


Designing for “challenge” in a large-scale adaptive literacy game for primary school children

Laura Benton¹ | Manolis Mavrikis¹  | Asimina Vasalou¹ |
Nelly Joye² | Emma Sumner³ | Elisabeth Herbert³ |
Andrea Revesz⁴ | Antonios Symvonis⁵ | Chrysanthi Raftopoulou⁵

¹UCL Knowledge Lab, UCL Institute of Education, University College London, London, UK

²Speech and Language Therapy, School of Health and Social Care, University of Essex, Colchester, UK

³Psychology and Human Development, UCL Institute of Education, University College London, London, UK

⁴Centre for Applied Linguistics, UCL Institute of Education, University College London, London, UK

⁵Department of Mathematics, School of Applied Mathematical and Physical Sciences, National Technical University of Athens, Athens, Greece

Correspondence

Laura Benton, UCL Knowledge Lab, UCL Institute of Education, University College London, London, UK.
Email: l.benton@ucl.ac.uk

Funding information

H2020 European Institute of Innovation and Technology, Grant/Award Number: 731724

Abstract

The use of learning games within the classroom is becoming increasingly common because of their potential to positively impact learning. Recent developments in adaptivity offer further possibilities to personalise learning by tailoring the game to an individual child's level or particular learning needs. However, designing an adaptive learning game is a complex process as many different game components have an impact on the provision of optimal challenge, crucial for maintaining player engagement, with limited prior work considering the multifaceted nature of this concept. This paper explores how to design for “challenge” within large-scale adaptive learning games through a case study focused on the design of a literacy game for three linguistically and cognitively diverse learner groups—novice readers, children with dyslexia and children learning English as a foreign language. In reflecting on our design process, we identify three key design tensions that arose: (a) supporting longer-term learning goals through game replayability; (b) fostering either replication or

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *British Journal of Educational Technology* published by John Wiley & Sons Ltd on behalf of British Educational Research Association

innovation in pedagogy through adaptivity rules; and (c) addressing diversity between learner groups. We present a set of design recommendations to guide researchers and designers in taking a multidimensional view of challenge when designing large-scale adaptive learning games.

KEYWORDS

adaptivity, challenge, children, design, games-based learning, literacy

Practitioner notes

What is already known about this topic?

- Adaptive learning games can have a positive impact on children's learning outcomes.
- Ensuring optimal challenge within games is important for maintaining engagement.
- Designing adaptive learning games is a complex process.

What this paper adds?

- Designing for optimal challenge within adaptive learning game should be considered as a multifaceted concept.
- Identification of key tensions related to optimising challenge that can emerge during the design of large-scale adaptive learning games.
- Recommendations for adaptivity researchers and learning game designers for how to address these tensions in adaptive learning game design.

Implications for practice and/or policy?

- We need a more systematic approach to adaptivity game design to ensure wider spread adoption.
- Learning game designers seeking to utilise adaptive components in designing for optimal challenge should consider a focus on learners who may require a more targeted approach.
- Adaptive learning games offer opportunities for pedagogical innovation in the classroom through exploiting innovative game features as well as large-scale data collection to support adaptive learning over time.

INTRODUCTION

The use of learning games is growing because of their potential to positively impact students' learning in several ways including cognitive outcomes (Clark et al., 2016; Zeng et al., 2020) and fostering motivation and persistence (Ke, 2016; Shute et al., 2013). Learning game designers are increasingly exploiting developments in adaptivity to provide a more personalised learning experience (Malva et al., 2020; Vanbecelaere et al., 2020), allowing games to be tailored to student differences in knowledge, abilities and disabilities as well as demographics, sociocultural background and affect (Shute & Zapata-Rivera, 2012). This enables a learning game to be targeted to diverse and/or multiple learner groups.

Achieving a state of “flow” is an overarching aim of games (Csikszentmihalyi & Csikszentmihalyi, 1988; Ke, 2016), and the provision of optimal challenge within the game is crucial to this end, through matching the game difficulty and the players' skill level (Denisova et al., 2017; Hamari et al., 2016; Kim & Ruipérez-Valiente, 2020). Although the terms challenge and difficulty can often be used interchangeably, they are interrelated yet distinct concepts—*difficulty* can be defined as the “probability of task failure” (Lomas et al., 2017), whereas *challenge* is a more nuanced and harder to define concept based on player perception of effort and experience ie, how difficult do they find the game, which depends on a player's particular skills and prior expertise (Denisova et al., 2020). Within the context of learning, tasks that are too difficult can result in cognitive overload whereas simple tasks may lead to feelings of boredom and disengagement (Hendrix et al., 2018; Shute et al., 2013). Player engagement within a game can be maintained by balancing game difficulty with player skill (Denisova et al., 2020); however, learning game design requires an in-depth consideration of optimal challenge. For example, the requirement for a particular interaction method like fast-tapping or dragging, the addition of a constraint such as a time limit, the choice of presentation format or order of the learning content can all impact the task difficulty, which raises the question: what should be adapted—the game mechanics, learning content or both?

As we review in more detail below, adaptivity has the potential to modulate various aspects of a learning game to provide an optimal challenge for players through making dynamic difficulty adjustments that foster learner engagement (Hamari et al., 2016). However, prior work rarely focuses on how difficulty is operationalised within these games (Klinkenberg et al., 2011) and is often limited to one-dimensional measures that lack nuance within difficulty adjustments (Gallego-Durán et al., 2018). Furthermore, decisions related to measuring progress and adjusting the difficulty level accordingly require the translation of (often ambiguous) learning theories and teaching practices (Lam, 2000; Nutley et al., 2003) into a concrete computational representation.

This paper examines how to design for optimal challenge within a large-scale adaptive learning game through a reflective case study focused on the design of a literacy game for three linguistically and cognitively diverse learner groups—novice readers, children with dyslexia and children learning English as a foreign language (EFL). These three groups were selected as they share a focus on learning to read in English but can also encounter different challenges during this process because of their varied skill levels and experiences. Whilst adaptivity offers the opportunity to modulate difficulty for diverse groups, this diversity can also introduce new questions as to how optimal challenge is operationalised. We describe our design approach and discuss the representation of challenge within this context, highlighting the different design tensions this raised. We conclude with a set of design recommendations for adaptivity researchers and learning game designers that provide guidance on how to utilise adaptive components to support optimal challenge for diverse groups of learners within large-scale learning games.

BACKGROUND

Defining adaptivity in learning games

While there is no agreed definition of adaptivity in educational technology, previous frameworks characterise adaptivity at different levels of abstraction depending on their aim or the community where they emerge from (Holstein et al., 2020). Most definitions of adaptivity consider *what* adaptive systems respond to and *when*, *how* and *why* they adapt ie, they refer to the extent a digital system responds to learners' actions and/or cognitive or other

psychological measures by providing appropriate modification or other forms of support (Aleven et al., 2017; EdSurge, 2016; Plass & Pawar, 2020; Scandura, 2014).

In positioning this work, we rely on Aleven et al. (2017) who expand the above definition of adaptivity by acknowledging that a system can be adaptive “by design” ie, responsive to the learning domain demands for different learner groups. EdSurge (2016) also offers a useful taxonomy in relation to *what* to adapt (ie, content, assessment and/or sequencing), which could be mapped to games especially if within sequencing we also consider the game mechanics.

