HUMAN TECHNOLOGY

ISSN: 1795-6889

www.humantechnology.jyu.fi Volume 15(2), June 2019, 226–255

PLAYERS' PROGRESSION THROUGH GRAPHOGAME, AN EARLY LITERACY GAME: INFLUENCE OF GAME DESIGN AND CONTEXT OF PLAY

Morten Njå
Norwegian Centre for Reading Education and Research
University of Stavanger
Norway

Abstract: *Researchers of serious games frequently investigate outcomes of play but overlook the underlying game-design components that drive those outcomes. In this paper, I aim to show how game design and context of play influence progression through GraphoGame, an early-literacy game. This is done by means of two intersecting studies. The first study shows how the game can be represented by a model that explicitly hypothesizes how the interaction between the player and the game drives progression. The second study explores user data generated by first graders (*N *= 137) who played the game over a period of 25 weeks as part of early literacy instruction. The juxtaposition of these two studies reveals factors that influence progression. I also highlight an underdeveloped area within the research field and point to the benefits that bridging game design and outcomes of play may hold for researchers, game developers, and educators.*

Keywords: *serious games, adaptive learning, game design, literacy games, GraphoGame*

DOI: https://doi.org/10.17011/ht/urn.201906123157

VERTAISARVIOITU
KOLLEGIALT GRANSKAD
PEER-REVIEWED
www.tsv.fi/tunnus

INTRODUCTION

Digital games are now an integral part of popular culture, and, in recent years, the presence of such games in educational settings has grown as well. The educational setting includes not only entertainment games repurposed for educational use but also games specifically designed to promote specific learning outcomes, which are commonly referred to as serious games (Wouters, van Nimwegen, van Oostendorp, & van der Spek, 2013). In these games, the entertainment factors typically associated with digital games are secondary to the instructional ones yet employed to motivate players to achieve better learning outcomes (Wouters et al., 2013). Although serious games do not necessarily increase players' motivation and engagement levels, these games may lead to positive learning outcomes, especially if used as part of other instructional activities (Boyle et al., 2016; Connolly, Boyle, MacArthur, Hainey, & Boyle, 2010).

To date, most researchers of serious games have investigated the outcomes of play (Boyle et al., 2016; Connolly et al., 2010), whereas less attention has been paid to how these games work (Gaydos, 2015; Lämsä, Hämäläinen, Aro, Koskinaa, & Äyrämö, 2018). Although it is quite possible to conduct such research without a comprehensive understanding of how these games work, this lack of attention to the workings of games means that the specific components driving outcomes of play may be undistinguishable from one another. Hence any insights pertaining to the game itself will be restricted to the game as a whole. Clark, Tanner-Smith, and Killingsworth (2014) argued that the design of digital tools, not the digital medium itself, is the strongest predictor of learning outcomes, and they call for research exploring how design choices in digital games influence learning outcomes. This view is shared by Lämsä et al. (2018), who called for research into how game design influences learning in children with learning difficulties.

This paper contributes to this underdeveloped area within research into serious games by investigating possible connections between game design and user progression in the Norwegian version of GraphoGame. This game is described as "a technology-enhanced learning environment for learning to read" for children who are "in the early stages of their formal education" (U. Richardson & Lyytinen, 2014, pp. 39-40). For my research of game design and user progression, I conducted two studies. Based on an investigation of the literature, the first study presented a theoretical model for how progression may occur as a result of playing the game, whereas empirical research in the second study generated user data from the gameplay. Juxtaposing the findings from those two studies makes it possible to identify some underpinning factors related to game design and the context of play that drive the outcome measured—in this case, progression through the game.

Researching Serious Games

Part of the reason for the predominance of studies investigating outcomes of play may be that serious games are designed to promote specific learning outcomes, and so researchers may be inclined to examine whether the games do what they are supposed to do. Further, stakeholders and decision makers in the educational sector often call for research that investigates "what works" and best practices in relation to digital learning technologies (Biesta, Edwards, & Allan, 2014; Eisenhart, 2006), which may further influence researchers' choice of focus. Because of this emphasis on measuring output, experimental research designs are commonly used in research into serious games. In the case of research into language learning, for example, the question asked may be whether a particular teaching strategy or activity enhances students' learning performance (Phakiti, 2014).

A systematic review of earlier research into the specific serious game studied here, GraphoGame, showed that most studies used experimental designs geared toward measuring reading outcomes from play (McTigue, Zimmer, Solheim, & Uppstad, 2019). As mentioned above, such research may be carried out without any scientific knowledge of the components that constitute the game because the object of study is the game itself, not its specific individual design elements. Hence the underpinning components that drive the output measured may remain unexplored.

A good metaphor for this is that of the black box, as explained by Latour (1987). This provides a useful framework when discussing how researchers deal with complex systems such as serious games. A black box metaphor is used here in the sense of an artifact for which knowledge exists about how to use it and about the outcome of using it, but for which there is a lack of understanding about how it works. Mark Richardson (2016, p. 661), in reference to Latour's (1987) use of the metaphor, explained that black boxes are found "where the outer skin [of an artifact] masks the inner workings and obstructs comprehension." The use of experimental research approaches may "thicken" this outer skin in that the game as a whole may be considered the object of study. What is more, in one common view, digital learning technologies are seen as tools designed to generate outcomes (Furberg & Lund, 2016). Hence, the tool itself will be of interest only in terms of the results generated by its use. Latour (1999, p. 304) touched on this phenomenon in his claim that the success of technology actually deters comprehension: "When a machine runs efficiently, when a matter of fact is settled, one need focus only on its inputs and outputs and not on its internal complexity. Thus, paradoxically, the more science and technology succeed, the more opaque and obscure they become."

Price, Jewitt, and Brown (2013) pointed out that the research approaches taken to digital technology are trailing because of the rapid changes seen in digital technology. The lack of effective investigative approaches capable of piercing the outer skin to unpack the components of a serious game is another reason why the scope of research may have to be restricted to observing the output (e.g., learning outcomes) from playing serious games. Researchers thus may be forced to use theoretical approaches not specifically designed with the affordances of digital technology in mind. For example, a vast amount of user data is generated by players' interaction with GraphoGame, but these data are rarely explored in-depth (McTigue et al., 2019). In this case, the sheer amount and complexity of the data, in and of itself, may obscure insight into the black box. Indeed, Saarela and Kärkkäinen (2017) pointed out that, even where high-quality data sets such as those from the Programme for International Student Assessment (PISA) are readily available, research making proper use of those data sets is surprisingly rare. However, in two related fields—learning analytics and educational data mining—approaches are being developed to facilitate the use of big data in educational settings, but the tools used have not become standard yet in researchers' toolboxes.

An Activity-Theory Perspective on Serious Games

Playing serious games involves a complex interplay of factors, such as patterns for players' interaction with the game and its components, cognitive processing during play, and factors relating to the social context of play. Each of these may influence the outcome of use to varying degrees. Activity theory (see, e.g., Engeström, 1999; Leontiev, 1977; Vygotsky, 1978) may provide a lens through which to explore such factors. Nardi and Kaptelinin (2006, p. 10) laid out the premise of this theory: "People act as subjects in the world, constructing and instantiating their intentions and desires as objects. Activity theory casts the relationship between people and tools as one of mediation; tools mediate between people and the world." This premise proposes that technology reaches beyond an isolated interaction between a person and an artifact (such as GraphoGame); rather, the use of technology, and by extension playing of serious games, embodies complex social dynamics. When it comes to the use of digital technology (e.g., serious games), this theory offers an approach enabling the coordination of various aspects of technology, such as physical interaction, conceptual interaction, and social-context interaction (Nardi & Kaptelinin, 2006).

The application of activity theory to research into serious games requires a change of perspective compared with typical experimental research. As Bannon and Bødker (1991, p. 241) pointed out regarding studying artifacts (e.g., GraphoGame), "We cannot study them as things, [*sic*] we need to look at how they mediate use." In other words, the focus of research must shift from the game itself to the actions taken when users interact with the game. The application of this theory creates a need to pinpoint exactly what players are doing in their interactions with a serious game. As a result, aspects influencing gameplay that otherwise tend to be hidden inside the black box may be revealed.

At the core of this theory is the concept of activity. Leontiev (1977) broke down activities into three interconnected hierarchical levels. At the top of this hierarchy is the *activity* that a person engages in (such as playing a serious game). *Actions* are segments consciously chosen by a person (the subject) to attain goals (the object) that typically relate to the motivation underpinning the undertaken activity (such as solving a task in the game), whereas *operations* are aspects of engagement whose performance does not require any conscious thought (such as clicking a mouse button or drawing on tacit knowledge).

