A Critical Examination of Feedback in Early Reading Games

Laura Benton[1], Asimina Vasalou[1], Kay Berkling[2], Wolmet Barendregt[3], Manolis Mavrikis[1]

[1]UCL Knowledge Lab, UCL Institute of Education, University College London, London, UK
{l.benton, a.vasalou, m.mavrikis}@ucl.ac.uk

[2]Department of Computer Science, Cooperative State University Baden Württemberg, Karlsruhe, Germany
berkling@dhbw-karlsruhe.de

[3]Department of Applied Information Technology, University of Gothenburg
Gothenburg, Sweden
wolmet.barendregt@ait.gu.se

ABSTRACT
Learning games now play a role in both formal and informal learning, including foundational skills such as literacy. While *feedback* is recognised as a key pedagogical dimension of these games, particularly in early learning, there has been no research on how commercial games available to schools and parents reify learning theory into feedback. Using a systematic content analysis, we examine how evidence-based feedback principles manifest in five widely-used learning games designed to foster young children's reading skills. Our findings highlight strengths in how games deliver feedback when players succeed. Many of the games, however, were inconsistent and not proactive when providing error feedback, often promoting trial and error strategies. Furthermore, there was a lack of support for learning the game mechanics and a preference for task-oriented rewards less deeply embedded in the gameplay. Our research provides a design and research agenda for the inclusion of feedback in early learning games.

Author Keywords
Feedback; learning games; children; reading.

ACM Classification Keywords
K.8.0 [Personal Computing]: Games – *General*; K.3.1 [Computers and Education]: Computer Uses in Education – *Computer-assisted instruction*.

INTRODUCTION
Recent years have seen a growth in the learning games market (also referred to as educational/serious games, games-based learning or games for learning), projected to reach $4.8M within the next two years [4]. The inclusion of these games within formal education has become more commonplace with many games prioritising the teaching of curriculum subjects such as literacy and numeracy [6-8]. The promise to support or even enhance learning through games has increased designers' responsibility to elucidate their design rationale. However, games researchers have often expressed concerns as to whether learning games effectively marry good game design and pedagogy [9-11].

Understanding the pedagogies expressed in current games requires engagement with theory that can subsequently guide critique, and also support the development of scientifically-based evaluation criteria [8]. In line with this view, Ke [16] analyses learning games literature to highlight that previous research has "predominantly focused on reporting the learning effectiveness of games without a detailed record of game design features and processes". She proceeds to recommend an increased focus on describing theoretical underpinnings and game design rationales. This paper seeks to contribute to this space by critically examining the design of *feedback* in early learning games.

Feedback plays a powerful role in raising achievement above and beyond other instructional interventions [19]. Games, it has been argued, are particularly apt in delivering *in-the-moment* feedback [20, 21], and are "feedback-rich environments that can provide many, often subtle, cues about player status" [22]. Previous research has begun to recognise the need to examine how feedback is designed in learning games [23]. This work has often, however, treated feedback at a high-level, for instance simply identifying the presence of appropriate feedback [8, 24-26], thus excluding a deeper analysis of how this feedback has been designed.

This paper seeks to understand and evaluate how feedback is currently represented in learning games for early learners through the critical case of the reading domain. A review of the empirical literature on feedback and learning games is used to inform a framework for content analysis that is subsequently applied to five popular early learning games for reading. Taking a theoretical and critical lens, we scrutinise the types of feedback present in these games to propose new opportunities for game design and research.

The primary contributions of this work are threefold. First, we iteratively refine a comprehensive analytic framework

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

ACM ISBN 978-1-4503-5620-6/18/04…$15.00
https://doi.org/10.1145/3173574.3173947

for guiding the design and evaluation of feedback types within learning games for children more generally. We foresee this framework not only being of value to learning games researchers and designers, but also supporting practitioners in critically appraising their learning game choices beyond basic high-level checks that feedback is: 'included', 'appropriate' and/or 'useful'. Second, our content analysis uncovers the prioritisations of current approaches to feedback design within popular early reading games and resulting classroom learning implications. Lastly, in scrutinising their design, we identify three design and research opportunities for early learning games more generally: (i) providing learning support beyond the content domain to also teach the mechanics of the game activity; (ii) supporting deep learning through elaborative feedback that allows learners to understand and correct their error; (iii) to move beyond task-oriented rewards by more deeply embedding learning content within the gameplay.

RELATED LITERATURE

Dimensions of Feedback

Feedback has been defined as information given by an agent (human or digital) to inform learners about their performance and understanding. Feedback is most powerful when it is proceeded with instruction, and hence learners who lack the required knowledge will benefit more from instruction than feedback [1]. In their seminal paper reviewing the evidence of the impact of feedback on achievement, Hattie and Timperley [1] set out four major levels for the focus of feedback. Below we introduce these levels and illustrate each level of feedback with examples from the domain of reading:

Task-level: corrective feedback or knowledge of results, provides information about how well the task has been performed. Intended to support surface-level learning in terms of the ability to acquire, store, reproduce and use knowledge. E.g. "Your answer is correct".

Process-level: feedback related to underlying processes used in the task as well as relating/extending to other tasks. Intended to support deeper-level learning in understanding, enabling the identification of relationships and transfer of knowledge to other contexts. E.g. "Remember the same sound in English can be written in different ways".

Self-regulation-level: feedback supporting self-evaluation, self-efficacy and self-beliefs. Enables students to become more effective learners through monitoring, directing and regulating their own learning strategies, for instance in addressing errors. E.g. "Try breaking down longer words into syllables to help you read the text more accurately".

Self-level: feedback directed at learner personal characteristics, such as praise used as a reinforcer/reward. It is differentiated from praise accompanied by task-focused information. E.g. "Well done, you are a good reader".

At each level three key questions underpin the successful application of feedback [21], which include:

- *Where am I going? (Feed Up)* Requires clear goals and success criteria to be defined.
- *How am I going? (Feed Back)* Requires the identification and communication of the learner's current strengths and weaknesses in relation to the goal/success criteria they are trying to achieve.
- *Where to next? (Feed Forward)* Requires guidance and scaffolds to enable the learner to know what to do in the future.

