Cipher in Classrooms: Evaluating Digital Game-Based Language Learning for Irish Vocabulary Acquisition

Liang Xu[1(✉)], Mark Andrade[2], Elaine Uí Dhonnchadha[2], and Monica Ward[1]

[1] Dublin City University, Dublin, Ireland
{liang.xu,monica.ward}@dcu.ie
[2] Trinity College Dublin, Dublin, Ireland
{andradem,uidhonne}@tcd.ie

Abstract. This study explores the educational outcomes of Cipher, a Digital Game-Based Language Learning (DGBLL) application, focusing on vocabulary acquisition in Irish language learning. Cipher integrates engaging game mechanics with pedagogical principles, with the aim of enhancing Irish learners' language skills. Using a double-baseline model, we assess learning gains among 45 primary school students. Results show vocabulary improvements in the experimental groups compared to the control group, underscoring Cipher's educational potential. Key factors contributing to its success include a pedagogical focus, cultural responsiveness, curriculum alignment, content co-creation and innovative game design. This study highlights the potential of DGBLL applications like Cipher to support language acquisition, particularly in less commonly taught languages.

Keywords: Digital Game-Based Language Learning · L2 Learning Outcomes · Vocabulary Acquisition · Indigenous Language · Intervention Evaluation

1 Introduction

The integration of digital technologies into education has increasingly reshaped traditional learning paradigms, offering innovative methods which can enhance student engagement and learning effectiveness. DGBLL combines the immersive and interactive aspects of gaming with educational content, making learning more enjoyable and effective [1, 2]. In the context of the Irish language, DGBLL presents a unique opportunity to address several educational challenges. Irish, an official language of Ireland, is spoken very well by only approximately 4% of the population [3]. Traditional language learning methods often fail to maintain student interest, leading to a decline in engagement and retention in the Irish context [4]. Cipher, a DGBLL for Irish, integrates game elements with educational content to enhance the learning experience for Irish language students. It incorporates storytelling, culturally responsive content, and interactive challenges to make language learning more appealing and effective. To evaluate the educational effectiveness of Cipher, we used the rigorous double baseline model, which

A. Schönbohm et al. (Eds.): GALA 2024, LNCS 15348, pp. 25–35, 2025.
https://doi.org/10.1007/978-3-031-78269-5_3

involved conducting two pre-tests to establish a comprehensive baseline before implementing a 5-week intervention period and a post-test. The research was conducted in an English-medium primary school, focusing on vocabulary acquisition as a key measure of language learning. The study compared the performance of control and experimental groups, assessing Cipher's impact on vocabulary learning.

2 Related Research

Digital Game-Based Language Learning (DGBLL) merges the realms of gaming and language education, utilising digital games as a medium to enhance language acquisition [1]. This approach leverages the engaging and interactive nature of games to create learning environments that can motivate learners and improve their language skills [2]. Research has shown the positive impact of DGBLL on language acquisition, particularly vocabulary development. A meta-analysis [6] of DGBLL studies reveal medium to large effect sizes for vocabulary acquisition, underscoring its effectiveness compared to non-gaming instructional methods. However, there is a need for more comprehensive analyses of DGBLL for less commonly taught languages, such as Irish. The current body of research on DGBLL for the Irish language is very limited, with only a few available studies [7–9]. One of the challenges in measuring the effectiveness of a learning intervention is making sure that observed improvements in vocabulary knowledge are due to the intervention rather than other factors. A significant issue arises from the fact that participants can potentially learn from the tests themselves. When the same or similar tests are used for both pretests and post-tests, there is a risk that improvements in scores are partially due to test familiarity or practice effect rather than the educational intervention. Additional challenges include the diversity of contexts, participants, methodologies and curriculum alignment. Many studies focus on English and adult learners and are not integrated into a structured classroom setting, leading to results that may not generalise to the learning experiences of schoolchildren [10]. Games designed purely for entertainment have been found to be more engaging and effective for language learning compared to those specifically developed for educational purposes [6]. However, this also means that curriculum integration and assessment is challenging.

