 | .27 | .24, 1.31 | | |

5. Discussion

Twelve articles examined on the effectiveness of immersive learning technologies in ELT were investigated by implementing random effects model and revealed that the overall effect size was .84, which implies a large effect size based on Cohen's classification. Namely, the lessons deploying immersive technologies produced better learning outcomes compared to the lessons without these technologies. Parallel to this finding, the effect of VR, AR and MR technologies on learning performances in general was examined, and it was reported that the overall effect size as .87 according to the analysis of 33 studies, also indicating a large effect size [27]. Although this relevant study did not focus on specific subjects such as English, it is still suggested to actively integrate immersive learning technologies in English education to increase the learning outcomes of the students.

Second, the mean effect size of immersive learning technologies in ELT based on dependent variables was reported to be effective in cognitive and affective domains. The effect size for both cognitive domain (.85) and affective domain (.98) presented large effect size. However, a small effect size for interpersonal domain (.39) was documented. A statistically significant difference was not observed among domains. Nevertheless, future studies need to put more emphasis on how to enhance the effectiveness of immersive learning technologies on these learning domains and set clear objectives regarding their implementation in advance. VR and AR were found to be effective in interpersonal relationship in a previous study conducted on mental health [44]. Reference [45] mentioned that a natural and entertaining surrounding is sought in EFL schools. Thus, these immersive technologies come into prominence to engage the students in motivating, interesting, enjoyable and anxiety-free environments, and encourage teachers to integrate in their classrooms. Hence, these immersive technologies should be employed to enhance affective aspects such as motivation and learning interests in addition to cognitive aspects.

Third, the mean effect size of immersive learning technologies in ELT based on moderator variables was investigated. It was revealed that degree of immersion was the only moderator variable that can explain the difference in the effect size among studies. Immersive contents and environments (1.28) produced higher learning outcomes in contrary to non-immersive contents and environments (.63), which was estimated before the analysis since immersive contents and environments engage learners more in the content contributing to their understanding and ultimately achievement. In the reviewed AR articles, the learners would observe 2D/3D images when they scan the markers. Most of the times, those pictures would be for illustrating vocabulary items, and sometimes explanatory text or pronunciation of vocabulary would be provided. This type of content might be exciting for learners at the beginning but lose its influence after a while. Therefore, AR contents should be comprised of more dynamic contents where 2D/3D animations or videos are employed instead of static contents to keep learners engaged. Furthermore, learners should be provided with supplementary devices in VR such as HMD that can immerse them even more and make them feel as if what they see is real. Several articles [4, 6, 45] included in this study designed or provided a virtual environment that can be experienced through computers and controlled with mouse, which was categorized as non-immersive. Thus, teachers are encouraged to find ways to access to supplementary devices such as HMD to increase the immersive experience of their learners. Because such devices are usually expensive, google cardboard can be recommended since it is very cheap and accessible.

Fourth, sample type, technology type, tool type, work type, program type, duration (sessions), instructional technique, and AR type were the moderator variables which did not show statistically significant differences in learning outcomes. Starting with sample scale, better learning outcomes can be produced in small sample size considering it is easier for teachers to monitor learners and facilitate their learning in smaller groups. Likewise, earlier studies found that both experimental sizes resulted in having large effect sizes with no statistically significant difference [27, 46].

When it comes to technology type, both VR and AR had a large effect size. In line with this finding, previous study [27] showed that AR and VR proved to be of importance with a large effect size by AR and medium effect size by VR. Therefore, their finding supports the results of this study with regards to the effectiveness of AR and VR technologies. Teachers are encouraged to get their students to experience these technologies to increase their learning performances.

Next, computers generated better learning outcomes compared to smartphones and tablets. It may be attributed to the fact that computers provide a much more immersive experience considering the size of the device. Also, they may be more convenient as the learners do not need to hold or carry the device in contrary to smartphones and tablets, which can tire the users. Notably, computers were generally employed for VR technology, whereas smartphones or tablets were used for AR technology. The device should be considered carefully for outside activities as learners would carry them all the time. Thus, it can be suggested that outside activities should not take long as previous studies also report that the duration for AR content should be restricted to 15 minutes at most [47-49]. On the other hand, this finding does not confirm the previous meta-analysis study on AR applications [46] who found that the integration of mobile devices showed the largest effect size, whereas the integration of webcam-based devices presented the smallest effect size indicating a statistically significant difference. The integration of the tools, therefore, might show variance depending on the context and objectives. Consequently, teachers are encouraged to choose the device that applies best to their conditions.

