

Comparing Digital Badges-and-Points with Classroom Token Systems: Effects on Elementary School ESL Students' Classroom Behavior and English Learning

Ryan Homer, Khe Foon Hew* and Cheng Yong Tan

The University of Hong Kong, Hong Kong // h1287891@connect.hku.hk // kfhw@hku.hk // tancy@hku.hk

*Corresponding author

(Submitted June 17, 2016; Revised October 3, 2016; Accepted November 9, 2016)

ABSTRACT

This paper reports the findings of a field experiment that gamified the classroom experience of elementary school ESL students by implementing digital badges-and-points which students could earn by achieving specific behavioral and learning goals. Altogether, 120 children in eight different classes participated in this study. Four of the classes (experimental group) used the digital badges-and-points available in ClassDojo, a free online classroom management system, while the other four classes (control group) employed a non-digital conventional school token point system. The results showed that digital badges-and-points afforded by ClassDojo significantly improved student learning in two classes (Grades 3 and 4) but not in Grades 1 and 2 classes. Overall, students reported enjoying using digital badges-and-points in the classrooms. Teacher observational data indicated that the digital badges-and-points group displayed more positive and on-task behaviors than the non-digital classroom token point system group.

Keywords

English as second language, Gamification, Engagement, Class token system, Digital badges-and-points, Speaking, Reading

Introduction

Gamification is a term usually used to denote the application of digital game mechanics, such as digital points, or badges in a non-game context to motivate behavior (Deterding et al., 2011). Points refer to tokens that can be collected by users, which can be used as progression indicators, and positive reinforcement (Richter et al., 2015). Badges refer to tokens that appear as icons or logos that signify accomplishments of a particular activity (Bunchball, 2010). Badges fulfill a person's need for acknowledgement and work as virtual status symbol (Sailer et al., 2013). Collectively, badges and points stimulate self-efficacy by measuring progression and providing feedback on an individual's own performance (Gnauk et al., 2012), as well as how one's performance compared with others. Although people prefer to assess themselves using nonsocial and objective standards, if these standards are not available, individuals will evaluate their abilities by comparing themselves with other people (Festinger, 1954).

Already firmly established in the commercial world, digital badges-and-points are widely viewed as a powerful strategy for building brand loyalty, and crowd-sourcing initiatives (Educause, 2011; Caponetto et al., 2014). However, the potential of digital badges-and-points in motivating people goes far beyond that of promoting business success (Caponetto et al., 2014; Lee & Hammer 2011). The use of digital badges-and-points might help improve ESL student classroom engagement and learning of English.

Traditionally, teachers in elementary or special education schools have widely used the classroom point system, a form of token economy, which consists of expectations for desired student behavior and learning, rules that govern how points are earned, and criteria for earning prizes such as stationary upon receiving a certain number of points (Donaldson et al., 2014). Although the use of such token economy has been reported to be effective in increasing appropriate behavior (Kazdin, 1982) and learning, schools today still face major problems around student engagement (Lee & Hammer, 2011). To address these problems, some educators have attempted other means to engage students. Digital badges-and-points may foster better student engagement because it makes the coursework look more like a game-like challenge rather than a chore (Educause, 2011).

Hitherto, a majority of previous studies focused on higher education, and can be characterized as qualitative case studies design in which practitioners describe their implementation of digital badges and points and report primarily user perceptions (Denny, 2013). For example, Chang and Wei (2016) investigated what digital game mechanics were perceived as engaging by MOOC learners. Analysis of 4,891 online survey responses revealed that digital badges were among the top five most engaging game mechanics. Although qualitative case studies are informative, they cannot offer causal explanation because they are seldom compared with a control.

There is a dearth of experimental studies examining the impact of digital badges-and-points on student learning and classroom behavior particularly at the elementary school context. For example, in a literature review of over 120 papers, Caponetto et al. (2014), found that only 3% of studies targeted the elementary school population.

According to Falkner and Falkner (2014), context is important in the system we gamify and the users who participate in it. Some of the questions that we wish to explore in this study include: Would young elementary school students of ages find digital badges-and-points engaging? Would the use of digital badges-and-points motivate them to exhibit certain desired classroom learning behaviors and promote learning? This paper makes a novel contribution by investigating the potential of digital badges-and-points on improving students' classroom behavioral engagement and learning of English among Hong Kong ESL elementary school children. Behavioral engagement refers to students participating in a classroom such as answering questions, and completing set work (Fredricks et al., 2004).

Research questions

The main purpose of this study was to compare the use of digital badges-and-points afforded by ClassDojo versus a non-digital conventional classroom token system on elementary school students' classroom behavior and English learning. The following questions were addressed in this study:

- To what extent does the use of digital badges-and-points afforded by ClassDojo have an impact on student learning when compared to a non-digital conventional classroom token system?
- To what extent does the use of digital badges-and-points influence student behavior when compared to a non-digital conventional classroom token system?
- How did the students perceive the use of digital badges-and-points afforded by ClassDojo?
- How did the teacher perceive the use of digital badges-and-points afforded by ClassDojo?

Method

Participants

The research was conducted during the English lessons at an elementary school in Hong Kong. Two classes from Grades 1, 2, 3 and 4 took part in the study that lasted about 16 weeks (Table 1).

Table 1. Study participants

Level	Experiment	Control
Grade One (P1) Reading	P1B (<i>n</i> = 18, 10 boys, 8 girls)	P1C (<i>n</i> = 14, 8 boys, 6 girls)
Grade Two (P2) Reading	P2B (<i>n</i> = 16, 9 boys, 7 girls)	P2C (<i>n</i> = 16, 9 boys, 7 girls)
Grade Three (P3) Speaking	P3B (<i>n</i> = 16, 10 boys, 6 girls)	P3A (<i>n</i> = 13, 7 boys, 6 girls)
Grade Four (P4) Speaking	P4B (<i>n</i> = 13, 9 boys, 4 girls)	P4A (<i>n</i> = 14, 9 boys, 5 girls)

The ages of the students ranged from 6 years old in P1 to 11 years old in P4. The eight classes were chosen because the same English teacher taught them throughout the entire research period. Of the two classes at each level, one class was randomly as the experiment class, and the other as the control class. All 63 experimental and 57 control students had Chinese as their mother tongue, while English was learned as a second-language.