Su, Feng, Hsu, and Yang (2013, p. 2577) stated that activity theory provides "a useful framework for conceptualizing technology as a dynamic mechanism that conditions and enables development and change in learners and in the mechanism itself." Thus, the perception of the object of study as "a whole" limits the research approaches taken. The activity theory-based model for analyzing serious games and conceptual design proposed by Carvalho et al. (2015) made a distinction among gaming, learning, and instructional activities, further highlighting how the process of playing the game involves actors other than the player, such as game designers and teachers. The player engages in gaming and learning activities, which have separate tools and objectives. Game designers influence the game intrinsically, whereas the teachers responsible for deploying the game influence it extrinsically. Hence, this model provides a way to break down the components of the serious game and identify how players, designers, and teachers engage with it. The premise of this approach—that human activities take place in social contexts—enables not only exploration of the various actors involved with the game (i.e., researchers, designers, and others) and their interaction with the player through the game's components but also exploration of the agents actually present during the playing of the game (i.e., teachers and students).

Learning and Entertainment in Serious Games

Balancing learning and entertainment in a game can be a complex prospect. Arnab et al. (2015) called attention to the challenge of incorporating established pedagogical approaches into serious games, that is, accounting for differences in perspectives among game designers and educators of what a learning game should be. With game designers leading the development process, games may be entertaining but lack essential processes for knowledge acquisition. However, when educators are in charge, the game may be efficient as a learning tool but not fun or motivating to play (Marne, Wisdom, Huynh-Kim-Bang, & Labat, 2012). Nothing prevents these two approaches from coexisting, but it would require stakeholders (i.e., game experts and pedagogical experts) to share a common language (Marne et al., 2012). Even so, as serious games are intended to promote knowledge, pedagogical approaches should underpin the design.

GraphoGame emphasizes the serious element of serious games but also makes use of game elements intended to engage and motivate players. A synthesis of research on Graphogame (McTigue et al., 2019) showed that GraphoGame may support users in developing sublexical skills and improving letter–sound knowledge and phonological processing. However, these authors found that the game supported better word reading only with strong adult interaction during play. McTigue et al. (2019) also provided an overview of Graphogame's theoretical grounding, which included the simple view (Gough & Tunmer, 1986) word reading (Ehri, 2005), psycholinguistic grain size theory (Ziegler & Goswami, 2005), and orthographic depth hypothesis (Katz & Frost, 1992).

The Present Studies

The present paper is based on two studies. The aim of the first study was to investigate how game design influenced players' progression through the game. The related research question was how the interaction between player and game was operationalized. The second study drew upon the exposition of game elements discovered in the first study and sought to investigate the direct consequences of this game design. The aim was to document any progress that became evident in 137 Norwegian first graders playing GraphoGame four times a week for 25 weeks and to relate this progress to the design elements of the game. The related research question regarded what differences in progress could be seen between a group of students initially identified as being at risk of reading and writing difficulties and a group initially identified as not being at risk of such difficulties.

User data were collected at 5-week intervals. They showed how far the students progressed through the game during the timeline of play. The division of students into groups was based on the results of a screening test administered at the onset of schooling. The at-risk group consisted of students identified as being in danger of developing reading difficulties (n = 17) whereas the regular (not at-risk) group consisted of the remaining students (n = 130). This division was made to explore the possibility of different progression trajectories based on students' starting point.

STUDY 1: OPERATIONALIZING GRAPHOGAME

The aim of this first study was to explore the influence of GraphoGame's design on user progression. Hence, the primary research question sought to articulate how the interaction between player and game could be operationalized. In response to that question, a conceptual model detailing how progression occurs through GraphoGame was developed. This model pinpoints central design elements of the game, identifies how players interact with the game, and demonstrates how this interaction drives progression. The main function of this model is to make explicit certain components that would otherwise be hidden in the black box.

Theoretical Grounding for the Model

Generally speaking, the ways in which information flows between the player and the game influences how the interaction ultimately drives progression through play. Earlier models exploring the interaction between computers and humans provided inspiration for the operationalization of GraphoGame. Schomaker et al. (1995) presented a model providing a bird's-eye view of the interaction information flow as it alternates between human and computer through the interface (Figure 1). In brief, the computer captures human output (e.g., touch, voice) from its input modalities (e.g., mouse, keyboard, touch screen, microphone), processes this information according to its programming (marked as "cognition" in the model), and produces output media. The user perceives those media through human input channels (e.g., visual, audible, and tactile) and then mentally processes that information, which leads to the next cycle of information flow. The computer, the interface, and the human can be considered separate spaces in this interaction. Although this is a simplistic model, it provided a starting point for laying out the various spaces and stages in the model presented in this study.

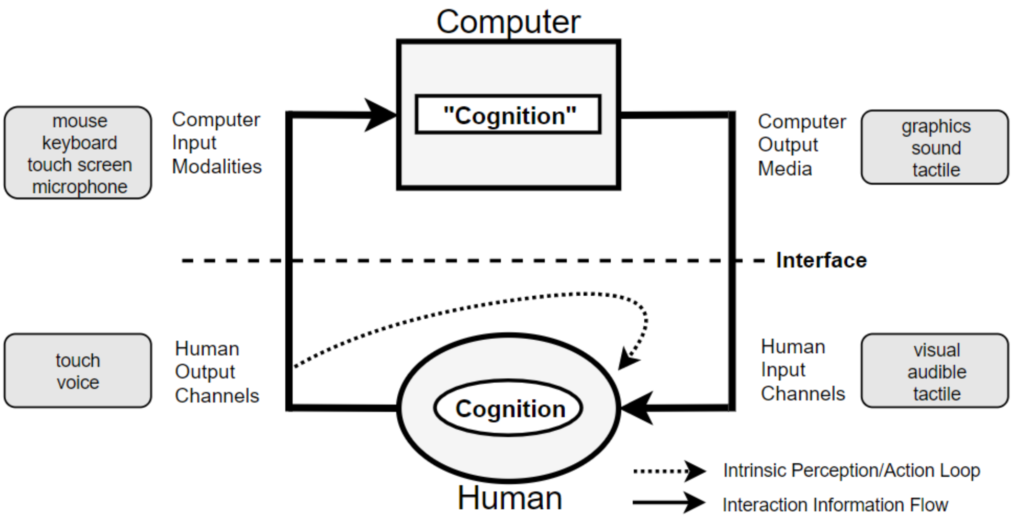

Figure 1. Basic model of human–computer interaction (adapted from Schomaker et al., 1995). This figure illustrates how information flows between computers and humans through various input and output channels, modalities, and media. Humans use their cognitive abilities to interpret the computer output and then take action by using their output channels. The computer processes this output through its programming ("cognition") and provides output media through the interface.

Bienkowski, Feng, and Means (2012) described the structure of an adaptive-learning system reflecting the continuous interaction between a student and that system (Figure 2). This interaction generates user data that are stored in a database. Those data are analyzed by a predictive model that aggregates the data, presents the aggregate to teachers through a dashboard, and feeds data about the student to its adaptation engine. This engine generates output content that suits the individual requirements of the student. (Some adaptive-learning systems, but not that of the present version of GraphoGame, also include an intervention engine that allows teachers or administrators to directly influence the content delivered to the student.)

Interactive Model of Progression Through GraphoGame

GraphoGame provides a learning environment where players may practice sound–letter correspondences. The content is presented as multiple-choice tasks, which represent the learning activity. A game starts with the player selecting one of of the nine possible game modes.[1] A game mode represents a series of tasks that share the same theme. Figure 3 shows an example of a task from the Balloon Game.