As for work type, teachers can consider arranging individual works while designing classes with immersive learning technologies. Reference [28] found that designing game individually compared to group work increased learner performance more. Nevertheless, group works are suggested [50] especially if interpersonal skills such as communication, collaboration, and interaction are aimed to be fostered.

Regarding program type, previous studies did not conduct a similar analysis. However, it can be implied that programs developed by researchers are as effective as the ones that are already available. Thus, researchers are encouraged to develop more programs that can meet learners' needs the best. Nevertheless, it is suggested that programs should be developed in a way that allows for editing and creation for other users as this might pose a challenge for teachers when they desire some changes in contents [6, 51]. Furthermore, teachers are encouraged to choose flexible contents so that they can easily add to, remove and change contents [48, 49, 52].

When it comes to duration (sessions), it can be claimed that these immersive technologies have a positive influence on learning outcomes whether few numbers of sessions or longitudinal sessions are held. On the contrary, an opposite relationship between the number of intervention sessions and learning outcomes for games was reported in the earlier study [28]. Their result suggests that games in VR environments may not be necessarily effective as the number of the session increases. Therefore, teachers are recommended to design the lessons to the minimum [48, 49, 52, 53].

As for instructional technique, it can be interpreted as observation is more effective compared to game as observation produced higher effect size. Although it was an unexpected result, it bears valuable implications. Observation addresses the way that learners receive information in a passive way. Game, however, refers to the studies implementing game elements. Notably, a majority of the articles examined vocabulary skills of the learners. Thus, it might be inferred that vocabulary learning requires a less active process. Hence, teachers are suggested to design their lessons in a way that learners can focus more on contents without having distracting activities. If learners are engaged in contents, their understanding might be enhanced. On the other hand, teachers are encouraged to pay attention to the mechanics of game so that they can design more effective games through VR or AR. Additionally, there is a need to provide further implications regarding ideal instructional techniques for other skills of language. As an illustration, role-play technique might be more effective for speaking skills, inquiry technique for writing skills, or game technique for listening skills.

Teachers are encouraged to choose AR type that applies best to their objectives. They can plan rather effective lessons via marker-less AR considering that it utilizes real objectives as markers. Hence, they can take advantage of objects in their surroundings to teach English, especially vocabulary. In this way, learners can associate displayed contents with real objects better, which eventually contributes to retention. It is worth noting that there was no location-based AR in the reviewed articles. Therefore, researchers and teachers are encouraged to employ location-based AR in their lessons as long as the environmental conditions are met.

6. Conclusions

Studies on learning outcomes of immersive learning technologies in ELT began being conducted starting from 2010 and it was necessary to verify these immersive technologies' learning outcomes objectively. Hence, this study verified the overall effect of immersive learning technologies in ELT and resulted in a large effect size (.84). Therefore, it was concluded that immersive learning technologies are effective in learning English. Moreover, it was inferred that immersive learning technologies ought to be employed in ELT to increase learning outcomes in cognitive, affective and interpersonal domains. Both affective and cognitive domains had a large effect size, whereas interpersonal domains had a small effect size. Thus, there is a need for instructional design strategies and application plans to increase the learning outcomes of the interpersonal domain. Interpersonal domains demonstrated both the lowest effect size and number of the study ($n=1$), which calls for further research on these aspects. This study illustrated that immersive learning technologies should be integrated into ELT for higher learning outcomes.

In this study, moderator variables were sample size, technology type, tool type, work type, program type, duration (sessions), degree of immersion, instructional technique, and AR type. Consequently, degree of immersion was the main element which can explain the difference in the effect size among studies. Accordingly, better learning outcomes were obtained when the degree of immersion is higher in VR and AR studies, which illustrates implications and strategies for future researchers.

This study presented instructional design guidelines to provide practitioners with assistance in designing lessons with immersive learning technologies in ELT. Design guidelines were provided in two categories as lesson-related and technological guidelines. First, treatment duration, work type, and instructional technique

should be considered in terms of lesson-related guidelines. Starting with duration, the number of the intervention sessions do not impact the learning outcomes according to the results of this study. Hence, it should be decided according to sample group, lesson objectives, and lesson subject for an effective and meaningful learning. However, a reverse relationship was found between the number of intervention sessions and learning outcomes for games implying that games in VR environments do not produce better outcomes in parallel with the increase in the number of the sessions [28]. Therefore, teachers are encouraged to design the lessons to the minimum [48, 49, 52, 53] in case of games in VR environments. Regarding work type, students are suggested to work individually to increase immersive experiences for each learner at most, which is in accordance with the finding by [28]. Groups works can be considered when the aim is to enhance communication, collaboration and interaction skills. In terms of instructional technique, teachers should refrain from distractive activities while designing their lessons so that learners can get more involved in contents displayed through immersive technologies for a better understanding. If they plan to design games, they are suggested to be cautious about game elements and design them according to the lesson objectives.

