GAMIFIED MICRO-LEARNING FOR INCREASED MOTIVATION: AN EXPLORATORY STUDY

Till Halbach and Ivar Solheim
Norwegian Computing Center, Oslo, Norway

ABSTRACT

This work investigates in how far gamification and micro-learning, implemented by the novel technologies H5P and xAPI, are suitable to increase the motivation and learning performance of pupils with cognitive and behavioral challenges. The context for the field trials is the Norwegian SOL framework (short for Systematic Observation of Reading) in 6th and 7th grade. The results show that, albeit there are technical deficiencies, the technology is well suited for crafting engaging learning experiences. The prototype developed in the course of the project is capable of assessing task fidelity and pupil performance. With the proper statistical analysis, it can be a valuable ingredient for useful teacher assets such as monitoring tools.

KEYWORDS

On Demand, Just in Time, Micro-Learning, Mobile App, Gamification, H5P, xAPI

1. INTRODUCTION

Recent numbers show that, in Norway, two out of three young pupils drop out of higher secondary school (Lillejord et al. 2015). The costs of a single dropout are estimated to be around EUR 90,000 (Falch et al. 2009), accumulating to huge sums for society. The goal of the PLA Project – Personalized Learning Arena (Norsk Regnesentral 2018) – was to address this problem and to investigate how the high number of dropouts can be reduced by developing digital educational tools targeting vulnerable groups, i.e., pupils with reading difficulties, including dyslexia, and those lacking motivation for attending school.

The project was organized between 2015 and 2018 according to the user-driven research-based innovation program of the Research Council Norway. As such, the participating SME Conexus, one of the largest EdTech companies in Norway, played a central role as technology provider, and with responsibility for implementation and integration. The research institute Norwegian Computing Center, along with the interest organization Dyslexia Norway and Gjesdal municipality, were in control of research-related aspects, methodology, and content strategy, as well as user trials and evaluation.

The main case in the PLA Project was the "learn to read" framework called SOL (translated to Systematic Observation of Reading), which is specific to Norway and currently implemented in a number of schools with a total reach of approximately 130,000 pupils in about 30% of all Norwegian municipalities (Gjesdal kommune 2011). SOL consists of 10 steps which cover a pupil's anticipated reading progress, from pre-alphabetical reading over phonological and orthographical reading to, finally, an adult's literary reading ability. SOL is quite an extensive framework, and thus we decided to focus on two topics on Level 6 as explained below. The basic idea is to provide tools and tasks that the pupils find challenging, to be used for additional training and exercise on reading topics.

This work contributes to the field of exploratory learning technologies in multiple ways: Firstly, it introduces the research project to a wider audience and explains the approach chosen, including the technologies mYouTime, H5P, and xAPI. The content strategy and production of content is briefly discussed afterwards. Then, the assessment of the solution is explained in detail, with experiment description, results, and discussion, before the conclusion is drawn at the end.

2.   CHOICE OF TECHNOLOGY

Our technology provider Conexus offered the application mYouTime combined with H5P (explained below) as the main platform for content management and distribution (Conexus as 2018b), both linked to their Engage platform for tracking and analysis (Conexus as 2018a).

mYouTime and Engage were considerably improved throughout the project as a result of extended requirements and preliminary testing. In particular, mYouTime had to be integrated in Engage for performance monitoring.

2.1 mYouTime & Engage

mYouTime is quite a generic application and hence suitable for a wide range of domains, but originally it targets micro-learning and just-in-time content delivery. The application exists as a web application and as apps for iOS and Android operating systems, i.e., for smartphones and tablet computers. This aspect alone contributes to an increased motivation with the target group due to, at the time of writing, the fascination these devices and apps have on children at that age, their "coolness" and "x-factor".

A user with content generation privileges can author plain texts and rich media, including hyperlinks and H5P, or record video, audio, stills, and combinations thereof. The majority of users, though, will only "consume" the content made by others. The app is based on learning units, or lectures, which in turn compound of slides of the aforementioned content types. Newly sent lectures arrive first in a user's inbox and trigger a smartphone notification, before they can be addressed and finally archived.

