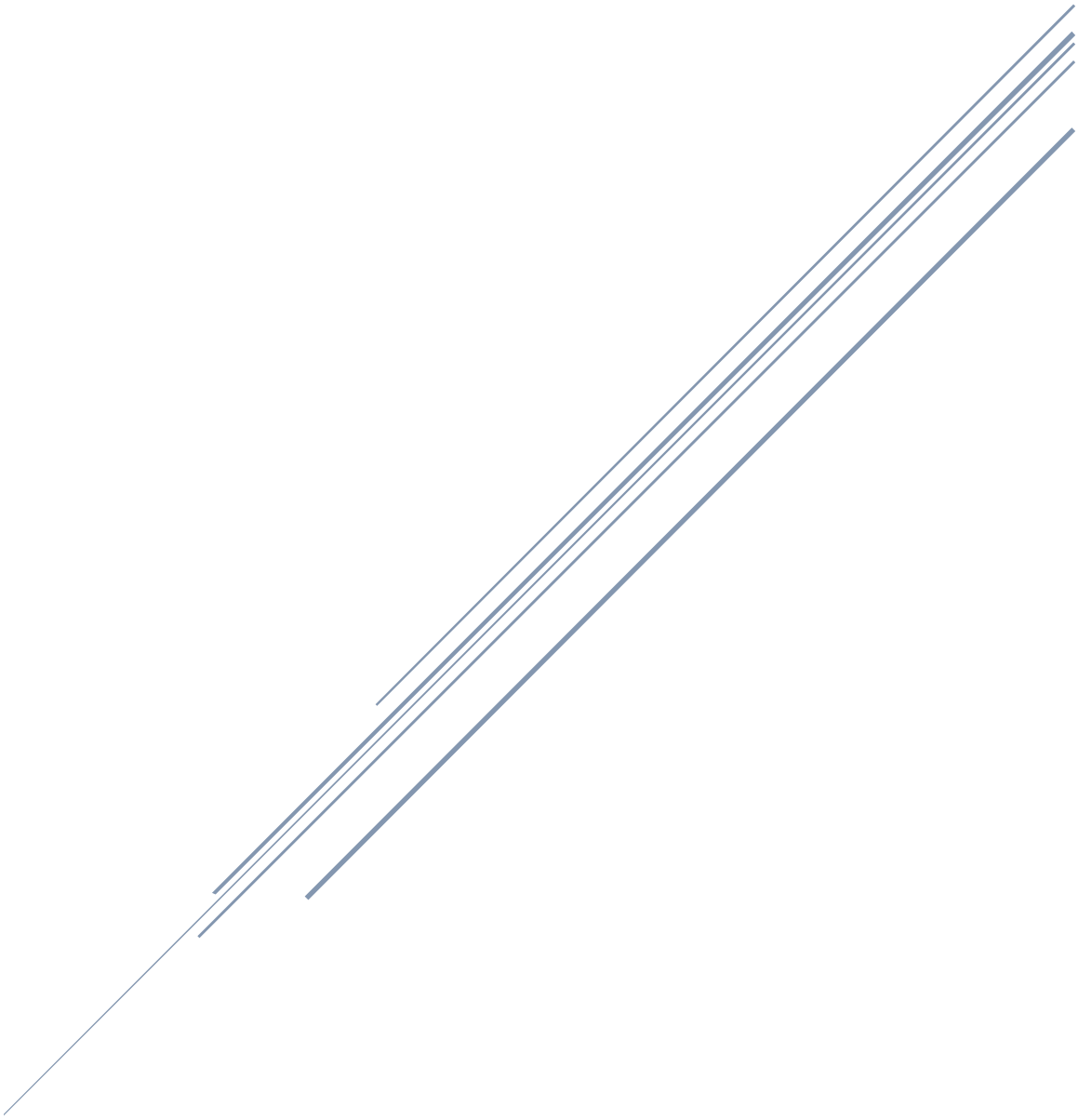
UNDERSTANDING THE IMPACT OF THE EMILE APP

A pilot study



Prepared by:
Ashmita Randhawa, Digital Catapult North East Tees Valley

Understanding the impact of the use of a Phonics app in the classroom

Executive Summary

In the summer of 2022, a short pilot study was conducted to assess the impact of the use of a Phonics app in the classroom. The app, known as Emile is the result of a collaboration between game developers and academics from Manchester Metropolitan University under a Knowledge Transfer Partnership funded by Innovate UK, Emile Education provides a range of resources for classrooms and homes on topics that are covered during numeracy and phonics. This pilot piece of research has been conducted on the phonics app only.

