論文

University EFL Classes in Japan Should Make More Use of Digital Learning

Christofer Bullsmith
(Atomi University)

Abstract: This paper investigates (1) what reasons EFL teachers at Japanese universities have for not making use of digital learning, and (2) whether those reasons are good ones. Five main reasons are identified and discussed one by one. Three small-scale trials (involving a total of 137 students, including 48 in two control groups) were used to examine the claims that students will not use digital learning and that digital learning is not effective. Results indicate that students will use digital learning if it is connected to their goals, and tend to spend more time on digital learning homework than on non-digital homework; and that digital learning is far more effective at raising TOEIC scores than traditional study methods. This licenses the conclusion that EFL teachers at Japanese universities do not have good reasons for not using digital learning, and should consider making use of it.

1.0 Introduction

In 2017, after 15 years working in the digital learning industry, I returned to fulltime classroom teaching by accepting a position at a university in Japan. One day I had workmates confident that digital learning was the best solution for most language-learning needs, and the next I had workmates generally disinterested in digital learning. I did a small informal survey of teachers at several universities in Japan to discover their attitudes and situation, discovering that none used digital learning in class or for homework, despite most of the textbooks they used in class having associated online contents (mostly accessed using scratch-to-reveal access keys), and despite most of them being aware that their university offered students access to some kind of digital learning contents. This motivated me to investigate (1) what reasons teachers have for not making use of digital learning, and (2) whether those reasons are good ones.

2.0 Reasons teachers have for not making use of digital learning

A paper questionnaire was individually administered to 40 fulltime English teachers at five different universities in Japan. The questionnaire was bilingual (English, then Japanese) with respondents asked to answer in whichever language they preferred, and was followed by a semi-structured interview. The questionnaire covered how much input they had on syllabi and materials, a number of background questions about their own computer and mobile device use and how they would go about learning a language, and then several questions about digital learning

(glossed in the questionnaire as ‘use of e-learning or online study materials; using a computer/smartphone/tablet to study, practice, or test’).

Of the 40 respondents, 23 were Japanese, and 17 non-Japanese. 19 had tenure; 11 were part-time lecturers with contracts at two or more universities; and 10 were private language school teachers. The key result is the answer (unprompted, free answer format with four blank lines provided for the answer, thereby allowing multiple answers, preceded by a question establishing whether they use digital learning) to ‘Why don’t you have your students use digital learning?’. Answers were processed to group similar items together (for example, ‘it’s not my decision’ answers included ‘I don’t set the syllabus’ and a Japanese version of ‘The text is set for us’, and so on), so the exact phrasing and number should not be considered important. Still, five clear themes were apparent in the answers, and each is given as a heading below and followed by discussion.

2.1 Reason 1: It’s not my decision

The most common response overall was some version of ‘It’s not my decision’. The 20 teachers giving this response were untenured, and all noted only having ‘no’ influence on syllabus or materials. In the interview, these teachers were otherwise mostly positive about digital learning, suggesting that it would be good to use it if it the price were right, or if it were easy to use as a teacher, or provided students actually used it (themes repeated in the responses discussed below).

Although investigating why untenured teachers were more positive about digital learning goes beyond the scope of this article, it may be relevant that untenured teachers were younger than tenured teachers (average age 44 versus 55), more confident using a computer (average scores across six Likert items, where 1 was not at all confident and 5 was very confident, 4.1 versus 2.2), more likely to have a smartphone (81% vs 37%), and if they had a smartphone spent longer on it per day (178 versus 107 minutes).

The remaining responses are therefore all from the 19 tenured teachers, all of whom noted ‘some’ or ‘complete’ influence on course syllabus and materials.

2.2 Reason 2: Students won’t use digital learning

The next-most common response was that students will not use digital learning. This is because, according to interviews: students are not motivated, or (presumably equivalently) will not do anything without a teacher standing over them; students don’t have time; some students don’t have a computer at home; or students simply don’t like e-learning.

To investigate these responses, a questionnaire was administered to 422 students (a mix of genders, years of study, and majors) at three universities. One lacked a smartphone, but she had a tablet; 14 did not have a computer at home, but all had used computers on campus both inside and outside class time; 31 noted they had a limited data plan for their smartphone and would be reluctant or unable to use it for study. Students spent an average of 487 minutes (just over eight hours) per day on their phones, though this figure is probably over-inflated by the iOS ‘Screen

Time' measure, which counts all time when the screen is on. Students were on average slightly unenthusiastic about trying 'e-learning' and 'CALL' (both 2.6 on a Likert scale, where 1 was not at all enthusiastic and 5 was very enthusiastic), commenting that it sounded old-fashioned. However, they were positive about 'digital learning' (4.0) and 'online learning' (4.1), and very positive about 'learning apps' (4.8), commenting that apps sounded fast, efficient, and fun. Overall, it seems clear that students *do* have time for digital learning, have not only computer access but also mobile device internet access, and are in fact favorably inclined towards trying digital learning provided it is presented as an 'app'. In the end, however, self-reported attitudes aside, the best test of whether students will use digital learning without a teacher standing over them is to give them a digital learning course and see (see section 3.0).

2.3 Reason 3: Digital learning doesn't give students certain important competences

The final three responses were all equally common. One was that digital learning does not give students certain important competences: according to follow-up interviews, these include presenting, complaining, turn-taking, making eye contact, and negotiating for meaning.

A useful concept here is that of *affordances*, or the possibilities of action furnished by an object or, in this case, the environment of learning (cars and motorbikes have different affordances: keeping you dry in the rain versus permitting access to narrow streets, for example). Traditional in-person classroom affordances include in-person social communicative modelling and practice. The concern is that digital learning, at least in the familiar form of self-study using text, audio, and perhaps some video, lacks these affordances. This seems true. Videos and audio allow modelling of social aspects, but realistic interactive practice is difficult given that the learner is alone with a smartphone or computer. For the moment, a virtual interlocutor using state-of-the-art facial and voice recognition would struggle to interact realistically.