In terms of *when* and *how* to adapt, while many definitions refer to real-time adaptivity (EdSurge, 2016), Aleven et al. (2017) propose that a system may also need to adapt over a long time span based on information maintained over a series of interactions. To achieve this, adaptive systems typically base diagnosis and adaptation selection procedures on several models including:

- a learner model (eg, cognitive or affective characteristics of novice/dyslexic/EFL readers);
- a knowledge domain model (eg, the reading skills to be acquired); and
- an instructional model (eg, reading skill acquisition theories and strategies that foster learning) (Zarraonandía et al., 2016).

This enables different learner profiles and characteristics to be taken into account. For example, the instructional model could be adapted to follow specialist dyslexia teaching strategies for learners with a dyslexia user profile, or the knowledge domain could exclude less relevant reading skills for learners with an EFL user profile, such as phonics-related skills that may have already been acquired in a native language.

Challenge and difficulty in learning games

Within adaptive learning games, the player learns a new concept or acquires a new skill through undertaking a series of learning tasks that progressively increase in difficulty (Plass & Pawar, 2020). This is achieved using several techniques, most notably some variation of statistical models that fall broadly under item response theory (Jones & Thissen, 2006). More recently, the Elo rating system originally used to rank and match chess players (Elo, 1978) and Bayesian Knowledge Tracing (Corbett & Anderson, 1994) are increasingly used in educational games (Hou et al., 2021; Klinkenberg et al., 2011; Vanbecelaere et al., 2020). For reviews and comparisons of these techniques, we refer the reader to McLaren and Nguyen (2021) Pelánek (2016) and Wauters et al. (2012). As a result of using such approaches, the sequence of games can be reorganised in response to the learner's specific needs. To this end, various game elements such as the mechanics/rules and narrative/characters as well as learning content and instructional supports each have the potential to be designed in an adaptive way to increase or decrease the task difficulty level (Lopes & Bidarra, 2011; Peirce et al., 2008; Streicher & Smeddinck, 2016) and therefore optimise the challenge experienced by the player.

However, in addition, the interaction *between* these game elements (Zarraonandía et al., 2016) can also impact the overall challenge. For example, the player may experience increased challenge if both the game mechanic and learning content are unfamiliar, rather than new learning content with a familiar mechanic. Therefore, it is important to consider both individual game elements as well as their interaction with one another when designing for optimal challenge (Sampayo-Vargas et al., 2013). Limited research has focused specifically on the broader learning task difficulty as part of the adaptation process, because

assessing and defining difficulty is inherently complex (Gallego-Durán et al., 2018). Because of this complexity, existing games often predefine difficulty as a series of discrete levels that the player moves up or down, or focus on adjusting the difficulty of a single game element (Gallego-Durán et al., 2018; Plass & Pawar, 2020; Yang et al., 2020), rather than independently and dynamically adapting distinct game elements, which give more flexibility to adapt to individual needs.

The choice of knowledge domain within an adaptive learning game can have an overarching impact on how best to optimise challenge within the game. Within the domain of literacy, reading is a cognitively demanding skill requiring the coordination of multiple interrelated processes, including word decoding, accessing vocabulary knowledge, and syntactic and semantic processing to aid comprehension (Oakhill & Cain, 2012; Torgesen, 2000; Verhoeven & Van Leeuwe, 2008). This entails initial isolated reading skills to be systematically practiced and measured to determine “mastery” (Afflerbach et al., 2008) before progressing to more advanced aspects of the reading process. However, the effort different learners require to undertake a particular learning activity can vary greatly depending on their specific characteristics (Yang et al., 2020) eg, a child with dyslexia can struggle with “easier” reading skills such as phonics but have greater proficiency in more advanced areas such as prefixes. A first step in establishing mastery is to align content with expectations for the age/stage of the learner and subsequently to set a specific assessment threshold for moving on in the game. However, although a general 80% success rate for progression thresholds has been proposed (Bloom, 1968; EEF, 2018), there is limited research investigating these thresholds as they apply to specific learner profiles.

Designing for optimal challenge

Carro et al. (2002) propose a methodology that supports the design of adaptive learning games. They highlight that separating the activities and the supporting game mechanics allows the learning environment to be better tailored to the individual user and also provides opportunities for reuse in other environments. However, while their design methodology supports a move toward operationalising difficulty, it focuses on the difficulty levels of individual learning activities within the game without recognising the impact of other components, such as the difficulty introduced by the content, or considering the overall challenge within the game.

Within the domain of literacy, there is a lack of empirical evidence that identifies difficulty in reading at a more granular level, such as individual letter–sound relationships (eg, reading the “c” as in “cat” vs. “c” as in “city”), and although teaching order is suggested in existing teaching schemes (eg, Letters and Sounds in the UK), the evidence informing this is unclear. For those learners who do not progress linearly following the pre-defined curriculum, these elements are all the more important to foster motivation and progress (Grigorenko, 2009; Rose, 2006). In addition, the required knowledge to specify item-level difficulty to follow statistical approaches such as item response theory, Elo and Bayesian Knowledge Tracing mentioned earlier is either in tacit form and difficult to obtain (Lam, 2000; Nutley et al., 2003) and for experts to externalise (Cooke & McDonald, 1986; Porayska-Pomsta et al., 2013) or requires previously collected data that may not be readily available (Wauters et al., 2012). This lack of empirical evidence as well as a need to consider difficulty at a more granular level for different learner groups suggests a need to draw more on pedagogical expert knowledge and knowledge elicitation (KE), which seeks to access, understand and represent the tacit knowledge of the domain (ie, *what*) and how it is taught (ie, *how* and *when*) (Porayska-Pomsta et al., 2013).

NAVIGO: DESIGNING A LARGE-SCALE ADAPTIVE LITERACY GAME

The focal game of this paper, *Navigo*, was designed and developed as part of a European-funded Innovation project over a 2-year period. The aim of the game is to support primary school children's acquisition of reading skills. Owing to the similar learning aims underpinning their reading acquisition, the game was designed for a range of learner groups: (a) children learning to read (ages 5–7); (b) older children (ages 8–11) with dyslexia; and (c) children who are outside of the UK learning EFL. To cater to the diverse language and cognitive profiles of each group, an adaptive approach was taken to the game to optimise challenge both within and across groups.

The game was collaboratively designed by several universities and a game design company. The academic team included linguistic experts, interaction designers and computer scientists. The game design team included graphics and animation artists, game developers and educational content experts. Additionally, we involved both teachers and students as informants during the design process to inform particular aspects of the game (Guha et al., 2013). This process included the following broad stages (discussed further in the following sub-sections):

1. Specification of the **learning content** (ie, reading skills to be learned within the game) by linguistic experts, involving development of a language domain model and associated content.
2. Iterative design process to develop appropriate and engaging **game mechanics and activities** to enable the practice of these reading skills, informed by linguistic experts, interaction designers, game designers and teachers/students.
3. Knowledge elicitation workshops with computer scientists and linguistic experts to define the **instructional approach** to teaching these skills through a set of adaptivity rules.

The remainder of this section describes in more detail the design rationale of *Navigo* to help contextualise our subsequent reflections on designing for optimal challenge within adaptive learning games.

