Received: 4 November 2020 | Accepted: 11 May 2021

DOI: 10.1111/bjet.13131

ORIGINAL MANUSCRIPT

British Journal of Educational Technology BERA

Algorithmic versus teacher-led sequencing in a digital serious game and the development of second language reading fluency and accuracy

Judit Serra | **Roger Gilabert**

Department of Modern Languages and Literatures and English Studies, Universitat de Barcelona, Barcelona, Spain

Correspondence
Judit Serra, Universitat de Barcelona, Barcelona, Spain.
Email: juditserra@ub.edu

Funding information
This project has been funded by Horizon 2020 programme by the EU (Grant Agreement No 731724).

Abstract

Recent years have seen a surge of calls for personalization of education. Automatised adaptivity in serious games has been advocated as a potential instantiation of such calls. Yet little is known about the extent to which personalised learning through automatised adaptivity poses an advantage for language learning over generalised teacher-led sequencing in digital, game-based learning environments. The goal of this paper is to address this question by comparing the learning outcomes in reading accuracy and fluency of didactic sequences designed by EFL teachers or by an adaptive algorithm. A total of 67 participants completed several proficiency and reading skills pretest and posttest and used the iRead system for 6 months. Results showed that all learners made progress in reading skills, but no significant differences were found between the two sequences in relation to the development of reading skills. It was also shown that adaptivity works best if it leads to increase in the number of games per feature. Results are discussed in the context of previous findings, and the role of adaptivity and sequencing is critically assessed.

KEYWORDS
adaptivity, game-based learning, literacy, reading skills development, serious games

Practitioner notes

What is already known about this topic?
- Serious games have the potential to aid learning but empirical research is needed.
- Findings about the efficiency of serious games are mixed.
- Current and reviewed versions of the Simple View of Reading constitute a suitable framework to measure reading acquisition.

What this paper adds?
- It contributes to the growing corpus of research on digital serious games.
- It provides empirical evidence on the use of an adaptive system in formal education.
- Comparing a teacher-led sequence to an algorithmic adaptive sequence on the same digital serious game has never been done before.
- The paper shows the need to obtain both system-internal and system-external data in order to capture the impact of gameplay on the development of L2 reading skills.

Implications for practise and/or policy
- It sheds some light on how certain game designs may actually help practise with different degrees of intervention by teachers.
- It is interesting for teachers to use an adaptive sequence that they can check and intervene in if needed.

INTRODUCTION

In unison with technological advances in society, the role of technology in education has undoubtedly become central in recent years. This key role has been further emphasised by learners' needs to confine temporarily in many parts of the world during the ongoing world pandemic. Out of a plethora of technological possibilities, a popular technological option both in class and beyond the class is happening through the use of digital serious games, which are aimed at learning specific educational content through digital gameplay.

Although digital serious games have been increasingly used in formal educational contexts and seem to provide a very exciting way to learn (Camacho & Esteve, 2018), a number of scholars (Hainey et al., 2016; Vanbecelaere, Van den Berghe, et al., 2020) have claimed that there is a lack of empirical evidence that supports digital game-based learning as a better classroom resource in comparison to more traditional methods. A fair question by many is whether games, their related algorithms and their potential for personalization pose any kind of advantage for learning over decision-making by teachers. Both the debate and an increasing interest in the issue have drawn considerable research attention in recent times. Within this overall focus, the goal of the study reported in this paper is to compare an algorithmic, adaptive sequence with a teacher-led sequence in the context of a serious game and to measure their potentially differential impact on second language (L2) reading development.

LITERATURE REVIEW

Digital serious games: The emergence of adaptivity in reading

Serious games are designed for a primary purpose other than pure entertainment, as defined by Djaouti et al. (2011). In the last few years, there has been an extraordinary surge in the use of serious games in formal education, particularly in the area of mathematics and literacy skills (Vaala et al., 2015). One of the components of digital serious games that has been claimed to have potentially an advantage over nondigital games is that of adaptivity. Recent serious games have begun to include an adaptive component that makes it possible to personalise the learning path of the game's user through complex algorithms. Beyond the ongoing debate about the exact definition of the more general concept of personalization (Vanbecelaere, Vasalou, et al., 2020), adaptivity is defined here as the ability of a system to adjust instruction based on learner abilities and/or preferences, at any particular point of the instruction process, with the goal of acting on identified learner characteristics and improving the efficiency and efficacy of learning (Oxman & Wong, 2014; Vandewaetere et al., 2011). Hence, it allows for the automatic adaptation of game elements such as content, user interfaces, game mechanics or game difficulty in order to customise or personalise the interactive experience (Holmes et al., 2018; Streicher & Smeddinck, 2016). Such customization and adaptivity, some scholars have claimed, have the potential to cause faster and more effective learning. The claim, however, has received only partial empirical testing and support.

Adaptivity has been applied differently in various games and contexts, which may explain why so far investigations on the issue have yielded mixed results. Vanbecelaere, Van den Berghe, et al. (2020) tested the effectiveness of a reading game in reading acquisition. The algorithm let readers progress at their own pace and mastery of any subgame was set at 80% before they could move to other more difficult subgames. They obtained mixed results when assessing primary school children that played that reading game on simple and complex word reading fluency and text reading fluency. Kartal and Terziyan (2016) also got mixed results after exposing some participants to a phonological awareness serious game. Participants who played the adaptive game outperformed the participants that did not play games in the dimensions of letter-naming and phoneme segmentation. This, however, was not the case for rhyming or syllable blending.

Kyle et al. (2013) defined the way in which a digital serious game, the GraphoGame, was adaptive. GraphoGame first introduced all grapheme-phoneme correspondence (GPC) and ordered them from frequent to infrequent, whilst also considering rhyming words. The students advanced through a series of graduated game paths, with multiple levels. The study concluded that participants playing two adaptive literacy serious games outperformed those not playing in reading, spelling and phonological skills in a pretest–posttest design. Van de Ven et al. (2017) focused on special needs children. They progressed through 12 increasingly complex levels, organised on the basis of feature frequency, and learners could order and replay the games they wished to play. They found an effect of serious game play in reading fluency and pseudo-word reading but no effect on decoding.

A recent study by Vanbecelaere, Van den Berghe, et al. (2020) revealed that participants playing an adaptive reading game or the same game without the adaptive component significantly improved from pretest to posttest on phonological awareness and letter knowledge. Furthermore, they concluded that prior knowledge, home language and socioeconomic status did not moderate the impact of learning. Nevertheless, Sampayo-Vargas et al. (2013) found that students who played an adaptive game that involved L2 cognates achieved significantly higher learning outcomes. Other studies suggested that groups that used adaptive systems achieved significantly better reading comprehension for both L1 and L2, particularly in low attainment children (Camacho & Esteve, 2018; Ching-Kun et al., 2013).