Methodology

How was the research conducted?

The research was conducted in two phases utlising multiple methods to 1) understand the needs of teachers for the teaching of phonics in the classroom , and 2) assess the effectiveness of the Emile app in the development of Phonics knowledge. This research was designed to only ever serve as a pilot, to understand the effect of such testing under real world conditions. This is was a randomised control trial, and was designed to very much fit in with the working of classrooms and families without significant disruption to their routines – effectively attempting to assess the impact of the Emile phonics app 'in the wild'.

The first phase of research involved running interviews with teachers across two schools to understand what their opinion of current teaching methods were in terms of Phonics, and what could be done to alleviate any pain points that they might have.

In the second phase, the app was used by a school as 'homework' that was assigned to the students in one school. . A second school was identified as a 'control' where the app was not used for the same 4 weeks. Both schools were at the same point in the Phonics curriculum at the start of the intervention, and were using or had used the same extra resources in the form of 'Read, Write Inc.' As noted above, the intervention was adopted by one school and not the other and the Phonics Screening check (old exam from 2018) was used to test students at the start and end of the research. Both schools tested students using the Phonic Screening check (PSC2018) to establish a baseline, with one school assigning homework on the Emile app every day for 4 weeks. Both schools would then administer the same PSC2018 test at the end of the 4 weeks.

Where was the research conducted?

Both schools are based in the North East of England, selected for their interest in using the Phonics app.

Results

Phase 1

Teachers at both the schools indicated that their approach to the teaching to Phonics was synthetic, although it is interesting to note that teachers at both schools indicated a desire to be utilising a whole book approaching to teaching Phonics, as it would allow them to focus on the allow them to focus on developing reading skills for students alongside Phonics.

Teachers at both the schools had not used any app or games-based approaches to the teaching of Phonics before this, although the teacher at school Daffodil had seen the use of such an approach in Mathematics and was keen to try it in Phonics. The teacher had observed that students performance had improved

over time in those Mathematics lessons and overall student enjoyed learning Mathematics much more because of the game that they were using; two things the teacher was hoping to see in Phonics as well.

Phase 2

Hierarchical multiple regression (also known as sequential multiple regression) was performed on the data that was obtained from the 2 schools to predict the effect of the use of the Emile app, gender, and ESL on student performance. Given students were tested at the end of 4 weeks of use, the dependent variable or outcome in this case was the difference in scores between Week 4 and Week 0.

It is to be noted that the first model in this analysis, which was to use the variable of using the Emile app or not only did result in a statistically significant model, $R^2 = 0.206$, $F_{(1,95)} = 58.467$, $p < 0.05$ (0.03); adjusted $R^2 = 0.186$. It also demonstrated that the use of the app allowed for the increase in scores from not using the app to using the app by two and a half points (mean = 2.5, stdev = 0.5).

What did teachers have to say about the use of the app?

The teachers at School Daffodil were asked their opinions of the app, and they indicated that the app was not only easy to use and to implement, but that the app was *"fantastic! It's engaging, it targets sounds the children need to consolidate and has helped the children in my class progress."*

When queried about the impact that the app had on the young pupils, the teachers indicated that they saw an increase in the amount of homework that was completed by setting the app as homework rather than other games or worksheets.

What is important to note that the teachers noted a change in the performance of the young pupils in using the app, as evidence by the quote below;

> *'Ultimately, children who were not expected to pass the Phonics Screen passed and the only real change was introducing Emile. I wish we had introduced it sooner!'*

Limitations of the research

It is of note that this research and analysis is very much a pilot and has been used to assess if this methodology can be utilized by the founders of the Emile Education app as a methodology to effectively evaluate the app with larger datasets.

While this is encouraging data to build on, using just the intervention itself to predict outcome is not the best way forward as it does not account for the realities of student performance such as English as a second language. It is also important to note that this analysis has been performed on a very small dataset of pupils, and to build more confidence in the model and to truly assess the impact of the realities of student backgrounds a larger dataset would be required.