Learning content

In developing *Navigo*, we first defined the set of linguistic phenomena and features required to become a fluent reader. This set, and the connections between its elements, is often referred to as a “domain model” and is one of the first steps in building adaptive, intelligent learning environments (cf., Mavrikis & Holmes, 2019; Plass & Pawar, 2020; Sottolare et al., 2013). Six linguistic levels were identified: *phonology*, *word recognition*, *orthography*, *morphology*, *morphosyntax* and *syntax*. A linguistic level is represented by several language phenomena or structures, called language *categories*, each including a set of specific instances ie, the *features*, with related word/sentence examples (see Table 1 for an overview). In addition to incorporating the knowledge required to learn to read, *Navigo*'s domain model incorporates a pedagogical model informing progression within the hierarchy modelled as a directed acyclic graph.¹

For each linguistic feature, two types of information are stored: (a) a numerical rating of relative difficulty (in relation to the other features in a linguistic level) defined by the linguistic experts based on existing curricula materials and their own expertise and (b) prerequisite feature(s)—ie, features to be mastered prior to making this feature available within the game. Table 2 presents an example of the prerequisites and difficulty for one language feature, the initial consonant blend “thr”, which was rated 2 (with 1 being the easiest) as it is a slightly

TABLE 1 Domain model levels, categories and example features with total number in parenthesis (see deliverable^a for full details)

Linguistic level (6)	Language category (26)	Example feature (279)
Phonology	Phonics	“s” as in sad
	Blends	“sl” as in “slap”
	Syllables	Chunking two-syllable words
Word recognition	Common sight words	Frequent words 1—Reception (“a”, “and”, etc.)
Orthography	Confusing letters	“d” and “b”
Morphology	Prefixes	“mono”, “multi”, etc.
	Suffixes	Quickly, loudly
Morphosyntax	Adverbs	More slowly
	Verb tenses	“s” as in “he plays”
Syntax	...	
	Adjectives	“nice” as in “a nice dress”
	Pronouns	“each other” as in “they like each other”
	...	

^aFor full details, see project deliverable: https://ireadprojecteu.files.wordpress.com/2018/04/iread_d4-2_final_incl_appendices.pdf.

TABLE 2 Feature prerequisites and difficulty

Linguistic level	Language category	Example feature	Difficulty of feature (1–4)	Prerequisites	Word-level complexity ^a (1–8)
Decoding words	Blends	“thr” as in “throw”	2	“t” as in “tap”; “h” as in “hat”; “r” as in “rat”	2

^aEach word in our dictionary is associated with a positive integer value (1–8), indicating the “word’s difficulty”. This classification is a combination of several attributes, such as the number of syllables or the frequency of a word in child texts. For example, words with difficulty 1 (the simplest) are monosyllabic words, words with difficulty 2 are two-syllable words with specific consonant–vowel patterns, etc.

more complex feature compared to eg, individual letter–sound features, within the linguistic level. The defined prerequisites were used by the adaptive sequencing to determine the next feature selection (see Instructional Approach). Finally, word-level complexity indicates the complexity of the words that should be selected to practise this feature.

Differences existed with respect to the relevance of the domain levels and categories for the three learner groups. Whilst novice readers follow a typical progression mapped out in various models such as the dual-route model (eg, Brown & Deavers, 1999; Ehri, 2017), children with dyslexia can experience a range of reading difficulties, including the reading of irregularly spelled words, novel or nonwords and making derivational errors (eg, reading “performing” as “performance”) or visual errors (eg, reading “perform” as “perfume”) (Harley, 2013). For EFL learners, oral and written skills are typically learned at the same time, which means that children cannot rely on the same strategies as native speakers. There are also cross-linguistic influences on this group depending on their first language. Given the pedagogical information integrated within the domain model, these differences were addressed by creating a unique language domain model for each learner group, defining initial mastery level pre-sets for individual language features based on the characteristics of each group.

Following this, content was designed for each domain model language feature. For features relating to word-level reading skills, a *child-appropriate word dictionary* was created containing 12,000 words. The dictionary was automatically tagged with the domain model language features (Vasalou et al., 2021). For features relating to sentence-level reading skills, a *child-appropriate set of sentences* was authored for each syntax language features, with 10 sentences per feature totalling 1000 sentences.

Game mechanics and activities

Within Navigo, the child is the central character, and the mission is to save his/her grandma who has become lost in a pyramid after a storm hit her village. The game is set in Ancient Egypt, chosen because of its association with archaeological adventure, the discovery of ancient treasures and the decryption of languages. The game aesthetics reinforce the focus on literacy, for example by depicting hieroglyphs in the environment.

Whilst moving between rooms within the pyramid, the child encounters a series of game activities structured around a range of mechanics. These activities and mechanics were developed following an iterative design process, which incorporated three broad phases (Figure 1). The first iteration involved a number of design workshops evaluating an early game prototype with both linguistic/interaction design experts and separately with five primary teachers. The experts focused on establishing which of the proposed mechanics could be used to practise the different skills within the domain model. After trialling the game activities, the teachers were asked a series of questions about the appropriateness of these activities and mechanics for the target user group as well as the classroom context.

The second iteration involved linguistic experts reviewing the mechanics to ensure full domain model coverage. When gaps were identified, new mechanics were proposed using existing paper-based learning activities as a starting point. Additionally, children (ages 6–8) evaluated a refined version of the game prototype following the problem identification picture cards method (Barendregt et al., 2008) to ensure that the game mechanics were fun and engaging.

The final iteration involved linguistic/interaction design experts and children playtesting the fully developed game prototype to identify usability bugs as well as content errors. The problem identification picture cards method was followed with the children.

The final version of Navigo incorporates 15 different Ancient Egyptian-themed game mechanics and is designed to accommodate content that draws from domain model features in both words and sentences. These mechanics include multiple choice, matching, splitting, sequencing, hit the target and fill in the gaps (see Figure 2 for examples). Combined with the 279 language domain model features and content, the mechanics are used across 900 potential unique game activities.

Instructional approach

The underlying instructional approach within the game was implemented through a series of adaptivity rules informing how Navigo chooses language features from the domain model, learning content and game mechanics.² Each rule is informed by a pedagogical rationale that emerged from our KE workshops with reading development experts and linguists and supported with existing empirical evidence. The rules make adaptations to the game to ensure:

- the learner is at an *appropriate level*, the *revision* and *reinforcement of content* is supported as well as encouraging *continued motivation* (eg, through ensuring appropriate level and variation in feature types);