Using a digital adaptive component contrasts with traditional teaching methods, where the teacher is in charge of building up a syllabus that determines what is to be taught (selection) and the order in which this will be done (sequencing), by usually applying increasing difficulty, which is based on curriculum guidelines and often textbooks, teacher training, intuition and experience. With such an approach, all learners need to go through the same linguistic items in a one-size-fits-all fashion, regardless of their previous acquired knowledge and no matter if they are cognitively ready to learn the new linguistic feature (Pieneman, 1989). There is a gap in research analysing in what ways teacher-led sequencing is beneficial or detrimental to learning and so the comparison between algorithmic and teacher-led sequencing has been extremely challenging. In general, the issue of sequencing in second language acquisition (SLA) is an unresolved one.

Camacho and Esteve (2018) have suggested that there is insufficient research dealing with the real direct impact of the technology in the learning process. There are very few well-designed empirical studies that support positive outcomes of Foreign Language (FL) learning (Golonka et al., 2014). Even though in the last decade there has been a raise of publications on technology-based reading programmes for elementary grades, few studies are devoted to evaluating the impact of tablet literacy apps on reading (Jamshidifarsani et al., 2019). Streicher and Smeddinck (2016) pointed out a lack of research in the field because of the high costs for design, authoring and the technical implementation of both the games and the adaptation system. As with other areas in SLA (see the unresolved issue of sequencing in SLA—Baralt et al., 2014), there is also insufficient research dealing with the issue of sequencing and the exact criteria that should guide sequencing.

The mixed range of findings reported on this section and the gaps identified so far strongly suggest a need to empirically investigate the impact of adaptive serious games and their sequencing on language development, in general, and reading development in particular. It should be noted that, rather than on the effectiveness of iRead as a serious game, this paper focuses on the issue of the players' learning sequence and its effect on reading development.

Adaptivity in the serious reading game iRead

The present investigation is based on an app designed by the iRead Horizon 2020 innovation project (grant agreement No 731724) which aimed to develop innovative learning tools for personalised learning that included an adaptive component. Although the system was designed for English, Spanish, Greek and German, we report on EFL in Spain here. The system consisted of three applications (games, e-reader and a teacher tool) designed to be used with tablets, and the system was embedded in regular classroom practise.[1]

The iRead system included automatic adaptive principles, operationalised in terms of content difficulty and competence showed in the games already played (Tsatiris & Karpouzis, 2020). All language features to be worked on by players in the game were described in a domain model for language difficulty and prerequisites, which was decided item by item by a panel of language experts who based their selection and sequence on the current literature. Therefore, features known to be easier for learners would appear earlier in the game and the player could only be presented at a given time with a limited number of open features with a similar level of difficulty. As each feature was mastered, the system unlocked the next more difficult features as per the domain model. The adaptive algorithm computed the player's performance on each mini game in terms of number of successes and failures and adjusted to the pace of each learner. This means some players could move on after playing three games correctly on a feature, whilst others could keep playing on the same feature until they mastered it. The adaptivity component took into account the

prerequites for each feature in order to present features in increasing difficulty during gameplay (eg, individual vowels were presented before diphthongs), and the level of feature mastery was set at 70% of success having played a minimum amount of three times on the same language aspect. The system made use of Exponential Moving Average (EMA), which utilises *n* past attempts with reduced weights (from starting value (5) → maximum (10) in three perfect games with a Maximum reduction = 1). Once this success rate was accomplished, more complex features would be unlocked.[2] This describes the adaptivity component that operated under the algorithmic sequence, with no teacher intervention. In the teacher-led sequence, however, the algorithmic sequence described above was not operational and it was teachers who set specific content and games for the whole class through the teacher tool, as they would do in a more traditional teacher-led sequence.

Reading acquisition in a FL

Since iRead was devised to promote reading skills, both FL acquisition and reading acquisition processes are addressed in this section. General agreement exists around the idea that reading acquisition is a basic process in our daily lives, and acquiring proper reading skills is essential for our personal development and normal functioning in society. In the educational system, it has often been shown that children who have not acquired proper reading skills are bound to face more obstacles in their overall learning process and success (Jiménez, 2009).

In terms of research in reading acquisition, Gough and Tunmer's (1986) Simple View of Reading (SVR) model is a well-accepted and empirically supported framework that stands out in terms of its explanatory power. This model states that reading ability is the product of decoding and comprehension abilities ($R = D \times C$). Decoding is defined as the "ability to read isolated words quickly, accurately and silently" (p. 7). Comprehension is understood as the way information, sentences and discourse are interpreted. In the original study by Gough and Tunmer, these two dimensions were measured by means of a pseudoword reading task and a listening comprehension task. There is a broad agreement on the importance of both skill sets explained in the SVR for the prediction of reading comprehension (eg, Catts et al., 2015; Nation, 2019).

Nevertheless, the simplicity of the framework has also been questioned and several studies have suggested some adjustments. Cain et al. (2015) analysed the relationship between word recognition (decoding in the SVR), listening comprehension and reading comprehension, adding also vocabulary as a variable, in a large sample of students in grades 1–3 for English as an L1. They confirmed that before grade 2 (around age 7), decoding has a central role, and after that, it is listening comprehension that becomes more prominent. This is due to word recognition skills becoming more automatic in older readers, which enables listening comprehension to play a greater role in the prediction of reading comprehension.

Word reading accuracy is a component of word recognition, defined as the ability of a reader to correctly generate a phonological representation of an encountered lexical item (Cain et al., 2015). Beginner readers struggle with decoding words and often use the strategy of sounding words out (Bernhardt et al., 2006). However, as accuracy improves (together with other aspects), reading proficiency increases and readers instantly begin to recognise whole words (Nation & Snowling, 1998), moving from a sublexical, word-by-word reading mode to a lexical, whole-word approach to reading. L2 learners tend to use mostly the former strategy since they do not feel comfortable with the FL, but they use the latter for frequent words (Bernhardt et al., 2006).