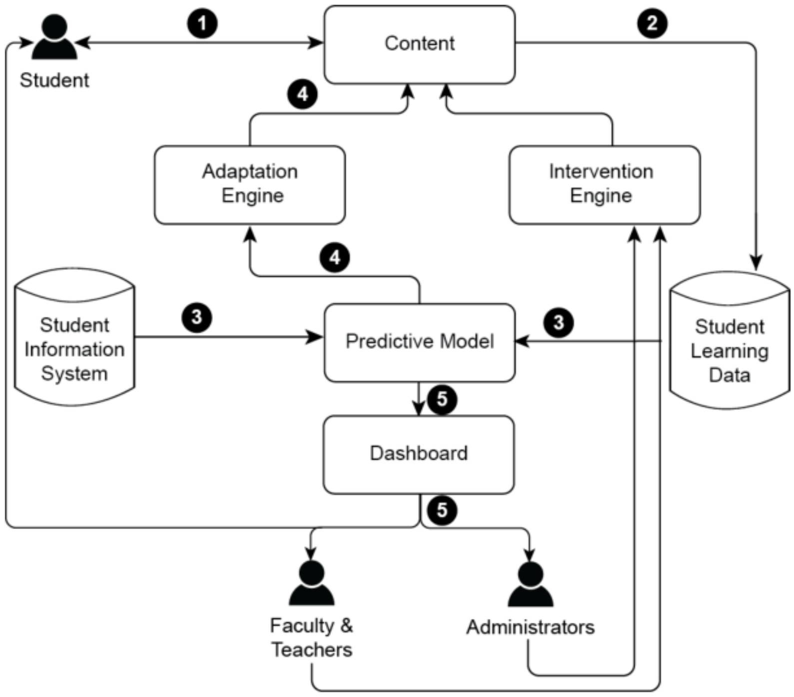

Figure 2. Diagram depicting data flow (the arrows) in a typical adaptive-learning system (ALS; Bienkowski et al., 2012). The model shows how students, administrators, and teachers interact with the various components of an ALS. The student interacts with the content (Line 1, bidirectional to emphasize the interaction aspect) and the outcome of this interaction is stored as student learning data (Line 2). These learning data, along with other student information, are funneled to the ALS's predictive model (Line 3). This combined information is used to create new content through the adaption engine (Line 4), which completes the cycle between student and ALS. In addition, this information can also be accessed by students, teachers, and administrators through a dashboard (Line 5), providing insight into students' performances. In addition, teachers and administrators may influence the content directly through the ALS's intervention engine.

As the player is presented with the graphical elements, as seen in Figure 3, the sound representing the word *gå* [walk] is played out loud. The player's task is to select the word that corresponds with this sound from three possible choices (i.e., one target and two distractors). Following this task, the player is presented with another task until all tasks in the game level have been completed. The yellow bar at the bottom of the screen indicates how far through this game the player has progressed. At completion, the player gets to choose a new game mode. The player may complete several game modes during a single play session.

In the following, I propose a model that details how the interaction between players and the game drives progression. The model produced during the first study (Figure 4) consists of five stages of a cycle that are positioned across three spaces: the *game space* represents the programming and data-access layers of the game, the *interface space* details the way players and the game interact, and the *mind space* deals with how the player processes the information presented through the user interface.

The five stages of interaction are distributed across three independent but interlinked spaces. The game state represents the values stored in the game to represent players' current progress through

Figure 3. An example of a multiple-choice task in Graphogame.

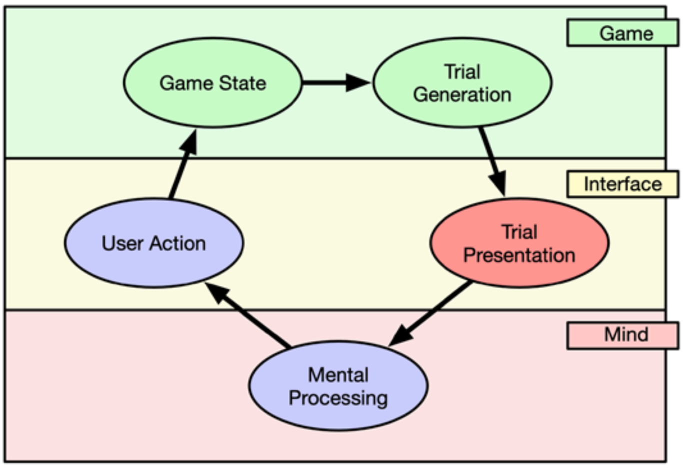

Figure 4. Five stages show the interaction between player and GraphoGame that drives progression through the game. The stages move from the program (game design), through the interface, and into the mind of the user, resulting in user reaction within the interface and a change in the game state. The cycle then repeats.

the game at any given time. Those values are used to generate trials (the learning task; *trial generation*), which are presented visually and aurally to the player through the user interface (*trial presentation*). The player then internally processes those visual and auditory cues (*mental processing*), which causes the player to take an action (i.e., a response to the perceived task) through interacting with the user interface (*user action*). Finally, the result of this action feeds and alters the game state, which in turn establishes the basis for the next trial, and so on. Progression through the game then takes place through continuous repetition of these stages of the model.

As an example of how these activity stages unfold, I present a case where the player first selects the Pirate Game on the selection screen shown between games. The first trial (Figure 5) is generated based on data relating to the player's current progression (i.e., based on prior play or from an initial level), taken from the game state. At the start of this trial, a recording of the sound /e/ is played and the letters *e*, *a*, *u*, *b* and *r* are shown on the screen. The player's task is to identify the letter (i.e., click *e*) that corresponds to the sound played. In this example, the player clicks on the correct item, and the game state is altered to reflect this. A new trial is then generated (Figure 6). The words *riv* [tear], *mur* [wall], *rot* [root] and *rim* [rhyme] are shown on the screen. The sound /riv/ is played and the player again makes the right choice by clicking on the balloon containing this word. After successfully completing the task, the player earns a reward (indicated by the coin appearing above the word *riv* after selecting it). The student then performs six more trials to complete the selected game mode. At this point, the player is presented with a new game-selection screen and may select a new game mode. In the following subsections, this example will be discussed in greater detail against the background of the individual stages of the model.

Game State

The game state represents the current state of various variables detailing a player's progression and performance in the game at any specific time during play. As the player interacts with the game, these values change to reflect the outcome of the player's actions. Two clusters of variables are discussed as part of the adaptive interaction cycle: *player knowledge* and *player performance*. The first of these clusters, player knowledge, includes variables that reflect how well the player knows the content of the game (a value assigned by the game that may or may not reflect a user's actual knowledge). These variables persist between game sessions. Each item in the game is

Figure 5. GraphoGame trial, player's initial trial.

Figure 6. GraphoGame trial, player's subsequent trial.

assigned a knowledge value, which is increased or decreased based on how the player responds in trials where that item is the target (i.e., the correct item). The amount of the incremental change is governed by the number of distractors (i.e., wrong items presented as options) in the trial. For example, the knowledge value of the target item *los* [pilot] will increase more after a successful trial if there were three distractors rather than two. Other variables are derived from the knowledge value of the items. For example, the game keeps track of how many items are known for each content type and for each group of items (dimension). Within the second cluster of variables, the game tracks player performance for each content type in the game (i.e., letter content, syllable content, and word content). Those values are based on performance in the last (up to) 10 trials, and they reset between play sessions. In other words, each time a player starts a new play session, his or her player performance needs to be re-established. The player's performance variable for each content type increases or decreases throughout a gaming session as the player makes correct or incorrect selections for that specific content.

The Norwegian-language version of GraphoGame includes a total of 730 items as targets and/or distractors that a student may encounter during play (see Appendix A). These items are divided into three content types: letter content (24 items), syllable content (272 items), and word content (434 items). The items are further organized into dimensions (i.e., item groups) in roughly an ascending order of difficulty. Letter Dimension 1 consists of the letters that are considered the easiest to learn, whereas Letter Dimensions 2 and 3 include more challenging letters. (Three letters of the Norwegian alphabet have been omitted: *c*, *q* and *z*; these letters appear infrequently in the Norwegian language and are typically not focused on at the onset of letter instruction.) Syllable content consists of two- or three-letter one-syllable words divided into 22 dimensions, each containing a median of 15 target items. The items included in a dimension were grouped together on the basis of specific criteria. For example, the items in Syllable Dimension 1 consist entirely of letters from Letter Dimension 1 (e.g., *er* [am/is/are], *is* [ice], *om* [about]); the items in Syllable Dimension 2 consist of letters from Letter Dimensions 1 and 2 (e.g., *at* [to], *av* [of], *en* [one]); and the items in Syllable Dimension 3 all start with a consonant (e.g., *ta* [take], *be* [pray], *fe* [livestock]). Word content consists of three- to six-letter words divided into 90 dimensions. A median of 12 targets are included in each dimension (excluding dimensions consisting of minimal pairs, i.e., words that differ with respect to a single sound, such as *ul* [howl] and *ull* [wool]). The first seven dimensions consists of three-letter words whereas Dimensions 8–19 include words with double consonants (representing a common spelling difficulty in Norwegian). Dimensions 20–84 consist of minimal pairs. Finally, Dimensions 85–90 include more difficult words including digraphs, that is, cases where several letters are used to represent a single sound.