However, there are plenty of other important skills that digital learning *can* give students. Digital learning has its own affordances, allowing students to study whenever and wherever they like, at a suitable pace and level, without the waiting common in classrooms, while receiving instant personalized feedback, detailed and computationally-demanding individual tracking (for example, taking care of the mechanics of spaced repetition), and so on. These affordances make digital learning useful for learning grammar, vocabulary, reading, and listening.

To put these in perspective, Nation argues that a language course should include four roughly equal strands: meaning-focused input, meaning-focused output, language-focused learning, and fluency development (see for example Nation & Newton 2009: 1-2). The suggestion is then that traditional classrooms *alone* can be good places to learn some interactive social aspects of the first two strands, at least for listening and speaking. However, digital learning at least affords (perhaps *better* affords) the latter two strands, and seems apt to support less social aspects of the former two. Therefore, the digital learning proponent may accept that digital learning is not suitable for practicing some social competences, while arguing that digital learning can provide most key

language competences. In practical terms, learners can work digitally on their grammar, vocabulary, reading, writing, general listening and speaking, and fluency; but need to deal with people to master some ineluctably social competences.

As an aside, such social competences may be less important to modern students than teachers from a previous generation might expect. 422 students asked about English study goals mentioned Instagram comments more often than conversation, voice assistants (Siri, Alexa etc.) more than the telephone, internet search skills more than speaking ability, and chatbots more often than emails.

2.4 Reason 4: I don't know how it works

Answers within this theme included not knowing how to use the contents, how to how to check on student progress, or how to demonstrate it to students or provide technical support.

While interesting as a statement of fact, and something that digital learning providers should presumably be more aware of (for example by better helping teachers learn, or by making it clearer that teachers don't have to know because students can be supported directly), this is not in itself a good reason for not making use of digital learning. That is, if digital learning would benefit students and there are no other compelling reasons for not using it, teachers who don't know how it works should learn or delegate.

2.5 Reason 5: Digital learning doesn't work

The final (third equal) reason given by teachers was that digital learning doesn't work, or isn't effective, or is not as effective as traditional alternatives (study with textbook and CD; study in a classroom).

Indeed, compelling evidence regarding the efficacy of digital learning seems rare. Golonka *et al* (2014) surveyed 350 studies of technology use in foreign language learning and conclude that 'evidence of efficacy is limited'. Singer & Ivory (2017) suggest in the context of classroom computer use that there is 'little rigorous evidence so far to indicate that using computers in class improves educational results', and that with large sums of money involved schools may be making purchase decisions based not on evidence but rather on freebies from educational technology companies, in a system reminiscent of doctors falling for pharma freebies.

This skepticism is understandable: educators have seen expensive specialist computer-based integrated learning systems in the 1980s, and various incarnations of analogue language labs through to the 1990s (all expensive to install and maintain and quickly outdated), and then in the early 2000s expensive Learning Management Systems intended to manage libraries of courses but sold without the actual courses.[1] Digital learning might look like simply the latest fad –

[1] In the author's experience, during this period, it was common for large corporate and university digital learning clients to be desperately hunting for course content to put on their expensive LMS and therefore make some sense of the cost before it became obsolete – even if porting an internet-based

the latest attempt by commercial interests (including school administrations) to apply production-line economic principles to education, in direct tension with teachers' engagement with individuals, and thereby devaluing teachers. Still, a healthy skepticism regarding new technology must still allow that some may be worthwhile, provided the burden of proof is properly shouldered. It should also help that modern technical infrastructure tends to reduce set-up costs. No custom lab is required when everyone has a smartphone and internet access, and no dedicated server on campus is required when apps and data are maintained in the cloud; as a result, modern digital learning courses can compete with traditional textbooks on price.

Another reason for skepticism regarding digital learning comes from philosopher Hubert Dreyfus, co-originator of the hugely influential five-level (novice, competent, proficiency, expert, mastery) model of skill acquisition (Dreyfus & Dreyfus 1980). Dreyfus argues that online learning, in the absence of in-person interactions with classmates and teachers (apt to trigger our innate emotional and social responses), can bring people to 'competent' at best, at which level they require some ongoing level of attention/cognitive load to deliberately decide on a plan/schema to respond to a situation, rather than being able to respond automatically (Dreyfus *et al* 1986).

In a response similar to that in section 2.3 above, the digital learning proponent may simply provisionally grant this, while continuing to hold that digital learning is worthwhile. That is, even if Dreyfus is right and social interaction is ultimately required for expertise, the possibility of becoming 'competent' in English via digital learning is enough to make it valuable. If 'competent' English corresponds to, say, CEFR B2 (upper intermediate) level, and digital learning offers some shortcut or alternative for what would usually require 500-600 hours of guided learning in a classroom, then it looks worthwhile.

Confusingly, however, Dreyfus' own examples of expertise include an expert traffic controller controlling planes using screens and radio, and an expert video gamer dodging virtual missiles. Presumably if limited voice interaction with visual prompts and 1980s computer games can trigger our innate emotional and social responses in such a way as to allow us to reach expert level in a skill, then modern digital learning can do likewise. Indeed, plain text and non-interactive audio can seemingly trigger strong emotional responses (literature and radio plays, for example), so Dreyfus' objection may come down to little more than the unsurprising historical fact that Dreyfus didn't find 1980s digital learning nearly as engaging as in-person teaching.

However, as late as 2008, Dreyfus was still highly skeptical, suggesting that '[i]t is now clear that distance learning has failed. The major universities have given up on it and consider their investments of hundreds of thousands of dollars as sunk costs' (2008: xi). If this is an appeal to the authority of major universities, it seems misguided: in 2002, 34.5% of US higher education providers offered complete online degrees, and 50% reported online education was 'critical to their long-term strategy'. This increased year on year to 62.4% and 66% respectively in 2013 (US

---

course that could be accessed from anywhere and had social and reporting functions onto a dated in-house LMS meant all those features were lost.