One common approach to ensure that observed improvements in vocabulary knowledge are due to the intervention rather than test familiarity is using pretests and post-tests with the same items presented in different orders [10, 11]. This method reduces the practice effect. Contextualised testing involves embedding test words in meaningful sentences or providing cues to ensure that students understand the context and meaning of the words. Including both experimental and control groups is one of the most common testing approaches. This allows researchers to distinguish the effects of the intervention from test familiarity [10], and allows for a clear comparison of vocabulary gains attributable to the intervention versus traditional learning methods. Delayed post-tests assess long-term retention, distinguishing between immediate recall and sustained learning. The double baseline model presents a novel approach to evaluating language learning effectiveness [5]. This model involves establishing two baseline measurements before implementing an intervention. Traditional pretest post-test designs often struggle with issues arising from test familiarity. By incorporating a second baseline, researchers

can better distinguish between genuine learning gains due to the intervention and those resulting from repeated exposure to test materials. If improvement 1 (i.e. the difference between pretest 1 and pretest 2 scores) is none or very small and improvement 2 (i.e. the difference between pretest 2 scores and post-test scores) is significantly greater than improvement 1, it shows that repeated exposure to the test words does help slightly but the learning intervention contributed more significantly to learning. Collectively, these methodological strategies aim to ensure that observed vocabulary gains are attributable to the educational intervention rather than simply due to repeated test exposure.

3 Methodology

Cipher is a DGBLL application that fosters language acquisition through an engaging game environment, where players solve language challenges to uncover stories and defeat villains. Developed with a pedagogically informed, learner-oriented approach, the game incorporates elements such as levels, power-ups, game characters and scoring systems to maintain player interest and motivation. The game content is culturally responsive, utilising Irish folklore and mythology to enhance relevance for learners. Additionally, it features AI-generated images and text-to-speech technology to support vocabulary acquisition, reading comprehension and writing [12]. Designing and developing a DGBLL game is both exciting and challenging. With this in mind, Cipher was designed with five key principles in mind. They are a green approach (i.e. reuse and repurpose), a focus on pedagogy, cultural responsiveness, curriculum alignment and co-creation. When designing a DGBLL resource, it can be tempting to focus on the game elements, but a pedagogical focus is important. The Cipher game uses storytelling to immerse learners in tasks that involve deciphering "magic stones" (ciphered words) and "enchanted scrolls," (ciphered texts). AI-generated images are integrated for context illustration of vocabulary and to enhance understanding (see Fig. 1). A culturally responsive approach to teaching involves taking into consideration cultural differences and focusing on what students can do [13]. Cipher was designed to provide learners with an opportunity to read more Irish, in a rich and visually appealing way which complements the school curriculum. Using co-creation with teachers, Cipher texts at an appropriate level of difficulty help students to focus on spellings and reading.

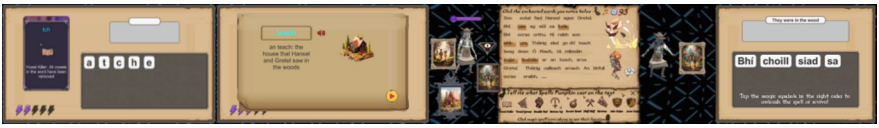

Fig. 1. Vocabulary, reading and writing.

Testing took place in a suburban, English-medium, all boys school, with mixed-ability classes. As our test methodology, we use a double baseline approach [5], conducting two pre-tests for both the experimental and control groups. The primary school had three 4th class classes (age range 9 - 12). One class was randomly selected as the

control group (Room 5), while Room 1 and Room 3 were the experimental groups having 24 students in each class. The control group attended regular Irish classes, similar to the experimental groups, but without Cipher's intervention. The experimental groups received the same standard instruction with the addition of Cipher as a supplementary tool. There were 45 participants in total as some students were exempt from Irish. Each teacher was given an overview of the Cipher game and understood the purpose of the experiments. The researchers led the initial session of the intervention to ensure the game was properly introduced and any technical issues were addressed. For the remaining sessions, the classroom teachers were able to facilitate the game sessions themselves without difficulty. This was important, as the game was designed to be integrated into regular Irish classes without requiring ongoing researcher involvement. Specific vocabulary tasks were developed in conjunction with teachers for the Cipher experiment. It was important to have a mix of easy and difficult words, and also a mix of target words (Cipher words) and non-target words (non-Cipher). The first task was a **vocabulary checking task** whereby students were asked to spell a word. There were 30 words in this task: 20 Cipher words (5 easy and 15 difficult) and 10 non-Cipher words (5 easy and 5 difficult). During the test, each word was spoken, then read in the context of a sentence and then the word was repeated again. The words were read out in blocks of 10 words with a short break in between each block. There were two unrelated practice words before the start of the experiment. The second task was a **vocabulary matching task**, which is generally easier for learners. Students are asked to match words in Irish (L2) to their meaning in English (L1). To reduce guessing, there were more English words than Irish words in the matcher task. The same list of vocabulary words was used for this test. The double baseline design involved a first pretest (pretest 1), a second pretest one month later (pretest 2), a 5-week game block, followed by a post-test. Additionally, a user satisfaction survey was conducted at the end of the intervention, but it was not included in the paper due to limited space. However, previous user satisfaction research for Cipher was conducted and can be found in study [14].