Feedback is most effective when it aims to move learners between levels from task to process to self-regulation [1]. However, a learner's level of knowledge impacts their feedback needs [21]. To our current interest, Hattie and Gan [27] suggest task-level feedback is particularly powerful for novices who need feedback to acquire content knowledge. Shute [18] explains task-level feedback "typically provides more specific and timely (often real time) information to the student about a particular response to a problem or task". As described above games are particularly suited to providing this form of feedback.

Achievement-Focused Praise

Praise is a core dimension of games and is typically expressed in the form of *rewards*, for example through feedback messages conveyed in pictures, sound effects and/or video clips [29]. Praise is also prevalent in teaching practice as teachers are often encouraged to employ praise in motivating their pupils [30]. Nonetheless, the value of praise on learner achievement, which is our focus in this paper, has received mixed empirical evidence.

Drawing from a synthesis of 12 meta-analyses (196 studies in total), Hattie and Timperley [1] show that self-level feedback, particularly praise directing learner attention away from the learning focus, is the least effective for achievement. Unpacking this conclusion, Skipper and Douglas [30] explain that praise can be delivered at the self-level or process-level, with process-level praise being shown to help learners deal with future failures [31, 32], but self-level praise having potential negative consequences such as undermining their achievement motivation [32] or promoting cheating behaviours [33].

Feedback in Learning Games

Instructional design in games includes the provision of feedback, enabling connections between gameplay and initial instructional objectives [13, 17], and informing the learner about their next step [16]. Johnson et al. [17] group the learning game feedback types into *outcome feedback* (relating to task-level feedback [34]) and *elaborative* or *explanatory feedback*[1] (relating to all levels of feedback

[1] Referred Johnson et al. as *process* feedback, but we use the alternative names to prevent confusion with Hattie and Timperley's notion of process-level feedback.

[34]). Outcome feedback includes information about the response correctness, error location and performance measures (e.g. via a numerical scoring system). Elaborative feedback includes specific task/topic information, corrective strategies, why a response is (in)correct or hints/prompts. These feedback types are not mutually exclusive – e.g. a game score could be combined with guidance on how to improve that score next time [17], but the specific use of elaborative feedback has been shown to be very effective in learning achievement [17, 18, 35].

In evaluating game feedback types, Moreno [36] found that novice college students learned more (in the context of botany) when provided with outcome-elaborative feedback than just outcome feedback. Mayer and Johnson [37] replicated these results in the context of electronic circuitry with the same profile of learners. Moreno [36] suggests elaborative feedback may reduce novice learners' cognitive load as they do not then spend time searching for a plausible explanation for their result. However the existing literature does not clearly outline how these findings would apply to early learners, who are considered novice learners in a large number of domains [38]. We suggest young children's elaborative feedback needs careful design to reflect their current levels of cognitive development and metacognitive capabilities within the specific domain.

Research Motivation and Goals

In summary, feedback is a core dimension of instructional support and considered a powerful mechanism for learning and achievement. While there is empirical evidence that demonstrates what feedback is most effective such as the use of outcome-elaborative feedback and task-level feedback for novice learners, there is also mixed evidence for some types of feedback such as the inclusion of praise within task and process-level feedback. Games, it has been argued, are ideal for operationalising *in-the-moment* feedback. However, the effectiveness of game feedback depends on sound pedagogical game design decisions informed by theory. This research presents a critical analysis of the *dimensions of feedback that early learning games for reading promote and exclude.* We concentrate on commercial, early learning games because of their underrepresentation in previous analyses of games [39, 40]. We argue that it is crucial to scrutinise the design of early learning games given their increasing use in classrooms [7].

METHODOLOGY

Following a critical case approach [41] (which allows "logical generalizations" to be made from a single case), we focus on early learning game feedback in the domain of reading. Reading is a foundational skill underpinning all children's learning and involves both low (e.g. phonics) and high level (e.g. inference) skills, thus representing different learning process levels also critical to other domains.

Game Selection

Five early reading games (comprising 35 mini games) were analysed. The games were identified in a series of interviews with eight primary school teachers (from four primary schools) in the UK. These interviews were undertaken as part of a larger project seeking to understand the opportunities of digital technologies for reading. To select our participants, we employed a maximum variation sampling strategy that sought to increase differences between schools to distil common patterns in games usage in the classroom [41]. We applied the following criteria to select a balanced representation: location (urban/rural), type (faith/state/independent), technology adoption (high/low) and socio-economic background (affluent/deprived).

During the interviews the teachers were asked about their current routine for teaching reading and what games they incorporated into this routine. This process led us to identify five commercial reading games used by the teachers (see Table 1). The selected games were designed to teach early reading skills to children aged 5 to 7 years, or to teach older children who are still struggling with acquiring these early reading skills. All of the games were available online, with two also available as apps. The games covered key reading areas such as phonics, vocabulary, fluency and comprehension. All games had a substantial user base across UK schools as well as in some cases worldwide, reinforcing the relevance of these games within primary school classrooms more broadly.

Each game comprised a series of learning activities, i.e. *mini-games*, but due to large numbers (100+) in some games it was not possible to include them all in the analysis. Therefore, we followed a maximum variation sampling approach deliberately maximising differences in both mechanics and reading areas [41] (see Table 1), with 35 mini-games selected. This approach to sampling allowed us to capture both variations in feedback design as well as shared patterns of game feedback across different games.

Content Analysis of Games

We employed a deductive content analysis approach (focused on the mini-games) similar to Roskos et al. [42] who (in the context of e-books) drew on prior literature to first develop a content analysis framework and then used empirical data to guide the qualitative content analysis.