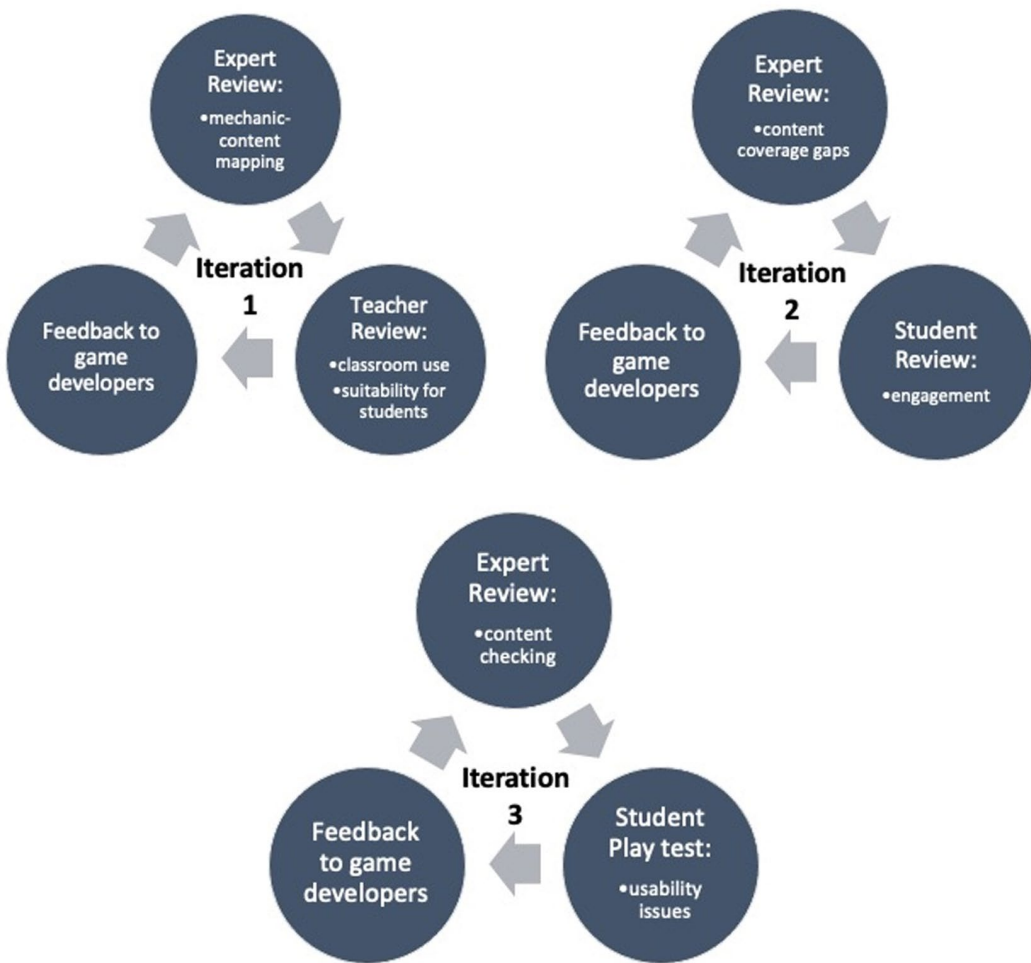


FIGURE 1 Overview of game activity and mechanic iterative design phases

- the *diversification of language content* as well as the *choice of words* and features within the game activities consider learner group difficulties to allow additional practice, support the gradual building of confidence and avoid additional cognitive load; and
- the selected mechanics support the move from *declarative to procedural* and then *automatised knowledge* to achieve reading fluency as well as mitigate additional cognitive load when *new knowledge* is introduced.

These rules were represented through an adaptive sequencing algorithm, which we present below to illustrate how the different algorithm phases were implemented. Note that we use and extend the EdSurge (2016) taxonomy to additionally include “game mechanic selection” under the analysis phase to cover the adaptive learning game context:

1. Collect data

- **Type:** academic performance data (% correct answers).
- **Granularity + Difficulty:** discrete skill (ie, language feature) and difficulty level (numerical rating).
- **History:** learner's profile over time (current mastery level and previously played features).



Perilous Paths

Multiple choice game mechanic

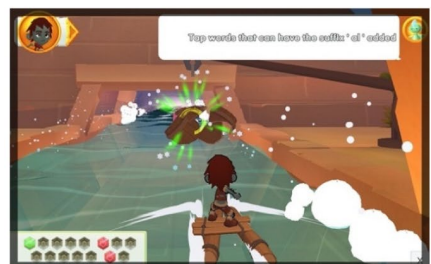
Choose the bridge containing the word that correctly completes the sentence.



Pillar Pusher

Sequencing game mechanic

Drag the pillars in the correct order to make the word that you hear.



Raft Rapid Fire

Hit-the-target game mechanic

Hit the barrels labelled with words that can have a particular suffix added to them.

FIGURE 2 Three examples of Navigo game mechanics

2. Analyse data

- **Learner analysis:** applying thresholds of mastery (note that in line with the expert suggestions from our KE workshop, feature mastery is updated based on performance across three games, to validate initial successes or failures (Tsatiris & Karpouzis, 2021)).
- **Skill selection:** number of options (prioritised by no. times feature has been practised multiplied by difficulty rating—then ordered with lowest priority rating selected first).
- **Game mechanic selection:** no. options (filtered by selected feature, prioritise low-difficulty mechanics for new skill/low mastery level and high-difficulty mechanics for higher mastery level).
- **Content analysis:** skill selection (content must contained selected feature), learner history (word-level content only—words are selected based on the language feature difficulty as well as learner familiarity and past performance with language feature eg, tricky words are given priority to practise again).

3. Adjust content

- **Delivery:** automatically assigns content to player.
- **Amount:** group of content (series of game activities).
- **Design:** independent (selects from a range of language categories based on feature availability).