Trial Generation

To generate trials, GraphoGame uses an adaptation engine. (This term encompasses all components of a game involved in the generation of trials.) Concretely, trials are generated by algorithms that use the current values of variables (i.e., the game state) as their starting point. The adaptation engine's task is to advance the player through the game while recalibrating the difficulty level if the player performs poorly. In other words, the adaptation engine predicts what will lead to the best outcome for the player and generates a trial based on this. Typically, it will tend to step up the difficulty level. However, if a player's performance drops below a certain threshold, the engine will generate easier trials to keep the player engaged. This accuracy threshold varies across content types; it is 85% for letter content, 75% for syllable content, and 65% for word content.

Trial generation is a two-step process. In the first step, the adaptation engine selects a content type for the trial (i.e., letters, syllables, or words) based on player knowledge and player performance (see the game state stage). If the player is deemed to know fewer than 40% of the letters, the content type is restricted to letter content.[2] Beyond this threshold, there is a likelihood for initiating other content types, and this likelihood increases drastically once a player knows all 24 letters. The basic likelihood of syllable content is inversely proportional to the likelihood of letter content. For example (assuming no likelihood of word content, as is the case at earlier stages of game use), if the likelihood of letter content is 35%, then that of syllable content is 65%. If a player is performing well on syllable content,[3] the likelihoods are recalculated[4] in favor of word content. The likelihood of word content is calculated based on the combined likelihood of letter content and syllable content.[5] For example, if the likelihood is 10% for letter content and 30% for syllable content, the likelihood of word content will be 60%.

The second step of trial generation is based on the outome of the first step: The adaptation engine assembles a list of items (one target and one or more distractors), all from a single dimension of the content type selected at the previous stage. If the player is performing poorly, easier trials are generated, whereas if the player is performing well, more difficult trials are created. For example, Figure 7 shows a trial consisting of one target and four distractors from Syllable Dimension 9. The adaptation engine first created a list of suitable dimensions based on the percentage of known items in each dimension. For easy trials, known items are preferred; these are selected from dimensions with a percentage of known items greater than the target percentage for the content type in question. The opposite is true for difficult items. For instance, if a player knows 85% of the syllables in Syllable Dimension 3 and 45% of those in Syllable Dimension 4, a difficult trial may have an unknown word from Syllable Dimension 4 as target whereas the target of an easy trial may be a known item from Syllable Dimension 3. For syllable and word content, the number of distractors is increased by one after a correct answer and decreased by one after an incorrect answer to a maximum of four. For letter content, the maximum number of distractors is six.

Trial Presentation

After a trial has been generated, it is presented to the player through the user interface in visual and auditory forms. The interface can be seen as a mediating tool with which the player interacts as part of playing GraphoGame. The game area is assembled from graphical elements of various shapes and functions. Some of these can be interacted with (e.g., the objects representing items in Figures 6 and 7), but most are purely visual (e.g., the background image).

GraphoGame presents nine different game modes, each with a different visual profile and minor variations in gameplay. Figure 7 shows the game area for the Balloon Game, and Figure 8 shows that for the Pirate Game. In the Balloon Game, the player avatar is displayed on-screen; in the Pirate Game, a pirate with a cannon and a pirate flying a red balloon are displayed. The tile shapes that represent the items are also different: circle-shaped or shaped like hot-air balloons.

As the examples show, the presentation of the tasks varies. There are slight differences between the nine game modes available, but the core gameplay remains the same across all

Figure 7. Balloon Game, three syllable items.

Figure 8. Pirate Game, five letter items.

game modes. For example, in the Balloon Game, a player receives auditory feedback if he or she selects the wrong item, but that item remains visible on the screen until the player correctly identifies the target. Consecutive unsuccessful trials in the Balloon Game will have one fewer distractors until there is only one item left, which ends the game. By contrast, if a player selects a distractor in the Worm Game, all distractors vanish, leaving only the target item on the screen. The player is then required to click on this item before proceeding to the next trial. Additional gameplay factors are programmed into the game, such as time limits, but they are not activated until a player has progressed quite far in the game.

GraphoGame also provides incentives to keep the players motivated. Players earn coins for each successful trial played, which they may use in the in-game shop to buy accessories with which to dress their avatar. Those accessories have no effect on gameplay but are included in the game as an incentive to keep on playing.

The game has a high degree of transparency, meaning a player who wants to engage the game finds few barriers to that goal. However, user agency is limited when playing the game. In fact, the only thing the player can do is to click on an object containing one of the items that are part of the trial.

Mental Processing

As the player perceives the visual and auditory presentation of the trials, the player's brain will process the incoming stimuli and prompt the player to interact with the game. To answer correctly without guessing, the player needs knowledge about the correspondence between the spoken letter or word and its visual representation. The player's perceptive faculties, both visual and auditory, are required to perform the task correctly or to expand knowledge after an incorrect answer. For each consecutive trial containing the same target, the player's knowledge should increase, making it more likely that the correspondence concerned is correctly identified.

These repetitions are intended to aid the player in mapping the spoken language to the written language, an important step toward learning to read (U. Richardson & Lyytinen, 2014, p. 43). GraphoGame applies a synthetic phonics approach, where speech sounds (letter content) are introduced first, which may then be used to decode words that include these sounds (U. Richardson & Lyytinen, 2014, p. 45). As players continuously process trials, their ability to read words is expected to improve steadily.

However, this view of knowledge expanding as a direct consequence of the interaction between player and game is not sufficient, as the knowledge that the players show when playing the game may also originate from outside of the game. Rather than restricting the perspective to include only the interaction between player and game and the context in which the game is employed as part of classroom play, it is preferable to take a broader language-learning perspective where the player of the game also will build skills of the type that are transferrable to other contexts. In addition to knowledge gained from playing GraphoGame, players also would strengthen their knowledge of phoneme–grapheme correspondence by participating in other teacher-guided activities aimed at developing letter knowledge and word recognition. Further, other language-learning strategies may be introduced outside of the phonics approach deployed by the game that may lead players to apply these strategies when playing the game. For example, some students may have developed orthographic reading skills, which is an applicable strategy when identifying word content.

Player Action

It may be assumed that a player's action is the outcome of the mental processes triggered by the visual and auditory input from the user interface. When players engage with the game through the user interface, they make choices by way of clicking on one of the items presented. This click results from a conscious or unconscious process occurring in the brain, something that can be observed only indirectly through actions taken. After mentally processing the trial, the player interacts physically with the user interface by selecting one of the items. This action, from the perspective of activity theory, represents the player's (i.e., the subject of the action) goal of selecting the correct item in a trial (i.e., the object of the action) by using the various tools available in GraphoGame.

The question of what lies behind the action of clicking on a specific item is complicated. In a somewhat simplistic reasoning, a player making the appropriate choice based on knowledge rather than luck must correctly identify the sound(s), must know the letter corresponding to each sound, and must properly match the sounds to the corresponding letters. These represent the necessary components to adequately decode the letters and words encountered in the game. This game-related outcome requires the player to possess an adequate level of knowledge, either learned from playing the game or from other sources outside of the game. The actions the student performs when faced with each new trial in the game will govern his or her progress through the game, as those actions adjust the current game state and hence change the premises for the next trial generation.

Model Summary

In the proposed model, five stages detail how the interaction between student and game takes place and how this interaction drives user progression. This operationalization of gameplay provides a way in which to interpret game data and to identify connections that might otherwise have remained hidden. Progress in a game is governed by the design of the content used and by the adaptation engine that facilitates the distribution of this content based on player interaction. Because this distribution of content is based on specific instructional methods (phonics), progression also may demonstrate the scaffolding of learning based on this method. The activity of playing the game creates a learning environment that exploits the affordances of digital technology: storage and retrieval of data occurs instantly, and the data are used to assemble new, customized tasks on demand and in real time.

STUDY 2: PROGRESSION THROUGH GRAPHOGAME

In this second study, I investigated the effects of game design on student progression. The research question focused on what differences in progress could be seen between at-risk and other students. The availability of two groups based on their school-starting evaluation made it possible to explore whether their progression through GraphoGame varied depending on their baseline skill levels. The purpose of this study was to identify implications of GraphoGame's design as observed through the user data recorded from the students' actions. This ties in with the overarching aim of exploring how design and context of play influence measured output, which in this study is represented as progression through the game.

Method

Deployment and Participants

The Norwegian version of GraphoGame was developed to be part of the On Track intervention study. This was a randomized controlled trial aiming to investigate the effect of an early-intervention strategy to prevent the development of reading difficulties (Lundetræ, Solheim, Schwippert, & Uppstad, 2017). The participants in the present study belong to an On Track subsample from two schools where GraphoGame was played by all students in the classes (*N* = 137), not only by those identified at risk of developing reading and writing difficulties. All students received ordinary reading and writing instruction supplemented with the element of playing GraphoGame for specified amounts of time. The students were all first-graders and 5 to 6 years old; the majority of them were native Norwegian speakers.