Framework development

The three broad dimensions proposed by Hattie and Timperley [1], *Feed Up* (where am I going?), *Feed Back* (how am I going?) and *Feed Forward* (where to next?), initially directed the construction of an analytic framework. We subsequently excluded *Feed Forward* because it was missing completely from two games (the mini-games were standalone) and where it did exist the logic was not always transparent, i.e. it was often not possible to infer how gameplay performance drove *Feed Forward*, putting at risk the reliability of our analysis. Using the remaining two dimensions, we identified and combined three frameworks previously developed in learning sciences and learning games research [1, 17, 29], each of which was informed by a thorough literature review on feedback. This combined

Game (mini-games sampled)	Overall learning goal	Reading area(s) (no. of mini-games)	Gameplay description	User base
Teach Your Monster to Read (TYMTR) (11)	Letters, sounds and single sentences	Phonics (8) Sight words (2) Comprehension (2)*	Online/app-based world with three game levels each containing a sequence of mini-games (some playable standalone)	Used by over 500,000 children [2]
Busy Things (BT) (8)	English Curriculum objectives	Phonics (6) Morphology (1) Syntax (1)	Online learning portal with standalone mini-games organised by age/reading area	4000+ schools subscribe [3]
Education City (EC) (5)	English Curriculum objectives	Phonics (3) Morphology (1) Comprehension (1)	Online learning portal with standalone mini-games organised by age/reading area	15,500+ schools, 70 countries [5]
Nessy Reading and Spelling (7)	Fundamental reading skills (for struggling readers)	Phonics (3) Sight words (1) Syntax (1) Comprehension (2)	Online learning program with 100 sequenced learning lessons (split into 10 'islands') each including mini-games	10,000+ schools worldwide [12]
Fonics (4)	44 initial sounds (phonemes)	Phonics (4)	Online/app-based mini-games which can be played in sequence or standalone	1,750+ schools, 72 countries [15]

Table 1. Overview of sampled games (*one mini-game covers two different areas)

approach provided us with a more comprehensive methodology to appraise how game feedback is designed.

Feed Up: *Feed Up* types were primarily informed by Hattie and Timperley [1]. In order for feedback to be effective first effective instruction needs to happen. We sought to establish whether this instruction occurred within the game - if it taught the literacy concept prior to gameplay and through what mode. Furthermore, to experience success within the game, the player also needs to learn the game play schema (including the games rules, underlying narrative and player interactions) [13] to master how to play the game [14]. We thus also examined the forms of support available for learning the game play mechanics. Next, we turned to how the games conveyed task expectations, identifying if the learning objective and success criteria for each mini-game were made explicit to the player [1].

Feed Back: *Feed Back* types were informed by the serious games framework set out by Johnson et al. [17], which captured both outcome and elaborative types of feedback. Given our coding scope on mini-games, we excluded aspects of their framework that related to feedback given outside the mini-game (e.g. percent accuracy). Furthermore, it was deemed necessary to account for rewards in the *Feed Back* dimension. While game rewards are a motivational tool [29], by rewarding successes the learner also gains knowledge of their results (both at task-level and self-level), thus facilitating the learners' understanding of their strengths. Wang and Sun's game reward system framework was used [29], but being a consequence of play across mini-games 'plots' and 'unlocking content' were excluded.

Application of framework

Our analysis was iterative involving three phases. In phase one, the games were coded by two authors of the paper with expertise in interaction design, reading and learning games. They divided the games between them and undertook the coding independently. They then discussed the coding outcome, highlighting representative examples of each code and adjusted the coding where there were application discrepancies to ensure consistent coding of all games[2]. The reasons for these discrepancies included: coding errors; undecided or differently interpreted codes. In light of these discrepancies the definitions within the coding framework were updated and an illustrative example for each code from the data was added to enhance its interpretation.

In phase two, the third author of the paper with expertise in reading and learning games, independently coded a subset of the mini-games (30%), deemed sufficient in previous work [43]. To establish inter-rater reliability (i.e. between the first/second coders and third coder – see Table 2) we used Cohen's Kappa which was κ=0.57 for this phase. This suggests a moderate agreement [44], due to still many discrepancies in the coding. A discussion of the disagreements revealed the following issues (codes appear in bold): different definitions of the game scope e.g. coding errors; not coding for *optional* support for the **gameplay mechanics**; undecided or differently interpreted codes e.g. viewing the **response specific** code as a sub-code of **topic specific** rather than applying these codes separately. During this second coding phase we also inductively identified one feature refinement (**try again**) and one new feedback feature (**punishment**) that our coding framework did not address fully, leading us to revise the framework (see Table 2). We split **try again** into three sub-categories that recognised the variability in the mini-games e.g. content changes (**same mechanics, new content**) and number of attempts (**limited** and **unlimited**). Furthermore it was observed that some games included **punishments** for errors

[2] Note inter-rater reliability is not relevant here as the coders looked at different games

	Type	Code	Description	Code Source	Inter-rater reliability κ
Feed Up	Learning Objective	Yes/No	Is the learning objective of the game clear?	[1]	1
	Success Criteria	Yes/No	Are the criteria that the player has to fulfil to achieve success clear?	[1]	0.5
	Learning Instruction	Visual/Verbal/ Model/None	Does the game introduce the learning concept prior to gameplay? In what mode(s)?	[1]	0.13 (0.64)
	Gameplay Mechanics	Visual/Verbal/ Model/None	Does the game provide any support for learning the gameplay mechanics? In what mode(s)?	[13, 14]	0.25 (1)
Feed Back	Outcome	Knowledge of Result	States that the answer is correct/incorrect	[17, 18]	0.67 (0.79)
		Knowledge of Correct Result*	Provides the correct answer	[17, 18]	0
		Try-Again (unlimited)*	Allows unlimited attempts with the same content	[17, 18] + inductive coding	1 (1)
		Try-Again (limited)*	Allows limited attempts with the same content (in terms of options or time)	[17, 18] + inductive coding	1 (0.75)
		Try-Again (new content)*	Allows player to try again with same mechanics but different content	[17, 18] + inductive coding	1 (0.71)
		Error Flagging*	Highlights where the error was made	[17]	0 (1)
	Elaborative	Topic Specific	Provides additional information about specific literacy concept	[17]	0.75
		Response Specific	Explains why answer is correct/incorrect	[17]	1
		Informational*	Gives information about how to work out correct answer or advance general understanding	[17]	1
		Hints, Prompts or Cues*	Guides player to correct answer (without providing answer)	[17]	1
	Rewards^	Score System	Uses numbers to represent performance	[28]	0.25
		Experience Points	Enhancement of player avatar abilities	[28]	1
		Item Granting System	Virtual items that can be used in the game	[28]	0.38 (0.58)
		Resources	Collectable valuables used in gameplay	[28]	1
		Achievement Systems	Collectable avatar/player titles	[28]	1
		Feedback Messages	Evoke praise through text, pictures, sound effects or video clips	[28]	0
	Punishments*	Removal	Temporary removal of rewards (re-gainable through game play)	inductive coding	(0.33)
		Loss	Loss of lives/points (not re-gainable)	inductive coding	(0.33)

Table 2. Final coding framework (*) applies to incorrect responses only (^) applies to correct responses only. Third coding phase inter-coder reliability in brackets.