In studies dealing with L2 reading acquisition, accuracy has been usually operationalised through word and nonword decoding tasks. Reading fluency has been defined by Fuchs

et al. (2001, p. 239) as: "… the oral translation of text with speed and accuracy". In our context it has usually been operationalised through a timed reading passage in which words per minute are calculated. Kim (2012) claimed that little attention has been paid to reading comprehension for L2 learners. In general terms, research in the field has also used the SVR as a model of reference, but the issue remains as to whether the SVR provides a sufficiently complete view to understand reading acquisition in an L2 (for a model adapting the SVR to the L2 and related empirical research see Proctor et al., 2005). In the present study, both a word test and a nonword test were used to tap into word decoding and a timed short passage is used to tap into learner's reading fluency.

From the literature available in the field, a number of gaps were identified. Those included the need for more empirical research measuring the impact of serious games on learning, the effects of sequencing on the development of L2 reading skills and the empirical testing of theoretical models of reading for L2 learning. This paper aims to contribute to filling those gaps and to the growing corpus of research into reading development in the context of serious games. It does so by exploring the impact of an innovative and adaptive digital serious game delivered through two difference sequences and its effect on the development of foreign reading skills as measured by decoding ability and fluency rate.

RESEARCH QUESTIONS

By considering all the factors mentioned in the previous section, the main objective of this paper was to investigate whether and how the adaptive component of the serious game iRead may contribute to reading skills acquisition in a second language in comparison to a traditional teacher-led sequence. To do so, three research questions were addressed:

> RQ1 Does the type of sequence (algorithmic vs. teacher-led) differentially impact the improvement of reading skills through games?
>
> RQ2 Does the number or variety of games affect the learning of reading skills?
>
> RQ3 Is the learning of reading skills mediated by decoding ability, vocabulary knowledge, or listening ability?

As it is exploratory in nature and no directional hypotheses were available, this study adopts the null hypothesis for all questions and makes no predictions as to how gameplay and the reading of e-reader texts may impact reading development.

METHOD

Participants

Two schools from Spain that participated in the iRead Project were selected for this study. This involved 107 students from 6th grade of Primary Education (11 and 12 years old) from four intact groups and three teachers who participated in the current study, although only 67 students completed all the tests in the study (Table 1). The schools were in middle-to-high socio-economic areas in Barcelona or its surroundings. Participants were L1 Catalan-Spanish speakers, and English was their L2. In both schools, English was introduced formally at the beginning of Kindergarten (age 3) and classes were rather traditional in nature, where a grammar-oriented textbook was followed. Data were collected from September 2019 until

TABLE 1 Participants' characteristics

EFL students (*n* = 67)				
School	1	1	2	2
Class	6A	6B	6A	6B
Number of students	26	26	27	28
Participants	16	6	22	23
Gender	8 males	4 males	11 males	13 males
	8 females	2 females	11 females	10 females
Sequence	Algorithmic	Teacher-led	Algorithmic	Teacher-led

March 2020 when system use was abruptly interrupted by the pandemic. Both schools were well equipped in terms of technological resources and were used to using technology in the classroom. Every student in the class would use a tablet to carry out iRead sessions, and the WIFI connection was stable and reliable.

The iRead tools

Navigo games (*Navigo: The Pyramid of the Lost Word*s): the player's customised avatar was taken on an adventure into a mysterious pyramid with the double goal of finding the main character's grandmother and saving as many villagers as possible. Inside, there were a range of different puzzles aimed at developing language skills related to reading. They included 15 different dynamics that involved choosing the bridge with the right option (phonological, morphological, syntactic), matching items to open doors, slicing words with a laser beam, blowing up the right word or pushing some pillars to create the right sentence amongst others. From the point of view of SLA and language curriculum design, they were traditional in their design (eg, multiple choice, matching, odd one out, amongst others), focused on isolated forms, and did not include an interactive component (eg, as in task-based games).[3] The choice of features for the games was the results of combining a *domain model* (which specified the features and their difficulty and sequence), with *a game specification* (which assigned specific game dynamics to specific features) and *a 22,000-word dictionary* that provided the words that contained the features that fed the games and the e-reader. The point of the games is to bring certain features and contrast to the attention of learners and to raise the learner's awareness of such features and contrasts in a playful environment. As confirmed by questionnaires and interviews, games were challenging and attractive to students. Language content included phonic skills, syllabification, common sight words, confusing letters, adjectives, pronouns, complex sentences, discourse anaphors, negation, passives, prepositions, determiners, adverbs, questions, modal verbs, prefixes and suffixes, adverbs and verbs. In all cases, features were not only played by the learners but also read out loud by the automatised voice system embedded in the game (two peninsular Spanish voices were used: a read female voice was used for the recording of individual sounds, and a female Google voice for automatised text-to-speech) that pronounced the sounds, words, phrases or sentences in the games. The audio was meant to aid with, for example, the lack of correspondence between graphemes and phonemes in English.

E-reader: it included hundreds of texts which could be adapted in terms of appearance (font type, font size, font colour, amongst others) to each child's preference. Learners could listen to the text being read by an automatised voice system. This was an in-built characteristic of the reader based on the idea that reading enhanced audio is more efficient than just reading or just listening (Chang & Millet, 2015). Every sentence that was read was also

highlighted simultaneously in order to try to keep the readers focused on the text. Learners could check the meaning of unknown words by clicking on them and getting a dictionary explanation and they could save in a list of difficult words ('tricky' words in the system) where they could listen and even upload a picture/drawing to be associated with each word. Moreover, every time learners accessed a text, they would find a prereading activity that consisted of an explanation of a language feature and some examples of it. Then, in the text, the feature would be highlighted. Its multimedia and enhancement functionalities were meant to maximise processing and facilitate comprehension. Finally, there was also a list of recommended books and a list of favourite books.

Teacher tool: the teacher tool gave teachers the autonomy to make their own choices regarding features, games and books that they wished to assign to each individual student or to the whole class. For the study presented here, it was the teacher in the teacher-led group who determined the sequence that everyone would follow, whereas in the algorithmic sequence teachers let the system choose accordingly with the adaptivity rules.

Design and procedure

All students from the four groups used the iRead system for approximately 6 months and used a 1-h English session per week. They started in October, after the pretests took place, and finished in March when schools in Spain closed due to COVID-19 measures. Learners in the schools involved in this study played 7896 games in total. Posttests took place online, and this produced a considerable loss of participants, which took place particularly in one of the schools.

Before collecting data, participants' guardians signed a consent form which informed them about all steps that would be taken in the study and about the anonymous and volunteering nature of their participation. Firstly, prior to the implementation of the system, the participants carried out some pretests, which included a background questionnaire and reading measures (fluency, vocabulary size and reading comprehension). The exact same reading skills measures were also tested as posttests. Thanks to the adaptive algorithm present in the system as well as the possibility of choosing what games and features to work on, it was possible to create the two treatment conditions.