The students underwent a set of diagnostic tests at the onset of schooling designed to identify those who might be at risk of developing reading difficulties. The tests were specifically developed for the On Track intervention study. Those screening tests provide an opportunity to differentiate GraphoGame players (as opposed to other interventions employed as part of the On Track program) and to explore the extent to which students' prereading skills influenced their progression through the game. This could provide insights into how learning may occur through the game based on existing knowledge at the start of the interventions; it also could yield information about how the adaptation engine adjusts content based on the players' previous knowledge. The screening tests included letter knowledge and rapid automatized naming as well as isolation and blending of phonemes (i.e., language sounds). Students scoring below the 30th percentile on any of these three tests were given one risk point for each test. They were also given one additional risk point if at least two close relatives reported having reading difficulties. Students obtaining three or more risk points in all were categorized as at risk. This yielded an at-risk group (*n* = 17) and a regular (not-at-risk) group (*n* = 120).

Data Collection

Progression data were segmented into five measuring periods of 5 weeks each (excluding holidays). Table 1 shows the time span for each period. The fourth and fifth periods were extended by one week owing to the winter holidays.

Two data sets were generated. The first was exported directly from the database containing students' user data for each of the five measuring periods and also for all periods combined. The exported data related to the number of days played, the number of trials played, and the time spent playing trials in the game. The second data set was manually collected from the GraphoGame website, where various types of aggregated data from the database are presented. This included game logs showing current progression in terms of items known at the onset of each trial played. The status of known items was given for the most recent play session for each student within each time period. In addition, known items were manually correlated to content type, as the list of known items did not show this information. If an item was found in more than one content type or dimension, the lowest one was used to indicate progress (see Appendix A for item lists).

Data Analysis

Progression was measured separately for the regular group and the at-risk group. Then the groups were compared to identify any significant differences. In testing the null hypothesis (H_0), I confirmed that no difference existed between the two groups in their progression as measured with the user data relating to known items that had been extracted. The significance threshold was set to .05.

Two approaches were taken to exploring the data, each yielding one block. The first block contained data for days played, trials played, and time spent playing. These data were presented for each measurement period as well as for the total timeline of play. Because of the difference in size and variance between the two groups, Welch's *t*-test (i.e., a test that does not require the assumption of equal variance) was used to test the validity of the null hypothesis.

The second block of data detailed players' overall progression and their progression for each content type. These data were presented as boxplots, which provide a clear picture of the distribution of player progression. Because boxplots use median values, the Mann–Whitney *U* test was used to test the null hypothesis for this block. Data that summarize a complete period

Table 1. Span of Time for Each of the Five Measuring Periods Used to Assess Progress.

Period	From	Until	Time span
P1	2014/10/13	2014/11/14	32 days
P2	2014/11/17	2014/12/19	32 days
P3	2015/01/05	2015/02/06	32 days
P4	2015/02/09	2015/03/20	39 days*
P5	2015/03/23	2015/05/05#	43 days*

Note. *The time span for the last two measuring periods were extended due to school holidays. #Students at School A ended gameplay 04/30 while students at School B ended gameplay 05/05.

are referred to as P1 … P5, whereas measurement points (snapshots) of progression are referred to as M1 … M5.

Results: Play Sessions, Trials Played, and Time Spent Playing

Table 2 shows that the regular group played more sessions (*M* = 83.7, *SD* = 7.1) than the at-risk group (*M* = 79.1, *SD* = 8.0). However, this difference was not significant (p = .039).

Table 3 shows the total number of minutes spent playing in the different measuring periods and in total across the timeline of play. The regular group (*M* = 516, *SD* = 82) played the game for slightly longer each session than the at-risk group (*M* = 479, *SD*=76), but this difference was not significant (p = .078).

Table 4 shows the number of trials completed by the students during each measuring period and in total. The regular group played more trials (*M* = 7260, *SD* = 1864) than the at-risk group (*M* = 5948, *SD* = 847), and this difference was strongly significant at $p < .001$. During the intervention timeline, the at-risk group actually kept pace with the regular group during the first two periods (P1: p = .557 and P2: p = .946), but the difference between the groups was strongly significant throughout the rest of the intervention (P3: p = .002, P4: $p <$ .001, P5: $p < .001$).

Table 5 shows the average number of minutes spent playing during each play session. The regular group (*M* = 6.15, *SD* = 0.72) and the at-risk group (*M* = 6.04, *SD* = 0.57) played for a similar duration per session; there is strong statistical support for this conclusion (p = .481).

Table 2. Number of Play Sessions in Each Measuring Period (P1 to P5) and in Total.

Group		P1	P2	P3	P4	P5	TOTAL
Regular	*M*	17.4	15.2	18.3	18.1	14.5	83.7
	SD	1.6	1.7	2.0	2.5	3.0	7.1
At Risk	*M*	16.8	15.8	17.3	16.7	12.6	79.1
	SD	2.5	1.3	2.1	3.0	4.1	8.0

Note. *M* = mean; *SD* = standard deviation

Table 3. Time Played, in Minutes, in Each Measuring Period (P1 to P5) and in Total.

Group		P1	P2	P3	P4	P5	TOTAL
Regular	*M*	114	92	115	109	86	516
	SD	18	17	21	27	29	82
At Risk	*M*	109	94	106	99	71	479
	SD	20	12	19	30	32	76

Note. *M* = mean; *SD* = standard deviation.

Table 4. Number of Trials Played in Each Measuring Period (P1 to P5) and in Total.

Group		P1	P2	P3	P4	P5	TOTAL
Regular	*M*	1648	1256	1567	1560	1229	7260
	SD	355	360	489	673	525	1864
At Risk	*M*	1598	1251	1298	1035	766	5948
	SD	323	268	272	330	310	847

Note. *M* = mean; *SD* = standard deviation.

Table 5. Time Played per Session, in Minutes, in Each Measuring Period (P1 to P5) and in Total.

Group		P1	P2	P3	P4	P5	TOTAL
Regular	*M*	6.51	6.02	6.25	5.98	5.86	6.15
	SD	0.80	0.88	0.93	1.04	1.28	0.72
At Risk	*M*	6.51	5.98	6.10	5.83	5.56	6.04
	SD	0.55	0.74	0.80	0.98	1.13	0.57

Note. *M* = mean; *SD* = standard deviation.

Results: Longitudinal Progression

Overall Progression

Overall progression (see Appendix B, Table B1) was measured in terms of the number of known items out of the total number of items in the game (N = 730) at each measuring point along the timeline. The boxplots shown in Figure 9 give an overview of total progression across the intervention period.

The boxplots indicate the percentages of known items (y-axis) for the regular group and the at-risk group at the five measuring points (x-axis). There were no outliers in either group. The regular group had progressed further than the at-risk group at all measuring points, and this difference was strongly significant at $p > .001$ in all cases. The interquartile range was fairly consistent along the timeline for both groups, with the at-risk group more closely clustered together. For the regular group as a whole, the range of progression covered almost the full inventory of items (especially towards the end of the intervention period), and a steady rise in median progression was seen throughout the period; the at-risk group progressed more slowly and had a more limited overall range of progression. Comparison of the two groups shows that regular group accelerated away from the at-risk group throughout the intervention: At M1, the average difference between the groups was 61 known items, but at M5 this had increased to 141 known items.

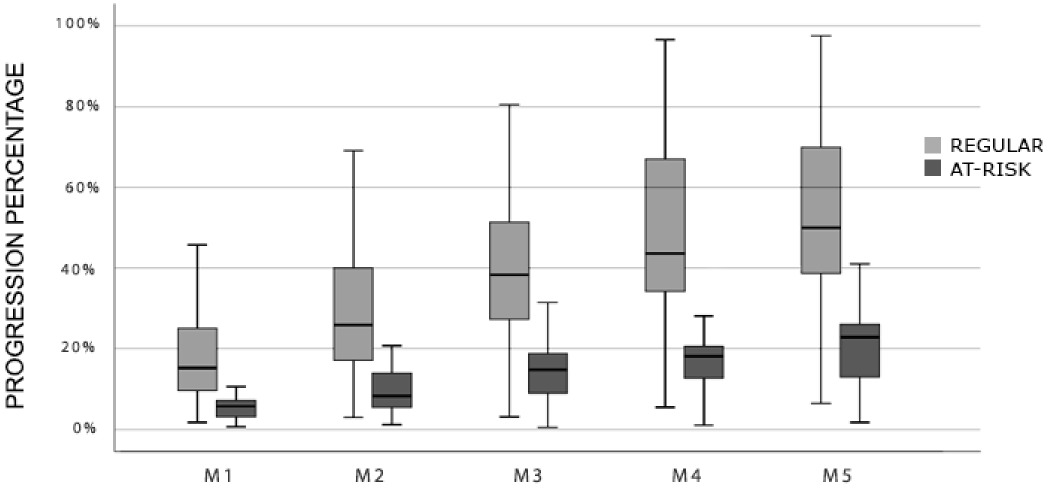

Figure 9. The figure shows boxplots of percentage overall progression for regular and at-risk students through GraphoGame at each of the five measuring points (M1 … M5).