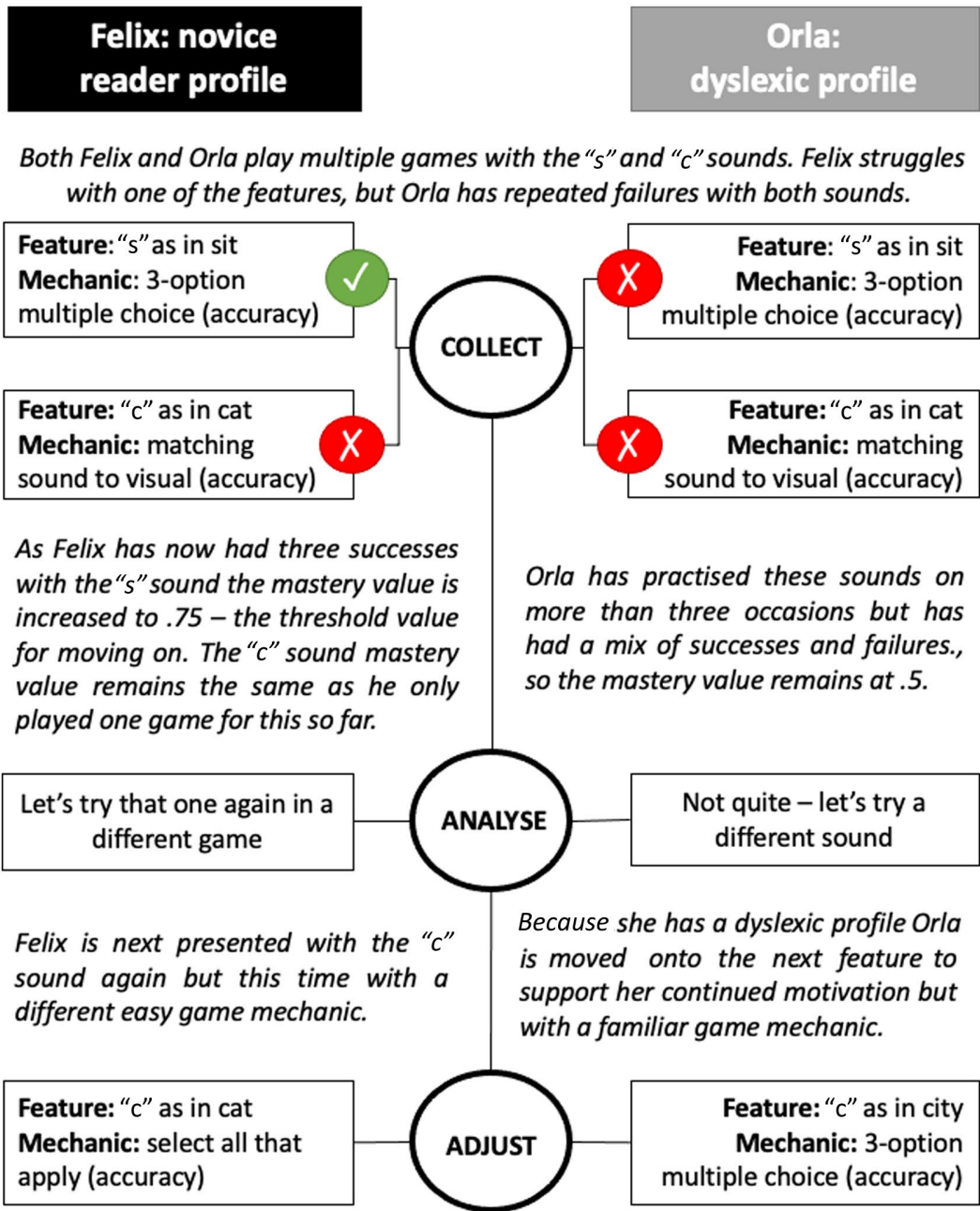


FIGURE 3 Example of an adaptive sequencing rule for novice reader and dyslexic profiles

This algorithm allowed for particular adaptations to be made for specific user profiles. For example, children with dyslexia could move on in the game without achieving full feature mastery to foster continued motivation, as continuous failures in achievements and an insistence on inadequacies can impact motivation (Bandura, 1982), thus providing a more flexible learning pathway that reflects their diverse profiles. Figure 3 demonstrates this through a comparison of the adaptive sequencing algorithm in action for two different learner profiles—novice reader and dyslexia (and based on EdSurge (2016) taxonomy). This

example illustrates how the game assesses a child's current mastery of each feature and subsequently makes decisions about which feature/game mechanic to present next. Note that the threshold to move on is lower for learners with a dyslexic profile, but they still need to achieve one to two successes to unlock proceeding features. They will also be given opportunities to revisit these earlier features to ensure they do not progress too far without acquiring foundational skills.

DESIGN TENSIONS

We faced several tensions through the process of designing for challenge within our adaptive game; below, we present our reflections around these tensions under three key themes.

Designing for replayability

Navigo is a large-scale game designed to be used over an extended period of time to improve children's reading performance. Therefore, it was important to have a sufficient variety in the learning activities to increase language exposure to foster reading fluency as well as provide novelty, which is important for maintaining engagement and motivation over an extended time period (Lomas et al., 2017). Moreover, compared to an equivalent non-adaptive game, in an adaptive game it is necessary to foster variations in learning pathways, thus requiring a larger amount and variety of game mechanics and language features, a goal we broadly refer to as “replayability”.

Given the high number of games and language features incorporated in Navigo, it was necessary to consider the optimisation of challenge within each playable game activity by taking a systematic approach to the definition of difficulty. In the domain model, difficulty was encoded at a fine-grained level within each reading skill (ie, *language feature*), which was used to define a learning progression across the 279 features contained in the model. Difficulty was also encoded in the *game mechanic* eg, a six-option multiple-choice mechanic was defined as more difficult than three options. Furthermore, to demonstrate mastery of a language feature, it is important to be exposed to it (thus practise it) multiple times and in different contexts to promote skill transfer (Carroll et al., 2011; Hattie, 2008; Perkins & Salomon, 2012) before moving on. Therefore, we systematically defined sequencing through selecting different *game activity* types (groups of similar mechanics defined as accuracy, building or automaticity game activities) for each language feature. This allowed progressive learning as well as provided an opportunity for children who struggle with a particular feature to replay the same feature without it feeling repetitive, nor impacting motivation. This modularised approach to optimising challenge enabled us to generalise our approach to the 900 game activities. In turn, the adaptivity rules prioritised language features and content from the available features in the learning progression. The same rules ensured that children always started with a familiar game mechanic when working on a new language category. Additionally, when a child began to play a new language feature, the adaptivity rules first chose an accuracy game activity, followed by a building game activity, before presenting an automaticity (timed) game activity once their mastery was secure. This reflects a common approach to adaptive learning game design, which is to integrate a series of mini-games that each focus on a specific skill or concept into an overarching narrative, allowing the challenge to be more easily optimised for each learner therefore fostering replayability (Zarraonandía et al., 2016).

When we started to combine *language features* from the domain model, and *game mechanics* to produce playable *game activities*, our modularised approach to optimising challenge revealed new considerations and tensions. Games that addressed language features

at a word-level (eg, decoding and morphology) drew content from the child-friendly 12,000-word dictionary we developed. This provided the advantage of exposing children to a diversity of content over existing literacy games, which frequently rely on short word lists. However, we quickly realised that some words were not suitable in light of the child's level in the learning progression (eg, too long/complex). This contributed to an additional encoding of word difficulty when specifying *game activity* parameters. Consequently, we adjusted the difficulty level of the target or distractor words in the game eg, varying word length/syllables (Morrison & Ellis, 1995), excluding more advanced language features (Carlisle & Stone, 2005) and ranking words based on their frequency (Hiebert et al., 2019). In summary, in contrast to previous approaches in adaptive games, which have viewed challenge one-dimensionally, modulated through discrete levels of difficulty (Gallego-Durán et al., 2018; Plass & Pawar, 2020; Yang et al., 2020) within Navigo challenge emerged to be a multifaceted construct described by an interaction between the extensive language features and game mechanics brought together in the game activity, which enabled fine-grained adjustments to balance player skill and difficulty as well as sustaining novelty. Below, we summarise the different dimensions that contributed to replayability:

- **Language feature:** Different language features were assigned a relative difficulty rating based on the curricula and linguistic expert assessment, enabling tailored progression through the learning content as well as supporting novelty by opening multiple features at the same difficulty level.
- **Game mechanic:** Different mechanics were also assigned difficulty ratings. For example, one easy mechanic was a simple three-option multiple choice whereas a more difficult mechanic was a drag and drop sequencing mechanic, which required six pillars to be correctly ordered and positioned. Challenge could be optimised by considering the familiarity of the language feature and combining it with an appropriate mechanic. This also maintained novelty by allowing similar mechanics to be interchanged for a given language feature providing additional practice opportunities.
- **Game activity:** Groups of mechanics were defined as Accuracy (eg, select a feature/word), Building (eg, construct a word/sentence) and Automaticity (eg, select a feature/word within a given time), with Accuracy mechanics providing the simplest learning activities and Automaticity mechanics the most difficult. This provided a way to optimise challenge through balancing player skill with the selected language feature with the difficulty of the overall learning activity.
- **Word-level content** (correct answers and distractors): Words selected as correct answers as well as distractors were adapted to provide easier words when introducing a feature by restricting the length eg, to less than six letters for Accuracy and Building game activities as well as selecting words without prefixes/suffixes and more frequent words. In contrast, for Automaticity game activities, length restrictions were removed and less frequent words were incorporated to increase the level of difficulty for students who had first demonstrated mastery in the Accuracy/Building game activities. This also maintained novelty in the language that the players encountered.