Background of the English lessons

The English teacher taught reading in double-period lessons each week to the P1 and P2 classes involved in this research. The aim of these lessons was to develop students reading skills. The lessons emphasized shared and guided reading by focusing on a number of big books. The lesson began with students sitting on the reading mat at the front of the classroom. First they sang some songs, followed by phonics practice, high frequency words, and big book shared reading. The students would then move to their group tables for either guided reading or an activity based on the big book's language structure. Students were expected to read and answer comprehension questions, and complete any set work.

The English teacher taught speaking in a single lesson each week to the P3 and P4 classes. The aim of these lessons was to provide students with the necessary skills to do well in the Hong Kong Territory-wide System Assessments (TSA), a nationwide examination. These lessons used questions from past examination papers to

practice and develop students' presentation and speaking skills, as well as improve their speaking confidence. Sitting on their usual chairs in their usual groups, the lessons began with students discussing the past week or any special events with the teacher. Next, either student presentations or sharing from the previous week's work would be undertaken, or a new question from a past examination paper would be examined in preparation for sharing and presentations for the following week. Students were expected to answer questions, share opinions, complete any set work, and present their work either as part of a group or individually.

Experimental lessons

This research utilized the flexibility of the ClassDojo platform to apply digital badges-and-points into the learning of the experimental classes. The participants were familiar with ClassDojo because they had been using it for about four months prior to this study. This makes the ClassDojo environment an unlikely novel experience for the participants. Using P3B's class home page as an example, Figure 1 (best seen in color) shows what the students saw during the lessons. The points were in green next to the students' name. The points were recorded on the students' profiles and could be viewed throughout the lesson via the class display page, or at home by the students or their parents using their assigned log in account.



Figure 1. An example of a class home page

During these lessons, ClassDojo was used to award points to students for achieving certain targeted behavior or learning objectives, which were tailored to each year level (see Figure 2). These points were recorded on the class's homepage on ClassDojo during the lessons and accumulated throughout the duration of the research. For the sake of consistency, all experimental and control reading classes (P1 and P2), as well as the speaking classes (P3 and P4) had the same categories of points.

1) Reading Classes	P1 and P2
Positive Behaviour -	
Plus 1 point - Answering Questions, Helping Others, Listening, Reading	
Plus 2 Points - Reading Very Well	
Negative Behaviour	
Minus 1 Point - Not Listening, Talking & Shouting	
2) Speaking Classes	P3 and P4
Positive Behaviour -	
Plus 1 point - Answering Questions, Reading Well, Speaking	
Plus 2 Points - Elaborating Answers, Imagination	
Negative Behaviour	
Minus 1 Point - Not Listening, Talking & Shouting	

Figure 2. Allocation of points

Using ClassDojo's customizable avatar option, a selection of digital badges was designed to be awarded to students who accumulated a certain amount of specified points (see Figure 3). All participants knew exactly how many points each badge worth. Once a student accomplished the targeted amount of points, the teacher awarded them with a badge in the form of a new avatar on their class page of ClassDojo. This can be seen in Figure 1, which shows the variety of badges that students achieved in a class. Upon achieving the highest amount of points, students received a physical prize, such as stationary.



Figure 3. Reward chart

Before the experiment began, the English teacher first opened a new ClassDojo account and set up a new class page for the reading classes (P1 and P2), and the speaking classes (P3 and P4). This involved entering student names and selecting an avatar for each student in each of the experiment classes. At the beginning of each lesson, the teacher used the overhead projector to display the class home page (Figure 1) on the whiteboard for all students to observe. The teacher then reminded the class how to earn points by referring to the reward chart in the classroom (Figure 2).

The teacher awarded points to individual students or groups of students if they achieved the behavior or learning targets outlined in Figure 2 via his cell phone. For example - if only seven students read a book, then these seven would be awarded a point each; or if only one student elaborated the answers while speaking, only he or she would be awarded a point; or if the whole class read well during shared reading, then everyone would win a point. Having the mobile application at hand allowed the teacher to award points instantaneously. This also allowed the teacher to award points when walking around the classroom observing the students. The teacher also deducted points from individuals who displayed the negative behavior outlined in Figure 2. A loud sound was played by the computer whenever a point awarded (high pitch) or deducted (low pitch), so therefore whether the class home page was on display or not, the students were always aware of their behavior and learning levels. Any points, added or deducted, were automatically recorded to the relevant students profile on the class home page.

Control lessons

The same teacher used the same lesson plans, teaching approach and materials as the experimental classes at each grade level. In other words, the control classes were planned to be exactly the same as the experimental classes except for the exclusion of ClassDojo. Instead, the teacher continued to use the non-digital conventional school points system that was used in all classes in the school (see Figure 4).

This worked by first dividing each control class into four groups of between 4 to 6 students each. The teacher used this group setting to implement the school award system (see Figure 4). The school based award charts were drawn and displayed on the class boards at all times. Points were won by individuals within the group or the group as a whole, and were awarded for the same number of reasons as in the experimental group (see Figure 2). For example - if two students in a group elaborated the answers while speaking, then the particular group would be awarded a point; or if the whole group read well during shared reading, then the particular group would win a point. The group with the highest points at the end of the week received a prize (e.g., a box of chocolates).

The teacher utilized the chart in a similar way to ClassDojo – referring to the number of points each group had at regular intervals throughout the lesson, highlighting and reinforcing how to win points, and deducting points for negative behavior.

<u>Group 1</u> 	<u>Group 2</u>
<u>Group 3</u> 	<u>Group 4</u>

Figure 4. School token point system chart

Measures

Pre and post tests

Both the experimental and control groups took exactly the same pre and post-test at each year level. The purpose of conducting the pre-tests was to establish group equivalence in terms of the students' prior ability. The pre-tests took place the week before the research started, between January 26th and the 30th. The post-tests took place after the research had been completed, between May 25th and 29th. The specific pre and post-test questions at each level were *different* from each other to eliminate any carryover learning effects, but were similar in terms of scope and difficulty to maintain test fairness.