Second, technology type, tool type, program type, and degree of immersion need to be considered regarding technological guidelines. Starting with technology type, both VR and AR produced positive outcomes. Therefore, both need to be explored further in ELT. As for tool type, computers produced the largest effects on learning outcomes. Thus, computers might be an effective way to provide immersive experiences considering its size compared to smartphones and tablets. However, computers might also pose challenges because they are not portable. In this case, smartphones and tablets offer more ubiquitous experiences especially for outdoor activities. Hence, practitioners are encouraged to choose the device that applies best to their conditions. When it comes to program type, it was concluded that programs that were developed for the research purposes were as effective as the existing ones encouraging researchers to design more programs. As for teachers, they are suggested to choose adaptable programs that easily allow for addition, removal, and change depending on the needs and objectives [48, 49, 52]. Finally, regarding degree of immersion, teachers are recommended to design AR contents in a more dynamic way involving videos and animations to enhance learners' immersion, which will ultimately contribute to higher learning outcomes. For VR contents, teachers should implement supplementary devices such as HMD supplementing computers. They may use google cardboards which are not only cheap and accessible but also appropriate for better immersion.

Finally, this study had some limitations. First, presenting an in-depth discussion was challenging due to a small number of studies even though the studies were collected from 4 major databases in the last 10 years. Thus, future studies might take different sources into account such as ProQuest, ScienceDirect, or JSTOR and broaden the scope of the research through other languages than English. Second, this study investigated variables that have an impact on the effect size via traditional meta-analysis methods and calculated the effect size for each variable. However, future studies should focus on conducting a multilayer-based meta-analysis since the method used in this study has limitations in calculating the effect size by considering various variables which might influence the learning outcomes at the same time [32]. Third, this study excluded some articles to conduct the meta-analysis to reveal the learning outcomes of immersive learning technologies in ELT as they did not provide enough statistical information.

Acknowledgments: This paper is a revision of some of the thesis of the first author.

Conflicts of Interest: The authors declare no conflict of interest.

References

- [1] R. A. Reiser and J. V. Dempsey, Trends and issues in instructional design and technology, (3rd ed.), Allyn & Bacon, 2012.
- [2] M. L. Liaw, "EFL Learners' Intercultural Communication in an Open Social Virtual Environment," *Journal of Educational Technology & Society*, vol. 22, no. 2, pp. 38-55, 2019, <https://www.jstor.org/stable/26819616>.
- [3] M. F. Urun, H. Aksoy, and R. Comez, "Supporting foreign language vocabulary learning through Kinect-based gaming," *International Journal of Game-Based Learning (IJGBL)*, vol. 7, no. 1, pp. 20-35, 2017, doi: <http://dx.doi.org/10.4018/IJGBL.2017010102>.