College Board 'Annual Survey of Colleges' figures, Allen *et al* 2014). On May 31 2019, one website (https://www.classcentral.com/) offered 9,834 online courses at 908 universities; the universities offering the most courses were MIT, Stanford, Michigan, Pennsylvania, and Harvard, while 43 fully online degrees were listed from schools including London University and the University of California. On a more populist note, four language-learning apps (Rosetta Stone, Duolingo, Babbel, and mango) in 2017 jointly claimed around 207 million active users. While universities may indeed have considered some digital learning costs as sunk (see the comments on LMSs above), they clearly don't think that distance learning has failed.

Returning, then, to the question of whether digital learning works, it is again open to digital learning proponents to grant that digital learning may be less effective than traditional alternatives, while arguing that there is little force to the objection because the alternative to digital learning in many cases is not traditional learning but *no* learning. That is, students who are unable or unwilling to study traditionally (for example because travel time to a classroom is prohibitive, or they are too tired in the evening to get out a textbook and CD) can use digital learning in short chunks of underutilized time (on the train, over coffee, and so on).

Granting every objection against digital learning, however, seems hasty. Evidence regarding the efficacy of digital learning may be rare, but can be gathered.

3.0 Research: three trials

Of the five reasons for not making use of digital learning identified, reasons 1, 3, and 4 were dismissed in the discussion above. The remaining two reasons require research to properly evaluate, and the remainder of this paper details three trials designed to achieve that.

Reason 2, 'students won't use digital learning'

Reason 5, 'digital learning doesn't work'.

3.1 Trial one: will students use digital learning? (Optional access.)

In the first week of class, a complete class of 16 students (second-year non-English majors taking a compulsory 15-week general English course with an Oxford University Press textbook and set curriculum) were given a 20-minute introduction to a digital learning course, including a quick demonstration of how to get started, find a lesson on a specific topic, and study the lesson. Students were then given login information and a list of digital lessons whose topics mirrored the set text, and the remainder of the course proceeded as normal with no mention of the digital course (though students who logged in and registered their email address received an automated weekly reminder email).

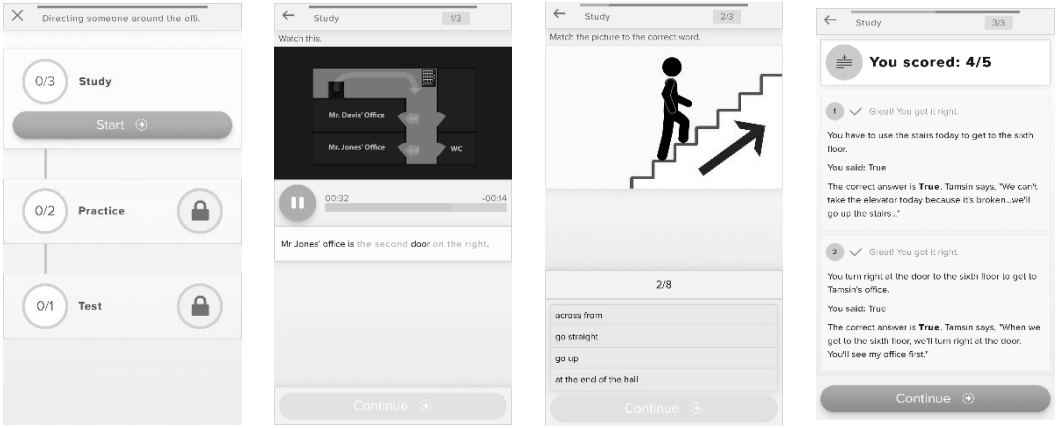
The digital learning course chosen was the general English course 'Practical English 7' from 'Reallyenglish' (Reallyenglish 2019; see figure 1 below). While other courses mentioned by teachers in interviews were considered, other options were ultimately discarded because of a lack of support for small groups (two provided only multi-year institution-wide pricing structures), a

lack of pre-made content, or a lack of an offline mobile study option. In addition, the author had previously worked extensively with Reallyenglish and was familiar with the platform and content, making the introduction to students easy.

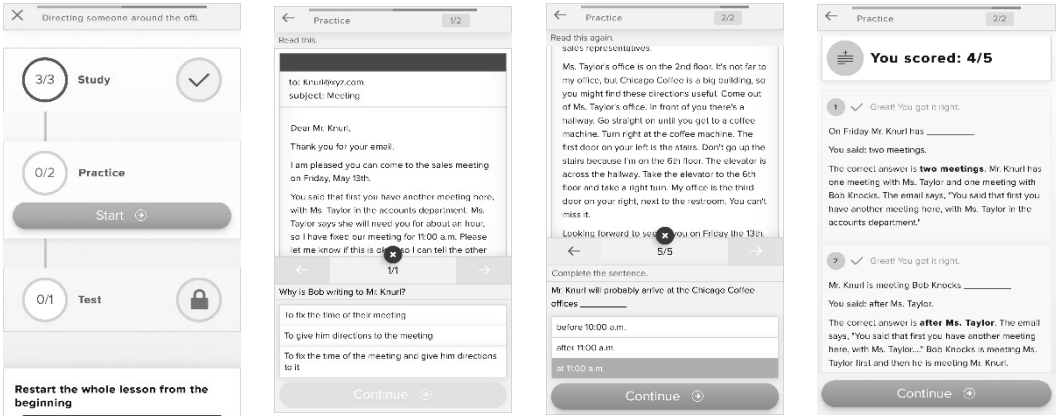
The course consists of 350 lessons split between reading, listening, and grammar, arranged by level and topic. Students can sort or search by keywords, or complete three short diagnostic tests to see a shortlist of recommended lessons which updates as lessons are completed. Each lesson takes around 30 minutes to complete but can be left and returned to at any point. Lessons can be downloaded while on a WiFi connection (for example, at school or a convenience store) and then studied offline. Each lesson starts with study, moves through two sections of practice, and finishes with a short test (see the sample screenshots below). Lessons are intended to reflect best classroom and self-study practice and as such should be familiar to students in terms of goals, content, format and so on. For example, a lesson in the class textbook on 'Describing people' was closely matched by a digital lesson (grammar, 'Describing people and places') and extended by another (listening, 'Describing co-workers').