A lecture may theoretically consist of an arbitrary number of slides (with an upper limit of 20). The time needed to finish a single lecture depends, in addition to the slide count, on factors like the duration of contained timed media, such as video and audio, the amount of text, the content's difficulty, and similar, and is of course influenced further by human factors such as reading pace, ability for rational thinking, and others. It is hence neither feasible nor desirable to quantify the optimal number of slides. However, keeping lectures short and the slide count low is of uttermost importance in order to exploit all advantages of the "micro" in micro-learning and for the pupils to stay motivated.

mYouTime is linked to another application in Conexus portfolio, Engage (previously Vokal). Engage is a tool for teachers. Its purpose is, among others, to track and show the performance and progress of pupils, and to enable comparisons across individuals and groups, such as classes, teachers, and schools. Engage is only available as a web application.

2.2 H5P & xAPI

H5P is a free and open-source JavaScript-based framework for implementing interactive content for the Web (H5P Consortium 2018). The name stands for HTML 5 Package, referring to the combination of HTML, CSS, and JavaScript in a single container for deployment in a suitable content management system. H5P supports different types of interactivity and content, such as questionnaires, quizzes, interactive videos, audio recordings, and more. At the time of writing, 39 different types are known. We employed a subset of these in this work. It should also be mentioned that the available content types offer plenty of gamification elements to boost motivation: interactivity (dragging, choosing, buttons, etc.), user control (previous and next buttons, play/stop, etc.), feedback (instant flagging of wrong answers, display of achieved vs. maximum point score, progress bar, etc.), continuous playing ("try again" button, etc.), loss avoidance (motivating messages even for low point scores), exploration possibilities (possible to terminate a lecture at any time without disadvantage), competition (for tasks with time constraints), and other. The effect of all these elements cannot be underestimated, as seen in the results.

In the H5P type "Mark the words", the pupils had to mark particular words in a text, in our case those that included a particular sound when read aloud. In "Memory Game", six (our choice) pairs of identical words had to be found. "Drag the words' was an exercise where 4-9 (our choice) given words had to be dragged to appropriate placeholders in the text. "Multiple-choice quiz" consisted of a series of questions, where the pupils partly had to mark the correct word among all available options, and partly where the correct answer (a word) had to be dragged upon an illustrating image. Finally, "Interactive video" was used like an audio book, where a voice read a piece of text out loud. In between the playback was paused several times, and upon each pause a multiple-choice quiz with one to three questions was shown.
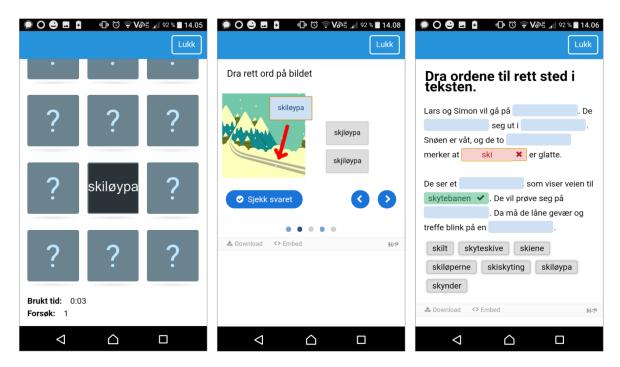
Figure 1. Illustration of H5P content types " memory game " (left), "quiz with drag the image" (middle), and "drag the words" with immediate feedback (right)

To sum up, an H5P content type is capable of supporting a single or multiple tasks and, in the prototype, the majority of lecture slides consisted of H5P content types. A lecture could thus be said to consist of a series of H5P types.

H5P generates xAPI data (ADL 2018), which are basically variable-value messages in JSON format. xAPI messages carry information about pupil activity, such as task completion, duration needed, points achieved, and similar. These messages are usually stored in databases dubbed Learning Record Store (LRS) for tracking and analysis purposes, a strategy we have followed in this work as well.