For the P1 and P2 classes, a pre and post reading test was designed using the content of one of the big books they had read that school year. The book content was condensed and printed out on a piece of A4 paper for the students to read. The pre-test and post-test had a similar word count. During the test, the teacher did not offer any help and would allow five seconds before moving the student on if they got stuck on a word. Students were graded between 1 and 15 marks on their clarity, fluency, pronunciation and ability to read. The test took 2 to 4 minutes per student to complete.

For the P3 classes, a pre and post-speaking tests were designed using questions from one of the TSA past examination papers for each year level. Both P3 tests were picture descriptions in which the students were shown four sequenced pictures and then asked five questions to describe what was happening. Students were graded between 1 and 15 marks on their ability to convey information clearly, fluently, and intelligibly with good pronunciation. Students who elaborated their answers and used imagination gained higher marks. The test took 2 to 4 minutes per student to complete. For the P4 pre and post tests, the students had to present their opinions on a set topic, such as eating habits, which was presented to students in the form of a mind map with hints on an A4 sheet of paper. Students had one minute to prepare and one minute to talk. Students were graded between 1 and 15 marks on the content, presentation language, grammar and pronunciation. Students who elaborated and used imagination gained higher marks.

Teacher observations

A behavior chart was used to record the teacher's observations of the student's behavior during the lessons of the experimental and the control classes (Table 2).

The inclusion of the specific behavioral indicators was informed by relevant literature, as well as the teacher's own expectations of the students during the lessons. For example, Tulley and Chiu (1995) found that disruption (e.g., talking out of turn) and inattention (not listening to teacher, being off-task) to be the most frequent

behavior problems. Therefore, in our present study, we developed specific indicators to observe these behaviors (e.g., whether students listen to teacher or other students, complete set work).

Table 2. Behavior chart

Behaviour	Nearly all of the students	Most of the students	Some of the students	Only a few students	One or two students	N/A
Listening to teacher or other students						
Reading						
Answering Questions in class						
Following instructions						
Sitting still						
Waiting for their turn to speak in class						
Staying on task						
Completing set work						

Six rating categories - *Nearly all of the Students (except 1 or 2)*, *Most of the Students (except 3 or 4)*, *Some of the Students (6-10 students)*, *Only a few Students (4 or 5 students)*, *One or Two Students*, and *Not Applicable* – were used to quantify the behavior by the whole class during lessons (class size from 13 to 18 students). The same chart was used for all the classes. The observations began from the initial implementation of the experiment and continued in every lesson until the end.

The teacher completed the charts by putting a tick or cross in the appropriate box. If a behavior was deemed not applicable to the lesson, the teacher ticked the N/A box. For reliability purposes, other teachers observed six lessons (P1B, P1C, P2B, P2C, P3B, and P4A) and completed a behavior chart in order to cross check and thus ensure that the one undertaken by the teacher was consistent. The overall inter-observer agreement was 80%.

Student surveys

In order to help gauge students' perceptions of the use of game mechanics in the lessons and also explore any differences in attitude and motivation towards the lessons, two student surveys were designed, one for all the experimental class students and one for all control class students. The students completed the surveys during the final lesson of the research.

The survey for the experimental classes consisted of four questions. The first question asked if they felt interested in the English lessons. The remaining three questions focused on the use of ClassDojo (e.g., whether students liked it and if it helped them participate more).

The survey for the control classes consisted of three questions. The first question asked if they felt interested in the English lessons. The remaining two questions focused on whether they participated in the lessons and how motivated they felt to do so.

Teacher reflection

In order to ascertain a teacher's perception on the use of game mechanics, he was asked to provide a written reflection on his experience. The reflection would focus on the teacher's experiences and opinions, both positive and negative, of using ClassDojo in the lessons, what actually happened during the lessons and the impacts it had on motivation and behavior of the students during the lessons.

Data analysis

Once collected and recorded, the test scores, behavior charts and student surveys were subjected to descriptive statistics and independent t-test analyses for each pair of experiment-control group in different grade level. To determine effect sizes, we calculated Cohen's (1988) *d* statistic. Once the teacher reflection data had been

collected, they were analysed and interpreted using an analytic strategy involving data reduction, pattern-matching, explanation-building, and conclusion drawing (Miles & Huberman, 1984).

Results

To what extent does the use of digital badges-and-points afforded by ClassDojo have an impact on student learning when compared to a non-digital conventional classroom token system?

Table 3 summarizes student performance on the pre- and post-tests. To test for initial group equivalence, we conducted *t* tests.

Table 3. Summary of pre-test and post-test results

Grade	Pre-test		Post-test	
	Mean	SD	Mean	SD
Reading				
Grade 1 _{experiment}	8.94	3.56	11.00	3.68
Grade 1 _{control}	9.43	2.74	11.14	1.61
Grade 2 _{experiment}	8.69	3.81	11.13	3.38
Grade 2 _{control}	10.81	2.51	11.25	2.65
Speaking				
Grade 3 _{experiment}	9.00	2.56	12.69*	2.09
Grade 3 _{control}	8.46	2.99	10.08	3.17
Grade 4 _{experiment}	10.15	2.64	12.00*	2.77
Grade 4 _{control}	8.36	3.71	9.14	4.04

Note. * $p < .05$.

Comparing students' pre-test performance: Reading

The results of the independent sample *t* test (for pre-tests) indicated that there was no significant difference in the students' prior knowledge or ability in the experimental and control groups for the reading levels (Grade 1: $M(SD)_{\text{experiment}} = 8.94(3.56)$, $M(SD)_{\text{control}} = 9.43(2.74)$, $t(30) = -0.421$, $p = .677$; Grade 2: $M(SD)_{\text{experiment}} = 8.69(3.81)$, $M(SD)_{\text{control}} = 10.81(2.51)$, $t(30) = -1.864$, $p = .072$). Therefore, all Grades 1 and 2 classes were considered equivalent in terms of their prior reading knowledge or ability.

Comparing students' pre-test performance: Speaking

There was also no significant difference in the students' prior knowledge or ability in the experimental and control groups for the speaking levels (Grade 3: $M(SD)_{\text{experiment}} = 9.00(2.56)$, $M(SD)_{\text{control}} = 8.46(2.99)$, $t(27) = 0.523$, $p = .605$; Grade 4: $M(SD)_{\text{experiment}} = 10.15(2.64)$, $M(SD)_{\text{control}} = 8.36(3.71)$, $t(25) = 1.438$, $p = .163$). Hence, all Grades 3 and 4 classes were considered comparable in terms of their prior speaking knowledge or ability.