Designing for pedagogical replication or innovation

Throughout the design process, we encountered tensions around which aspects of reading pedagogy to replicate as adaptivity rules. During the KE workshops, the pedagogical experts often struggled with the mental effort and abstraction required to keep in mind how various components of the game were interwoven. They also struggled to see the relationships, priorities and overlaps between different rules. Recognising this design problem and

also the need to maintain a manageable implementation, Plass and Pawar (2020) highlight the importance of considering the specific learning objectives of the game and focusing on a small number of variables to adapt backed up with empirical evidence demonstrating beneficial impact on learning outcomes. This tension, therefore, resulted in a theory-driven focus to identify available evidence that could inform adaptations in difficulty to provide optimal challenge in the game and beneficial learning outcomes for each group.

A further tension was the lack of specific expertise available per se in, for example, adaptive game sequencing. Therefore, the “knowledge” required to inform such sequencing needed to be acquired from pedagogical experts and adapted through an iterative process of implementation and testing. This problem is known as the “bottleneck” of knowledge engineering in the KE process. Within the context of education, the issue may be exacerbated because of teaching practice often relying on intuition and hard to codify a priori (Hewitt et al., 2003; Lin et al., 2005). To overcome this, we developed a paper-based template to scaffold an interdisciplinary discussion that promoted (a) the need for connecting pedagogical rationale to empirical evidence and (b) the importance of designing computational representations of pedagogical rationale that relied on the logic of the Navigo games and underpinning technical infrastructure, allowing us to define difficulty within the context of the game activity.

Although our design approach had an initial focus on replicating existing pedagogical practices, in a second step large-scale data collection from adaptive game use could offer the opportunity to provide insights and greater nuances into designing optimal challenge for different learners, enabling innovation in pedagogy.

Designing for different diverse learner groups

Targeting three linguistically and cognitively diverse learner groups presented tensions in how we viewed challenge within the game from several perspectives. The breadth of the language domain model allowed us to address a wide range of ages and skills. However, it also introduced the requirement to place children at an appropriate starting point. As such, we faced the common AI problem of “cold-start” (Drachler et al., 2008) ie, to make informed recommendations data should have been collected in advance, and even if that were the case (eg, if our game had been used extensively), any new child would not be known to the system. A common approach to address this problem is an initial placement assessment (eg, a dynamic quiz when a student signs up); however, the cost and complexity of assessing each child led us to represent children’s existing knowledge within their user model through pre-set mastery values for each language feature. Reflecting on this approach, it is worth highlighting that it was more straightforward in some cases than others. For novice readers, our decisions were guided by the detailed English National Curriculum, which indicated what skills and knowledge children were expected to have had exposure to in a given year group (DfE, 2014), and therefore the language features associated with these areas were set at the mastery threshold. Conversely, for EFL, a lack of empirical research led us to rely on linguistic experts within the project. We also found the need to reflect the learner group profiles in how mastery was used to open new language features in the progression. Given their learning differences, children with dyslexia often make smaller atypical increments in progress. Highlighting their consistent failure and focusing on inadequacies by revisiting the same learning content can negatively impact motivation (Bandura, 1982) as well as self-concept (Coleman & Hendry, 1999; Vygotsky, 1980). This led us to implement alternative triggers that would allow these children to access more advanced features even if they had not fully mastered “easier” features in the sequence. Future adaptive learning games may wish to consider a placement test for a more fine-grained approach to mastery, with the

precise thresholds at which different learners should move on an important area for further development.

Furthermore, in relation to the game mechanics, although we considered the overall challenge within the context of the language feature eg, choosing one out of three words was easier than selecting all words that apply, there was also the impact of the mechanic interaction mode eg, drag/drop or fast taps, which could increase the overall challenge for younger children whose motor skills are still developing (Hourcade, 2008). Another consideration was the choice of vocabulary, particularly within the sentence-level games, as it was necessary to ensure that the selected sentences were understandable without additional context by both native and non-native English speakers. Including these considerations within our approach to optimising challenge, which was already proving to be complex, was beyond the time and resource constraints of the project. Zarranondia et al. (2016) highlight the tension of maintaining game coherence in adapting too many different game features. Therefore, in the cases where adaptivity was not feasible or peripheral to the central learning goal, we followed the design principle of “designing for the baseline”, where we ensured that the game component was appropriate for the group that would experience the most challenge.

CONCLUDING THOUGHTS

This paper has reported reflections from the design process of a literacy game for three linguistically and cognitively diverse learner groups. We have identified tensions related to designing for optimal challenge in adaptive learning games, which include: (a) supporting longer-term learning goals through game replayability; (b) fostering either replication or innovation in pedagogy through adaptivity rules; and (c) addressing diversity between learner groups. Below, we highlight the implications of this work for future large-scale adaptive learning game projects and present a set of recommendations for researchers and game designers that provide key considerations when optimising challenge in adaptive learning games:

Conceptualise task difficulty

Within a large-scale multi-learner adaptive game, challenge is multifaceted, with the balance between player skill and task difficulty one aspect of this. Difficulty can be considered in relation to the target domain feature/skill, game mechanic, game activity and learning content (including both target and distractor items). Designers should be aware of these different facets and focus the adaptivity design on the subset most relevant to their central learning goal to ensure the manageability of the implementation (Zarraonandía et al., 2016).

Address complexity of multiple adaptations and interactions between them

Our modularised approach to optimising challenge within each game element was in keeping with our aim to produce a game at scale. This enabled us to cover the target age range and curriculum scope and introduce children to a wider range of word content as well as support novelty. However, despite this approach, the complexity still became hard to manage at times, indicating a need for an authoring tool to visualise the relationships between various game elements.

Designing for the baseline

As we have seen, it is not always possible to adapt multiple game components. Therefore, in adaptive learning games attempting to address diverse learner groups, we suggest for each distinct component identifying the group that will typically start with the lowest skill level in relation to this and to target the design to their level. This should ensure optimisation of challenge for the broadest population.

Target learners with atypical profiles/learning challenges

Our design work has highlighted the particular importance of adaptive technologies for learners with less predictable learning profiles, who would struggle more with common classroom technologies based on predefined sequences and limited nuance within their conceptualisation of challenge (Plass & Pawar, 2020). This indicates an opportunity for adaptive learning game designers to consider the needs of these populations in designing the adaptive sequencing algorithm, in particular, to maintain engagement and learning motivation eg, by providing additional opportunities to practise the same skill in a different way or varying mastery thresholds required to move on.

Identify opportunities for pedagogical innovation

Replication of existing pedagogy is not always possible or desirable. Adaptive games offer new approaches for optimising challenge within learning in ways that would not be an option within traditional classroom teaching as well as potentially identifying areas for improvement within existing curricula. For example, instant tailored feedback can be provided on any learner action through adaptive content (EdSurge, 2016). Similarly, game mechanics offer new ways of interacting and understanding learning content, which could impact the overall challenge in unexpected ways. Furthermore, designers may consider how dynamic mastery assessment should be and whether it should rely on children's learning history. This paper evidences the lengthy effort involved in developing the initial footprint of an adaptive system. Large-scale data can provide designers with new insights into difficulty and progression for different learner profiles, which can then be used to update initial assumptions within the adaptivity design as well as foster innovation in pedagogy. This highlights the need for long-term research that enables the consideration of adaptivity design in learning games over time (Aleven et al., 2017).

In conclusion, we argue that the identified tensions and recommendations presented above are transferable to future design, development and evaluation of other large-scale adaptive learning games. As adaptive learning games are becoming increasingly mainstream, we need a more systematic approach to adaptivity design to ensure wider spread adoption. We hope that this work will support this call by guiding adaptivity researchers and game designers to take a multidimensional view when designing for optimal challenge within adaptive learning games.