Due to the implementation of this study in a realistic context, randomization of subjects was not possible and so intact groups were used. Teachers in both the A groups were always asked to let system follow the algorithm's adaptivity parameters without teacher intervention. Teachers in B groups were asked to choose the content and the games in their preferred, most logical sequence for every weekly iRead session. Teachers could overcome the default adaptivity of the system by choosing specific features or games themselves through the Teacher Tool.

All in all, the experience by leaners was similar in every way: they played games for 1 h a week and read texts from the e-reader, the only difference being the sequence of presentation, either adaptive or teacher-led, which the students were not aware of. Therefore, in the two teacher-led groups all students from a class would follow the same learning sequence, which was chosen by the teacher. In the algorithm-led groups, each student would follow a different sequence that was the result of their performance. The students' learning sequences were monitored by the researchers, and they absolutely followed the type of treatment they were assigned (either totally adapted by the algorithm or totally adapted by the teacher). Learners used iRead for 1 h every week in their regular EFL class with the assistance and under the supervision of their teacher. Learners worked individually with their tablets and used headphones, and the teacher did not intervene in class unless there were connectivity issues. Classroom observations took place, and there were no big differences

amongst groups or amongst schools in terms of methodology and the role of the teacher during the use of iRead.

Instruments

Both system-external features for proficiency and reading skills and system-internal features for gameplay and texts read were used. Amongst others which are not reported here, system-external features measured reading skills development at pretest and posttest as well as proficiency:

- A word and nonword reading test measuring decoding skills. The word reading test was taken from BAS 3 (Elliot & Smith, 2011) and included a list of words to be read aloud accurately in a minute. The nonword reading test was carried out in a similar way. The tests were adapted by iRead partners at University College London.
- An oral reading fluency test (State of Florida Department of Education, 2009) that consisted of the OPM Oral Reading Fluency. In this test, participants had to read a text aloud accurately. Scores were calculated by subtracting the errors from the total words attempted in story in 60 s.
- PET test was used to test listening comprehension and reading comprehension (as a measure of EFL proficiency).
- A test of vocabulary knowledge. It was the Picture Vocabulary Size Test-PVST (Anthony & Nation, 2017 adapted by Plumiège & Peters, 2019) and consisted of a receptive vocabulary size test in which the participant had to choose the most accurate picture describing a word in a context. The score was obtained out of 96 items.
- A background questionnaire was filled in by the participants. It was designed specifically for this study and it included personal data such as date of birth and gender.

The analytics function of the iRead system automatically recorded several internal measurements to minute detail, which included the total number of games played, games played per feature, the number of different features and the number of errors amongst others. General measures were taken from all features and all games played by learners. As for specific measurements, the number of games per feature and the number of errors per feature were calculated on 12 features that are specifically problematic of L1 Spanish/Catalan learners of English and it included vowers, consonant and consonant clusters (eg, '-ee' as in 'see' which requires a /i:/ sound, or initial 'kn' which is pronounced as /n/ as in 'knife'). Internal system data were handled by means of both Tableau and python functionalities.

Statistical procedures

Statistics were carried out with the use of SPSS software. All variables were checked for normality by means of the Shapiro Wilk test, which is considered more powerful than Kormogorov–Smirnov. For mean comparisons related to proficiency, one-way ANOVAs were used. Whenever variables were not normally distributed, Kruskal–Wallis test were used. In order to check any improvements in performance from pretest and posttest performance, repeated-measures ANOVA was used. Repeated-measures ANOVAs were also used for measuring gains in reading accuracy and fluency as well as for potential interactions with other variables. Whenever any measures were not normally distributed, nonparametric Friedman and Wilcoxon Signed Rank Tests were used. Pearson and Spearman tests were

TABLE 2 Descriptive statistics of proficiency scores in PET reading, PET listening and vocabulary knowledge test

Sequence	*N*	Min	Max	*M*	*SD*	Asym	Curtosis
Algorithmic							
Reading	38	6	28	15.63	4.978	.251	−.413
Listening	38	3	16	10.32	2.798	−.552	.376
Vocab	38	28	70	51.58	9.953	−.561	.170
Teacher-led							
Reading	29	7	27	14.34	5.287	.780	−.276
Listening	29	3	18	9.24	3.356	.798	1.131
Vocab	29	25	69	49.28	11.364	−.310	−.802

used to check correlations between proficiency and performance and between performance in the games and book and gains in reading accuracy and fluency.

RESULTS

Results are presented by research question and divided into general and specific measures. First, the results for general measures are presented that refer to all features offered by the system, and then the results of specific measures were analysed that refer to 12 GPC features that were selected because they are well-known to be particularly difficult for L1 Spanish speakers learning English as an L2. Preliminary descriptive statistics (Tables 2 and 3) and ANOVAs (Table 4) did not show any significant differences in proficiency (as measured by the PET listening and PET reading sections) or in vocabulary size between the two sequences. This rendered groups comparable before the study started.

> Research question 1: Does the type of sequence (algorithmic vs. teacher-led) play a role in the improvement of reading skills through games?

General measures

Preliminary analysis showed that the word, nonword and WPM test were highly correlated, which suggests the tests worked well as a measurement of accuracy and fluency rate. Participants that read more words and nonwords correctly also produced more correct words per minute in the passage. This was also true for the posttest measures. Repeated-measures ANOVAs tests (Tables 3 and 4) showed that accuracy and fluency gains from pretest to posttest were all significant for all reading skills performance measures, that is, for individual words, nonwords and word per minute (WPM). No interaction was found between gains in accuracy and fluency and sequence, except for nonwords where learners in the teacher-led sequence read significantly more correct nonwords than learners in the algorithmic sequence. It can be, therefore, concluded that as far as general measures are concerned, learners in both sequences improved their overall accuracy and fluency of their reading from pretest to post-test. As an answer to the first research question, the type of sequence did not differentially affect gains in reading skills.