3. EVALUATION

The app, mYouTime and H5P combined with backend Engage, was evaluated at a school in Gjesdal municipality in Western Norway during 10 days in March 2018. It turned out that it was difficult to recruit pupils from the target group for this research project, but eventually we had six pupils with reading challenges, all of them in 6th and 7th grade. They had various complex cognitive and behavioral challenges. Four of them were male, two female. The evaluation was approved in advance by the Norwegian Center for Research Data. In addition, we had asked the pupils' parents for their consent.

The entire evaluation was carried out in a browser in Google Chromebook, i.e., utilizing mYouTime's web interface. The pupil had to go to mYouTime's website, which basically starts the app, and then to login with the national login solution Feide, which requires entering of the social security number and a password. After the pupils had received a demonstration of the app, they were given a schedule with five to seven lectures for every day of the evaluation, corresponding to an estimated entire duration of 15-20 minutes. Most of the tasks were new, but some were also repetitions. Tasks could also be repeated if all other (scheduled) tasks were solved before the stipulated duration. Additionally, the pupils were encouraged to try the app at home, too.

All data reported from mYouTime and H5P were gathered and analyzed in the aforementioned Engage tool. As an additional measure, the data were stored in a suitable LRS.

3.1 Content Production

In total 26 lectures were crafted for the evaluation, each consisting of two to five tasks, or three to five slides. A task was implemented by means of H5P as mentioned above; see the screen dumps in Figure 1 for how some of the H5P content looked like. The lectures covered two topics on Level Six of the aforementioned reading learning program SOL concerning automating word recognition. The choice of topics, selection of words, and tasks was carried out by the reading experts who also have central responsibility for the maintenance and development of SOL.

Topic 1 was about the spelling and pronunciation of the [ʃ] sound, which can be written as sj, skj, and sk in Norwegian. The topic was presented as 10 lectures with similar structure but different texts and words. Each lecture was based on a short simple text with roughly 50 words on the average, four to nine of which included the sound of concern. Text and words were repeated throughout up to five different tasks, each implemented as an H5P type. The following H5P types were utilized for Topic 1: "Mark the words", "Memory Game", "Drag the words", "Interactive Video", and "Multiple-choice quiz".

Topic 2 covered the spelling of the 400 most popular Norwegian words and was implemented as 16 different lectures with similar structure but varying words. One lecture was compound of three slides. On the first slide, the pupils had to read and memorize 25 words. Both next slides showed the same set of words but with (five) hidden misspellings, respectively, where the task was to find them.

The basic idea behind all these tasks is repetition of a limited set of words over a given time span to train the pupils' automatic recognition of words, and to vary the way this is done to maintain engagement and avoid boredom.

4. RESULTS & DISCUSSION

During the evaluation, it became clear quickly that Engage was not working as expected due to technical difficulties with regard to the integration of mYouTime and H5P data in Engage. It was, however, possible to extract the data from the LRS and analyze them manually.

In total 5216 xAPI messages were generated and stored in the LRS during the trial. Most H5P content came with the correct descriptor, other were just given a very generic "other" value, which basically could mean anything. Not surprisingly, this complicated the interpretation of LRS data significantly. The basic problem here was the lack of a unique identifier that allowed to associate an LRS entry with a particular piece of content. This identifier lack also made it impossible to reliably associate LRS data with a particular task in case of multiple identical content types in the same lecture, say two "mark the words".
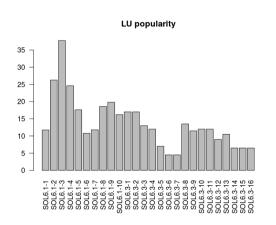
Another problem was that the "interactive video" type contained other H5P types; in this case a "multiple choice" type. However, as detailed above, also native "multiple choice" type content had been crafted, and sadly it was hence not possible to differentiate both in the LRS. In addition, the multiple-choice content type erroneously did not track a pupil's duration correctly, which rendered this content type useless.