Comparing students' post-test performance: Reading

The results of an independent sample *t* test (for post-tests) indicated no between-subject difference in post-test scores for the reading groups – Grades 1 and 2 students (Grade 1: $M(SD)_{\text{experiment}} = 11.00(3.68)$, $M(SD)_{\text{control}} = 11.14(1.61)$, $t(30) = -0.135$, $p = .893$; Grade 2: $M(SD)_{\text{experiment}} = 11.13(3.38)$, $M(SD)_{\text{control}} = 11.25(2.65)$, $t(30) = -0.116$, $p = .908$).

Comparing students' post-test performance: Speaking

The results of an independent sample *t* test (for post-tests) revealed that the post-test scores for Grade 3 students in the experimental group, $M(SD)_{\text{experiment}} = 12.69(2.09)$, was significantly higher than for students in the control group, $M(SD)_{\text{control}} = 10.08(3.17)$, $t(27) = 2.661$, $p = .013$, $d = 0.972$ at the 0.05 level of significance with a large effect size (Cohen, 1988). Similarly, Grade 4 students in the experimental group, $M(SD)_{\text{experiment}} = 12.00(2.77)$

also scored significantly higher with a large effect size than the students in the control group, $M(SD)_{\text{control}} = 9.14(4.04)$, $t(25) = 2.128$, $p = .043$, $d = 0.83$.

The box plots of post-test scores for all groups are shown in Figure 5. The box plot is a useful technique to present a visual summary of the distribution of a dataset. Specifically, the box plots in Figure 5 showed the spread of all the post-test data points for the reading (P1 and P2), and speaking (P3 and P4) classes. For example, comparing the P4 control and experimental classes, the following observations can be made: (a) the post-test scores of the P4 experimental class ranged from 9 to 15, excluding one outlier (id number 6), while those in the control class ranged from 0 to 14; and (b) the median post-test score of the P4 experimental class was 13 compared to 10 of the control class. These observations suggest that overall students in the P4 experimental class did better in their speaking test than those in the control class.

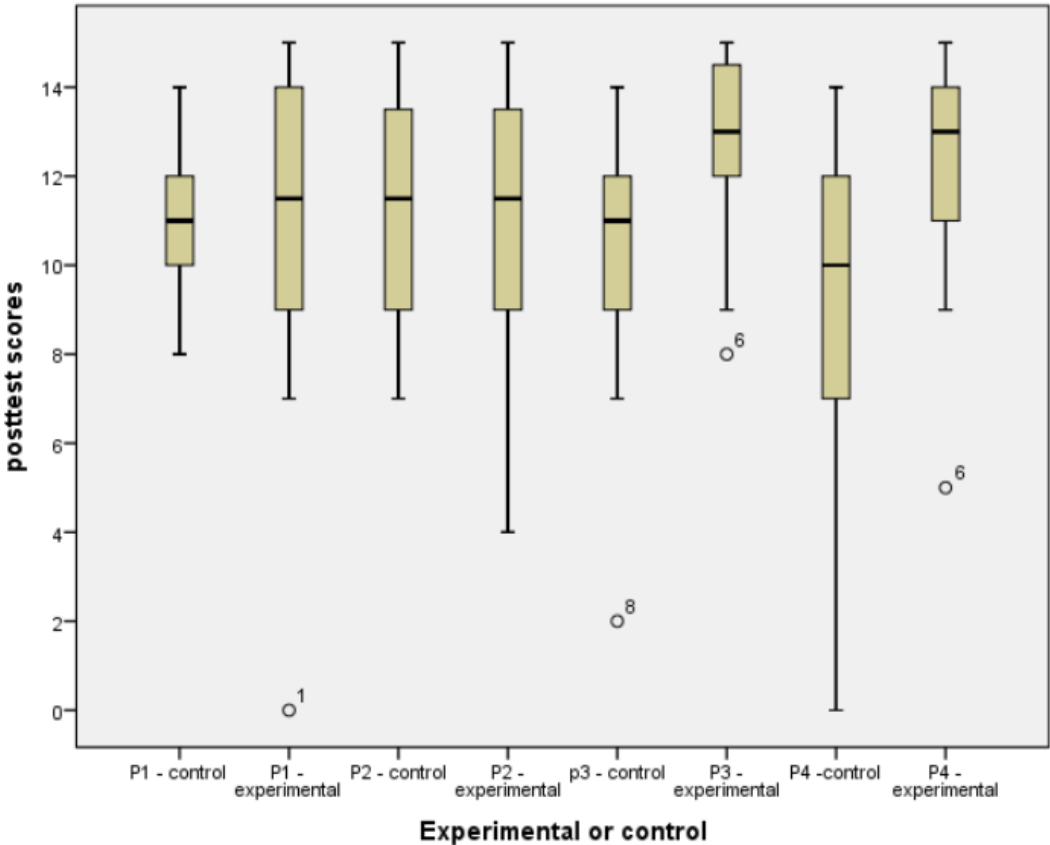


Figure 5. Box plots of post-test scores

To what extent does the use of digital badges-and-points influence student behavior when compared to a non-digital conventional classroom token system?

Reading (P1 and P2 classes)

The reading experimental group (P1 and P2 classes) had a total of 16 lessons observed, while the control group had 17. In the tables, each particular behavior types has two figures under each of the six measurements. For example, the behavior *Staying on task* in Table 4 (experimental group) displays 10 and 62.5% under the measurement *Most of the students*. The first figure shows the number of lessons that the experimental classes achieved that measurement, while the second number is this number of lessons expressed as a percentage of the total number of lessons (16). So in about 63% of the classes, *Most of the students* in the experimental group stayed on task during the English reading lessons.

Table 4 shows that the majority of students in the experimental classes behaved considerably well in all behavior categories. The experimental groups achieve a score of over 80% in six categories (A, B, D, E, G, H). For example, in 81% of the lessons most or nearly all of the students read and stayed on task, in about 88% of the lessons most or nearly all of the students sat and completed the set work, and in 93% of the lessons most or nearly all of the students successfully followed the teacher’s instructions.