		TYMTR (11)	BT (8)	EC (5)	Nessy (7)	Fonics (4)
Learning Objective		82%	100%	100%	100%	75%
Success Criteria		18%	100%	100%	100%	25%
Learning Instruction	*Visual*	45%	13%	80%	86%	-
	Verbal	64%	13%	80%	86%	-
	Model	64%	13%	80%	86%	-
	None	36%	88%	20%	14%	100%
Gameplay Mechanics	*Visual*	-	-	80%	100%	-
	Verbal	91%	100%	80%	100%	50%
	Model	-	-	-	100%	-
	None	9%	-	20%	0%	50%

Table 3. Summary of Feed Up coding (mini games coded)

and therefore we added codes to recognise rewards that were **removed** or **lost**. After this final phase, the inter-rater reliability was recalculated using Cohen's Kappa, which was κ=0.75 (Table 2 shows updated reliability in brackets at code-level) suggesting a substantial agreement [44].

RESULTS

In this section we present the numerical findings from our analysis, illustrated with examples from the mini-games. It should be noted that as we selected a diverse sample of games (with respect to domain and mechanics) rather than all mini-games from each game, the reported results are not absolute but rather a proportion of the sampled mini-games. Codes from the framework appear in bold text.

Feed Up (Where am I going?)

Table 3 summarises the outcomes of the *Feed Up* analysis. Our findings show that **learning objectives** were found in all games. However, some of the mini-games within TYMTR and Fonics did not present learning objectives, showing an inconsistency in design of mini-games within the same game. For example, within Fonics one mini-game explicitly highlighted the learning objective for a specific phoneme (/ll/) by stating "Can you find the /ll/ sound". By contrast, another mini-game simply asked the child to "Drag the words to the correct picture" without describing the objective (of reading comprehension).

Success criteria were included consistently in three of the five games. Within two of these (EC and Nessy) this criterion was made very clear, with the target number of successes displayed visibly on the screen. In addition, at the start of each Nessy mini-game the child was given the number of correct answers required to "pass". Within BT the success criteria were more implicit – there were numbers of options present in the design, but the game did not present an explicit target (such as number of correct answers) and the child could typically try again an unlimited number of times. TYMTR and Fonics did not include success criteria consistently across mini-games. TYMTR mini-games contained a progress bar, but this did not express how many correct trials were required to be successful in the game. Similarly in the Fonics game only one mini-game indicated how many questions were required to be answered correctly.

The majority of the games appeared to be consistently providing **learning instruction** for the concept that the mini-game was focused upon. These games used a variety of modes, often in combination, mirroring existing pedagogies such as multisensory learning [45]. However, there were some games (BT and Fonics) that did not include these teaching elements consistently or at all, suggesting their value would be predominantly for practicing familiar concepts. The results also revealed that one game, TYMTR, took a different approach, incorporating a combination of both teaching and practice-focused mini-games.

The majority of the mini-games reviewed provided support for the **gameplay mechanics**. Two games, EC and Nessy, used a combination of different modes to achieve this with other games relying solely on the verbal mode (which was also re-playable if the child wanted to repeat the instructions). From these, Nessy taught the mini-game mechanics the most consistently and comprehensively by providing a tutorial video for each mini-game. This explained the game mechanics whilst demonstrating the mini-game being played. However, the child was required to explicitly select this and had the option to go straight to playing the game. Despite the inclusion of some game play support in all games, three games (TYMTR, BT, Fonics) featured a subset of mini-games that required the use of intuition to work out how to play. Many mini-games reinforced the overall gameplay schema through following a common narrative (e.g. helping a monster with a specific task) or consistent interactions (e.g. tapping on one of four options), which once learned could be applied to subsequent mini-games.

		TYMTR (11)	BT (8)	EC (5)	Nessy (7)	Fonics (4)
Correct Feedback	*Knowledge of Result*	100%	100%	100%	100%	100%
	Topic Specific	9%	25%	60%	57%	100%
	Response Specific	-	-	40%	-	-
Knowledge of Results: Rewards	*Score System*	-	13%	100%	100%	-
	Item Granting	55%	-	-	100%	-
	Achievement Systems	9%	-	-	14%	-
	Feedback Messages	91%	100%	100%	100%	100%
	None	-	-	-	-	-

Table 4. Summary of Feed Back coding for correct response

Feed Back (How am I going?) – Correct Response

Table 4 provides the results of the *Feed Back* coding for a correct response. These results highlight that in all cases there was **knowledge of result** feedback if the child got a correct response, which was communicated in several ways: sound effects, colour changes, and variety of rewards. **Topic specific** feedback was also found in all games, but was used inconsistently across the individual mini-games within a given game. Fonics was the exception, consistently highlighting, sounding out phonemes and (where relevant) reading aloud the whole word for correct responses, which in turn reinforced the letter-sound mappings within words.

Looking across all games, we found that the EC mini-games incorporated the most varied and detailed feedback for correct responses. Not only did it provide the most comprehensive feedback regarding the topic going beyond simply reading aloud and highlighting, but also some mini-games provided more information about the specific sound being focused on and included illustrative images within the feedback. Furthermore, EC was the only game that included **response specific** feedback explaining why the response was correct, e.g. "Words like 'surprise' help us to imagine how someone may have felt".