After 15 weeks (one semester, the duration of the classroom course) the trial results show that students did an average of 5.3 lessons (slightly over two and a half hours of study) each during the 15 week period. However, the average is a misleading statistic in this case. With the standard deviation 15.3, half of the students didn't register, and half registered and completed one or more lesson, with one outlier completed 64 lessons (33 hours). It is probably more informative to give the median (0.5). While the outlier is of some interest (the student in question asked if they could buy continuing access to the digital course, commenting that for the first time they had found a way to study English that allowed them to study at their own pace and didn't trigger social anxiety), the *prima facie* conclusion to be drawn from the trial is that most students indeed won't use digital learning.

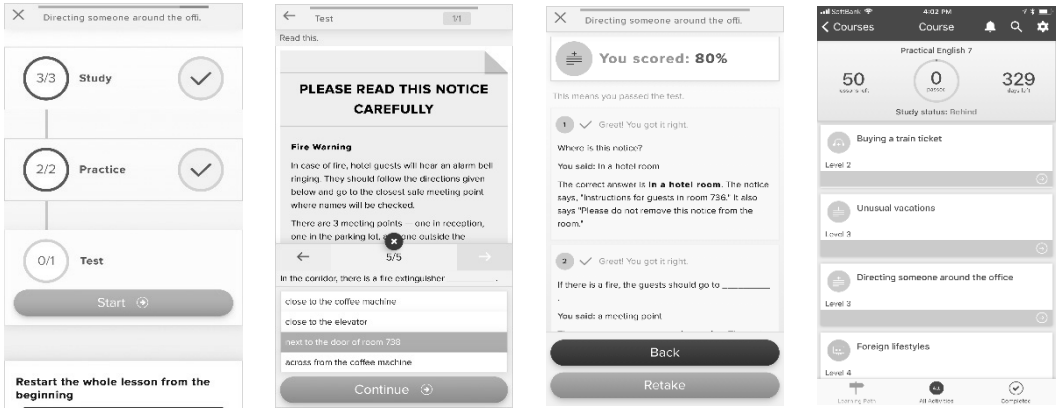
Fig 1. Reallyenglish 'Practical English' lesson (sample overview)



Lessons start with 'Study', an introduction of key language (here, eight lexical items, 'across from', 'go straight' and so on) using audio, text, and animation; then have a short quiz, displaying marks and explanations (shown in Japanese for students).



In 'Practice', students practice with the language in context (here, a 'Reading' lesson, as a short text; other lessons use conversations etc.). Here, they read once for gist (a single question), and then again for detail (several questions), again seeing marks and explanations. This entire procedure is then repeated for a slightly longer text.



Lessons end with 'Test'. Students can repeat any lesson or lesson part as many times as they like; distractors are shuffled each time. The final screen shows a return to the course menu; students can search for a particular topic and level, or take the next recommended lesson.

However, this seems hasty. The 16 students were non-English majors whose main motivation was getting course credit, and the e-learning was not clearly or explicitly linked to their goals (credit and course assessment). Other possible conclusions, equally licensed by the result but with different implications for our investigation of digital learning, are that students won't do work that isn't linked to their goals, irrespective of whether that work is digital or not; or that many students are reluctant to engage with any unfamiliar study style.

3.2 Trial two: will students use digital learning? (Linked to assessment.)

A second trial was run with different students with the goal of having a control 'non-digital' group and a treatment 'digital' group, and of making sure that in both cases homework was closely tied to the students' main goal (course credit). Participants were first-year non-English majors on 15-week courses, with assessment based on in-class tests and class participation.

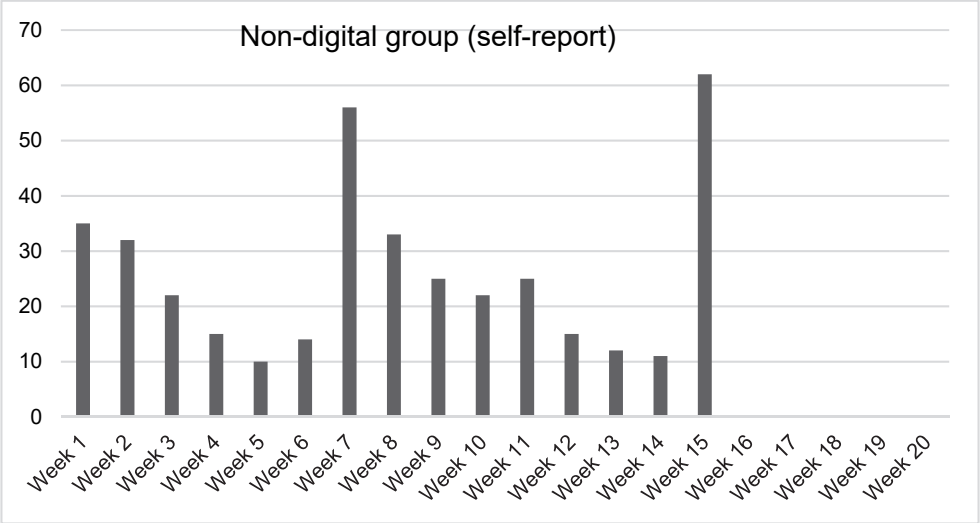
Each week, a 'non-digital' control class of 32 students were given materials useful for review and preparation (mostly additional printed materials from the textbook publisher's website, plus some 'extra practice' sections from the back of the book).