Concerning xAPI values, only the statement types "answered" and "completed" were found in the LRS, meaning that a pupil's progress could only be tracked for completed tasks, as there were no "has started"-type messages. Finally, the xAPI's timestamp field could not be used to pinpoint events exactly in time as it turned out that the messages had been buffered and sent in bundles to the LRS in order to save network capacity. As a result, for instance, bundled "answered" messages had timestamps with only millisecond differences, which of course could not have been generated in real life. The order of generation, though, was maintained by the system, which allowed to track a pupil's progress over time.

The most important fields of xAPI messages stored in the LRS were those carrying values for duration, score, and score range. Together with a field for the topic and a field for the pupil/student/user, this could be exploited to derive detailed statistics as explained in the following.

Simply counting the LRS entries with a particular topic, optionally normalized by the number of tasks and number of pupils, gives the popularity of lectures, see Figure 2, left side. Such information can be used to optimize task giving and to rule out particular unpopular tasks. Related to this is the comparison of task fidelity for a particular topic to the task fidelity for all similar topics, i.e., topics of the same nature but with different content/values, computed over all pupils. This is shown on the right side of Figure 2. The plot shows the task fidelity in terms of duration and score for the task "SOL6.3-1" ("mark misspelled words") compared

to the entire series "SOL6.3" with tasks of the same type but different values/words. The scatter plot also shows the smoothed trend (regression line) and confidence interval (gray). It can be seen that the durations achieved in "SOL6.3-1" lay much more scattered than those for the series. This corresponds with the lectures' statistical values as shown in Table 1, where the average duration accomplished in "SOL6.3-1" is roughly 76% higher than that of "SOL6.3". At the same time, the average relative score of "SOL6.3-1" is 22% lower than that of "SOL6.3". It can be concluded that with this lecture, the pupils both needed more time to succeed, and they achieved lower rates, which places this particular task among the more difficult tasks in the entire series. That is, the amount and possibly choice of words should be reconsidered.
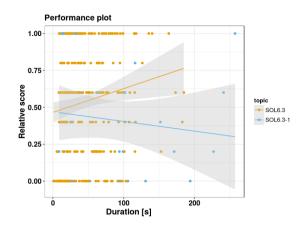

Figure 2. Lecture fidelity in terms of task finalization count (left), and score-duration comparison including regression curve and confidence interval (right)

Table 1. Comparison of score and duration of all pupils for the specific task SOL6.3-1 against all tasks SOL6.3 of the same type (but different values)

Lecture	Score	Duration [s]
SOL6.3-1	0,42	79
SOL6.3	0,54	45

The LRS data also give answers to aspects regarding specific task characteristics. For instance, we included numerous variants of the H5P task "Drag the words", where a set of words had to be put into appropriate placeholders in the text. It was possible to configure the task to give instant feedback in terms of a green background and a check mark as soon as a word had been dropped over a placeholder, but we were not sure if this was the right decision. Maybe this meant too much help and too little independent work? The data's verdict, however, is clear: With instant feedback enabled, the maximum score was unachieved in 12% of all trials, happening to 5 out of 6 pupils. The numbers were computed by filtering LRS data for the case that score and maximum score were unequal, and then simply counting the occurrences and names of pupils. The fact that 80% of all pupils did not achieve maximum score despite instant help advocates for that this feature does not help too much.

The main objective with xAPI messages, though, is to track and evaluate individuals. We propose to measure a pupil's performance as the product of score and duration. Given a particular topic and task, the performance of one specific pupil can be compared to peers, the entire class, or any other group of interest. As an example, "Girl2" is compared to the other participants in the evaluation for the task "Mark the words" for the topic "SOL6.3-2". The score here is normalized to the maximum score possible. Each point in the scatter plot in Figure 3 represents a pupil's termination of a task. As can be seen, the majority of trials by "Girl2" are in the lower part of the plot (and below the smoothing line), meaning her scores are below average, and all her registered durations are moderate to high (none are low). This corresponds to the mean scores as given in Table 2, but when it comes to the mean duration, it turns out that "Girl2" uses approximately as much time as the others.
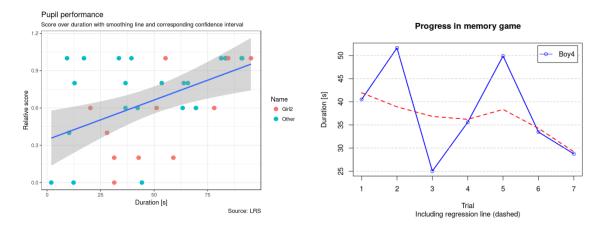
Figure 3. Pupil performance in terms of score-duration points including regression curve and (gray) confidence interval for the lecture "SOL6.3-2" (left), and with progression in duration over time for the task "Memory game" (right)