ACKNOWLEDGEMENTS

We would like to acknowledge the pivotal roles that our iRead partners from fishinabottle (Drew Wilkins, Adam Kaye, Tom Tsagkatos and Dawn Elliott) and the National Technical University of Athens (Kostas Karpouzis and Georgios Tsatiris) played in designing the Navigo game. We also thank the BJET reviewers for their helpful feedback and suggestions on earlier versions of this paper.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

DATA AVAILABILITY STATEMENT

No data are available in association with this paper.

ETHICS STATEMENT

This paper does not report on work involving human subjects.

ORCID

Manolis Mavrikis  <https://orcid.org/0000-0003-0575-0823>

ENDNOTES

- ¹ For full details, see project deliverable: https://ireadprojecteu.files.wordpress.com/2018/04/iread_d4-2_final_incl-appendices.pdf.
- ² For more details, see project deliverable: https://ireadprojecteu.files.wordpress.com/2020/03/iread_d4.4_user-adaptation-component.pdf.

REFERENCES

- Afflerbach, P., Pearson, P. D., & Paris, S. G. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher*, *61*, 364–373. <https://doi.org/10.1598/RT.61.5.1>
- Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2017). Instruction based on adaptive learning technologies. In *Handbook of Research on Learning and Instruction* (pp. 522–560). Routledge.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, *37*, 122–147. <https://doi.org/10.1037/0003-066X.37.2.122>
- Barendregt, W., Bekker, M. M., & Baauw, E. (2008). Development and evaluation of the problem identification picture cards method. *Cognition, Technology & Work*, *10*, 95–105. <https://doi.org/10.1007/s10111-007-0066-z>
- Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment*, *1*, 1–12.
- Brown, G. D., & Deavers, R. P. (1999). Units of analysis in nonword reading: Evidence from children and adults. *Journal of Experimental Child Psychology*, *73*, 208–242. <https://doi.org/10.1006/jecp.1999.2502>
- Carlisle, J. F., & Stone, C. A. (2005). Exploring the role of morphemes in word reading. *Reading Research Quarterly*, *40*, 428–449. <https://doi.org/10.1598/RRQ.40.4.3>
- Carro, R. M., Breda, A. M., Castillo, G., & Bajuelos, A. L. (2002). A methodology for developing adaptive educational-game environments. In *International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems* (pp. 90–99). Berlin, Heidelberg: Springer.
- Carroll, J. M., Bowyer-Crane, C., Duff, F. J., Hulme, C., & Snowling, M. J. (2011). *Developing language and literacy: Effective intervention in the early years*. John Wiley & Sons.
- Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research*, *86*, 79–122. <https://doi.org/10.3102/0034654315582065>
- Coleman, J. C., & Hendry, L. B. (1999). *The nature of adolescence*. Psychology Press.
- Cooke, N. M., & McDonald, J. E. (1986). A formal methodology for acquiring and representing expert knowledge. *Proceedings of the IEEE*, *74*, 1422–1430. <https://doi.org/10.1109/PROC.1986.13643>
- Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-adapted Interaction*, *4*, 253–278.
- Csikszentmihalyi, M., & Csikszentmihalyi, I. S. (1988). *Optimal experience: Psychological studies of flow in consciousness*. Cambridge University Press.
- Denisova, A., Cairns, P., Guckelsberger, C., & Zendle, D. (2020). Measuring perceived challenge in digital games: Development & validation of the challenge originating from recent gameplay interaction scale (CORGIS). *International Journal of Human-Computer Studies*, *137*, 102383.
- Denisova, A., Guckelsberger, C., & Zendle, D. (2017). Challenge in digital games: Towards developing a measurement tool. In *Proceedings of the 2017 Chi Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2511–2519). New York, NY: Association for Computing Machinery.
- DfE. (2014). *National curriculum in England: English programmes of study*. Department for Education. <https://www.gov.uk/government/publications/national-curriculum-in-england-english-programmes-of-study>
- Drachler, H., Hummel, H. G., & Koper, R. (2008). Personal recommender systems for learners in lifelong learning networks: The requirements, techniques and model. *International Journal of Learning Technology*, *3*, 404–423. <https://doi.org/10.1504/IJLT.2008.019376>

- EdSurge. (2016). *Decoding adaptive*.
- EEF. (2018). *Mastery learning*. Teaching and Learning Toolkit. Education Endowment Foundation. <https://educationendowmentfoundation.org.uk/pdf/generate/?u=https://educationendowmentfoundation.org.uk/pdf/toolkit/?id=156&t=Teaching%20and%20Learning%20Toolkit&e=156&s=>
- Ehri, L. C. (2017). *Reconceptualizing the development of sight word reading and its relationship to recoding*. Reading Acquisition (pp. 107–143). Oxfordshire: Routledge.
- Elo, A. E. (1978). *The rating of chessplayers, past and present*. Arco Pub.
- Gallego-Durán, F. J., Molina-Carmona, R., & Llorens-Largo, F. (2018). Measuring the difficulty of activities for adaptive learning. *Universal Access in the Information Society*, 17, 335–348. <https://doi.org/10.1007/s10209-017-0552-x>
- Grigorenko, E. L. (2009). Dynamic assessment and response to intervention: Two sides of one coin. *Journal of Learning Disabilities*, 42, 111–132. <https://doi.org/10.1177/0022219408326207>
- Guha, M. L., Druin, A., & Fails, J. A. (2013). Cooperative inquiry revisited: Reflections of the past and guidelines for the future of intergenerational co-design. *International Journal of Child-Computer Interaction*, 1, 14–23. <https://doi.org/10.1016/j.ijcci.2012.08.003>
- Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior*, 54, 170–179. <https://doi.org/10.1016/j.chb.2015.07.045>
- Harley, T. A. (2013). *The psychology of language: From data to theory*. Psychology press.
- Hattie, J. (2008). *Visible learning*. Routledge.
- Hendrix, M., Bellamy-Wood, T., McKay, S., Bloom, V., & Dunwell, I. (2018). Implementing adaptive game difficulty balancing in serious games. *IEEE Transactions on Games*, 11(4), 320–327.
- Hewitt, J., Pedretti, E., Bencze, L., Vaillancourt, B. D., & Yoon, S. (2003). New applications for multimedia cases: Promoting reflective practice in preservice teacher education. *Journal of Technology and Teacher Education*, 11, 483–500.
- Hiebert, E. H., Scott, J. A., Castaneda, R., & Spichtig, A. (2019). An analysis of the features of words that influence vocabulary difficulty. *Education Sciences*, 9, 8. <https://doi.org/10.3390/educsci9010008>
- Holstein, K., Alevan, V., & Rummel, N. (2020). A conceptual framework for human–AI hybrid adaptivity in education. In *International Conference on Artificial Intelligence in Education* (pp. 240–254). Cham: Springer.
- Hou, X., Nguyen, H. A., Richey, J. E., Harpstead, E., Hammer, J., & McLaren, B. M. (2021). Assessing the effects of open models of learning and enjoyment in a digital learning game. *International Journal of Artificial Intelligence in Education*, 1–31.
- Hourcade, J. P. (2008). Interaction design and children. *Foundations and Trends® in Human-Computer Interaction*, 1, 277–392.
- Jones, L. V., & Thissen, D. (2006). 1 A history and overview of psychometrics. *Handbook of Statistics*, 26, 1–27.
- Ke, F. (2016). Designing and integrating purposeful learning in game play: A systematic review. *Educational Technology Research and Development*, 64, 219–244. <https://doi.org/10.1007/s11423-015-9418-1>
- Kim, Y. J., & Ruipérez-Valiente, J. A. (2020). *Data-driven game design: The case of difficulty in educational games* (pp. 449–454). Springer International Publishing.
- Klinkenberg, S., Straatemeier, M., & van der Maas, H. L. (2011). Computer adaptive practice of Maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education*, 57, 1813–1824. <https://doi.org/10.1016/j.compedu.2011.02.003>
- Lam, A. (2000). Tacit knowledge, organizational learning and societal institutions: An integrated framework. *Organization Studies*, 21, 487–513. <https://doi.org/10.1177/0170840600213001>
- Lin, X., Schwartz, D. L., & Hatano, G. (2005). Toward teachers' adaptive metacognition. *Educational Psychologist*, 40, 245–255. https://doi.org/10.1207/s15326985ep4004_6
- Lomas, J. D., Koedinger, K., Patel, N., Shodhan, S., Poonwala, N., & Forlizzi, J. L. (2017). Is difficulty overrated? The effects of choice, novelty and suspense on intrinsic motivation in educational games. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 1028–1039). New York, NY: Association for Computing Machinery.
- Lopes, R., & Bidarra, R. (2011). Adaptivity challenges in games and simulations: A survey. *IEEE Transactions on Computational Intelligence and AI in Games*, 3, 85–99. <https://doi.org/10.1109/TCAIG.2011.2152841>
- Malva, L., Hooshyar, D., Yang, Y., & Pedaste, M. (2020). Engaging Estonian primary school children in computational thinking through adaptive educational games: A qualitative study. In *2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT)* (pp. 188–190). Tartu, Estonia: IEEE.
- Mavrikis, M., & Holmes, W. (2019). Intelligent learning environments: Design, usage and analytics for future schools. In *Shaping Future Schools with Digital Technology* (pp. 57–73). Singapore: Springer.
- McLaren, B. M., & Nguyen, H. (2021). AIED Digital Learning Games: Where we are and how we might make a bigger impact. In B. du Boulay, A. Mitrovic, & K. Yacef (Eds.), *Handbook of Artificial Intelligence in Education*. Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 116.
- Nutley, S., Walter, I., & Davies, H. T. (2003). From knowing to doing: A framework for understanding the evidence-into-practice agenda. *Evaluation*, 9, 125–148. <https://doi.org/10.1177/1356389003009002002>