TABLE 3 Descriptive statistics of gains in words, gains in nonwords and gains in words per minutes (WPM) per sequence

Sequence	*N*	Min	Max	*M*	*SD*	Asym	Curtosis
Algorithmic							
Words at pretest	38	18	63	42.61	9.313	−.134	.788
Nonwords at pretest	38	26	57	43.53	9.267	−.249	−1.091
WPM at pretest	38	40	128	89.47	20.707	−.357	.123
Words at posttest	38	25	69	47.42	9.205	−.272	.516
Nonwords at posttest	38	30	58	45.53	6.306	−.293	−.253
WPM at posttest	38	57	136	106.05	17.574	−.656	.793
Teacher led							
Words at pretest	29	18	61	40.90	12.373	−.034	−.915
Nonwords at pretest	29	12	58	37.76	11.701	−.162	−.371
WPM at pretest	29	42	126	83.90	23.651	−.007	−1.103
Words at posttest	29	25	76	49.41	12.266	.068	−.232
Nonwords at posttest	29	22	71	44.52	10.016	.138	.935
WPM at posttest	29	48	145	96.97	24.472	−.047	−.524

TABLE 4 Repeated-measures ANOVAs for gains in words, gains in nonwords and gains in words per minute (WPM)

Origen		Type III sum of squares	*gl*	Mean square	*F*	Sig.	Partial Eta squared
Gains in words	Assumed sphericity	1.461.957	1	1.461.957	45.026	.000	.409
Gains in nonwords	Assumed sphericity	630.882	1	630.882	15.835	.000	.196
Gains in WPM	Assumed sphericity	7.228.781	1	7.228.781	89.746	.000	.580
Gains in words *sequence	Assumed sphericity	112.673	1	112.673	3.470	.067	.051
Gains in nonwords *sequence	Assumed sphericity	186.225	1	186.225	4.674	.034	.067
Gains in WPM *sequence	Assumed sphericity	101.318	1	101.318	1.258	.266	.019

Specific measures

As for specific measures, 12 particularly difficult GPC features were selected. These were both in the pretest/posttest and were specifically addressed by the system. Data showed that the algorithmic sequence group worked on the features to a significantly higher extent than the teacher-led sequence. Since the data from the algorithmic group were not normally distributed, the Kruskal–Wallis test was used to check if the sequence had an impact on gains from pretest to posttest on the 12 features, concluding that the result was not significant ($p = 0.244$). As far as specific measures are concerned, and as an answer to the first research question, the type of sequence did not differentially affect gains in reading skills, and any gains by the participants in the two sequence can be considered equal.

TABLE 5 Descriptive of number of games, different features, errors, books and tricky words as per specific measures

Sequence	*N*	Min	Max	*M*	*SD*	Asym	Curtosis
Algorithmic							
Games	38	63	174	115. 08	26.109	.268	−.274
Features	38	37	74	52.42	8.452	.509	.047
Errors	38	22	181	80.71	38.084	1.002	.656
Books	38	16	136	69.94	26.175	.172	.165
Tricky words	38	0	130	20.36	29.155	2.063	-4.672
Teacher-led							
Games	29	45	215	118.55	38.387	−.107	.374
Features	29	13	38	21.00	6.453	1.594	1.867
Errors	29	27	235	117.03	58.062	.273	−.704
Books	29	19	128	55.03	21.501	1.494	3.813
Tricky words	29	0	105	24.55	25.187	1.665	2.816

Research question 2: Does the number or variety of games affect the learning of reading skills?

In terms of general measures, descriptive statistics (Table 5) and one-way ANOVA results showed that both groups were comparable in the number of games they played (df = 66, F = 0.194, p = 0.661). Kruskal–Wallis tests yielded significant differences, however, in the range and variety of features they played in those games (p = 0.007). Post-hoc analyses revealed that learners in the algorithmic sequence played a wider range of different features but mostly within the dimensions of GPC. On the contrary, learners in the teacher-led sequence played a smaller range of different features that were spread across dimensions (phonology, morphology, morphosyntax and syntax). Students in the teacher-led sequence committed significantly more errors (p = 0.000) than students in the algorithmic sequence. Readers in the algorithmic sequence read significantly more books (df = 66, F = 6.208, p = 0.015) but selected a similar number of tricky words (p = 0.100) as they were reading.

Spearman correlations showed that there was no correlation between any of the variables related to gameplay and gains in decoding or reading fluency abilities. The only exception to that was the number of different features, which showed a week negative correlation with gains in words (r = −0.270*; p = 0.037) and nonwords (r = −0.231; p = 0.030), suggesting that the fewer features they played with, the more likely they were to show gains in accuracy. As for the specific measurements on the 12 selected features, one-way ANOVAs showed that the algorithmic sequence group played a significantly higher number of games related to the 12 features. In terms of gains, it was checked feature by feature and playing more games on features did not impact gains of that feature.

In sum, as far as specific measures are concerned and as an answer to research question 2, playing more games did not result in gaining decoding or reading fluency skills. Playing a narrower variety of features, however, resulted in more gains in decoding skills than playing a wider variety of features.

Research question 3: Is learning of reading skills mediated by decoding ability, vocabulary knowledge, or listening ability?

Briefly put, initial decoding ability was shown to impact gains in decoding and reading fluency abilities by the end of the project, with less of a role for listening and no role for vocabulary knowledge. This can be broken down into three main findings related to words, nonwords and fluency: firstly, a weak-to-moderate but significant correlation ($r = -0.393$, $p = 0.000$) was found between initial word decoding and gains in word decoding, suggesting that the fewer words they could read correctly at the start of the project, the higher gains they had. This was true for both sequences. Neither initial listening ability nor vocabulary knowledge seemed to impact gains in decoding or reading fluency skills; secondly, a moderate but significant correlation was found ($r = -0.634$, $p = 0.000$) between nonword decoding ability and gains in nonword decoding, again suggesting that learners with an initial lower decoding ability showed greater gains. This is true for both sequences, but it was higher for the algorithmic sequence ($r = -0.748$, $p = 0.000$) than for the teacher-led sequence ($r = -0.423$, $p = 0.000$). Learners with an initial lower listening ability also showed more gains ($r = -0.211$, $p = 0.043$); thirdly, initial reading fluency impacted on reading fluency gains moderately ($r = -0.338$, $p = 0.003$) again suggesting that the less fluent they were at the beginning of the project, the greater number of words per minute they gained. Nevertheless, when considering the two sequences separately, it was only the algorithmic sequence that showed such an effect. Listening abilities moderately correlated with fluency gains ($r = -0.402$, $p = 0.006$) only for the algorithmic group whilst vocabulary was not correlated with gains in words per minute in either sequence.

In short, results regarding research question 3 suggested that whilst vocabulary knowledge did not predict gains in decoding or reading fluency abilities, participants with lower initial decoding abilities ended up with higher decoding abilities than those with initial higher abilities. This effect was significantly stronger for the algorithmic group. Initial listening also partially predicted learning in decoding and reading fluency skills at the end of the project.