As described earlier, game rewards are an alternative expression of **knowledge of result**. The most common form was the use of praise through **feedback messages**. This included phrases like "Well Done" or cheering, positive sound effects/music and animated celebrations from game characters. Two games incorporated a **score-based** reward system, which in Nessy subsequently translated into earning a certain number of 'nuggets'. Some TYMTR mini-games provided a chance to collect **items** by cashing in stars earned when making good progress.

Feed Back (How am I going?) – Incorrect Response

In contrast to the correct responses, there was significantly less consistency in how feedback was designed for incorrect answers (see Table 5). Compared to how **knowledge of result** was designed for correct responses, there was also less prominence given to this for incorrect answers. The child was made aware of an error more implicitly through the game being reset and being required to **try again**. Various manifestations of **try again** were identified across the games and also within the mini-games, although each game tended to favour a particular type. TYMTR allowed **unlimited attempts** to try again for incorrect answers, as did many of the BT mini-games, whereas the EC mini-games typically gave a **limited number** of attempts before providing the correct answer (**knowledge of correct results**). Fonics also provided a **limited number** of attempts before the game was over, but did not inform the child of the correct response. Nessy used an alternative strategy, providing the correct answer each time and letting the child try again with **new content** but using the same game mechanic. Nessy was therefore the only game that explicitly discouraged a trial and error approach.

In probing whether the games provided elaborative feedback about the error, we found there was a relatively low number of mini-games that supported the child to understand the particular error made. Some of the Nessy mini-games provided **topic specific** feedback such as showing the word within the sentence, providing the context of use, or reading aloud the target sound and word it was used within, reinforcing the individual sound and how it is blended into a word. Similar to our observations about its feedback design for correct responses, EC mini-games also provided the most varied feedback for incorrect responses, incorporating **topic specific** feedback **(**e.g. reminding the purpose of apostrophes) and providing **hints, cues or prompts** (e.g. giving a strategy for choosing the correct answer/highlighting the sentence part to focus on).

As with the reinforcing role of rewards in **knowledge of results** for correct responses, punishments are an alternative way of expressing **knowledge of results** during errors. Punishments were generally avoided in three games. However, within Nessy giving incorrect answers **lost** the child the possibility of gaining a nugget, and too many incorrect answers meant insufficient nuggets were available to pass the level. Within Fonics each incorrect answer resulted in the **removal** of a life (if three lives are lost the game is over), but there was an opportunity to regain these lives by answering correctly.

		TYMTR (11)	BT (8)	EC (5)	Nessy (7)	Fonics (4)
Incorrect Response	*Knowledge of Results*	45%	88%	80%	86%	25%
	Knowledge of Correct Results	-	25%	80%	86%	-
	Try-Again (Limited Attempts)	-	25%	80%	-	75%
	Try-Again (Unlimited Attempts)	100%	63%	20%	14%	-
	Try-Again (New Content)	-	13%	-	86%	25%
	Error Reporting	-	13%	20%	-	-
	Topic Specific	-	-	40%	43%	-
	Response Specific	9%	-	-	-	-
	Hints, Cues or Prompts	9%	-	60%	-	-
Knowledge of Results: Punishments	*Removal*	-	13%	-	-	100%
	Loss	-	-	-	100%	-
	None	100%	88%	100%	-	-

Table 5. Summary of Feed Back coding for incorrect response

DISCUSSION

This research set out to specifically examine *what dimensions of feedback early learning games for reading promote and exclude* in order to unpick these design decisions critically. To achieve this, we iteratively designed a new analytic framework for the content analysis of learning game feedback, informed by feedback theory and improved through our reflexive use of the framework on a sample of five games (comprising 35 mini-games). Below we discuss the design and research implications from the analysis undertaken. Analytic codes from the framework are referenced in parentheses and appear in bold.

Broadly our analysis highlighted the presence of two types of games, *learning* and *practice* games. Feedback is powerful only when it builds on prior instruction [1], however two of the five games were predominantly or completely missing the teaching of the learning concept (**learning instruction**; Table 3). We would thus characterise these two games as practice games that assume concepts have been introduced *prior* to game play. The relationship between learning and practice was best reflected in TYMTR where learning games were followed by games that practiced the skills taught earlier. This finding broadly suggests that teachers using practice games in the classroom must ensure their pupils have already been taught the appropriate concepts. Yet, the two practice games included in the analysis were identified as 'games for learning' by the primary school teachers interviewed in the research, potentially questioning their scrutiny for how these games were designed and used with their pupils.

Promoted Feedback Dimensions

Theory-led Game Design Exemplars

Previous work in the domain of learning games has sought to develop design patterns that can marry game mechanics with evidence-based instruction [46]. Taking a theory driven perspective in our analysis of the five games, we identify three new exemplars of game feedback shown in past work to increase learning and achievement.

Echoing the importance of setting the child's learning expectations to know where they are going [1], most games posed a clear goal (**learning objective**; Table 3). Learning objectives were introduced by referring to the literacy objective contextualised in the task mechanics (e.g. 'put all in the sheep in the /s/ pen'). Further enhancing the *Feed Up* dimension, three games included criteria that clearly showed what a child needed to achieve to be successful in the game (**success criteria**; Table 3). Criteria for success were either implicit in the task (e.g. by posing one game round with a clear set of options), or on screen through a quantified target (e.g. a set number of stars that needed to be acquired). With the exception of the two practice games, *Feed Up* was also proceeded by first introducing and teaching the key literacy concept addressed in the game (**learning instruction**; Table 3). Reflecting a multisensory approach to reading instruction, all the analysed games used in tandem visual, verbal, and modelling modes for instruction [45] reinforcing sounds, letters and meaning.