- Oakhill, J. V., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*, 16, 91–121. <https://doi.org/10.1080/10888438.2010.529219>
- Peirce, N., Conlan, O., & Wade, V. (2008). Adaptive educational games: Providing non-invasive personalised learning experiences. In *2008 Second IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning* (pp. 28–35). Banff, AB: IEEE.
- Pelánek, R. (2016). Applications of the ELO rating system in adaptive educational systems. *Computers & Education*, 98, 169–179. <https://doi.org/10.1016/j.compedu.2016.03.017>
- Perkins, D., & Salomon, G. (2012). Knowledge to go: A motivational and dispositional view of transfer. *Educational Psychologist*, 47, 248–258. <https://doi.org/10.1080/00461520.2012.693354>
- Plass, J. L., & Pawar, S. (2020). Toward a taxonomy of adaptivity for learning. *Journal of Research on Technology in Education*, 52, 275–300. <https://doi.org/10.1080/101015391523.2020.1719943>
- Porayska-Pomsta, K., Mavrikis, M., D'Mello, S., Conati, C., & Baker, R. S. (2013). Knowledge elicitation methods for affect modelling in education. *International Journal of Artificial Intelligence in Education*, 22, 107–140.
- Rose, J. (2006). *Independent review of the teaching of early reading*. Department for Education and Skills.
- 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 & Education*, 69, 452–462. <https://doi.org/10.1016/j.compedu.2013.07.004>
- Scandura, J. (2014). Adaptive learning: How it is learned or what is learned. *Technology, Instruction, Cognition, and Learning (TICL)*, 9, 237–239.
- Shute, V. J., Ventura, M., & Kim, Y. J. (2013). Assessment and learning of qualitative physics in Newton's playground. *The Journal of Educational Research*, 106, 423–430. <https://doi.org/10.1080/00220671.2013.832970>
- Shute, V. J., & Zapata-Rivera, D. (2012). Adaptive educational systems. *Adaptive Technologies for Training and Education*, 7, 1–35.
- Sottolare, R. A., Graesser, A., Hu, X., & Holden, H. (2013). *Design recommendations for intelligent tutoring systems: Volume 1-learner modeling*. US Army Research Laboratory.
- Streicher, A., & Smeddinck, J. D. (2016). Personalized and adaptive serious games. In *Entertainment Computing and Serious Games* (pp. 332–377). Cham: Springer.
- Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice*, 15, 55–64. https://doi.org/10.1207/SLDRP1501_6
- Tsatiris, G., & Karpouzis, K. (2021). Developing for personalised learning: The long road from educational objectives to development and feedback. *ArXiv*, abs/2101.11333.
- Vanbecelaere, S., Van den Berghe, K., Cornillie, F., Sasanguie, D., Reynvoet, B., & Depaepe, F. (2020). The effectiveness of adaptive versus non-adaptive learning with digital educational games. *Journal of Computer Assisted Learning*, 36, 502–513. <https://doi.org/10.1111/jcal.12416>
- Vasalou, A., Sumner, E., Joye, N., Benton, L., Herbert, E., Raftopoulou, C., & Symvonis, A. (2021). Developing a digital annotated primary-level word dictionary for use within reading apps. In *British Dyslexia International Conference 2021*. British Dyslexia Association. <https://readprojecteu.files.wordpress.com/2021/06/bda-poster.pdf>
- Verhoeven, L., & Van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 22, 407–423. <https://doi.org/10.1002/acp.1414>
- Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wauters, K., Desmet, P., & Van Den Noortgate, W. (2012). Item difficulty estimation: An auspicious collaboration between data and judgment. *Computers & Education*, 58, 1183–1193. <https://doi.org/10.1016/j.compedu.2011.11.020>
- Yang, Q.-F., Chang, S.-C., Hwang, G.-J., & Zou, D. (2020). Balancing cognitive complexity and gaming level: Effects of a cognitive complexity-based competition game on EFL students' English vocabulary learning performance, anxiety and behaviors. *Computers & Education*, 148, 103808. <https://doi.org/10.1016/j.compedu.2020.103808>
- Zarraonandia, T., Díaz, P., & Aedo, I. (2016). Modeling games for adaptive and personalized learning. In *The future of ubiquitous learning* (pp. 217–239). Berlin, Heidelberg: Springer.
- Zeng, J., Parks, S., & Shang, J. (2020). To learn scientifically, effectively, and enjoyably: A review of educational games. *Human Behavior and Emerging Technologies*, 2, 186–195. <https://doi.org/10.1002/hbe2.188>

How to cite this article: Benton, L., Mavrikis, M., Vasalou, A., Joye, N., Sumner, E., Herbert, E., Revesz, A., Symvonis, A., & Raftopoulou, C. (2021). Designing for “challenge” in a large-scale adaptive literacy game for primary school children. *British Journal of Educational Technology*, 52, 1862–1880. <https://doi.org/10.1111/bjet.13146>