Table 4. Behavioral chart data for the experimental reading classes ($n = 16$ lesson observations)

Code	Assigned points	5	4	3	2	1	0
	Behaviour	Nearly all of the students	Most of the students	Some of the students	Only a few students	One or two students	N/A
A	Listening to teacher or other students	4 25%	10 62.5%	2 12.5%	0 0%	0 0%	0 0%
B	Reading	8 50%	5 31.3%	2 12.5%	0 0%	0 0%	1 6.3%
C	Answering questions in class	2 12.5%	10 62.5%	4 25%	0 0%	0 0%	0 0%
D	Following instructions	6 37.5%	9 56.3%	1 6.3%	0 0%	0 0%	0 0%
E	Sitting still	1 6.3%	13 81.3%	2 12.5%	0 0%	0 0%	0 0%
F	Waiting for their turn to speak in class	1 6.3%	8 50%	6 37.5%	1 6.3%	0 0%	0 0%
G	Staying on task	3 18.8%	10 62.5%	3 18.8%	0 0%	0 0%	0 0%
H	Completing set work	7 43.8%	7 43.8%	0 0%	0 0%	0 0%	2 12.5%

It can be seen that students in the control classes behaved less well in almost all behavior categories (see Table 5). The percentage of lessons in which nearly all the students achieved the behavioral targets A to H was less than the experimental group, with the highest percentage being about 6%. When combining the measurements *Nearly all of the students* and *Most of the students* together, it can be seen that only one of the targeted behavioral goals (Completing set work) were achieved in more than 50% of the lessons by most of the students or more. Instead, the majority of the targeted behavioral goals in the control group were achieved by *Some of the students* and *Only a few students*.

Table 5. Behavioral chart data for the control reading classes ($n = 17$ lesson observations)

Code	Assigned points	5	4	3	2	1	0
	Behaviour	Nearly all of the students	Most of the students	Some of the students	Only a few students	One or two students	N/A
A	Listening to teacher or other students	1 5.9%	6 35.3%	9 52.9%	1 5.9%	0 0%	0 0%
B	Reading	1 5.9%	7 41.2%	9 52.9%	0 0%	0 0%	0 0%
C	Answering questions in class	0 0%	4 23.5%	8 47.1%	5 29.4%	0 0%	0 0%
D	Following instructions	0 0%	7 41.2%	6 35.3%	4 23.5%	0 0%	0 0%
E	Sitting still	0 0%	2 11.8%	9 52.9%	3 17.6%	3 17.6%	0 0%
F	Waiting for their turn to speak in class	0 0%	2 11.8%	8 47.1%	5 29.4%	2 11.8%	0 0%
G	Staying on task	0 0%	4 23.5%	10 58.8%	2 11.8%	1 5.9%	0 0%
H	Completing set work	1 5.9%	11 64.7%	4 23.5%	1 5.9%	0 0%	0 0%

In order to better illustrate the differences in behavior between the two groups, a weighted mean score was obtained for each of the behavior types. To do this, each of the six measurements were assigned points – 5 points - *Nearly all of the Students*, 4 points - *Most of the Students*, 3 points - *Some of the Students*, 2 points - *Only a few Students*, 1 point - *One or Two Students*, and 0 points for *Not Applicable*. The higher the weighted mean score, the more positive the behavior of the class. The weighted mean score was calculated by multiplying the number of lessons each behavior measurement was recorded in a particular behavior by the assigned points (sum), and dividing this figure by the total number of lessons (count). For example, the mean score of 4.7 for experimental

group in the behavior *Listening to Teacher or Other Students* is calculated by $(4 \times 5 + 10 \times 4 + 2 \times 3 + 0 \times 2 + 0 \times 1 + 0 \times 0) / 16 = 4.1$.

Figure 6 illustrates the differences in behavior between the experimental and controlled groups. Overall, the experimental group behaved better in all behaviors than the controlled group. It scored a weighted mean score of 4 or more (equivalent to most of students showing a particular behavior in class) in four behaviors, compared to the controlled group that failed to score a mean of 4 or more in any of the behaviors.

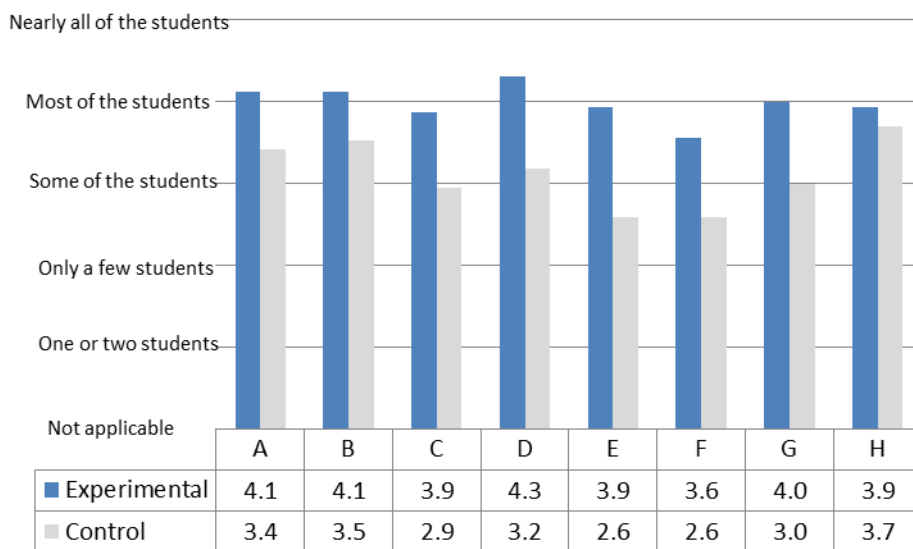


Figure 6. Weighted mean scores for the reading group

Speaking (P3 and P4 classes)

The experimental speaking group (P3 and P4 classes) had a total of 20 lessons observed. The experimental groups achieve a score of 85% or more in all eight categories (Table 6). For example, in 95% of the lessons most or nearly all of the students stayed on task, and completed the set work.