Full Report

Context

As part of their research report, Wyse & Bradbury (2022) argue that the approach to phonics and research teaching in England is 'not sufficiently underpinned by research evidence' (ibid), with the authors recommending that there be an overhaul of national curriculum policy. Emile education stands as a potential disruptor to the delivery of Phonics teaching, and therefore the founders were keen to build a research base on the effectiveness of the app in the teaching of Phonics.

Funded by Innovate UK's Edge Catapult Grant, Digital Catapult NETV explored the use of the Emile Education Phonics app in a real word setting to understand and highlight the potential impacts and /or effectiveness of its use.

Founded as a collaboration between game developers and academics from Manchester Metropolitan University under a Knowledge Transfer Partnership funded by Innovate UK, Emile Education provides a range of resources for classrooms and homes on topics that are covered during numeracy and phonics. This pilot piece of research has been conducted on the phonics app only.

Research Design

This research has employed multiple methods to further understanding the teaching of Phonics in the primary classrooms and to understand the potential effectiveness of the Emile app in the teaching of Phonics. Utilising a two-phased approach, researchers sought to establish a baseline understanding of teachers approached the teaching of Phonics, understand the appeal to the use of an app in the teaching of phonics, and ultimately assess the effectiveness of a games-based app in its use in the classroom.

This research was designed to only ever serve as a pilot, to understand the effect of such testing under real world conditions. This is not a randomised control trial, and has been designed to very much fit in with the working of classrooms and families without significant disruption to their routines – effectively attempting to assess the impact of the Emile phonics app 'in the wild'.

The first phase of this research involved interviews with the phonics teachers at the recruited schools to help understand the approach to Phonics teaching in those schools. This also enables us to gain understanding of other resources that are currently in use (apps, games) in schools and why those have been selected.

The second phase of this research involved the use of the app by a school as 'homework' that was assigned to the students in one school. Two schools in the North East of England were selected for this study; these two schools were selected because of 1) their willingness to participate in the research, and 2) they both were at the same phase of teaching of Phonics in Year 1 and they both were using or had used the same extra resources in the form of 'Read, Write Inc.' The intervention was adopted by one school and not the other and the Phonics Screening check (old exam from 2018) was used to test students at the start and end of the research. Both schools tested students using the Phonic Screening check to establish a baseline, with one school assigning homework on the Emile app every day for 4 weeks. Both schools would then administer the same PSC2018 test at the end of the 4 weeks.

All attempts have been made to maintain the anonymity of the schools, and all student data collected was anonymised by the teachers themselves before sharing with the researcher.

Recruitment and Sample

Given the struggles that schools have experienced with COVID-19, recruitment proved tough and was delayed due to lack of availability of teachers or schools that were willing to participate in the research. Recruitment emails were sent by teams at both Emile Education and Digital Catapult NETV ; 14 schools responded, but only 2 were willing to participate in the research. This intervention The 2 schools are based in the North East of England – in different parts of the region and have a different student body profile and testing results in national phonics screening checks, however, the decision was made to continue the research as this was indeed as 'real world' a scenario as could be obtained. Profiles of the two schools and its students can be found below.

School 1: Daffodil

Only 1 teachers was involved in the delivery of Phonics lessons and was therefore interviewed for Phase 1 of the research.

School 1 is a primary school that is located in a large market town in a large combined authority in the North East region. It has been rated as 'Good' in its most recent inspection and is a community mixed school. The total pupil population in 2018/19 was 221, with an equal gender ratio, approximately 26% of the population requiring SEN support , approximately 4% of the population whose first language is not English and over 70% of the pupils eligible for free school meals during the past 6 years.

Data provided on the Year 1 students participating indicates that of the 19 students in the classroom, 12 were female, and that 2 students were ESL students – one whose first language was Russian, and the other Arabic. Data was not provided on free school meals or on age.

This is the school that the Emile app was used as homework.

School 2: Marigold

Two teachers were involved in the delivery of Phonics lessons for Year 1, and so they were both interviewed for Phase 1 of the research.