Praise and punishment

Within the existing literature there was mixed evidence for the role of praise in raising achievement. All of the games included task-level praise in the form of feedback messages (**feedback messages**, Table 4) for getting the answer correct. However, much of the praise was implicit e.g. positive sound effects and animations, and not directed to the self, e.g. generic phrases such as "Great job". There were only a few examples of self-level praise within TYMTR e.g. "The Tricky is so amazed with *you* she will come with you on the journey". Overall the use of praise in the games was quite limited. Within the praise that was included we observed a predominant focus on task-level, which has been effective in the learning of novices [27], as well as an avoidance in most games of the self-level praise, which has been discouraged within the literature [1, 32, 33].

Additionally, we identified the presence of various forms of punishment within a subset of games, but also in some cases opportunities to recover from these punishments through subsequent successes (**removal**; Table 5). Punishments are a commonly found feature within games [47], with game designers typically making "the failure consequences interesting, and fun" [48] e.g. through your avatar 'dying', or returning to the beginning of the game. However, in light of findings within the wider feedback literature that negative feedback can impact younger children's subsequent learning more significantly than other learner groups [49], there is a risk that without careful design, punishments could negatively impact young children's motivation or engagement with the game.

Excluded Feedback Dimensions

Alongside identifying the strong congruence between theory and game design, our analysis also found gaps in game feedback design. In analysing current game design limitations, we have identified three design and research opportunities in the space of games for early learning.

Need to Support Learning Mechanics as well as Content

In contrast to the uniform inclusion of effective teaching principles for reading in all of the games, with the exception of Nessy, the remaining four games reflected less effort in supporting learning of the game mechanics (**gameplay mechanics**; Table 3). Typically in games the player develops an understanding of the game play schema through experiencing failures at various points in the game and then trying again [50]. However, within learning games it is difficult to separate failure due to the game mechanic or failure due to a gap in understanding the learning content. Previous work has shown when children experience breakdowns during learning games they may need support with both the learning content and with working out the game mechanics [51]. This need for support has been found to increase in pace with the complexity of game mechanics [47, 52]. Plass et al. [50] recommend in learning game design the choice of game mechanics should not introduce these unnecessary confounds. Whilst the reviewed games mainly utilised more familiar multiple choice mechanics, given the young learner group we argue that they will still need opportunities to become familiar with the broader game play schema prior to focusing on new learning content. The most appropriate form for this support remains an open research question.

Deep Learning Comes from Elaborative Feedback

All the games reviewed partially included the *Feed Back* phase by communicating the child's strengths and weaknesses in relation to the learning goal [1]. During *successful* game performance, this was primarily achieved through a clear indication that the correct option was chosen using sound and colour to indicate success (**knowledge of result**; Table 4).

However, while knowledge of one's performance is a critical part of feedback, feedback is most effective when it is elaborated [37], for instance, by reinforcing attributes of the target concept (**topic specific**; Table 4) or building upon topic specific feedback to explain why it is correct (**response specific**; Table 4) [17, 18, 36, 37]. The games we analysed presented some topic specific feedback for successful game performance, albeit not consistently across all mini-games. Additionally, only one game presented response specific feedback. In further probing the games, elaborative feedback was technically attainable within some of the domain areas they covered. For instance, TYMTR and the other phonics-focused mini-games reviewed used narration to introduce letters and words at the start of each mini-game. Thus, the mechanics of highlighting letter/word attributes to deliver topic specific feedback were within existing technical capabilities [48]. Although designers should keep in mind *Feed Back* needs careful integration to ensure it does not interfere with game play [53].

Turning our attention to game feedback during *unsuccessful* game performance, our research showed a clear and consistent gap in game design practice. An incorrect response was often communicated implicitly by asking a child to try again, indicating that their previous attempt was not correct (**knowledge of result**; Table 5). While try again was the primary response to error, its role in learning was not clear. Most of the games allowed a child to repeatedly make the same mistake (although within some games the number of attempts was limited e.g. in Fonics the player has a set number of lives) without providing them with elaborative feedback or even the correct answer to allow them to learn from these failures, mirroring findings by Blair [38] in maths games for young children. Moreover, very few of the games included elaborative feedback to support the child to understand their error. One exception and exemplar of good practice on both try again and elaborative feedback was Nessy: upon an error the correct answer was immediately explained giving the child a chance to apply this knowledge in the same context, but with new content.

In summary, our findings highlight a broad orientation in the games toward informing the child's understanding of their current performance, and providing opportunities to correct an error. These games did not capitalise on the value of feedback for deep learning by supporting the child to understand why they did well or did not succeed. This uncovers an opportunity for further design work to enable young children to build on successes and learn from errors through elaborative feedback targeted at their cognitive development and metacognitive capabilities.

Restricted Forms of Game Reward

When gameplay is successful, game rewards come in the form of new acquisitions and features in the game world

[29]. Our analysis showed that many typical game rewards were absent in the five games we analysed, such as, for example, the ability to enhance the skills of one's avatar (**experience points**; Table 2). Game rewards were primarily expressed as short feedback messages after each task praising the child (**feedback messages**; Table 4). This highlights an increased emphasis on short learning tasks, and less focus on additional game tropes that are often woven into larger narratives connecting learning content with game play. This design practice can be interpreted to restrict opportunities for more immersive playful approaches to learning. However, this conclusion is mainly drawn from an analysis of mini-games and should be interpreted in line with our analytic focus. Given the exclusion of *Feed Forward* in the analysis we did not take into account rewards that occurred outside of the mini-game for the two games that posed a narrative (TYMTR, Nessy). Nonetheless, anecdotally, our full teacher interviews showed that maximising (explicit) learning time was a desirable game feature posing a constraint on how learning games should be designed for the classroom. This finding highlights a potential tension in designing games for classroom use: short task-oriented games that mirror classroom learning may satisfy educators and parents that learning time is maximised, but the types of game rewards excluded highlight restrictions in 'intrinsic integration' where learning content is more deeply embedded (but potentially less explicitly identifiable) within the gameplay [14]. This restriction may impact a child's motivation and engagement with the game, with potential negative implications for their overall achievement.

CONCLUSION

This paper identified the importance of evaluating how *feedback* is currently designed in early learning games. To achieve this, we consolidated, refined and applied a comprehensive game content analysis framework, to analyse several widely-used reading games for young children. This descriptive analysis was followed by a critical evaluation of existing early learning game feedback decisions, drawing from the games and learning literature to problematize them.