Table 6. Behaviour chart data for experimental speaking classes ($n = 20$ lesson observations)

Code	Assigned points	5	4	3	2	1	0
	Behaviour	Nearly all of the students	Most of the students	Some of the students	Only a few students	One or two students	N/A
A	Listening to teacher or other students	15 75%	4 20%	1 5%	0 0%	0 0%	0 0%
B	Reading	10 50%	7 35%	0 0%	0 0%	0 0%	3 15%
C	Answering questions in class	11 55%	9 45%	0 0%	0 0%	0 0%	0 0%
D	Following instructions	16 80%	4 20%	0 0%	0 0%	0 0%	0 0%
E	Sitting still	15 75%	4 20%	1 5%	0 0%	0 0%	0 0%
F	Waiting for their turn to speak in class	11 55%	7 35%	2 10%	0 0%	0 0%	0 0%
G	Staying on task	14 70%	5 25%	1 5%	0 0%	0 0%	0 0%
H	Completing set work	16 80%	3 15%	1 5%	0 0%	0 0%	0 0%

Table 7 shows the behavioral chart of the control speaking group. The control group had also the same number of lessons observed. The percentage of lessons in which nearly all the students achieved the behavioral targets A to H was less than the experimental group, with the highest percentage being 20%. When combining the

measurements *Nearly all of the students* and *Most of the students* together, it can be seen that only two of the targeted behavioral goals (listening to teacher or other students and sitting still) were achieved in at least 85% of the lessons by most of the students or more. Instead, the majority of the targeted behavioral goals in the control group were achieved by *Some of the students* and *Only a few students*.

Table 7. Behavioral chart data for control speaking classes ($n = 20$ lesson observations)

Code	Assigned points	5	4	3	2	1	0
	Behaviour	Nearly all of the students	Most of the students	Some of the students	Only a few students	One or two students	N/A
A	Listening to teacher or other students	3 15%	15 75%	1 5%	1 5%	0 0%	0 0%
B	Reading	0 0%	10 50%	7 35%	0 0%	0 0%	3 15%
C	Answering questions in class	0 0%	0 0%	12 60%	8 40%	0 0%	0 0%
D	Following instructions	0 0%	6 30%	12 60%	2 10%	0 0%	0 0%
E	Sitting still	4 20%	13 65%	3 15%	0 0%	0 0%	0 0%
F	Waiting for their turn to speak in class	1 5%	9 45%	8 40%	2 10%	0 0%	0 0%
G	Staying on task	0 0%	6 30%	9 45%	5 25%	0 0%	0 0%
H	Completing set work	1 5%	4 20%	11 55%	3 15%	0 0%	1 5%

In order to better illustrate the differences in behavior between the two groups, a weighted mean score was also obtained for each of the behavior types (Figure 7). Overall, the experimental group behaved better in all of the behaviors than the controlled group. It scored a weighted mean score of 4 or more (equivalent to most of students showing a particular behavior in class) in seven categories of behaviors, compared to the controlled group that scored a mean of 4 or more in only two of the behaviors.

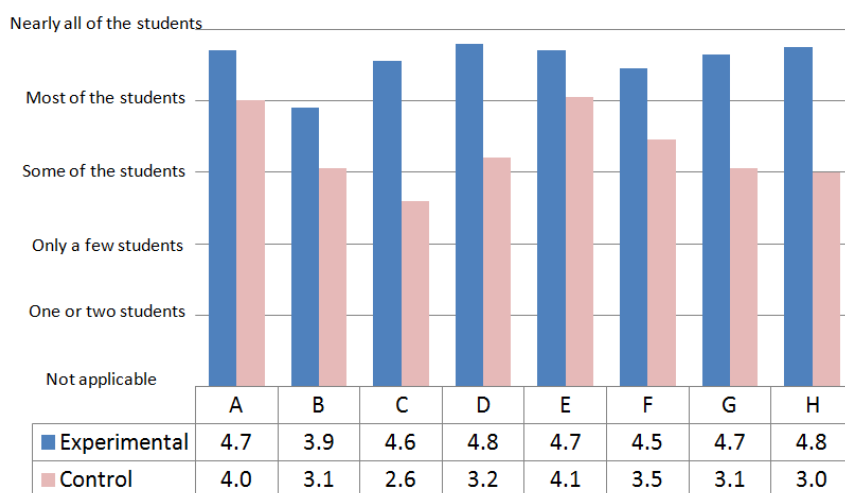


Figure 7. Weighted mean scores for the speaking group

How did the students perceive the use of digital badges-and-points afforded by ClassDojo?

Reading (P1 and P2 classes)

More than 88% of students felt interested in English lessons (agree + strongly agree) (Table 8). 85% of students liked using ClassDojo in lessons and the same percentage agreed that ClassDojo made them participate more. Finally, about 74% of students agreed that ClassDojo helped them work harder in the lessons. It can be seen that the student's responses in the experimental reading group were generally very positive. The results suggest that

students not only enjoyed using the digital badges and points in the classrooms, but also perceived the digital badges and points enhanced their motivation and participation towards learning reading.

Table 8. Survey results of the experimental reading classes

Experiment classes	Total - 34 students				
	Strongly Disagree	Disagree	Neither Disagree or Agree	Agree	Strongly Agree
I feel interested in the English Lessons	0.0%	0.0%	11.8%	23.5%	64.7%
ClassDojo makes me participate more in the lessons	2.9%	5.9%	5.9%	23.5%	61.8%
ClassDojo helps me work harder in the lessons	0.0%	5.9%	20.6%	14.7%	58.8%
I like using ClassDojo	0.0%	2.9%	11.8%	8.8%	76.5%

With regard to the control group, Table 9 shows that 83% of students felt interested in the teacher's English lessons. However, a lesser figure of 50% of students answered that they participated in the lessons with only 60% agreeing that they felt eager to participate in the lessons. These results suggest that although students enjoyed the teacher's lessons, many did not feel motivated enough to participate in the learning. Overall, it can be seen that the student's responses in the control group were generally not as positive as those in the experimental group.

Table 9. Survey results of control reading classes

Control classes	Total - 30 Students				
	Strongly Disagree	Disagree	Neither Disagree or Agree	Agree	Strongly Agree
I feel interested in the English Lessons	3.3%	0.0%	13.3%	23.3%	60.0%
I participated /joined in the lessons	10.0%	6.7%	33.3%	10.0%	40.0%
I am eager to participate in the lessons	10.0%	10.0%	20.0%	23.3%	36.7%

Speaking (P3 and P4 classes)

About 93% of students felt interested in the English speaking lessons (agree + strongly agree) (Table 10). 86% of students liked using ClassDojo in lessons and the same percentage agreed that ClassDojo made them participate more. The results suggest that students not only enjoyed using the digital badges and points in the classrooms, but also perceived the digital badges and point enhanced their motivation and participation towards learning and practicing speaking.