- [4] Y. L. Chen, "The effects of virtual reality learning environment on student cognitive and linguistic development," *The Asia-Pacific Education Researcher*, vol. 25, no. 4, pp. 637-646, 2016, doi: <https://doi.org/10.1007/s40299-016-0293-2>.
- [5] S. T. Magar and H. J. Suk, "The Advantages of Virtual Reality in Skill Development Training Based on Project Comparison (2009-2018)," *International Journal of Contents*, vol. 16, no. 2, pp. 19-29, 2020, doi: <https://doi.org/10.5392/IJoC.2020.16.2.019>.
- [6] Z. W. Hong, Y. L. Chen, and C. H. Lan, "A courseware to script animated pedagogical agents in instructional material for elementary students in English education," *Computer Assisted Language Learning*, vol. 27, no. 5, pp. 379-394, 2014, doi: <https://doi.org/10.1080/09588221.2012.733712>.
- [7] C. Qu, Y. Ling, I. Heynderickx, and W. P. Brinkman, "Virtual bystanders in a language lesson: examining the effect of social evaluation, vicarious experience, cognitive consistency and praising on students' beliefs, self-efficacy and anxiety in a virtual reality environment," *PloS one*, vol. 10, no. 4, pp. 1-26, 2015, doi: <https://doi.org/10.1371/journal.pone.0125279>.
- [8] T. K. Arslantaş and S. T. Tokel, "Anxiety, motivation, and self-confidence in speaking English during task based activities in Second Life," *Kastamonu Education Journal*, vol. 26, no. 2, pp. 287-296, 2018, doi: <http://dx.doi.org/10.24106/kefdergi.3639889>.
- [9] R. W. Chen and K. K. Chan, "Using Augmented Reality Flashcards to Learn Vocabulary in Early Childhood Education," *Journal of Educational Computing Research*, vol. 57, no. 7, pp. 1812-1831, 2019, doi: <https://doi.org/10.1177/0735633119854028>.
- [10] C. C. A. Tsai, "Comparison of EFL Elementary School Learners' Vocabulary Efficiency by Using Flashcards and Augmented Reality in Taiwan," *Stanisław Juszczak*, pp. 53-65, 2018, doi: <http://dx.doi.org/10.15804/tner.2017.50.4.04>.
- [11] A. Taskiran, "The effect of augmented reality games on English as foreign language motivation," *E-Learning and Digital Media*, vol. 16, no. 2, pp. 122-135, 2019, doi: <https://doi.org/10.1177/2042753018817541>.
- [12] E. Solak and R. Cakir, "Investigating the role of augmented reality technology in the language classroom," *Online Submission*, vol. 18, no. 4, pp. 1067-1085, 2016, doi: <http://dx.doi.org/10.15516/cje.v18i4.1729>.
- [13] H. Lee and C. H. Cho, "Is Augmented Reality Advertising a Cure-all? An Empirical Investigation of the Impact of Innovation Resistance on Augmented Reality Advertising Effectiveness," *International Journal of Contents*, vol. 15, no. 3, pp. 21-31, 2019, doi: <https://doi.org/10.5392/IJoC.2019.15.3.021>.
- [14] A. H. Safar, A. A. Al-Jafar, and Z. H. Al-Yousefi, "The effectiveness of using augmented reality apps in teaching the English alphabet to kindergarten children: A Case Study in the State of Kuwait," *Eurasia Journal of Mathematics, Science & Technology Education*, vol. 13, no. 2, pp. 417-440, 2017, doi: <https://doi.org/10.12973/eurasia.2017.00624a>.
- [15] C. D. Vazquez, A. A. Nyati, A. Luh, M. Fu, T. Aikawa, and P. Maes, "Serendipitous language learning in mixed reality," *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 2172-2179, 2017, doi: <https://doi.org/10.1145/3027063.3053098>.
- [16] M. Borenstein, L. V. Hedges, J. P. T. Higgins, and H. Rothstein, *Introduction to meta-analysis*, Hoboken, NJ: Wiley, 2009.
- [17] R. T. Azuma, "A survey of augmented reality," *Presence: Teleoperators & Virtual Environments*, vol. 6, no. 4, pp. 355-385, 1997, doi: <https://doi.org/10.1162/pres.1997.6.4.355>.
- [18] Y. L. Chen, "The effects of virtual reality learning environment on student cognitive and linguistic development," *The Asia-Pacific Education Researcher*, vol. 25, no. 4, pp. 637-646, 2016, doi: <https://doi.org/10.1007/s40299-016-0293-2>.
- [19] E. Dolgunsöz, G. Yildirim, and S. Yildirim, "The effect of virtual reality on EFL writing performance," *Journal of Language and Linguistic Studies*, vol. 14, no. 1, pp. 278-292, 2018.
- [20] M. Park, "Innovative assessment of aviation English in a virtual world: Windows into cognitive and metacognitive strategies," *ReCALL*, vol. 30, no. 2, pp. 196-213, 2018, doi: <https://doi.org/10.1017/S0958344017000362>.
- [21] J. C. Yang, C. H. Chen, and M. C. Jeng, "Integrating video-capture virtual reality technology into a physically interactive learning environment for English learning," *Computers & Education*, vol. 55, no. 3, pp. 1346-1356, 2010, doi: <https://doi.org/10.1016/j.compedu.2010.06.005>.
- [22] K. H. Cheng and C. C. Tsai, "Affordances of augmented reality in science learning: Suggestions for future research," *Journal of science education and technology*, vol. 22, no. 4, pp. 449-462, 2013, doi: <https://doi.org/10.1007/s10956-012-9405-9>.