The 'digital learning' class of 27 students had the same textbook and general syllabus but were given access to the same digital course used in trial one. Each week, they were given the name of a lesson useful for review and a lesson useful for preparation.

Students were prompted each week to complete a very short journal (essentially a record of time spent studying). For both groups, each class started with a short quiz which counted towards the final grade and clearly reflected the material pointed out to students as relevant (variously non-digital and digital) plus the previous class.

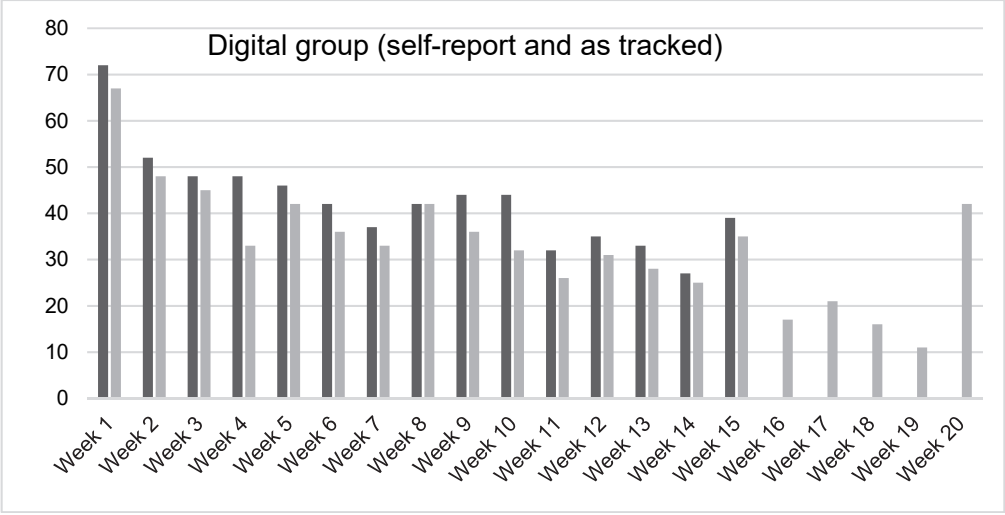
Trial data is summarized in figures 2 to 4 below.

***Fig 2*. Average minutes of homework per week (non-digital group, n=32), as self-reported.**



Students in the non-digital group reported doing an average of 389 minutes of homework, which represents an average of 26 minutes per week over the 15-week course. Substantial spikes coincided with mid-term and end-of-term tests.

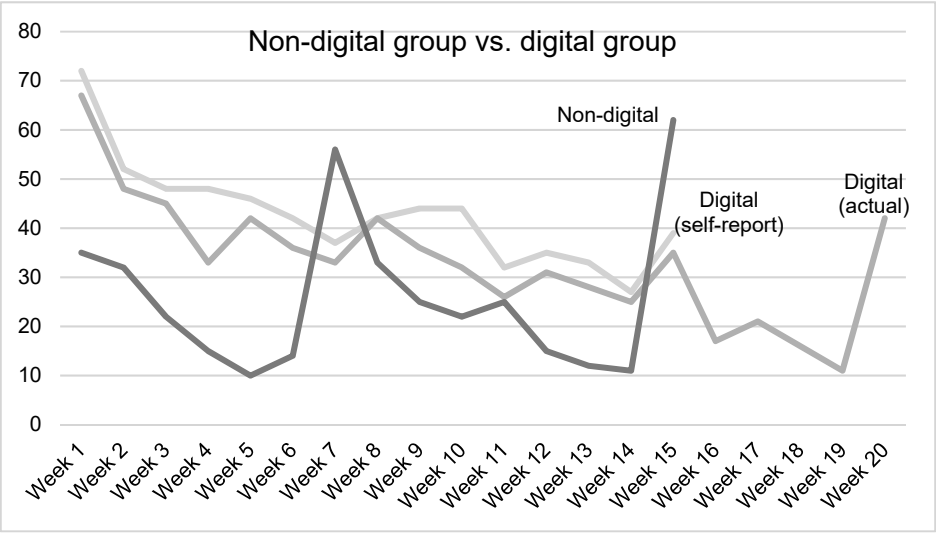
***Fig 3*. Average minutes of homework per week (digital group, n=27), as self-reported, followed by study time as recorded by the digital learning course.**



Students in the digital group reported doing an average of 641 minutes of digital homework during the course, or 43 minutes per week. No spikes appeared at mid-term and end-of-term (the cause of this difference is unknown: unfortunately, the course had finished before it was noticed). They reported doing no homework using the textbook or other materials. Study time as tracked by the digital learning system shows an average total of 559 minutes total during the 15-week course, which seems to suggest that while students are remarkably honest (correlation with reported study time 0.94), they over-report their study time by around 15%. However, the digital system study time tracks only time in lessons – time getting ready to study, or on the menu page finding lessons, or looking at study reports, is not counted. Accordingly, while there may be some over-reporting, it is remarkably low. Students reported studying mostly while on buses/trains and late at night.

The digital learning system also revealed that about half of students continued studying after the end of the classroom course (despite assessment being already finalized), in most cases studying more per week than they had during the course. On average across all students, they studied for 107 minutes in weeks 16 to 20 (when the online course ended).

***Fig 4*. Average minutes of homework per week: non-digital group (n=32, self-reported) versus digital group (n=27, self-reported and actual).**



Comparing digital and non-digital groups, digital learning students spent about 65% more time on homework than non-digital students did during the course (comparing self-reports, 641 minutes versus 389 minutes), and in many cases continued studying after the course finished (after assessment was finalized).