This research makes three main contributions. Firstly, in consolidating and reflexively applying a holistic framework for evaluating feedback in early learning game design we offer a new *methodological* tool. This tool was based on three existing broad frameworks, and unifies codes for feedback in learning and learning games as well as game rewards. This methodology supports a fine-grained level analysis of game feedback, serving designers and practitioners who may want to use it as a guide for design or as an evaluation tool for games in reading and other domains. In applying this tool to games within the reading domain we refined and added new codes. We encourage others utilising this tool in different domains and game genres to similarly iterate its dimensions in order to further widen its scope and relevance. Secondly, we provide a *characterisation of the broad game genres for early reading*. We uncovered a prioritisation of task-oriented learning over intrinsically integrated learning content within the games, with more playful and immersive features of games such as rewards and praise as well as punishments limited. This focus on task-oriented learning, as opposed to learning through play, problematizes the kind of learning taking place, the limited space for more immersive games and the tension of including games within tightly packed curricula. Additionally, our analysis showed an important division between games for learning and games for practice. The lack of instruction in some games calls for practitioners to carefully evaluate their game choices to ensure pupils have the necessary knowledge prior to game play. Finally, and most importantly, our research allows us to identify *strengths and weaknesses,* as well as *open questions*, for the future design work in the area of game feedback. Specifically, while most games supported a clear direction on where the player was going (*Feed Up*), there were weaknesses in how the provision of feedback during game play (*Feed Back*) was delivered. This gap was especially visible in how little support children were offered to recover from their errors.

Given our focus and scope, our analysis was limited to the reading domain, and within this we recognise that many of the games focused on phonics. However, by examining the reading domain in depth, and articulating our methodological process, we allow other researchers to conduct similar analyses, ascertaining transferability to other domains [54], for instance to mathematics. Future research can build upon our findings to consider if the domain and its complexity, e.g. low-level skills such as phonics or basic arithmetic compared to high-level-skills such as comprehension or interpreting statistics, impact on how game feedback is designed. Equally, this work could be extended to other learning games types (beyond mini-games) with more complex narratives such as immersive games where further aspects such as feedback timing may have greater importance [23].

In closing, we hope that our work will shine a spotlight on the importance of well-designed feedback for early learning games, carving out priorities that direct game design developments in this area, whilst providing a guide for practitioners to evaluate the plethora of existing learning games toward ensuring meaningful learning experiences for young learners.

ACKNOWLEDGEMENTS

We thank teachers from Stormont School, Holy Cross Catholic Primary School and Whitchurch Primary School for helping to identify our critical case games. This work forms part of the iRead project which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731724.

REFERENCES

1. John Hattie and Helen Timperley. 2007. The power of feedback. *Review of educational research*, 77, 1: 81-112.
2. Teach Your Monster to Read. 2017. About the game. Retrieved 15 September 2017 from https://www.teachyourmonstertoread.com/about-the-game
3. Busy Things. 2017. About Us. Retrieved 15 September 2017 from https://www.busythings.co.uk/about-us
4. S Adkins. 2015. The 2014-2019 Global Edugame Market. *Whitepaper, Ambient Insight*.
5. Education City. 2017. International Overview. Retrieved 15 September 2017 from http://v1.educationcity.com/uk/international/international-overview
6. Sarah Vaala, Anna Ly and Michael H Levine. 2015. *Getting a Read on the App Stores: A Market Scan and Analysis of Children's Literacy Apps. Full Report*. ERIC.
7. Lori M Takeuchi and Sarah Vaala. 2014. Level up Learning: A National Survey on Teaching with Digital Games. *Joan Ganz Cooney Center*.
8. Stamatios Papadakis, Michail Kalogiannakis and Nicholas Zaranis. 2017. Designing and creating an educational app rubric for preschool teachers. *Education and Information Technologies*: 1-19.
9. Konstantin Mitgutsch and Narda Alvarado. 2012. Purposeful by Design? A Serious Game Design Assessment Framework. In *Proceedings of the International Conference on the foundations of digital games*. 121-128.
10. Brian M. Winn. 2008. The design, play and experience framework. *Handbook of research on effective gaming in education*: 1010-1024.
11. Jonas Linderoth. 2012. Why gamers don't learn more: An ecological approach to games as learning environments. *Journal of Gaming & Virtual Worlds*, 4, 1: 45-62.
12. Nessy. 2017. Nessy Reading & Spelling. Retrieved 15 September 2015 from https://www.nessy.com/uk/product/nessy-reading-spelling/
13. Craig A Lindley and Charlotte C Sennersten. 2006. A cognitive framework for the analysis of game play: tasks, schemas and attention theory. In *Proceedings of the Workshop on the cognitive science of games and game play, the 28th annual conference of the cognitive science society*.
14. MP Jacob Habgood and Shaaron E Ainsworth. 2011. Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *The Journal of the Learning Sciences*, 20, 2: 169-206.
15. Fonics. 2017. Where children learn to read. Retrieved 15 September 2017 from https://www.readwithfonics.com/
16. Fengfeng Ke. 2016. Designing and integrating purposeful learning in game play: a systematic review. *Educational Technology Research and Development*, 64, 2: 219-244.
17. Cheryl I Johnson, Shannon KT Bailey and Wendi L Van Buskirk. 2017. Designing Effective Feedback Messages in Serious Games and Simulations: A Research Review. In *Instructional Techniques to Facilitate Learning and Motivation of Serious Games* Pieter Wouters and Herre Van Oostendorp (eds.). Springer, 119-140.
18. Valerie J Shute. 2008. Focus on formative feedback. *Review of educational research*, 78, 1: 153-189.
19. John Hattie. 2008. *Visible Learning*. Routledge.
20. Marc Prensky. 2003. Digital game-based learning. *Computers in Entertainment (CIE)*, 1, 1: 21-21.
21. John AC Hattie and Gregory CR Yates. 2014. Using feedback to promote learning. *Acknowledgments and Dedication*: 45.
22. Erik Harpstead, Christopher J MacLellan, Vincent Aleven and Brad A Myers. 2014. Using extracted features to inform alignment-driven design ideas in an educational game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. 3329-3338.
23. Melissa Sommerfeld Gresalfi and Jacqueline Barnes. 2016. Designing feedback in an immersive videogame: supporting student mathematical engagement. *Educational Technology Research and Development*, 64, 1: 65-86.
24. Pieter Duysburgh, Karin Slegers, Karen Mouws and Marije Nouwen. 2015. Playful Sounds From The Classroom: What Can Designers of Digital Music Games Learn From Formal Educators? In *Proceedings of the Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. 1467-1472.
25. Laila Shoukry, Christian Sturm and Galal H Galal-Edeen. 2015. Pre-MEGa: a proposed framework for the design and evaluation of preschoolers' mobile educational games. In *Proceedings of the Innovations and Advances in Computing, Informatics, Systems Sciences, Networking and Engineering*. 385-390.
26. Vincent Aleven, Eben Myers, Matthew Easterday and Amy Ogan. 2010. Toward a framework for the analysis and design of educational games. In *Proceedings of the Third IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning (DIGITEL '10)*. 69-76.
27. John Hattie and Mark Gan. 2011. Instruction based on feedback. In *Handbook of research on learning and instruction* (eds.). 249-271.
28. Feng Wang and Michael J. Hannafin. 2005. Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, 53: 5-23.