DISCUSSION

As shown by the significant gains in decoding and reading fluency between pretest and posttest, learning took place in such a way that by the end of iRead project intervention, learners became significantly more accurate and more fluent when reading in the L2 than they were at the beginning. This development cannot be necessarily and exclusively attributed to iRead, though. We could speculate that since reading in all groups was reduced almost exclusively to their work in iRead, the use of system may have been the explanation behind their progress. But given that the intervention was embedded in a larger educational context, a control group would have been needed to confirm such an assertion. Nevertheless, it should be noted that analysing the efficacy of iRead was not the goal of this study. Central to the goals of the study described here was that progress was shown regardless of whether participants had followed iRead with a teacher-led sequence designed by teachers or an adaptive sequence designed by an algorithm. The results obtained in this study suggest that the sequence in which the contents were presented to the learners did not have a differential impact on how much they learned, since no significant differences were found in terms of gains between the two conditions under study. The results, however, raise interesting questions about adaptive algorithms, the length of the treatment and the issue of sequencing in SLA in the context of games.

Research question 1

As far as the issue of adaptive algorithms is concerned, results in this study are in line with Vanbecelaere et al. (2019) and against Kyle et al. (2013), since the expectation that the

learners in the adaptive condition would learn more than in the nonadaptive condition was not borne out. The first explanation may come from the strength of the distinction between the algorithmic sequence and the teacher led sequence, which may not have been large enough. The algorithm used by the iRead system took basic performance information by each player to be fed into the adaptivity algorithm. It took into account the prerequisites for each feature in order to present features in increasing difficulty during gameplay (eg, individual vowels were presented before diphthongs), and the level of mastery was set at 70%. After each feature was done successfully three times, the system selected the next more difficult feature. Nevertheless, the algorithm did not quite compute learning objectives, teaching strategies and end-user requirements, and it did not accommodate temporal queries (eg, the number of days since a specific feature was played). A more sophisticated algorithm may have helped to achieve a larger difference from the teacher led sequence and, perhaps, impacted learning to a higher degree. As suggested by one of the reviewers, an algorithm that incorporated item response theory combined with self-correcting elo rating would have been more appropriate, since responses to each item could have been given a relative weight rather than all responses to items being considered equal. From a SLA point of view, another way to improve the algorithm would have been to use overall learner performance data on each item that could have been fed back into the algorithm through, for example, Rasch modelling. That way the initial difficulty-based sequence of linguistic items built by experts in the domain models could have been adjusted on the basis of learners' performance results in the games, resulting in a dynamic rather than in a static domain model. Such an improvement can certainly be achieved by an even closer collaboration effort by engineers and language acquisition experts.

A second explanation may come from the length of the treatment, which was cut short by the world pandemic. It was designed for 7–8 months and it lasted 5–6 months. A full application of iRead may have shown a larger difference between the two conditions. This argument converges with the idea by Van de Ven et al. (2017) that the intervention may not have been long enough. Also related to length, playing 1 h a week for 4 months, even if it entailed playing hundreds of games by each student, may simply not have been sufficient for the treatment to have more obvious effects. The cyclic nature of the iRead game which would regularly recover already mastered items for their consolidation in the learner's repertoire may not have been fully achieved because of the shorter treatment.

Concerning the 12 specific features studied, they showed that any gains in the accuracy of those 12 features cannot be attributed to the sequence. This means that even if adaptivity worked and the algorithm offered learners features based on their increasing difficulty and they progressed up the difficulty scale in accordance with their performance on the game, adaptivity did not impact development in any particular way. Students in the teacher-led sequence, who did not work specifically on those features, also showed progress. There are no previous studies against which the kind of adaptivity iRead used could be compared, but the results in this study suggest that adaptivity needs to be carefully tuned (eg, by offering fewer features but with more games on each feature) if a clear advantage is to be seen.

Research question 2

Results related to research question 2 showed that learners in both conditions played a similar number of games. Whilst quantity was comparable, variety was higher for the algorithmic sequence, which pushed learners to deal with a wider variety of features. As seen from the results, most of the features were related to GPC. This would suggest that adaptivity played a role, since the system offered whichever feature was open as per the description and prerequisites of each feature in the domain model. Also, learners in the teacher-led sequence

made significantly more errors. This may be attributable to the fact that in the teacher-led sequence, the features learners had to deal with different dimensions (eg, morphology, morphosyntax, syntax) that had been defined in the domain model as more difficult features. As for the correlations between gameplay and gains, neither the number of games itself nor the number of errors were related to any gains.

The fact that the number of different features was negatively correlated with gains would suggest that adaptivity worked to the advantage of the teacher-led sequence. Learners in the teacher-led sequence were offered a narrower range of features that learners in the algorithmic sequence. These features were also more difficult (ie, morphological features or syntactic features as opposed to GPC features), and consequently, they caused learners to make more errors and therefore get more games on the same features until they learned them. This may explain in turn why sequence was not related to gains as seen in the answer to research question 1. Both groups showed a similar amount of gains even if they followed two different paths, one with lots of features but fewer games on each feature (algorithmic feature) and another with fewer features but more games on each feature (teacher led sequence). Again, it is an issue whether a narrower algorithm that offered fewer features but more games per feature would have shown an advantage for the algorithmic sequence over the teacher-led sequence.

As for specific measures, it was shown that playing more games on those features did not lead to more learning. This may have at least two interpretations. Firstly, that the algorithmic group, which played a wider variety of features but fewer games on each feature, did not play enough times on each feature for such play to have posed an advantage. It is well-known in psycholinguistic studies related to word learning (eg, Pellicer-Sánchez & Siyanova-Chanturia, 2018) that a considerable number of exposures to a new word in context are necessary for new word to be learned (some studies claiming 13 exposures to a new word in context). The same should be expected of other morphological or syntactic features and so exposure through only a limited number of games on each feature may have not been sufficient to learn the new feature. A second interpretation is that even if the features were not directly played by the learners in the teacher-led sequence, this may have been compensated by the fact that the features did not appear exclusively in the words chosen to play those 12 features but also in other words which were targeting other linguistic features, so theoretically some incidental, implicit learning may have happened. It is again an issue whether adaptivity would have played out differently had leaners had all hours of system use till the end of the academic year.