4 Results and Discussion

4.1 Data Analysis

The data analysis is conducted according to a double-baseline model [5]. The data was separated into two overall categories for further classification: the scores achieved in the vocabulary checker scores exercises, which checked for the student's ability to spell the Irish words, and the vocabulary matcher scores exercises, which checked their understanding of the words, i.e. their ability to match the Irish words to their English translations. In the vocabulary checker task, for each word was scored 0 or 1. A score of 0.5 was given where the word was correct except for incorrect accent marks (*síneadh fada*) or missing/unnecessary double consonant. In the vocabulary matcher exercise, a 0 or 1 score was given, depending on whether the words were matched correctly. In both exercises each student was given a percentage score. Within the two task groups, the data was split using the student's room number and the testing round. All testing and game-playing were anonymous but each student was assigned a unique user ID. Students were asked to write their user ID on their test sheets. However, some issues

arose that made it difficult to track students across the three rounds. Some students did not include their user ID on their test sheets, or their ID did not match their user IDs from the game. A final problem occurred where multiple students used the same ID in the same test round, presumably as they had shared a game device and user ID or may have used the same ID as a friend. Additionally, the decision was made not to merge the two experimental groups in the analysis due to variations in teacher involvement. Specifically, in one experimental group (Room 1), the regular teacher was absent for several sessions, and substitute teachers covered the class. The lack of continuity may have been an issue, as it usually takes a substitute teacher some time to become familiar with the students and their progress in each subject. This variation may have influenced the results. Such variability reflects the challenges of conducting real-world educational experiments and was accounted for in the analysis.

To gain an initial understanding of the performance of the three classrooms, each student's scores were split between their performance on target words, those that the student encountered in-game, and non-target words. Both lists had a mixture of easier and more difficult words. Running the Shapiro-Wilk test on each of the thirty-six data sets (target and non-target words, in each of the three rooms, across three rounds, in both forms of tests) established whether they were normally distributed. Half of the vocabulary checker tests (9/18) had a normal distribution, as did three-quarters of the vocabulary matcher tests (15/18). Specifically, regarding the data for the third round (post-test), most of the data was also normally distributed. Both the vocabulary checker and vocabulary matcher target word scores are normally distributed for all students with p-values of 0.106 and 0.109 respectively. While the vocabulary matcher non-target word scores were normally distributed (p-value of 0.235), the same wasn't true for the vocabulary checker non-target word scores ($0.0003 < 0.05$). This limited the scope of statistical analysis that could be performed as discussed in greater detail below. To graph the performance of each classroom, the mean and standard deviations of each of the 36 sets of scores were calculated and then plotted using Google Sheets. The data is shown in Table 1 and the graphs are shown in Fig. 2.

The performance of Room 1 for **vocabulary checker target words** exhibits an almost uniform increase from pre-test 1 to the post-test. The pre-test 2 to the post-test incline is slightly greater than between the two pre-tests ($m_{2-3} = 0.05 > m_{1-2} = 0.04$). This learning gain is less than had been expected, and maybe somewhat explained by the fact that Room 1 had several different teachers during the Cipher testing period. Room 3's performance is closer to the expected (and hoped for) outcome, as the slope between rounds 2 and 3 for target words is greater than that between the two pre-test rounds, implying the existence of an influence from the game. Room 5, the control group, also fits the model as very little increase in performance is seen between rounds 2 and 3. However, there is a strong increase between the two pre-test rounds. This may be explained by the teacher's strong focus on Irish in the classroom, especially through the use of interactive tools such as Duolingo. However, it seems that this progress flattens out in these cases during the intervention period. The equivalent scores for the **vocabulary checker non-target words** also give differing results. In Room 1 there's slight progress from round 1 to 2, which is then lost in the post-test. This could be due to teaching disruption, where the exposure to Irish outside of the game might have been reduced.