So in total, "Girl2" has likely more severe reading challenges than the others, but the trend can be said to be weak only. It is apparent, though, that it does not suffice simply to calculate and compare the average of some values for accurate tracking, as the scatter plot clearly illustrates.

Table 2. Comparison of average score and average duration of a particular groups for the lecture "SOL6.3-2"

Group	Score	Duration [s]
Girl2	0,53	52
Other	0,73	49

The memory game is quite useful for tracking a pupil's progress in duration, as all reported scores were equal to one, meaning all cards were correctly paired in the long run. I.e., duration was the only parameter that changed over several trials. This is depicted on the right hand in Figure 3 for the task "Memory game" in the lecture "SOL6.1-10" for "Boy4". As seen, the boy had seven trials, and his performance was quite flunky. However, when we smoothen the points by means of a locally-weighted polynomial regression, we can witness a 25% performance improvement, i.e., more than 10 s reduction in duration over this few number of trials. This shows not only the appropriateness of the task, but also the boy's ability (and willingness) to learn.

To conclude, the data generated by H5P as stored in the LRS can nicely be used to derive statistics in order to improve tasks by looking for high durations and low scores, but tracking progress is limited to completed tasks only. With a sufficient number of data points reported by H5P, it is possible derive meaningful trends in pupil performance as well. It is stressed that, even though the number of trial participants was low, the volume and variety of tasks elegantly allowed to arrive at meaningful results, rendering the solution and its inherent technology fit for the purpose as anticipated.

4.1 Supplemental Interviews with Pupils and Teachers

So far, we have shown how task quality and pupil performance could be sufficiently achieved and controlled. However, the main objective for the solution was to improve the pupils' motivation and eventually reduce the number of dropouts. This is a very ambitious goal and was not possible to measure given the scale of the trial and the limited time frame. Still, we were able to derive some trends by combining the above statistical analysis with an additional source of data in a qualitative approach. More precisely, pupils and teachers were interviewed after the trial about their experience, challenges, and perceived learning benefits. The interviews confirm and elaborate on the findings so far as follows.

The pupils clearly found the H5P tasks both motivating and useful for learning. Some of them were in the risk zone for dropping out of school, but experienced the trial as positive and engaging due to the strong gamification elements. This neatly confirms other research, for instance (Peirce 2013; Nolan & McBride 2014), which states that educational games can be highly useful in a learning context. In particular relevant

and motivating feedback is crucial. E.g., the memory game provides the score and congratulates when answers are correct. Also the time used on specific tasks is made visible for the pupils once the task is finalized – a motivating factor in itself. It becomes important for the pupils to solve these tasks as quickly as possible in competitive time comparisons. A corollary from this is that even pupils with motivation challenges can keep their attention over a longer period of time because they become motivated to solve more tasks to show both the teacher and for themselves that they can do this both fast and in a correct manner. The statistics showing a 25% improvement over time can be explained by the factors mentioned here.

What the above numbers do not show: It was further crucial that the teachers were able to provide personalized assistance during the trial. Although Engage suffered from various technical difficulties, teachers could follow the pupils' progress by looking at the log, and thus the teachers had a sufficient overview of the pupils' progress at any time. It is clear that thereby also teachers benefit from the developed solution. Notable events could be caught and followed up quickly, for example when a pupil for various reasons did not do a particular task she was told to do. Such technological enablers are particularly important for the given target group with pupils often struggling with lack of motivation and sense of failure at school.