Research question 3

The fact that an initial lower word and nonword decoding ability actually predicted higher gains is not an unusual finding in SLA studies. Students at the lower end of reading abilities made significantly more errors which, through adaptivity, provided them with more games to overcome those errors. In line with the findings by Camacho and Esteve (2018) and Ching-Kun et al. (2013), this also suggests that the use of the system, whether with an adaptive sequence or a teacher-led one, may have helped the weaker learners to higher extent that the more advanced learners. This same impact was found for listening abilities, since poorer listeners gained the most, whereby listening ability has been proven to be a crucial factor in the SVR (Cain et al., 2015). Interestingly, fluency rate measured by words per minute showed higher gains for initially less fluent learners, who benefitted the most from the algorithmic sequence. A second explanation may come from multimedia learning (Mayer, 2001; Paivio, 1986). The fact that learners where exposed to the items and the words that contained through two different channels (ie, the visual channel exposing them to the text, and

the auditory channel exposing them to the pronunciation of sounds and words) may have been particularly beneficial for those students with initially lower decoding and listening activities. The potential advantages of multimedia learning would certainly benefit from further research in the context of serious games.

CONCLUSIONS

This article has shown that although all learners displayed significant progress in reading skills from the start to the end of the iRead project (although not necessarily as a consequence of the iRead project), the sequence they followed in the games, either algorithmic or teacher-led, did not seem to have an impact on how much learning took place. This study has shown that learners took different routes to learning, with learners in the algorithmic sequence playing fewer games on a larger number of features and learners in the teacher-led sequence playing more games on a narrower set of features.

A number of general conclusions and implications can be drawn from the implementation and the results in this study. The first conclusion to be drawn from the results obtained in this study is that the possibility of choosing an algorithmic or a teacher-led sequence as teachers did through the teacher tool is without a doubt an innovative aspect of serious games design which allowed for the creation of two sequences. Secondly, this study has shown the potential of an adaptive component impacting the path of language development in general and readings skills in particular. This was shown to be particularly the case for learners who started off with a lower level of reading skills and who ended up showing higher gains. Adaptivity, however, seems to require careful design in order to get enough sustained and intense practise for development reading skills. Thirdly, the issue of sequencing remains an unresolved one in the context of both SLA and game-based learning. Future algorithms will need to include a more complex conceptualization and operationalization of feature difficulty, to be determined not only by initial expert decisions but also by data of actual gameplay to be fed back into the system and more sophisticated item response analyses. As for limitations, the main one was the lack of a control group that, even if beyond the goals of this study, would have provided information about the effectiveness of not just the sequence but of the system itself. Clearly, a larger number of participants would have given this study a lot more statistical power and may have emphasised any potential differences between the two sequences.

Finally, this study and the tool contained within it have shown that content in serious games can be delivered by both algorithms and teacher interventions and by a combination or alternative use of the two, with teachers having the opportunity to check and intervene in their students' individual paths. In line with Vanbecelaere, Van den Berghe, et al. (2020) and Holmes et al. (2018), we believe this has enormous potential for personalization of L2 learning.

ACKNOWLEDGEMENTS

We would like to acknowledge the contribution to this project of the participating schools Sant Nicolau in Sabadell and Franciscanes Poblenou in Barcelona. We would also like to acknowledge the invaluable help of Matthew Pattemore with the handling and extraction of system-internal data as well as Dyonisis Panagiotopoulos from the National Technical University of Athens and the whole NTUA team, the team at UCL for their help with decoding tests, as well as the whole iRead team. Finally, we would like to thank the GRAL (Language Acquisition Research Groups) at the University of Barcelona for their support.

CONFLICT OF INTEREST

To our knowledge, there is no conflict of interest since the iRead project has not been developed into a commercial set of tools and has only been used within an educational context for innovation and research purposes.

DATA AVAILABILITY STATEMENT

Our data would be available for inspection by other researchers and/or replication.

ETHICS STATEMENT

Ethical guidelines as following University of Barcelona as well as EU protocols guaranteed data protection standards were met. Anonymised code names were used were used for data recording, extraction and analysis.

Legal guardians of the participants signed a consent form in order to become a part of the study which described all procedures and ethical considerations.

ORCID

Judit Serra https://orcid.org/0000-0002-1601-6723

ENDNOTES

[1] iRead's Project Website: https://iread-project.eu.

[2] It should be noted that the authors of this article did not directly participate in the creation of the domain model for the study reported here or on the design of the algorithm.

[3] Again it should be pointed out that the authors of this study had no voice in the design of games, which were provided by the game developer in the project.

REFERENCES

Anthony, L., & Nation, I. S. P. (2017). *Picture vocabulary size test (Version 1.2.0) [Computer Software]*. : Waseda University. http://www.laurenceanthony.net/software/pvst

Baralt, M., Gilabert, R., & Robinson, P. (2014). An introduction to theory and research in task sequencing and instructed second language learning. In M. Baralt, R. Gilabert, & P. Robinson (Eds.), *Task sequencing and instructed second language learning* (pp. 1–34). Bloomsbury Academic.

Bernhardt, E. B., Kamil, M., & Brown, K. (2006). The state of second-language reading research. In K. Brown (Ed.), *Encyclopedia of languages and linguistics* (pp. 88–95). Elsevier.

Cain, K., Catts, H., Hogan, T., & Lomax, R. (2015). Learning to read: Should we keep things simple? *Reading Research Quarterly*, *50*, 151–169. https://doi.org/10.1002/rrq.99

Camacho, M., & Esteve, F. (2018). El uso de las tabletas y su impacto en el aprendizaje. Una investigación nacional en centros de Educación Primaria. *Revista de Educación*, *379*, 170–191. https://doi.org/10.4438/1988-592X-RE-2017-379-366

Catts, H. W., Herrera, S., Nielsen, D. C., & Bridges, M. S. (2015). Early prediction of reading comprehension within the simple view framework. *Reading and Writing*, *28*(9), 1407–1425. https://doi.org/10.1007/s11145-015-9576-x

Chang, A.-C.-S., & Millet, S. (2015). Improving reading rates and comprehension through audio assisted extensive reading for beginning learners. *System*, *52*, 91–102. https://doi.org/10.1016/j.system.2015.05.003

Ching-Kun, H., Gwo-Jen, H., & Chih-Kai, C. (2013). A personalized recommendation-based mobile learning approach to improving the reading performance of EFL students. *Computers & Education*, *63*, 327–336. https://doi.org/10.1016/j.compedu.2012.12.004

Djaouti, D., Alvarez, J., & Jessel, J.-P. (2011). *Classifying serious games: The G/P/S model. Handbook of research on improving learning and motivation through educational Games: Multidisciplinary approaches*. IGI Global. https://doi.org/10.4018/978-1-60960-495-0.ch006

Elliot, C. D., & Smith, P. (2011). *British ability scales 3*. GL Assessment.