Table 10. Survey results of the experimental speaking classes

Experiment classes	Total - 29 students				
	Strongly Disagree	Disagree	Neither Disagree or Agree	Agree	Strongly Agree
I feel interested in the English Lessons	3.4%	3.4%	0.0%	34.5%	58.6%
ClassDojo makes me participate more in the lessons	6.9%	0.0%	6.9%	31.0%	55.2%
ClassDojo helps me work harder in the lessons	10.3%	0.0%	20.7%	20.7%	48.3%
I like using ClassDojo	6.9%	3.4%	3.4%	13.8%	72.4%

Table 11. Survey results of the control speaking classes

Control classes	Total - 27 Students				
	Strongly Disagree	Disagree	Neither Disagree or Agree	Agree	Strongly Agree
I feel interested in the English Lessons	7.4%	7.4%	25.9%	14.8%	44.4%
I participated /joined in the lessons	3.7%	7.4%	29.6%	44.4%	14.8%
I am eager to participate in the lessons	3.7%	14.8%	44.4%	18.5%	18.5%

In contrast, only 59% of students in the control classes felt interested (agree + strongly agree) in the teacher's English lessons (Table 11). Just slightly half (59%) of students answered that they participated in the lessons

with only 37% agreeing that they felt eager to participate in the lessons. Overall, it can be seen that the student's responses in the control group were generally less positive than in the experimental group.

How did the teacher perceive the use of digital badges-and-points afforded by ClassDojo?

The teacher suggested that ClassDojo was very effective as a behavioral and classroom management system. Not only were students rewarded for basic good behavior, they were given points if they read well or elaborated on their answers, therefore helping to constantly reinforce and focus the students on achieving the targeted learning objectives, and thus, in a sense, developing valuable reflective learning skills.

According to the teacher, the experimental classes were far better behaved than the control classes, and much easier to manage at all year levels. Seeing their points increase and upgrading to a new badge really gripped the students' attention. On the whole, they were more willing to participate. Whereas five hands might go up to answer a question in the control classes, almost all the hands went up in the experimental classes. When classes got off task, the teacher rarely needed to raise his voice, and instead just displayed the point boards and then awarded points for the required behavior. Those who did not behave were soon brought back on task by their peers with the threat of whole class (implicating every student) point deductions by the teacher.

It took much more effort to achieve and maintain similar behavior in the control classes. Even though the behavior was not unsatisfactory, it did not match the experimental classes in terms of consistency in paying attention, following instructions, and sitting still. Furthermore, in the control classes only the more academically competent students would try to tackle reading difficult words, or elaborate when speaking. On the other hand, the majority of the students in the experimental classes were attempting to do so. The teacher was pleasantly surprised to see students who might usually sit back and observe now not only participating, but also pushing themselves further.

Not all students were so enthusiastic about winning badges or about ClassDojo, but the general feeling was extremely positive. In most classes, there were a majority of students who were engaged in the project and eager to gain points. This influence brought an element of peer pressure to classes with students working harder to keep up and achieve what their friends or the majority of the rest of the class had.

Discussion

The aim of this study was to explore if the behavior and learning of ESL students at an elementary school was influenced by introducing digital badges-and-points afforded by ClassDojo into the lessons. A total of 120 children in eight different classes participated in this study. Four of the classes (experimental group) utilized the digital badges-and-points, while the other four classes (control group) employed a non-digital conventional school point system. The participants in the experimental classes were familiar with ClassDojo because they had been using it for about four months prior to this study. This makes the use of digital badges and points an unlikely novel experience for the participants.

The use of digital badges-and-points afforded by ClassDojo significantly improved the *oral* post-test scores of Grades 3 and 4 students compared to their counterparts who utilized the non-digital conventional school point system. One explanation of the better performance of the experimental group students was that they were more engaged such as answering questions in class, staying on task, and completing set work.

However, we found no significant difference in *reading* post-test scores between the experimental and control groups for Grades 1 and 2. The reason for is currently not clear. It is possible that a combination of young age and circumstances of these classes had a part to play. As noticed by the teacher, students in the four P1 and 2 classes were still consumed with other happenings in the classroom, such as who they were sitting next to, in addition to the excitement of going to different classrooms. Moreover, the nature of the P1 and P2 English reading curriculum focused mainly on *reviewing* and *reinforcing* what they were expected to know at this level, rather than enforcing new skill or knowledge; this may have influenced the lack of significant differences in test scores. The curriculum probably was easy enough to follow without the need for additional motivational tools (e.g., digital badges), hence progression in both groups being similar and relatively good.

Our results indicated that the majority of students in the experimental classes behaved considerably better than their counterparts in the control classes who used the non-digital conventional token system. The qualitative

results also showed that both the teacher and the students perceived the use of gamification in lessons as having a generally positive impact on behavior and motivation.

We offer three plausible explanations for the generally favorable results concerning classroom management in the experimental group. First, the use of different badges (see Figure 3) gives students a sense of progression. According to the self-determination theory of motivation (Ryan & Deci, 2000), users seek competency. The element of progression shows users where they are in their learning and how far they are from reaching the goals; it motivates users to move toward completion and feelings of competence. Reaching a checkpoint or milestone, and earning new badges stimulates endorphin release in learners in a similar way that exercise, excitement, or love does (Wroten, 2014). This helps promote a feeling of well-being, and thus motivates learners to do try harder. This element of progression was not evident in the conventional school token point system.

Second, the use of the online ClassDojo helped create a sense of pervasiveness in the learning environment. Students can view the class home page which shows each student's achievements inside as well as outside class. The school token point system that was implemented in the control group can only be viewed in-class. Third, the online ClassDojo group tended to focus on individual-based achievement, as opposed to the group-based achievement practiced in the conventional school token system. It is possible that the use of individual-based achievement gave each learner a more personal responsibility for managing their own learning and behavior.