3.3 Trial three: does digital learning work?

A third trial was run on two classes, with a smaller class acting as a control (traditional homework assigned) and the other as a treatment group (digital learning homework assigned). The groups were a similar mix of third- and fourth-year Japanese 'Communication and Culture' majors with a smattering of management and human studies majors; none were majoring in English, and none had any concurrent English courses or reported any significant English learning or practice outside the work required for these courses. The 15-week courses were run using English materials but were not specifically EFL courses: the control class was on 'Animation & Comics', and the treatment group on 'International Youth Issues'.

Students were given a level check/pre-test using an assessment tool available through the Reallyenglish mobile app: a one-hour 100-question reading and listening test, in a format and style designed to mimic the two-hour 200-question TOEIC Listening & Reading test and yield statistically reliable estimates of candidates TOEIC scores. Listening and reading scores are each from 5 to 495, with the total score from 10 to 990.[2]

[2] The TOEIC 'measures the everyday English listening and reading skills needed to work in an international environment' (ETS 2019), and as such might not be an ideal measure of students' readiness for university courses on animation and youth issues. Still, the assessment tool was

Class	Treatment	n	Scheduled class hours	Pre-test average scores		
				Listening	Reading	Total
Animation & Comics	Control (paper-based homework)	16	22.5	224	140	364
Youth Issues	Treatment (digital learning)	46	21	214	138	352

No significant difference was observed between the average pre-test scores for control and treatment groups. The treatment class showed scores with a greater range, but this is unsurprising since it is a much larger group. Total scores of 225 to 550 are considered to correspond to CEFR A2 level ('Elementary'), so the observed scores here fit well with expectations from university entrance and placement tests.

Since students at this level will struggle with unsupported English materials (whether animations and comics or video clips and slides about youth issues), but class time is not supposed to center on language-teaching as such, homework chosen to support the next class was set each week. For the control group, this was paper-based homework (sections from textbooks; reading, listening, and vocabulary; checked briefly in class, counted towards the grade for the course); for the treatment group, this was the same digital learning course used in trials one and two above (one required lesson per week, and another lesson 'if you feel like it').

Class	n	Completers	Class hours (actual, average)	Homework hours (average)
Control	16	13	20	7.2 (self-report)
Treatment	46	46	19.9	11 (self-report) 8.8 (as tracked)

The control group reported doing an average of 29 minutes of homework per week, while the digital learning group reported doing an average of 44 minutes per week. The digital learning system tracked 35 minutes per week, though as per comments in section 3.2, this slightly underreports the time students actually spend, so what looks like a 25% overreport from students is actually far less. These results confirm those from trial two, showing that students will use digital learning, and that they tend to spend more time on digital learning than on paper-based homework (comparing self-reported figures, 53% more).

At the end of each course, students were given an exit test. This was another one-hour 100-question mock TOEIC test, with different content but in the same style and through the same mobile app. The initial intention was that this test would simply mark the end of the course and

convenient and yielded a result which could be easily interpreted by the instructor; further, nearly all of the students knew what TOEIC was and wanted to know their own TOEIC score, and most of the students were heading into a period of job-hunting. Note also that scores in the pre- and post-tests (estimated TOEIC scores) were not used in assessment for the course.

remind students of their estimated TOEIC scores just as many were heading into a period of job hunting, while the trial results might serve as a confirmation of trial two (by seeing if students in this different course would also use digital learning). After all, significant gains in TOEIC score (let alone in a estimation of TOEIC score obtained by a half-length mock test) cannot usually be expected at small timescales.

***Fig 5*. Estimated hours of study needed to progress between TOEIC levels (adapted from Prolingua 2000).**

TOEIC 200	TOEIC 300	TOEIC 400	TOEIC 500	TOEIC 600	TOEIC 700	TOEIC 800
→	100 hrs	250 hrs	450 hrs	700 hrs	1000 hrs	1350 hrs
	→	**150 hrs**	350 hrs	600 hrs	900 hrs	1250 hrs
		→	200 hrs	450 hrs	750 hrs	1100 hrs
			→	250 hrs	550 hrs	900 hrs
				→	300 hrs	650 hrs
					→	350 hrs

According to this chart, at TOEIC 300 level, a 100-point increase requires 150 hours of study, or about 1.5 hours per point. Though introduced as 'rough approximations', and pre-dating some changes to the TOEIC test, the general points that significant improvement in English proficiency overall (as against in some small subdomain such as reading train timetables or ordering drinks) takes a lot of time, and progressively more time at higher levels of proficiency, are well taken. Given this, after 20 hours of class we might expect an increase of about 13 TOEIC points – or less, say 10 points, since the classes were not focused on TOEIC or general workplace English. Add in seven or nine hours of homework for another five or six points, and we can expect a gain of about 15 or 16 points.

However, a difference of this size is too small to be reliably distinguished by TOEIC. While figures on this are surprisingly difficult to find given the scale of the TOEIC industry, an obsolete TOEIC user guide (ETS 2007) gives the Standard Error of Measurement (SEM) as 25 for each of the listening and reading components. This is an estimate of the variability (that is, the standard deviation) expected in observed scores: if a test taker were to take a large number of different versions of the test, we would expect 68% of their scores to cluster within 25 points of the mean, which is their 'true score'. In practical terms for a single test-taker, this means that you can be only moderately confident (probability 0.68) that your test score is within 25 points of your true score: a student with a true score of 200 in listening has a 68% chance of getting a score between 175 and 225, and a 32% chance of scoring outside this range. The same ETS guide gives a Standard Error of Difference (SED, the errors of measurement associated with two test administrations) of

35 for each component.[3] For our immediate purposes, this means that even if your 'true TOEIC score' has improved by 16 points between pre- and post-test, this is under half the expected variability expected in retest scores even when true score is held constant. For large groups, standard errors will tend to average out in such a way that even small improvements in true score can be detected with some confidence; but for individuals, retest scores may go down (probability > 0.16).