School 2 is a primary school that is located in another market town in one of the largest counties in the North East. It has been rated 'Good' in its most recent inspection by Ofsted and is also a community mixed school. In 2018/19 , it had 95 students on roll, with an equal gender ratio ; 9.5% of the pupils required SEN support, approximately 1% of the student population's first language was not English, and approximately 1% of the population was eligible for free school meals during the past 6 years.

Data provided on the Year 1 students indicates that of the 22 students in the classroom 8 were female, and there were no students who were ESL students. For parity's sake, no data on age or free school means has been utilised in the modelling.

Phase 1 results

Teachers from both schools (Daffodil and Marigold) indicated that they were experienced teachers in Phonics, having 20 years of experience between themselves collectively. They taught Phonics across both Year 1 and Year 2, and indicated that the teaching of Phonics began not only in reception, but sometimes in the pre-schools that were a part of their schools; essentially indicating that the children at their schools started learning Phonics at a very young age.

Teachers at both the schools indicated that their approach to the teaching to Phonics was synthetic, although it is interesting to note that teachers at both schools indicated a desire to be utilising a whole book approaching to teaching Phonics, as it would allow them to focus on the allow them to focus on developing reading skills for students alongside Phonics, rather than having to focus on just Phonics 'drills' (Daffodil school teacher). Teachers at both the schools also highlighted that Phonics was part and parcel of the everyday timetable with blocks in the timetable devoted entirely to the teaching of Phonics. Various schemes of work had been trialled at both the schools including Read Write Inc which had been in use at School Marigold until recently and was currently in use at School Daffodil.

Testing was a constant feature of the approach to Phonics teaching with teachers at both schools highlighting that they used testing (including old Phonics screening checks) to check students' progress on a very regular basis.

Teachers at both the schools had not used any app or games-based approaches to the teaching of Phonics before this, although the teacher at school Daffodil had seen the use of such an approach in Mathematics and was keen to try it in Phonics. The teacher had observed that students performance had improved over time in those Mathematics lessons and overall student enjoyed learning Mathematics much more because of the game that they were using; two things the teacher was hoping to see in Phonics as well.

Phase 2 results

Hierarchical multiple regression (also known as sequential multiple regression) was performed on the data that was obtained from the 2 schools to predict the effect of the use of the Emile app, gender, and ESL on student performance. Given students were tested at the end of 4 weeks of use, the dependent variable or outcome in this case was the difference in scores between Week 4 and Week 0.

Rational for hierarchical multiple regression

In standard multiple regression, all the independent variables are entered into the regression equation at the same time. By contrast, hierarchical multiple regression enables you to enter the independent variables into the regression equation in an order of your choosing. This has a number of advantages (Cohen et al., 2003), such as allowing you to: (a) control for the effects of covariates on your results; and (b) take into account the possible causal effects of independent variables when predicting a dependent variable.

How is the model built?

As stated above, hierarchical multiple regression is effectively the comparison of multiple regression models. The 3 models that were built in this analysis (as seen in Table 1 below), which are labelled Model "1", Model "2" and Model "3" in the "Model" column. For each model, the variables that are added to the previous model are shown in the "Variables Entered" column. Therefore, for Model 1, the variable that was considered was the presence or the absence of the Emile app (i.e. the use of it as homework); in Model 2, English as a second Language is added to the previous model, and final in Model 3, Gender is added to Model 2.

Table 1: Regression models the hierarchical modelling

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Emile[b]	.	Enter
2	English as a second language[b]	.	Enter
3	Gender[b]	.	Enter

a. Dependent Variable: Final intervention score

b. All requested variables entered.

The three models can essentially visualised as below:

$$Model\ 1: Emile\ (independent\ variable) = Final\ Intervention\ Score\ (difference)$$

$$Model\ 2: Emile + English\ as\ a\ second\ language\ (independent\ variables) = Final\ Intervention\ Score\ (difference)$$

$$Model\ 3: Emile + English\ as\ a\ second\ language + Gender\ (independent\ variables) = Final\ Intervention\ Score\ (difference)$$

When this model was built, care had to be taken that certain assumptions were met (a total of 8). The first two of these had to do with the study design itself, in that the dependent variable was continuous in nature (in this case , test scores fit this criteria), and there were at least 2 or more independent variables (in this case 3 were considered). The other six assumptions relate to how the data fits the regression model and were tested using SPSS Statistics.