We would like to highlight a certain point of interest. Some skeptics may argue that digital badges and points mainly act as extrinsic rewards which could undermine a user's intrinsic motivation (e.g., Nicholson, 2012). Such criticisms, however, remain questionable and speculative. First, it is not conclusive that extrinsic rewards will always interfere with intrinsic motivation. Several recent studies, for example, found that extrinsic rewards did not negatively affect the participants' intrinsic motivation needs (Ledford et al., 2013; Mekler et al., 2013). Second, focusing only on intrinsic motivation is not a practical strategy for schools. As Deci and Ryan (2000, p. 55) stated, "Frankly speaking, because many of the tasks that educators want their students to perform are not inherently interesting or enjoyable, knowing how to promote more active forms of extrinsic motivation becomes an essential strategy for successful teaching."

Conclusion

In addressing the limitations of past research, the overall aim of this paper was to explore the impact of digital badges-and-points afforded by ClassDojo on behavior and learning of students' at the elementary school level. We acknowledge that digital badges-and-points are not a universal solution to all motivational shortcomings; however this study suggests that they had a positive impact on students and teacher, considerably improving learning in some of the classes involved in the research, and positively stimulating many of the behaviors expected of student's during lessons in all of the classes involved in the research. No adverse effect on student learning or behavior was found.

For further research, a larger study in which students are exposed to digital badges-and-points for longer periods would be beneficial in assessing the longevity of their impact on learning and behavior. Further studies should also focus on the impacts of digital badges-and-points in a number of subject areas and not just English. Classroom dynamics change depending on the subject and it would be interesting to see the impact digital badges-and-points may have. We also suggest that a similar research project takes place in other schools that have a more diverse mix of students such as students in an international elementary school setting.

References

Bunchball. (2010). *Gamification 101: An Introduction to the use of game dynamics to influence behavior*. Retrieved from <http://www.bunchball.com/gamification/gamification101.pdf>

Caponetto, I., Earp, J., & Ott, M. (2014, October). *Gamification and education: A Literature review*. In *8th European Conference on Games Based Learning (ECGBL)* (pp. 50-58) (Berlin, Germany). Retrieved from <http://www.itd.cnr.it/download/gamificationECGBL2014.pdf>

Chang, J-W., & Wei, H-Y. (2016). Exploring engaging gamification mechanics in massive online open courses. *Journal of Educational Technology & Society*, *19*(2), 177-203.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge Academic.

- Deci, E. L., & Ryan, R. M. (2000). The “What” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry*, 11(4), 227-268.
- Denny, P. (2013). The Effect of Virtual Achievements on Student Engagement. In W. E. Mackay, S. Brewster, & S. Bodker (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 763-772). New York, NY: ACM Press.
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining gamification. In A. Lugmayr, H. Franssila, C. Safran, & I. Hammouda (Eds.), *Proceedings of the 15th international Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 9-15). New York, NY: ACM Press.
- Donaldson, J. M., DeLeon, I. G., Fisher, A. B., & Kahng, S. W. (2014). Effects of and preference for conditions of token earn versus token loss. *Journal of Applied Behavior Analysis*, 47, 537-548.
- Educause. (2011). *7 things you should know about gamification. Scenario, 2007* (June 15th). Retrieved from <https://library.educase.edu/resources/2011/8/7-things-you-should-know-about-gamification>
- Falkner, N. J. G., & Falkner, K. E. (2014). “Whither, badges?” or “wither, badges!?”: A Metastudy of badges in computer science education to clarify effects, significance and influence. In P. Kinnunen (Eds.), *Proceedings of the 14th Koli Calling International Conference on Computing Education Research* (pp. 127-135). New York, NY: ACM Press.
- Festinger, L. (1954). A Theory of social comparison processes. *Human Relations*, 7(2), 117–140.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. (2004). School engagement: Potential of the concept: State of the evidence. *Review of Educational Research*, 74, 59–119.
- Gnauk, B., Dannecker, L., & Hahmann, M. (2012). Leveraging gamification in demand dispatch systems. In D. Srivastava & I. Ari (Eds.), *Proceedings of the 2012 Joint EDBT/ICDT Workshops* (pp. 103–110). New York, NY: ACM Press.
- Kadzin, A. E. (1982). The Token economy: A Decade later. *Journal of Applied Behavior Analysis*, 15(3), 431-445.
- Ledford, G. E., Gerhart, B., & Fang, M. (2013). Negative effects of extrinsic rewards on intrinsic motivation: more smoke than fire. *WorldatWork Journal*, 22(2), 17-29.
- Lee, J. J., & Hammer, J. (2011). Gamification in education: what, how, Why Bother? Definitions and uses. *Exchange Organizational Behavior Teaching Journal*, 15(2), 1–5
- Mekler, E. D., Bruhlmann, F., Opwis, K., & Tuch, A. N. (2013). Do points, levels and leaderboards harm intrinsic motivation? An Empirical analysis of common gamification elements. In L. E. Nacke, K. Harrigan & N. Randall (Eds.), *Proceedings of the First International Conference on Gameful Design, Research, and Applications* (pp. 66-73). New York, NY: ACM Press.
- Miles, M., & Huberman, A. (1984). *Analyzing qualitative data: A Sourcebook of new methods*. Beverly Hills, CA: Sage.
- Nicholson, S. (2012). *A User-centered theoretical framework for meaningful gamification*. Retrieved from <http://scottnicholson.com/pubs/meaningfulframework.pdf>
- Richter, G. Raban, D. R., & Rafaeli, S. (2015). Studying gamification: The Effect of rewards and incentives on motivation. In T. Reiners & L. C. Wood (Eds.), *Gamification in Education and Business* (pp. 21-46). Cham, Switzerland: Springer International Publishing.
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25, 54-67.
- Sailer, M., Hense, J., Mandl, H., & Klevers, M. (2013). Psychological perspectives on motivation through gamification. *Interaction Design and Architecture(s) Journal*, 19, 28-37.
- Wroten, C. (2014). *Four tips: Gamification, according to endorphins by Christie Wroten: Learning solutions magazine*. Learning Solutions Magazine. Retrieved from <http://www.learningsolutionsmag.com/articles/1414/four-tips-gamification-according-to-endorphins>