Room 3's scores on both of the pre-tests have a marginal difference but do improve after playing the game. It's not immediately clear what causes the difference, but it can be posited that the game increased their interest in learning Irish [14]. The slope of Room 5's graph is almost identical to that in the target words, albeit starting and ending at higher points.

Table 1. The mean (μ) and standard deviation (σ) of each data set across the two exercises.

Vocab Checker		**Target words**		**Non-Target words**			
	R1 (μ σ)	R 2 (μ σ)	R3 (μ σ)		R1 (μ σ)	R 2 (μ σ)	R3 (μ σ)
Room 1	11% 7%	15% 8%	20% 12%	**Room 1**	18% 12%	23% 15%	17% 17%
Room 3	8% 6%	14% 9%	24% 12%	**Room 3**	13% 11%	13% 11%	20% 15%
Room 5	7% 7%	17% 9%	17% 10%	**Room 5**	14% 12%	22% 14%	23% 17%
Vocab Matcher		**Target words**		**Non-Target words**			
	R1 (μ σ)	R 2 (μ σ)	R3 (μ σ)		R1 (μ σ)	R 2 (μ σ)	R3 (μ σ)
Room 1	21% 10%	24% 12%	29% 12%	**Room 1**	29% 13%	25% 10%	22% 12%
Room 3	22% 6%	26% 12%	39% 15%	**Room 3**	24% 11%	26% 6%	29% 11%
Room 5	21% 15%	30% 7%	30% 14%	**Room 5**	27% 14%	25% 12%	27% 11%

Turning to the results for the **matching exercise target words** we see similarly shaped slopes in the target word scores to those in the vocabulary checker. Again, the difference for Room 1 in the increase between pre-test 1 and the post-test appears almost uniform, but there is a slightly greater incline between rounds 2 and 3. For Room 3, the increase between rounds 2 and 3 is greater than that of Room 1 for the **matching exercise target words**, as well as the increase in their score from round 2 to 3 for **the vocabulary checker target words**. Room 5 once again progressed between the two pre-tests, and then maintained the latter score in the third round. The final set of charts for the **matching exercise non-target words** is the most ambiguous and difficult to analyse. It's not immediately clear why Room 1's trajectory is worse here than in the vocabulary checker, even if it does start and finish at higher scores. Room 3's score again shows steady progress, which aligns well with the hypothesis of learning through testing. Room 5's score in rounds 1 and 3 are the same but appear to have a slight drop in round 2.

In summary, for the target words, the pattern was observed in both the vocabulary checker and matching exercise in both experimental groups, where the increase from pre-test 2 to the post-test was sharper than between the two pre-tests. This demonstrates that the game had a positive effect on vocabulary acquisition, even considering that participants learned from the tests. This pattern was not observed in the control group, nor for the non-target words. Additionally, in terms of the performance of the three classes at each data point for each task (e.g., vocabulary checker), comparative analyses were conducted across three rounds, with a total of 36 comparisons made. The majority of the differences between the classes are not statistically significant (21/36). Of the remaining differences, eleven showed a medium effect size, and four had a large effect size. The differences with a medium effect size (Cohen's d between 0.35 and 0.65) were

fairly evenly distributed among the three comparisons (Room 1 vs. Room 3, Room 1 vs. Room 5, and Room 3 vs. Room 5). In Round 2, there are three comparisons with a large effect size. In the vocabulary checker for non-target words, there is a statistically significant difference between Room 1 and Room 3, and between Room 3 and Room 5. The difference between Room 1 and Room 3 can be attributed to the improvements made by Room 1 and Room 5 due to learnings from the test, with no difference between the first two rounds for Room 3. In the vocabulary matcher for the targeted words, there is a large effect size between Room 1 and Room 5. Although all rooms improved from the test, Room 5 showed a more significant improvement compared to Room 1. The final significant difference is between Room 1 and Room 3 in the final round, for target words in the matcher exercise. This is also the most significant difference (Cohen's d of 0.76). This difference is due to Room 1 not improving as much as Room 3 between the second and third rounds. This demonstrates that the rooms do not have statistically different baseline scores, which could be problematic if there are large variations, as starting from a lower baseline typically allows for greater improvement in the post-test.