Letter Content Progression

Letter content (see Appendix A, Letter Items, and Appendix B, Table B2) is the first type of content encountered in the game, and progression to other types of content is held back until a player knows at least 40% of the letters. Given that M1 occurred after approximately 18 play sessions, the specific point where the students progressed from only letter content to letter and syllable content was not captured. The boxplots in Figure 10 show the median and percentile progression for known letters throughout the intervention period.

Most students in the regular group knew all letters at M1 and onwards (*Mdn* = 24). The at-risk group progressed more slowly, with a median of 20 known letters (83.3%) at the first measuring point (M1) and 23 (95.8%) at M2. Further along the timeline, most at-risk students knew all the letters (*Mdn* = 24). It should be noted that the outliers who can be seen in the regular group represent a small group of students (less than 5%) who progressed significantly more slowly than the rest of that group. There was a significant difference between the groups at all measuring points along the timeline ($p < .001$).

Syllable Content Progression

Syllable content (see Appendix A, Syllable Items, and Appendix B, Table B3) is the second content type encountered in the game. The boxplots in Figure 11 show the different patterns for the two groups' progression with regard to syllable content.

At M1, the regular group knew a median of 90 syllables (33.1%). This increased to 169 known syllables (62.1%) at M2 and to 251 known syllables (92.3%) at M3, at which point the regular group had thus largely mastered all of the syllable content. After this there was a plateau during the rest of the intervention period. For the at-risk group, the progression curve is rather linear throughout the intervention period. At M1, the median was 23 known syllables (8.5%). At M2, this had increased slightly to 37 known syllables (13.6%), and further along the timeline there

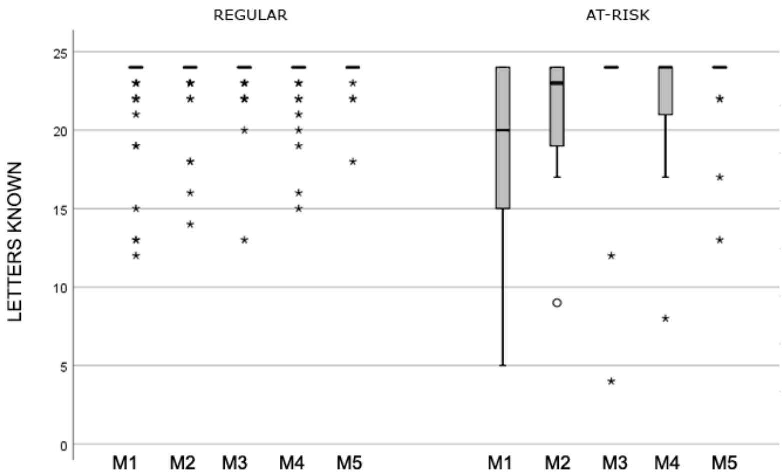

Figure 10. The figure shows boxplots of letter–content progression for regular and at-risk students through Graphogame at five measuring points (M1 … M5). The y-axis shows how many letters were known out of a maximum of 24 letters.

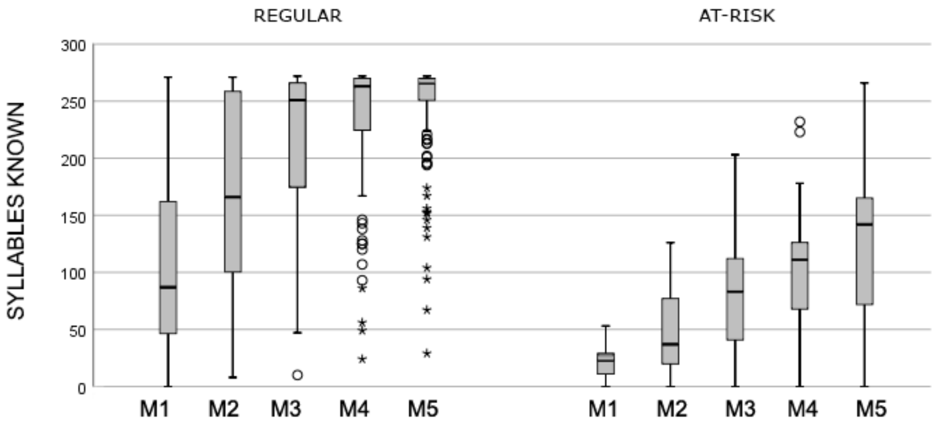

Figure 11. The figure shows boxplots of syllable-content progression for regular and at-risk students at five measuring points (M1 … M5). The y-axis shows how many syllables were known out of a maximum of 272 syllables.

there was a steady increase: 83 known syllables (30.5%) at M3, 111 known syllables (40.8%) at M4, and 142 known syllables (52.2%) at M5. The interquartile range narrowed in the regular group, largely as a result of the plateau, but remained similar across the intervention period in the linearly developing at-risk group. Notably, syllable content was the content type where the members of the at-risk group spent most of their game-playing time, whereas most members of the regular group progressed past the syllable stage during the third measuring period. There was a significant difference between the groups in all five periods ($p < .001$). In contrast to letter content, which most players learned during the first period, progression with respect to syllable content was spread out across the periods to a greater extent.

Word Content Progression

Word content (see Appendix A, Word Items, and Appendix B, Table B4) is the last type of content encountered. This content type also accounts for most of the content in GraphoGame, which has 60% more word items than syllable items. The boxplots in Figure 12 show that only the regular group made any significant progress with word content.

At the first two measurement points, only a few students in the regular group (the outliers) had made any significant progress in terms of words known. At M3, the median for the regular group was 6 known words (1.4%), whereas at M4 the median had increased to 35 known words (8.1%). However, at that point, the upper quartile was already at 197 known words (45.8%). At M5, the median was 76 known words (17.6%) and there had been a slight increase in the upper quartile to 214 words (49.8%). The at-risk group made only negligible progress with respect to word content throughout the intervention period.

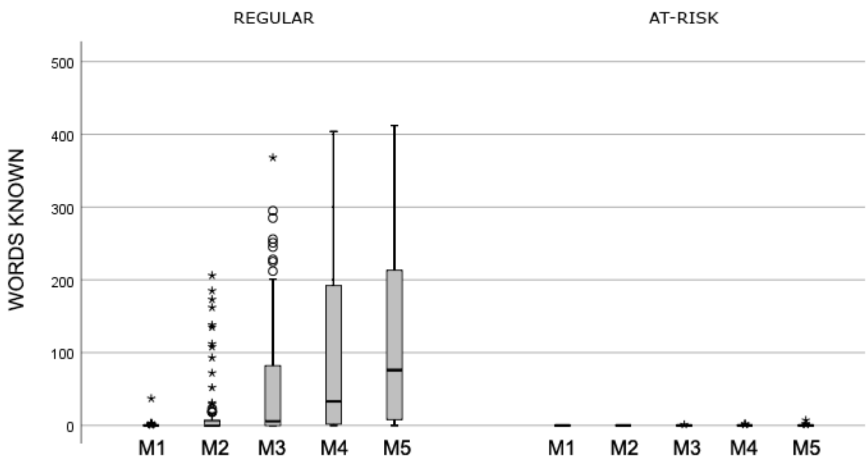

Figure 12. The figure shows boxplots of word-content progression for regular and at-risk at five measuring points (M1 … M5). The y-axis shows how many words were known out of a maximum of 434 words.

GENERAL DISCUSSION

Players' progression through GraphoGame is inherently tied both to features of game design and to the context of play: The game design establishes the premises for progression, whereas playing the game realizes those premises. In the following subsections, I discuss the three overlapping factors that influence progression, namely content design, play time, and adaptation design. These factors were identified as a result of applying the knowledge obtained from Study 1 in order to contextualize the findings made in Study 2.

In reading the analysis below, I must emphasize that I do not make any strong claims about learning outcomes as a result of this research. It is impossible to know exactly, from the data collected, what the participants have internalized as a result of playing even though progression is operationalized using the measurements of the number of known items (out of all 730 items available). Further, the factors that emerged from the juxtaposition of the two studies are not intended as a definite exploration of possible factors.