1. There was an independence of observations as assessed by the Durbin-Warson statistic of 1.6 (close to 2)
2. There is enough of a linear relationship between the outcome data and each of the independent variables.
3. The data shows homoscedasticity as assessment by the visual representation of a plot of the the studentized residuals (SRE_1) against the unstandardized predicted values.
4. The data do not show collinearity (correlation value <0.7 and the No tolerance Value <0.1)
5. There are no significant outliers, high leverage points or highly influential points based on a visual inspection of the data.
6. The assumption of normality was met as assessed by the Q-Q plot.

Each model has measures that show how well that particular model fits the data (i.e., how well it explains the dependent variable), and these are presented in the first half of the Model Summary table, as highlighted below in Table 2:

Table 2: Summary of fit of data in the three models of the hierarchical regression modelling

Model Summary[d]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.454[a]	.206	.186	2.403	.206	10.121	1	39	.003	
2	.476[b]	.226	.186	2.404	.020	.997	1	38	.324	
3	.490[c]	.240	.179	2.414	.014	.684	1	37	.413	1.581

a. Predictors: (Constant), Emile

b. Predictors: (Constant), Emile, English as a second language

c. Predictors: (Constant), Emile, English as a second language, Gender

d. Dependent Variable: Final intervention score

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	58.467	1	58.467	10.121	.003[b]
	Residual	225.289	39	5.777		
	Total	283.756	40			
2	Regression	64.227	2	32.113	5.559	.008[c]
	Residual	219.529	38	5.777		
	Total	283.756	40			
3	Regression	68.213	3	22.738	3.903	.016[d]
	Residual	215.543	37	5.825		
	Total	283.756	40			

a. Dependent Variable: Final intervention score

b. Predictors: (Constant), Emile

c. Predictors: (Constant), Emile, English as a second language

d. Predictors: (Constant), Emile, English as a second language, Gender

The multiple regression was run to determine if the addition of English as a second language, and then Gender of pupils improved the prediction of Phonics score improvement over and above the use of the Emile app alone. Table 1 above provide full details on each regression model. The full model of the use of the Emile app, English as a second language, and Gender is found to be not statistically significant, R^2 = 0.240 , $F_{(3,95)}$= 22.738, p >0.05 (0.016); adjusted R^2 = 0.179.

It is to be noted however, that the first model in this analysis, which was to use the variable of using the Emile app or not only did result in a statistically significant model, R^2 = 0.206, $F_{(1,95)}$= 58.467, p < 0.05 (0.03); adjusted R^2 = 0.186.

It was also appropriate to calculate the average of the difference in the scores between the start and the end of the intervention. This can be found in the table below:

Table 3: Average of difference in scores

	Average of difference in scores	Std Deviation
School Daffodil	2.5	0.5
School Marigold	1.5	0.05

Limitations of the research and recommendations for future research

It is of note that this research and analysis is very much a pilot and has been used to assess if this methodology can be utilized by the founders of the Emile Education app as a methodology to effectively evaluate the app with larger datasets.

While this is encouraging data to build on, using just the intervention itself to predict outcome is not the best way forward as it does not account for the realities of student performance such as English as a second language. It is also important to note that this analysis has been performed on a very small dataset of pupils, and to build more confidence in the model and to truly assess the impact of the realities of student backgrounds a larger dataset would be required.

It is also important to note that while the teacher at School Daffodil set using the Emile app as part of homework everyday for the students, not all students used it consistently everyday – a more consistent intervention would be to consider utilizing the app in the classroom as part of the teaching mechanism. The fact that students used the app so intermittently indicates that this significance of using the app needs to be considered with an abundance of caution, as other factors seem to be (previous learning perhaps) contributing that have not been measured or the low numbers of data mean that any contribution from the added variables is not being picked up.

It is also important to note that this research and data was only collected after 4 weeks of use – it would be important to run this research for longer to see if statistical significance can be obtained considering all factors.

References

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Wyse, D., & Bradbury, A. (2022). Reading wars or reading reconciliation? A critical examination of robust research evidence, curriculum policy and teachers' practices for teaching phonics and reading. Review of Education, 10, e3314. https://doi.org/10.1002/rev3.3314