Of course, this discussion has been about the SEM and SED of TOEIC, while the pre- and post-test in this trial is a half-size test developed to estimate TOEIC score, and which we can expect has a higher SEM and SED.

Accordingly, students were told briefly about SEM and shown an example of how test scores can go down even when true score has gone up ('even when you have worked hard and achieved well in the course'). (Fortunately, and perhaps because of this problem, the assessment tool only shows students scores as bands, while the raw precise scores are visible only to administrators.)

Class	**Post-test average scores (change from pre-test value; significance level)**		
	Listening	**Reading**	**Total**
Control (n=13)	225 (+1; p=69.999)	142 (+2; p=70.000)	367 (+3; p=43.626)
Treatment (n=46)	238 (+24; p=0.011)	168 (+30; p=0.0043)	406 (+54; p=0.000)

For the control (paper-based homework) group, average total scores rose 3 points. Though this was lower than expected, given the discussion above and the small group, it is not surprising. For the treatment (digital homework) group, average total scores rose 54 points. This was higher than expected, especially since the group was larger and SEM/SED effects should thereby be reduced. To test for significance, p-values (t-Test paired two-tailed test) were calculated for each change. Taking $p<0.05$ as the threshold for 'significant' (a 95+% confidence that the change is not due solely to chance) and $p<0.001$ for 'highly significant', there was *no* significant change in scores for the control group, and there was a *highly* significant change in each score for the treatment group.

This average gain is rather startling. If we take the classroom study as irrelevant for purposes of score gain, given that the control group showed no significant gain (the design intention is that the only relevant difference between the control and treatment group is the digital

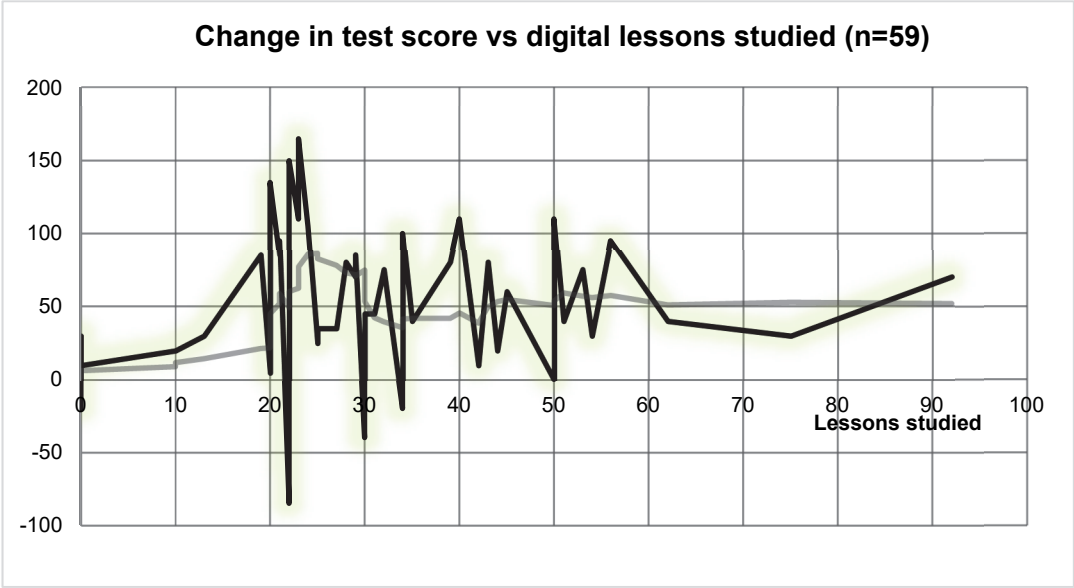
[3] Prolingua (2000) notes that ETS claims a SEM of 25 for the TOEIC and suggests that 'various independent studies have indicated that the TOEIC SEM more accurately lies in the range of 35 to 47'. The 25 figure claimed by ETS is actually for each of the listening and reading components, not for the overall score as Prolingua suggests. Further, note that the more relevant variability measure in the context Prolingua discusses (pre- and post-testing) is the SED (35 for each of the listening and reading components, according to ETS), not the SEM. These points may serve to explain the disparity Prolingua claims between ETS's figures and the independent studies.

treatment), then we have a result of a 54 point gain over 8.8 hours of digital study. This corresponds to a one point gain for every ten minutes of study, a *nine-fold* speed increase over the standard estimation in figure 5.

The natural response is to assume that some effect other than just increase in English proficiency must be partially responsible for what would otherwise be an incredible level of effectiveness. One possibility is that the digital treatment group do better in the post-test in part because they have become familiar with the test format, both in the sense of how it is presented in the digital app interface, and in the sense of being used to the style of TOEIC questions. That is, one encounter 15 weeks previous is not enough to create this effect for the control group; but that encounter plus an average of 33 digital lessons incorporating short TOEIC-style lesson tests over 15 weeks is enough to create some significant effect.

An exhaustive investigation of what portion of the 54 point gain is due to test familiarity effects and what portion is due to gains in English proficiency is unfortunately beyond the scope of this article and of the data available. While the data does show that test score gain is correlated with the number of digital lessons studied, this result does not distinguish proficiency gains from familiarity effects.

Fig 6. **Change in test score (pre- to post-test) versus digital lessons studied. This graph includes both control and treatment groups (the control group all studied 0 digital lessons). The gray line is a trend line (a visualization ignoring some outliers).**



Only a few students studied high numbers of lessons, so the rightmost third of the graph should be viewed with some caution, but the general trend is clear (Pearson Product-Moment Correlation Coefficient 0.30, p=0.020, ‘statistically significant’).