29. Hao Wang and Chuen-Tsai Sun. 2011. Game Reward Systems: Gaming Experiences and Social Meanings. In *Proceedings of the Fifth international conference of the Digital Research Association (DIGRA '11)*.
30. Yvonne Skipper and Karen Douglas. 2012. Is no praise good praise? Effects of positive feedback on children's and university students' responses to subsequent failures. *British Journal of Educational Psychology*, 82, 2: 327-339.
31. Melissa L Kamins and Carol S Dweck. 1999. Person versus process praise and criticism: implications for contingent self-worth and coping. *Developmental psychology*, 35, 3: 835.
32. Carol S Dweck. 2000. *Self-theories: Their role in motivation, personality, and development*. Psychology Press.
33. Li Zhao, Gail D Heyman, Lulu Chen and Kang Lee. 2017. Telling young children they have a reputation for being smart promotes cheating. *Developmental Science*.
34. Fabienne M van der Kleij, Theo JHM Eggen, Caroline F Timmers and Bernard P Veldkamp. 2012. Effects of feedback in a computer-based assessment for learning. *Computers & Education*, 58, 1: 263-272.
35. Richard E Mayer. 2014. *Computer games for learning: An evidence-based approach*. MIT Press.
36. Roxana Moreno. 2004. Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional science*, 32, 1: 99-113.
37. Richard E Mayer and Cheryl I Johnson. 2010. Adding instructional features that promote learning in a game-like environment. *Journal of Educational Computing Research*, 42, 3: 241-265.
38. Kristen Pilner Blair. 2013. Learning in critter corral: evaluating three kinds of feedback in a preschool math app. In *Proceedings of the 12th International Conference on Interaction Design and Children (IDC '13)*. 372-375.
39. Elizabeth A Boyle, Thomas Hainey, Thomas M Connolly, Grant Gray, Jeffrey Earp, Michela Ott, Theodore Lim, Manuel Ninaus, Claudia Ribeiro and João Pereira. 2016. An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education*, 94: 178-192.
40. Douglas B Clark, Emily E Tanner-Smith and Stephen Killingsworth May. 2013. *Digital games for learning: A systematic review and meta-analysis*. Menlo Park, CA.
41. Michael Quinn Patton. 1990. *Qualitative evaluation and research methods*. SAGE Publications, inc.
42. Kathleen Roskos, Jeremy Brueck and Lisa Lenhart. 2017. An analysis of e-book learning platforms: Affordances, architecture, functionality and analytics. *International Journal of Child-Computer Interaction*, 12: 37-45.
43. Matthew Lombard, Jennifer Snyder-Duch and Cheryl Campanella Bracken. 2002. Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human communication research*, 28, 4: 587-604.
44. Steve Stemler. 2001. An overview of content analysis. *Practical assessment, research & evaluation*, 7, 17: 137-146.
45. Jim Rose. 2009. *Identifying and Teaching Children and Young People with Dyslexia and Literacy Difficulties*.
46. Sebastian Kelle, Roland Klemke and Marcus Specht. 2011. Design patterns for learning games. *International Journal of Technology Enhanced Learning*, 3, 6: 555-569.
47. Conor Linehan, Ben Kirman, Shaun Lawson and Gail Chan. 2011. Practical, Appropriate, Empirically-Validated Guidelines for Designing Educational Games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. 1979-1986.
48. Marc Prensky. 2001. *Digital Game-Based Learning*. McGraw-Hill.
49. Anna CK Van Duijvenvoorde, Kiki Zanolie, Serge ARB Rombouts, Maartje EJ Raijmakers and Eveline A Crone. 2008. Evaluating the negative or valuing the positive? Neural mechanisms supporting feedback-based learning across development. *Journal of Neuroscience*, 28, 38: 9495-9503.
50. Jan L Plass, Bruce D. Homer, Charles Kinzer, Jonathan M. Frye and Ken Perlin. 2011. Learning mechanics and assessment mechanics for games for learning. *G4LI White Paper*, 1: 2011.
51. Asimina Vasalou, Rilla Khaled, Wayne Holmes and Daniel Gooch. 2017. Digital games-based learning for children with dyslexia: A social constructivist perspective on engagement and learning during group game-play. *Computers & Education*: 175-192.
52. Matthew Peter Jacob Habgood. 2007. *The effective integration of digital games and learning content*. University of Nottingham.
53. Deanne M Adams and Douglas B Clark. 2014. Integrating self-explanation functionality into a complex game environment: Keeping gaming in motion. *Computers & Education*, 73: 149-159.
54. Andrew K Shenton. 2004. Strategies for ensuring trustworthiness in qualitative research projects. *Education for information*, 22, 2: 63-75.