Progression Through GraphoGame

Content Design

Most students did not encounter the full range of content available during the play period, yet even so, progression unfolded differently in the regular group as compared to the at-risk group. At the last measuring point, after 25 weeks of play, the overall progression of the regular group fell just short of the halfway point in terms of items known (49.7%), with a mean of 7,292 trials played across 85 play sessions. For the at-risk group, progression at that point was less than a quarter (22.7%), with a mean of 5,986 trials played across 79.8 play sessions.

A closer look at progression for the individual content types provides a more nuanced picture. The transitions between content types show how quickly players moved from letter content via syllable content (two- or three-letter words) to word content (four- to six-letter words and minimal pairs). Most students in the regular group knew all of the letter content at the first measuring point (i.e., after an average of approximately 17 play sessions across 5 weeks), whereas many students in the at-risk group were still engaged with letter content at the second measuring point. In other words, it took the average at-risk student more than 30 play sessions across 10 weeks to progress past the letters. For syllable content, the regular group saw steady progression from the start, leading to a plateau, reflecting the fact that the majority knew most or all syllable content at the third measuring point (about 50 play sessions across 15 weeks). By contrast, the at-risk group manifested a linear progression throughout the play period, with a median progression just past the halfway point for syllables at the end of the 25-week period. When it comes to word content, only the regular group achieved any noticeable progression, and it was not until the third measuring point that the majority had made any significant progress. At the end of the play period, the median proportion of known word items was 17.6% for the regular group and zero for the at-risk group.

One possible interpretation of these results is that there is too much content in the game for the average pupil to encounter during an extended play period. However, other factors also need to be considered. The time spent playing and the workings of the adaptation engine cannot be separated from content design when discussing progression. In reality, both at-risk and regular students should be able to complete the game, given enough time, and progression might have been faster if the adaptive algorithms had been adjusted. That said, faster progression does not equal a better game. Even though no measurement specifically pertains to learning, the content of GraphoGame is designed and distributed in line with established theories of language learning that suggest that a slower pace of progression may be better in regard to long-term learning outcomes. In other words, content design is not just about the amount of content included, but also about instructional aspects leading to mastery of fundamental elements essential to reading.

The time factor can be seen as consisting of two aspects: how much time is spent playing and how the adaptation engine adjusts the measure of player performance between game sessions. In this case, playing GraphoGame was an activity carried out as part of regular classes, with 10 minutes allocated to play and 5 minutes for the teacher to start up the activity. The data collected show that of their 10 play minutes, players devoted approximately 6 minutes to engaging in trials. The data do not show how much time students devoted to other activities in the game, such as browsing the in-game store or customizing their avatar, but this should account for a significant part of the remaining four minutes. Adding a single minute of effective gameplay to each play session would increase the total number of trials played during the entire play period by 800–1200, if all other factors remained the same.

The short bursts of play a few times a week are engulfed in large chunks of time when the students are not playing. During that time, knowledge may regress, meaning that what was known at the end of the last play session may not have been retained over time. As mentioned above, although the game stores knowledge about each item between play sessions, player performance resets at the start of each new session. Although player performance is not the only factor influencing content generation, this does mean that a player will have to perform a few trials before attaining his or her previous performance level. The likelihood of being presented with easier content from lower dimensions and content types is thus higher at the beginning of a new session. Considering that a typical play session involves only around 70–90 trials, this design choice may play a significant role for the pace of progression.

Adaptation Design

The adaptive algorithm in GraphoGame is calibrated to let players encounter items and correctly identify them enough times for it to be reasonably certain that the item in question has been mastered. In order to progress through the game, a player needs to know most items from lower dimensions and to perform better than the target percentage for each content type. This means that if a player makes a few wrong choices, the adaptation engine will provide already-known content from lower dimensions. This slows down the pace of progression but may result in better long-term learning outcomes as the adaptive algorithm ensures that the player knows all of the letters that make up items in the syllable and word content. This principle may reflect a good design choice, as it ensures, for example, that players' progression is not overextended. However, the adaptation engine needs to serve two distinct purposes. Besides preventing students from progressing too fast, it must also correctly and effectively identify items as known to ensure a steady rate of progression and avoid students being held back (which may cause them to become bored). GraphoGame has a single algorithm for all players, which means that this serves as a significant factor in progression.

The players' progression as observed over the 25-week play period opens up interesting inquiries, such as how much content should be included in relation to the total play period and how long and how frequently should the students play the game. The discrepancy found here between the regular group and the at-risk group brings additional questions to the fore, such as whether, and if so how, the adaptation engine could/should be altered to promote better learning in players of varying skill levels, or whether the learners' progression would be better served by developing multiple versions of that engine. Although answering those questions falls outside the scope of this paper, my surfacing the questions has resulted from the juxtaposition of the study of game design and content with the study of outcomes of use, suggesting that the overall design of the present paper, with two studies of different types, has been a fruitful one.

The Activity of Playing GraphoGame

Activity theory establishes that human activity occurs in social contexts. From this perspective, the activity of playing GraphoGame may expand beyond the direct interaction between students and the game. On the one hand, the teachers who used GraphoGame as part of early reading instruction extrinsically influenced the activity, as did the researchers who determined how GraphoGame should be engaged (e.g., 10-minute play sessions 3-4 times a week). On the other hand, GraphoGame's developers and content designers intrinsically influenced the activity through the design of the game (e.g., adaptation design, trial presentation, content organization). In other words, some actors influenced the immediate environment of play, whereas others extended their influence through GraphoGame's code. A shared motive between the influencing actors may be that playing GraphoGame helps students on their path toward learning to read by strengthening their knowledge of letter–sound correspondences. As the students act with the game by following their own objectives (e.g., solving tasks, earning coins) and using the various tools the game offers, they, perhaps inadvertently, work towards this motive.

CONCLUSIONS

The median progression through GraphoGame in terms of items known was approximately 50%, with the at-risk group progressing significantly more slowly than the regular group. The juxtaposition of the two studies carried out yielded three factors that may provide part of the explanation for this finding: (a) the amount and structure (types and dimensions) of the content established the boundaries for maximum progression, whereas (b) the time spent playing the game and (c) the way the adaptation engine worked influenced how far through the game the students progressed. The factors discussed are not weighted in terms of their influence on progression. This does not mean that they are of equal influence, but rather that their relative importance is not considered here. Further, the progression as measured is not intended to reflect the quality of the factors discussed.

This approach to research is important yet underrepresented in the research field, where the prevailing approach is to measure output from use without regard to the underpinning factors behind this output. Drawing connections between specific game-design components and output can be beneficial in many ways. For teachers, insights of such connections may inform them of better ways of deploying the game in the classroom. For example, knowing that at-risk pupils progress significantly slower may be used to provide them with additional play time or adult supervision during play. For game developers, there are insights that may inform of ways to improve the game. For example, the content in the game may benefit from being restructured and the possibility of different versions of the game to better suit different players may also be considered.

To draw such connections, game design needs to be researched and understood with equal emphasis as on designing appropriate research means to capture data about players' interaction with specific game components. As this research has demonstrated, such connections can reveal important and actionable data that may be implemented to improve similar interventions. Further, this research revealed knowledge that may lead to better informed decisions by those who use GraphoGame, for example, as part of language instruction in the classroom or when designing studies.

Although various factors influencing progression have been brought to light, much still resides unknown in the black box. For instance, what activities are the students engaged with in the 4 minutes during each play session when they are not actively engaged with trials? Further, does the knowledge GraphoGame holds in its data banks align with knowledge measured outside of the game? Last, what are the reasons for the discrepancy in progression between the regular and at-risk students? These are questions where the answers are still obscured, and thus warrant further research.

IMPLICATIONS FOR RESEARCH

This paper provides an example of how studies of game design and outcome from use may be juxtaposed. This research also highlights how such an approach may yield insights about a game that are beneficial to researchers, game developers, and educators. Although my data highlight the benefits for GraphoGame, other serious games could benefit as well. Similar studies should be carried out to establish a broader understanding of which components in a game are linked to outcome as measured. Even mainly experimental studies can be expanded in scope to include knowledge that may feed back into design or be applied in educational settings. Researchers should strive to develop methodologies specifically designed to enable the drawing of parallels between game design and outcomes from use.