- [23] M. A. Castañeda, A. M. Guerra, and R. Ferro, "Analysis on the gamification and implementation of Leap Motion Controller in the IED Técnico industrial de Tocancipá," *Interactive Technology and Smart Education*, vol. 15, no. 2, pp. 155-164, 2018, doi: <https://doi.org/10.1108/ITSE-12-2017-0069>.
- [24] M. H. Wu, "The applications and effects of learning English through augmented reality: a case study of Pokémon Go," *Computer Assisted Language Learning*, pp. 1-35, 2019, doi: <https://doi.org/10.1080/09588221.2019.1642211>.
- [25] B. Mei and S. Yang, "Nurturing Environmental Education at the Tertiary Education Level in China: Can Mobile Augmented Reality and Gamification Help?," *Sustainability*, vol. 11, no. 16, pp. 1-12, 2019, doi: <https://doi.org/10.3390/su11164292>.
- [26] N. Hockly, "Augmented reality," *ELT Journal*, vol. 73, no. 3, pp. 328-334, 2019, doi: <https://doi.org/10.1093/elt/ccz020>.
- [27] M. Yoo, J. Kim, Y. Koo, and J. H. Song, "A meta-analysis on effects of VR, AR, MR-based learning in Korea," *The Journal of Educational Information and Media*, vol. 24, no. 3, pp. 459-488, 2018, doi: <http://dx.doi.org/10.15833/KAFEIAM.24.3.459>.
- [28] Z. Merchant, E. T. Goetz, L. Cifuentes, W. Keeney-Kennicutt, and T. J. Davis, "Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis," *Computers & Education*, vol. 70, pp. 29-40, 2014, doi: <https://doi.org/10.1016/j.compedu.2013.07.033>.
- [29] J. Garzón and J. Acevedo, "A Meta-analysis of the impact of Augmented Reality on students' learning effectiveness," *Educational Research Review*, vol. 27, pp. 244-260, 2019, doi: <https://doi.org/10.1016/j.edurev.2019.04.001>.
- [30] Z. Turan and B. Akdag-Cimen, "Flipped classroom in English language teaching: a systematic review," *Computer Assisted Language Learning*, pp. 1-17, 2019, doi: <https://doi.org/10.1080/09588221.2019.1584117>.
- [31] S. Wood and E. Mayo-Wilson, "School-based mentoring for adolescents: A systematic review and meta-analysis," *Research on social work practice*, vol. 22, no. 3, pp. 257-269, 2012, doi: <https://doi.org/10.1177/1049731511430836>.
- [32] B. Cho, "Verification of the Effect of Flipped Learning," Ph.D. dissertation, Dept. Edu. Tech., Ewha Womans University, Seoul, Republic of Korea, 2018.
- [33] A. J. Viera and J. M. Garrett, "Understanding interobserver agreement: the kappa statistic," *Fam med*, vol. 37, no. 5, pp. 360-363, 2005.
- [34] A. Henderson, N. Korner-Bitensky, and M. Levin, "Virtual reality in stroke rehabilitation: a systematic review of its effectiveness for upper limb motor recovery," *Topics in stroke rehabilitation*, vol. 14, no. 2, pp. 52-61, 2007, doi: <http://doi.org/10.1310/tsr1402-52>.
- [35] M. B. Ibáñez and C. Delgado-Kloos, "Augmented reality for STEM learning: A systematic review," *Computers & Education*, vol. 123, pp. 109-123, 2018, doi: <https://doi.org/10.1016/j.compedu.2018.05.002>.
- [36] R. Wojciechowski and W. Cellary, "Evaluation of learners' attitude toward learning in ARIES augmented reality environments," *Computers & Education*, vol. 68, pp. 570-585, 2013, doi: <https://doi.org/10.1016/j.compedu.2013.02.014>.
- [37] L. V. Hedges and I. Olkin, *Statistical methods for meta-analysis*, Orlando, FL: Academic Press, 1985.
- [38] W. Thalheimer and S. Cook, "How to calculate effect sizes from published research: A simplified methodology," *Work-Learning Research*, vol. 1, pp. 1-9, 2002.
- [39] S. Oh, *Theory and practice of meta-analysis*, Konkuk University Press, Seoul, 2007.
- [40] M. W. Lipsey and D. B. Wilson, *Practical meta-analysis*, Sage Publications, Inc., 2001.
- [41] J. A. Sterne and R. M. Harbord, "Funnel plots in meta-analysis," *The stata journal*, vol. 4, no. 2, pp. 127-141, 2004, <https://doi.org/10.1177/1536867X0400400204>.
- [42] M. S. Rosenberg, "The File-Drawer Problem Revisited: A General Weighted Method for Calculating Fail-Safe Numbers in Meta-Analysis," *Evolution*, vol. 59, no. 2, pp. 464-468, 2005, <https://doi.org/10.1111/j.0014-3820.2005.tb01004.x>.
- [43] A. Liberati, D. G. Altman, J. Tetzlaff, C. Mulrow, P. C. Gøtzsche, J. P. Ioannidis, and D. Moher, "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration," *Journal of clinical epidemiology*, vol. 62, no. 10, pp. e1-e34, 2009, doi: <https://doi.org/10.1016/j.jclinepi.2009.06.006>.
- [44] S. Han, "An Integrative Review on Augmented Reality/Virtual Reality Simulation Programs in the Mental Health Area for Health Professionals," *International Journal of Contents*, vol. 15, no. 4, pp. 36-43, 2019, <https://doi.org/10.5392/IJoC.2019.15.4.036>.