This graph also makes it easier to visually identify some data points which may be considered outliers that should be ignored: in particular, a cluster of eight students who studied around 20 to 25 lessons showed particularly large gains, and checking shows that these were all and only those students who spent significantly longer on the second test (an average of 48 minutes for the nine students, versus a first test average of 26 minutes). With gains from 100 to 165, one credible interpretation is that the high gains are partially due to them simply trying harder in the second test (taking more care in reading, checking answers, and so on). Another student in the same cluster (22 lessons studied) showed the largest drop in score, but apparently tried far harder in test one (46 minutes) than in test two (21 minutes). If all nine students with mismatched test times are ignored, the correlation between lessons studied and change in test score (average 44, down from 54) rises to 0.41. The new average of 44 points of score gain, now for an average of 9.0 (up from 8.8) hours study as tracked by the digital learning system, represents an average gain of 5 score points per hour study, suggesting that digital learning is 7.5 times (down from 9.2) as effective as regular classroom study at raising scores.

Finally, students in trial three were given a questionnaire asking about their experience in the course. Treatment group students were generally more positive about digital learning than control group students were about paper-based homework (average scores across three Likert items, where 1 was very negative and 5 was very positive, 4.3 versus 3.4), and wanted to use digital learning again (1 was definitely no and 5 definitely yes, average score 4.3).

3.4 Conclusions drawn from the three trials

In conclusion, then, trial one showed that most students will not leap into digital learning when it is not clearly linked to their goals. Trials two and three showed that students will use digital learning if it is linked to their goals. In fact, students spent *more* time on digital learning homework than they spent on traditional homework, even continuing to use it after course assessment was finalized (trial two) or averaging over twice the required level of work (trial three). So, teachers' concern about digital learning, that students will not use it, are unfounded.

The results from trial three suggest that EFL digital learning can be around 9 times as effective as traditional study (or, ignoring outliers, 7.5 times as effective). While it seems likely that some part of this impressive gain is due to test-format familiarity effects, some part is presumably due to proficiency gains. And even if the test-format familiarity effects are responsible for the greater part of the gains, this result is still evidence of digital learning 'working' to do something that students are likely to consider useful (namely, raising TOEIC scores). So, teachers' concern about digital learning, that it doesn't work, are unfounded.

4.0 Conclusion

The research questions were (1) what reasons teachers have for not using digital learning, and (2) whether those reasons are good ones. In answer to question one, five reasons were identified (section 2). To answer question two, each reason was discussed individually, and in two cases empirical research was carried out.

'It's not my decision' and 'I don't know how it works'. These two concerns were dismissed as essentially irrelevant to the issue of whether digital learning should be used, though the latter concern is a useful reminder that many people (teachers and students alike) require support in starting to use digital learning.

'Students won't use digital learning'. The concern is a reasonable one insofar as making sure students study is an issue irrespective of study method – that is, as trial one showed, digital learning is not a magic bullet, and study methods that are not clearly linked to students goals will be poorly used. For university non-English-majors who are motivated primarily by course credit, this means linking digital learning to assessment. When digital learning is linked to students goals, students in fact *will* use it, as shown by trials two and three. Indeed, students are seemingly apt to spend more time on it than they are on traditional study, indicating that this reason is not a good one.

'Digital learning doesn't give students certain important competences'. This can be provisionally granted without implying that digital learning should not be used, as such competences are only a small part of overall language competence. The concern is a reasonable one insofar as it serves to remind us that a digital learning course might, depending on the goals, be usefully complemented by a classroom component; or that full expertise or fluency might only be available after real-world practice.

'Digital learning doesn't work, or is less effective than traditional study methods'. However, trial three demonstrated that digital learning *does* work. While the concern is a reasonable one insofar that teachers and curriculum developers need to continuously evaluate the efficacy of materials and methods as their context changes, digital learning seems significantly more effective (nine or 7.5 times as effective by time) than traditional study at improving performance in a standardized English test, and more popular with students.

The overall conclusion is therefore that teachers' reasons for not using digital learning are *not* good ones. While acknowledging that different digital learning courses used in different contexts are likely to yield different results, EFL digital learning can work well and be popular with students, and EFL teachers in Japan ought to consider making use of it. Indeed, if larger scale studies in a wider range of contexts produce similar results to this one, the burden of proof may shift to non-digital learning.

References

Allen, I.E. & Seaman, J. (2014). Grade Change: Tracking Online Education in the United States. Babson Survey Research Group. Retrieved from www.onlinelearningsurvey.com/reports/gradechange.pdf

Dreyfus, S.E. & Dreyfus, H.L. (1980). A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition. Washington, DC: Storming Media. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/a084551.pdf

Dreyfus, H.L. & Dreyfus, S.E. & Athanasiou, T. (1986). *Mind over machine: the power of human intuition and expertise in the era of the computer*. New York: The Free Press.

Dreyfus, H.L. (2008). *On The Internet*. New York: Routledge.

ETS (2007). *TOEIC User Guide*. Princeton: Educational Testing Service

ETS (2019). The TOEIC Listening and Reading Test. Retrieved from https://www.ets.org/toeic/listening-reading

Golonka. E.M, Bowles, A.R., Frank, V.M., Richardson, D.L. & Freynik, S. (2014). Technologies for foreign language learning: a review of technology types and their effectiveness. *Computer Assisted Language Learning*, 27:1, 70-105, DOI: 10.1080/09588221.2012.700315

Prolingua.co.jp. (2000). Prolingua :: Executive Language Services. Retrieved from http://www.prolingua.co.jp/toeic_e.html

Nation, I.S.P. & Newton, J. (2009). *Teaching ESL/EFL Listening and Speaking*. New York: Routledge.

Reallyenglish. (2019). Practical English 7. Retrieved from https://www.reallyenglish.com

Singer, S. & Ivory, D. (2017, November 3). How Silicon Valley Plans to Conquer the Classroom. *New York Times*. Retrieved from https://www.nytimes.com