The trial also provided valuable results for the design of digital learning materials in the optimal way. Pupils with reading difficulties and dyslexia will usually try to avoid reading longer texts and sections, if they can avoid it. Some of the tasks we crafted were more verbose than others, and several pupils actually skipped reading crucial parts of it, e.g., where to draw the words. For "draw the words" tasks, we expected the main text (including dropzones) to be read first in order to be able to quickly understand where words should be placed correctly. To our surprise, however, most pupils used a more time-consuming "trial and error" strategy where they would try random words and see if they fit. This will usually take much longer time. Nevertheless, this should *not* lead to the conclusion that texts of a certain length should always be avoided; pupils need to practice this, too, but the trial shows that this challenge should be addressed when designing interactive tasks for reading training.

Concluding, the trial provides several leads for the appropriate design of learning tasks for this target group.

5. CONCLUSION

This paper explores the possibilities that lay in gamification and micro-learning, in this solution instantiated by novel technologies like H5P and xAPI, to increase the motivation and learning performance of pupils with cognitive and behavioral challenges. We describe the technologies involved and the setup of the field trials in detail, including content production, and the results and the opportunities and limitations given by this solution are discussed afterwards.

H5P is very powerful and promising, but it also is quite novel still, and as such some of its aspects are simply not mature enough for a production-level system. This applies in particular to partly insufficient and partly inconsistent messaging, as well as to some missing or erroneous values. Another drawback is that learning units are currently not tracked in real time, so reports on progress are inherently somewhat delayed. Apart from that, H5P-generated xAPI messages can with advantage be stored in an LRS and statistically analyzed. We have succeeded with deriving detailed statistics to assess task fidelity, and to identify particularly difficult or unpopular tasks. The paper also describes how a pupil's performance in terms of score and duration can be compared with other individuals and groups.

Interviews with both pupils and teachers have confirmed that the gamification elements provided by H5P have a positive effect on the learning performance of pupils with cognitive and behavioral challenges, and they are capable of increasing the pupil's motivation considerably. As a consequence, the high number of dropouts of school could be reduced, even though a tight project schedule and budget have not allowed to validate this claim. The trials have also shown that close monitoring of the pupils' learning progression is vital, and that teachers need to have the right analysis tools to be able to quickly follow up deviations from the learning schedule.

All in all, with a little improvement H5P is well suited for engaging learning experiences for the given target group as a minimum, and potentially also for other pupils.

ACKNOWLEDGEMENT

REFERENCES

Conexus as, 2018a. Conexus - For a smarter learning planet. Available at: http://www.conexus.net/ [Accessed May 25, 2018].

Conexus as, 2018b. mYouTime. Available at: http://myoutime.com/ [Accessed May 25, 2018].

Falch, T., Johannesen, A.B. & Strøm, B., 2009. Kostnader av frafall i videregående opplæring. *Senter for Økonomisk Forskning*.

Gjesdal kommune, 2011. SOL - Systematisk Observasjon av Lesing. Available at: http://www.sol-lesing.no/ [Accessed November 16, 2015].

H5P Consortium, 2018. H5P – Create and Share Rich HTML5 Content and Applications.

Lillejord, S. et al., 2015. Frafall i videregående opplæring: En systematisk kunnskapsoversikt. *Oslo, Norway: Kunnskapssenter for utdanning*.

Nolan, J. & McBride, M., 2014. Beyond gamification: reconceptualizing game-based learning in early childhood environments. *Information, Communication & Society*, 17(5), pp.594–608.

Norsk Regnesentral, 2018. PLA ― Personlig (mobil) læringsarena | Norsk Regnesentral. Available at: https://www.nr.no/nb/projects/pla-―-personlig-mobil-læringsarena [Accessed May 25, 2018].

Peirce, N., 2013. Digital game-based learning for early childhood. *A state of the art report. Dublin, Ireland: Learnovate Centre*.

The Advanced Distributed Learning Initiative (ADL). n.d. "Experience xAPI." Accessed July 26, 2018. https://www.adlnet.gov/research/performance-tracking-analysis/experience-api/.