- [45] J. C. Yang, C. H. Chen, and M. C. Jeng, "Integrating video-capture virtual reality technology into a physically interactive learning environment for English learning," *Computers & Education*, vol. 55, no. 3, pp. 1346-1356, 2010, doi: <https://doi.org/10.1016/j.compedu.2010.06.005>.
- [46] M. Ozdemir, C. Sahin, S. Arcagok, and M. K. Demir, "The effect of augmented reality applications in the learning process: A meta-analysis study," *Eurasian Journal of Educational Research*, vol. 74, pp. 165-186, 2018, doi: <http://doi.org/10.14689/ejer.2018.74.9>.
- [47] B. E. Shelton and N. R. Hedley, "Using augmented reality for teaching earth-sun relationships to undergraduate geography students," *The First IEEE International Workshop Augmented Reality Toolkit*, IEEE, pp. 1-8. 2002, doi: <http://doi.org/10.1109/ART.2002.1106948>.
- [48] L. Kerawalla, R. Luckin, S. Seljeflot, and A. Woolard, "'Making it real': exploring the potential of augmented reality for teaching primary school science," *Virtual reality*, vol. 10, no. 3-4, pp. 163-174, 2006, doi: <http://doi.org/10.1007/s10055-006-0036-4>.
- [49] S. Han, "Developmental Study on Augmented Reality Based Instructional Design Principles," Ph.D. dissertation, Dept. Education, Seoul National University, Seoul, Republic of Korea, 2019.
- [50] D. Liu, K. K. Bhagat, Y. Gao, T. W. Chang, and R. Huang, "The potentials and trends of virtual reality in education," *Virtual, augmented, and mixed realities in education*, Springer, Singapore, pp. 105-130, 2017, doi: https://doi.org/10.1007/978-981-10-5490-7_7.
- [51] N. Hockly, "Augmented reality," *ELT Journal*, vol. 73, no. 3, pp. 328-334, 2019, doi: <https://doi.org/10.1093/elt/ccz020>.
- [52] F. Liarokapis and E. F. Anderson, "Using augmented reality as a medium to assist teaching in higher education," 2010.
- [53] S. Cuendet, Q. Bonnard, S. Do-Lenh and P. Dillenbourg, "Designing augmented reality for the classroom," *Computers & Education*, vol. 68, pp. 557-569, 2013, doi: <https://doi.org/10.1016/j.compedu.2013.02.015>.

Appendices

| No | Author & Year | Level of Education | Language Skill |
|----|-------------------------------------|--------------------|---------------------------------------|
| 1 | Castañeda, Guerra, and Ferro (2018) | Primary school | Vocabulary |
| 2 | Chen and Chan (2019) | Pre-school | Vocabulary |
| 3 | Chen Y. L. (2016) | University | Vocabulary |
| 4 | Hao and Lee (2019) | Secondary school | Vocabulary |
| 5 | Hong, Chen, and Lan (2014) | Primary school | Speaking, listening, reading, writing |
| 6 | Kruk (2014) | High school | Grammar |
| 7 | Lantavou and Fesakis (2018) | Primary school | Vocabulary |
| 8 | Solak and Cakir (2016) | Primary school | Vocabulary |
| 9 | Tsai (2018) | Primary school | Vocabulary |
| 10 | Urun, Aksoy, and Comez (2017) | University | Vocabulary |
| 11 | Yang, Chen, and Jeng (2010) | Primary school | Vocabulary |
| 12 | Yang and Liao (2014) | University | Culture learning |



© 2020 by the authors. Copyrights of all published papers are owned by the IJOC. They also follow the Creative Commons Attribution License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.