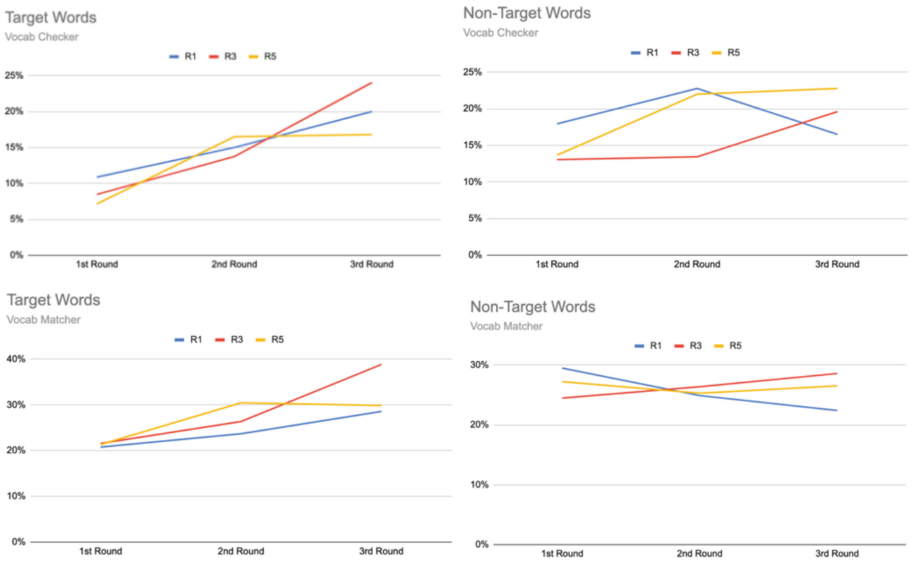

Fig. 2. Four graphs showing the trajectory of each class across the three different testing periods, for both the targeted and non-target words.

4.2 Application of Statistical Methods

As not all the data was normally distributed, non-parametric testing was used. The **Kruskal-Wallis** test is used to compare the means of three or more independent groups, in this case the three different classrooms. Due to the marking scheme there were several cases where multiple students had the same score and thus were ranked with the same value. As we see in Table 2 no p-value fell below the 0.05 significance level, with

the closest score coming from the second round of the vocabulary matcher for target words, where the mean scores differences nearly reach the significance level. This can be explained by low progress made by Room 3, even less progress made by Room 5 at Round 2.

Table 2. Kruskal-Wallis H- and p-values from tests (3 classrooms, each round, 2 exercises).

Vocabulary Checker							
	Target words				Non-Target words		
	Round 1	Round 2	Round 3		Round 1	Round 2	Round 3
H	3.85	1.25	2.22	**H**	1.62	4.97	2.10
p-value	0.146	0.535	0.330	**p-value**	0.445	0.083	0.351

Vocabulary Matcher							
	Target words				Non-Target words		
	Round 1	Round 2	Round 3		Round 1	Round 2	Round 3
H	0.30	5.91	4.23	**H**	1.72	0.83	1.27
p-value	0.861	0.052	0.121	**p-value**	0.422	0.659	0.531

Parametric tests were also carried out on the data where possible. **ANCOVA** (analysis of covariance) was carried out using classroom number and testing round as factors, with the covariate '*Cipher*', a binary variable added to indicate whether that class play the game or not. Results can be found in Tables 3 and 4. It was only possible to conduct the test on three of the four Round 3 datasets, as the data for non-target words in the vocabulary checker was not normally distributed. Table 3 shows that playing the game did have an impact on the student's performance. The *Round* value shows the overall difference between rounds, and so it is reasonable to say that the improved performance of all three classes, for different reasons, led to this difference. Regarding the scores for the vocabulary matcher, the results from the target words also appear to back up the findings from the group trajectories in Fig. 2. Both variables and the covariate had a significant impact on the scores of the classrooms at each stage. The non-target words result once again paints a mixed picture. The low p-value of *Room* can be justified given the previously explained decline in Room 1, but the reason for the covariate's low p-value appears far less clear. While it can be posited that the influence of playing the game increased the student's interest for the Irish language overall [14], it appears very significant due to the low p-value. Clearly, more research in this area is needed.

Table 3. ANCOVA findings on the 3 classrooms' scores on vocabulary checker target words.