Fuchs, L., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*(3), 239–256. https://doi.org/10.1207/S1532799XSSR0503_3

Golonka, E., Bowles, A., Frank, V., Richardson, D., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning*, *27*, 70–105. https://doi.org/10.1080/09588221.2012.700315

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading and reading disability. *Remedial and Special Education*, *7*(1), 6–10. https://doi.org/10.1177/074193258600700104

Hainey, T., Connolly, T., Boyle, E., Wilson, A., & Razak, A. (2016). A systematic literature review of games-based learning empirical evidence in primary education. *Computers & Education*, *102*, 202–223. https://doi.org/10.1016/j.compedu.2016.09.001

Holmes, W., Anastopoulou, S., Schaumburg, H., & Mavrikis, M. (2018). *Technology-enhanced personalised learning: Untangling the evidence*. Stuttgart: Robert Bosch Stiftung GmbH.

Jamshidifarsani, H., Garbaya, S., Lim, T., Blazevic, P., & Ritchie, J. M. (2019). Technology-based reading intervention programs for elementary grades: An analytical review. *Computers & Education*, *128*, 427–451. https://doi.org/10.1016/j.compedu.2018.10.003

Jiménez, J. E. (2009). *Manual para la evaluación inicial de la lectura en niños de educación primaria*. Oficina de Desarrollo Económico, Agricultura y Comercio (EGAT/ED).

Kartal, G., & Terziyan, T. (2016). Development and evaluation of game-like phonological awareness software for kindergarteners. *Journal of Educational Computing Research*, *53*, 519–539. https://doi.org/10.1177/0735633115608397

Kim, Y. S. (2012). The relations among L1 (Spanish) literacy skills, L2 (English) language, L2 text reading fluency, and L2 reading comprehension for Spanish-speaking ELL first grade students. *Learning and Individual Differences*, *22*(6), 690–700. https://doi.org/10.1016/j.lindif.2012.06.009

Kyle, F., Kujala, J., Richardson, U., Lyytinen, H., & Goswami, U. (2013). Assessing the effectiveness of two theoretically motivated computer assisted reading interventions in the United Kingdom: GG rime and GG phoneme. *Reading Research Quarterly*, *48*(1), 61–76. https://doi.org/10.1002/rrq.038

Mayer, R. (2001). *Multimedia learning*. Cambridge University Press.

Nation, K. (2019). Children's reading difficulties, language, and reflections on the simple view of reading. *Australian Journal of Learning Difficulties*, *24*(1), 47–73. https://doi.org/10.1080/19404158.2019.1609272

Nation, K., & Snowling, M. J. (1998). Semantic processing and the development of word-recognition skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language*, *39*(1), 85–101. https://doi.org/10.1006/jmla.1998.2564

Oxman, S., & Wong, W. (2014). *White paper: Adaptive learning systems*. Integrated Education Solutions.

Paivio, A. (1986). *Mental representations*. Oxford University Press.

Pellicer Sanchez, A., & Siyanova-Chanturia, A. (2018). Eye movements in vocabulary research: Current findings and future directions. *ITL - International Journal of Applied Linguistics*, *169*(1), 5–29.

Pieneman, M. (1989). Is language teachable? Psycholinguistic experiments and hypotheses. *Applied Linguistics*, *10*(1), 52–79. https://doi.org/10.1093/applin/10.1.52

Proctor, C. P., August, D., Carlo, M., & Snow, C. (2005). Native Spanish-speaking children reading in English: Toward a model of comprehension. *Journal of Educational Psychology*, *97*(2), 246–256. https://doi.org/10.1037/0022-0663.97.2.246

Puimège, E., & Peters, E. (2019). Learners' English vocabulary knowledge prior to formal instruction: The role of learner-related and word-related variables. *Language Learning: A Journal of Research in Language Studies*, *69*(4), 943–977. https://doi.org/10.1111/lang.12364

Sampayo-Vargas, S., Cope, C. J., He, Z., & Byrne, G. J. (2013). The effectiveness of adaptive difficulty adjustments on students' motivation and learning in an educational computer game. *Computers and Education*, *69*, 452–462. https://doi.org/10.1016/j.compedu.2013.07.004

State of Florida Department of Education. (2009). *Florida assessments for instruction in reading: Ongoing progress monitoring oral reading fluency grades 1-5*. Blackline Master.

Streicher, A., & Smeddinck, J. (2016). *Personalized and adaptive serious games. Entertainment computing and serious games*. Springer. https://doi.org/10.1007/978-3-319-46152-6_14

Tsatiris, G., & Karpouzis, K. (2020). *Developing for personalized learning: The long road from educational objectives to development and feedback*. Paper presented at the workshop 'Technology-mediated personalized learning for younger learners: Concepts, methods and practice' at the IDC Conference, London.

Vaala, S., Ly, A., & Levine, M. H. (2015). *Getting a read on the app stores: A market scan and analysis of children's literacy apps*. The Joan Ganz Cooney Center at Sesame Workshop.

van de Ven, M., de Leeuw, L., van Weerdenburg, M., & Steenbeek-Planting, E. G. (2017). Early reading intervention by means of a multicomponent reading game: Intervention affects reading game. *Journal of Computer Assisted Learning*, *33*(4), 320–333. https://doi.org/10.1111/jcal.12181

Vanbecelaere, S., Berghe, K., Cornillie, F., Sasanguie, D., Reynvoet, B., & Depaepe, F. (2019). The effectiveness of adaptive versus non-adaptive learning with digital educational games. *Journal of Computer Assisted Learning*, *36*(4), 502–513. https://doi.org/10.1111/jcal.12416

Vanbecelaere, S., Van den Berghe, K., Cornillie, F., Sasanguie, D., Reynvoet, B., & Depaepe, F. (2020). The effects of two digital educational games on cognitive and non-cognitive math and reading outcomes. *Computers & Education*, *143*, 103680. https://doi.org/10.1016/j.compedu.2019.103680

Vanbecelaere, S., Vasalou, A., Cornillie, F., Mavrikis, M., Depaepe, F., & Gilabert, R. (2020). *Technology-mediated personalized learning for younger learners: Concepts, methods and practice*. Workshop at the IDC conference, London, UK.

Vandewaetere, M., Desmet, P., & Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior*, *27*(1), 118–130. https://doi.org/10.1016/j.chb.2010.07.038

How to cite this article: Serra, J., & Gilabert, R. (2021). Algorithmic versus teacher-led sequencing in a digital serious game and the development of second language reading fluency and accuracy. *British Journal of Educational Technology*, 52, 1898–1916. https://doi.org/10.1111/bjet.13131