Vocabulary Checker – Target Words				
	Sum Sq	df	F	PR(>F)
Room	20.14	2	1.241	0.381
Round	205.56	2	12.671	**0.019**
Cipher	62.77	1	7.739	**0.050**
Residual	32.44	4	NaN	NaN

Two types of t-tests were conducted: independent and paired t-tests. The results from the **independent t-tests** did not return any significant information, while the **paired t-tests** did. As with ANCOVA, the tests were conducted on the Round 3 (post-test) results**.**

Table 4. The findings from ANCOVA, 3 classrooms' scores on all vocabulary matcher words.

Target Wds	Sum Sq.	Df	F	PR(>F)	Non target Wds.	Sum Sq.	df	F	PR(>F)
Room	299.35	2	11.563	**0.022**	**Room**	440.61	2	23.604	**0.006**
Round	195.89	1	7.451	**0.045**	**Round**	2.67	2	0.143	0.871
Cipher	149.03	1	11.513	**0.027**	**Cipher**	330.88	1	35.452	**0.004**
Residual	51.78	4	NaN	NaN	**Residual**	37.33	4	NaN	NaN

Cohen's d for each test was grouped into three ranges, low effect size (less than 0.35), high effect size (greater than 0.65), and a medium effect size. In Table 5, there is a medium effect size between Room 3 and Room 5 in the target words, both in the vocabulary checker and matcher, which is to be expected by comparing the experimental and control groups. The most significant difference from the data appears to be in the vocabulary matcher scores between Room 1 and Room 3, in the target words but also to some extent in the non-target words. It was only possible to conduct the paired t-tests on Room 3, for target and non-target words, as the rest of the data was not normally distributed for both word lists. For the vocabulary matcher, the **paired t-tests** compared the values that the three classes had achieved in the third round, again in the target and non-target words. The calculation returned both Cohen's d and p-values, with the former categorised in the three categories and p-values being compared against a 0.05 significance level. As the other two room's datasets were not normally distributed, Room 3's results from the paired t-tests do not give a lot of information (Table 6). The interpretation for the vocabulary matcher appears clearer, and again supports the argument that playing Cipher did improve the student's ability to match the Irish words to their English translations (Table 7) The low p-values of 0.01 and 0.02 are quite significant, as are the Cohen's d of 0.76 and 0.73. This shows a significant difference in the student's scores in the target word list compared to the non-target words. While it is not as clear in Room 1, for Room 3 it does show the positive difference that playing the game had on the student's ability in Irish.

Table 5. Cohen's d of comparisons between classrooms for target words in the vocabulary checker (left) and vocabulary matcher (right).

Vocabulary Checker	
Target Words	
	Cohen's d
Room 3 / Room 1	0.32
Room 1 / Room 5	0.29
Room 3 / Room 5	0.63

Vocabulary Matcher			
Target Words		Non-Target Words	
	Cohen's d		Cohen's d
Room 3 / Room 1	**0.73**	Room 3 / Room 1	0.50
Room 1 / Room 5	0.09	Room 1 / Room 5	0.34
Room 3 / Room 5	0.60	Room 3 / Room 5	0.18

In summary, at first glance, the data does not present straightforward evidence for the hypothesis that playing Cipher did improve the student's Irish, and most of these

Table 6. Paired t-test between target and non-target words (Room 3 vocab. Checker).

Vocabulary Checker					
	Cohen's d	Mean of diff	Std. Dev. Diffs	t-statistic	p-value
Room 3	0.39	4.615	11.948	1.393	0.19

Table 7. Paired t-test between target and non-target words (all rooms, vocab. Matcher).

Vocabulary Matcher					
	Cohen's d	Mean of diffs	Std dev. Diffs	t-statistic	p-value
Room 1	0.76	6.143	8.056	2.853	**0.01**
Room 3	0.73	10.286	14.030	2.743	**0.02**
Room 5	0.29	3.333	11.566	1.223	0.24

discrepancies (e.g., the vocabulary matcher scores between Room 1 and Room 3) arose from real-world factors (e.g., the absence of the regular teacher in Room 1) outside the control of this study. However, there are many positive indicators to be found. It appears clear that Cipher can work well when paired within a regular classroom context without additional assistance from researchers, as shown by the performance of Room 3 across the three rounds of testing. In the case of Room 1, where there had been some disruption to regular teaching the results were more mixed. The distribution of the data restricted the testing that could be performed. Independent t-tests showed the positive effect size between Room 3 and Room 5, the control group. Carrying out paired t-tests on the vocabulary matcher showed the large effect size between target and non-target words in the classes that used Cipher, and the smaller effect size in the control group.

5 Conclusion

Our findings indicate weak but observable evidence that Cipher positively influences language learning, particularly in vocabulary acquisition. The double-baseline model represents an advancement in the context of DGBLL, which provides a robust method for assessing the learning gains associated with the use of educational games. It highlighted that students in the experimental groups showed improvements in vocabulary checker and matcher tasks compared to the control group, demonstrating Cipher's educational potential. The integration of pedagogical principles, cultural responsiveness, and curriculum alignment contributed to its success. Future research should address the limitations of this study, such as the small sample size and lack of diversity in participant demographics. Conducting similar experiments in different educational settings and with larger, more diverse cohorts could provide a more comprehensive understanding of Cipher's effects. This study focuses on vocabulary acquisition as a starting point for assessing the effectiveness of Cipher. Further research could explore the impact of Cipher on other language skills, such as reading comprehension and writing proficiency.

Addressing these areas will help refine DGBLL applications and their implementation in educational contexts, ultimately enhancing their pedagogical value.

Acknowledgments. This research was supported by the Science Foundation Ireland Centre for Research Training in Digitally-Enhanced Reality (d-real) under Grant No. 18/CRT/6224. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

1. Chowdhury, M., et al.: Digital game-based language learning for vocabulary development. Computers and Education Open **6**, 100160 (2024)
2. Janebi Enayat, M., Haghighatpasand, M.: Exploiting adventure video games for second language vocabulary recall: A mixed-methods study. Innov. Lang. Learn. Teach. **13**(1), 61–75 (2019)
3. Census of Population 2022 - Summary Results, https://www.cso.ie/en/releasesandpublications/ep/p-cpsr/censusofpopulation2022-summaryresults/educationandirishlanguage/, last accessed 2024/07/25
4. O'Sullivan, K., Bird, N., Burns, G.: Students' experiences of the teaching and learning of Irish in designated disadvantaged schools. International Journal of Education, Culture and Society **4**(5), 87–97 (2019)
5. Üstün-Yavuz, M.S. Supporting struggling readers in later primary years: current practice in schools and a multicomponent reading comprehension intervention for year 3 children, Doctoral Thesis, University of Sheffield (2024)
6. Dixon, D.H., Dixon, T., Jordan, E.: Second language (L2) gains through digital game-based language learning (DGBLL): A meta-analysis. Lang. Learn. Technol. **26**(1), 1–25 (2022)
7. Collins, N., Vaughan, B., Cullen, C.: GaeltechVR: measuring the impact of an immersive virtual environment to promote situated identity in irish language learning. Journal of Virtual Worlds Research **12**(3). (2019)
8. Dalton, G., Devitt, A.: Irish in a 3D world: engaging primary school children. Lang. Learn. Technol. **20**(1), 21–33 (2016)
9. Chiaráin, N. N., Chasaide, A. N.: The Digichaint interactive game as a virtual learning environment for Irish. In CALL communities and culture–short papers from EUROCALL 2016, pp. 330–336. (2016)
10. Noroozloo, N., Ahmadi, S.D., Mehrdad, A.G.: The effect of using a digital computer game (sims) on children's incidental english vocabulary learning. Cumhuriyet Üniversitesi Fen Edebiyat Fakültesi Fen Bilimleri Dergisi **36**(3), 1991–2000 (2015)
11. Alfadil, M.: Effectiveness of virtual reality game in foreign language vocabulary acquisition. Comput. Educ. **153**, 103893 (2020)
12. Xu, L., Uí Dhonnchadha, E., Ward, M.: Harnessing the power of images in CALL: AI image generation for context specific visual aids in less commonly taught languages. In EUROCALL 2023 Short Papers, 92–96 (2023)
13. Samuels, A.J.: Exploring culturally responsive pedagogy: teachers' perspectives on fostering equitable and inclusive classrooms. Srate Journal **27**(1), 22–30 (2018)
14. Xu, L., Dhonnchadha, E. Uí., Ward, M.: User experience study of" Cipher: Faoi Gheasa", a digital educational game for language learning and student engagement. In Proceedings of the 2nd Workshop on Games Systems, pp. 5–8 (2022)