ENDNOTES

1. The nine game modes are balloon game, fishing game, flow game, basic game, ladder game, pirate game, race game, star game, and basic GraphoGame. Each game mode has slight variations in game play and how the multiple-choice tasks work.
2. The likelihood of letter content is between 33% and 100%. If the player knows all the letters, the likelihood is between 5% and 100%.
3. If the player performance is higher than 0.7, the likelihood of each type of content is recalculated in favor of word content. Further, if the performance is even better (higher than 0.9), the likelihood of syllable content is capped at 30%.
4. The recalculated formula for the likelihood of words takes into consideration the percentage of known syllables and the current performance with syllable content. Higher values indicate higher probability for word content.
5. The chance for word content equals 1 minus the chance for letters and syllables.

REFERENCES

Arnab, S., Lim, T., Carvalho, M. B., Bellotti, F., Freitas, S., Louchart, S., Suttie, N., Berta, R., & De Gloria, A. (2015). Mapping learning and game mechanics. *British Journal of Educational Technology*, *46*, 391–411. https://doi.org/10.1111/bjet.12113

Bannon, L. J., & Bødker, S. (1991). Beyond the interface: Encountering artifacts in use. In J. M. Carroll (Ed.), *Designing interaction: Psychological theory at the human-computer interface* (pp. 1–32). New York, NY, USA: Cambridge University Press.

Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief*. Washington DC, USA: SRI International. Retrieved from https://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf

Biesta, G., Edwards, R., & Allan, J. (2014). Introduction: The theory question in education and the education question in theory. In G. Biesta, J. Allan, & R. Edwards (Eds.), *Making a difference in theory: The theory question in education and the education question in theory* (pp. 1–8). London, UK: Routledge.

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Ribeiro, C., & Pereira, J. (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. https://doi.org/10.1016/j.compedu.2015.11.003

Carvalho, M. B., Bellotti, F., Berta, R., De Gloria, A., Sedano, C. I., Hauge, J. B., Hu, J., & Rauterberg, M. (2015). An activity theory-based model for serious games analysis and conceptual design. *Computers & Education*, *87*, 166–181. https://doi.org/10.1016/j.compedu.2015.03.023

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. (2014). *Digital games, design, and learning: A systematic review and meta-analysis* [executive summary]. Retrieved from http://www.sri.com/sites/default/files/publications/digital-games-design-and-learning-executive_summary.pdf

Connolly, T. M., Boyle, E., MacArthur, E., Hainey, T., & Boyle, J. M. (2010). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, *59*, 661–686. https://doi.org/10.1016/j.compedu.2012.03.004

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, *9*(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4

Eisenhart, M. (2006). Qualitative science in experimental time. *International Journal of Qualitative Studies in Education*, *19*(6), 697–707. https://doi.org/10.1080/09518390600975826

Engeström, Y. (1999). Activity theory and individual and social transformation. In Y. Engeström, R. Miettinen, & R. Punamäki (Eds.), *Perspectives on activity theory* (pp. 19–38). Cambridge, UK: Cambridge University Press.

Furberg, A., & Lund, A. (2016). En profesjonsfaglig digitalt kompetent lærer? Muligheter og utfordringer i teknologirike læringsomgivelser [A professional digital competent teacher? Possibilities and challenges in technology rich learning environments]. In R. J. Krumsvik (Ed.), *Digital læring i skole og lærerutdanning* (pp. 26–43). Oslo, Norway: Universitetsforlaget.

Gaydos, M. (2015). Seriously considering design in educational games. *Educational Researcher*, *44*(9), 478–483. https://doi.org/10.3102/0013189X15621307

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6–10.

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), *Advances in psychology: Orthography, phonology, morphology, and meaning* (pp. 67–84). Amsterdam, the Netherlands: Elsevier North Holland Press. https://doi.org/10.1016/S0166-4115(08)62789-2

Lämsä, J., Hämäläinen, R., Aro, M., Koskimaa, R., & Äyrämö, S. M. (2018). Games for enhancing basic reading and maths skills: A systematic review of educational game design in supporting learning by people with learning disabilities. *British Journal of Educational Technology*, *49*(4), 596–607. https://doi.org/10.1111/bjet.12639

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, MA, USA: Harvard University Press.

Latour, B. (1999). *Pandora's hope: Essays on the reality of science studies*. Cambridge, MA, USA: Harvard University Press.

Leontiev, A. (1977). Activity and consciousness. In *Philosophy in the USSR: Problems of dialectical materialism* (pp. 180–202). Moscow, Russia: Progress Publishers.

Lundetræ, K., Solheim, O. J., Schwippert, K., & Uppstad, P. H. (2017). Protocol: "On Track," a group-randomized controlled trial of an early reading intervention. *International Journal of Educational Research*, *86*, 87–95. https://doi.org/10.1016/j.ijer.2017.08.011

Marne, B., Wisdom, J., Huynh-Kim-Bang, B., & Labat, J.-M. (2012). The six facets of serious game design: A methodology enhanced by our design pattern library. In A. Ravenscroft, S. Lindstaedt, C. D. Kloos, & D. Hernández-Leo (Eds.), *21st century learning for 21st century skills* (pp. 208–221). Heidelberg, Germany: Springer.

McTigue, E., Zimmer, W., Solheim, O., & Uppstad, P. H. (2019). Critically reviewing GraphoGame across the world: Recommendations and cautions for research and implementation of computer-assisted instruction for word reading acquisition. *Reading Research Quarterly*. Advanced Online Publication. https://doi.org/10.1002/rrq.256

Nardi, B., & Kaptelinin, V. (2006). *Acting with technology: Activity theory and interaction design*. Cambridge, MA, USA: MIT Press.

Phakiti, A. (2014). *Experimental research methods in language learning.* New York, NY, USA: Bloomsbury Publishing.

Price, S., Jewitt, C., & Brown, B. (2013). Introduction. In S. Price, C. Jewitt, & B. Brown (Eds.), *The SAGE handbook of digital technology research* (pp. 1–5). Los Angeles, CA, USA: SAGE Publications.

Richardson, M. (2016). Pre-hacked: Open design and the democratisation of product development. *New Media & Society*, *18*(4), 653–666. https://doi.org/10.1177/1461444816629476

Richardson, U., & Lyytinen, H. (2014). The GraphoGame method: The theoretical and methodological background of the technology-enhanced learning environment for learning to read. *Human Technology*, *10*(1), 39–60.

Saarela, M., & Kärkkäinen, T. (2017). Knowledge discovery from the programme for international student assessment. In A. Peña-Ayala (Ed.), *Learning analytics: Fundaments, applications, and trends—A view of the current state of the art to enhance e-learning* (pp. 229–267). Cham, Germany: Springer International Publishing. https://doi.org/10.1007/978-3-319-52977-6_8

Schomaker, L., Nijtmans, J., Camurri, A., Lavagetto, F., Morasso, P., Benoît, C., Guiard-Marigny, T., Le Goff, B., Robert-Ribes, J., Adjoudani, A., Defée, I., Munch, S., Hartung, & K., Blauert, J. (1995). Taxonomy of multimodal interaction in the human information processing system (A Report of the Esprit Project 8579 MIAMI – WP 1 – February). Retrieved from https://www.researchgate.net/publication/245104638_A_Taxonomy_of_Multimodal_Interaction_in_the_Human_Information_Processing_System

Su, Y., Feng, L., Hsu, C.-H., & Yang, C.-C. (2013). Technology as an activity system for self-directed, expanding learning: An analysis based on activity theory. *Procedia - Social and Behavioral Sciences*, *106*, 2575–2582. https://doi.org/10.1016/j.sbspro.2013.12.296

Vygotsky, L. S. (1978). *Mind and society: The development of higher psychological processes*. Cambridge, MA, USA: Harvard University Press.

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, *105*(2), 249–265. https:/doi.org/10.1037/a0031311

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3

Author's Note

All correspondence should be addressed to
Morten Njå
Norwegian Centre for Reading Education and Research
University of Stavanger
Professor Olav Hanssens vei 10, N-4036 Stavanger, Norway
morten.nja@uis.no

Human Technology: An Interdisciplinary Journal on Humans in ICT Environments
ISSN 1795-6889
www.humantechnology.jyu.fi

APPENDIX A: LEARNING CONTENT

These tables show the items and dimensions for each content type in the Norwegian version of GraphoGame.

Letter Items (3 dimensions)

1. i, l, s, o, e, a, m, r
2. u, t, b, f, n, v, k, å
3. h, p, d, g, æ, y, ø, j

Syllable Items (22 dimensions)

1. er, is, om, or, os, la, sa, le, se, li, ri, lo, ro
2. at, av, en, et, ik, ul, ur, ål, år, ås ~ er, is, om, or, os, la, sa, le, se, li, ri, lo, ro
3. ta, be, fe, te, ni, ti, vi, bo, mo, to, ru, lå, må, rå, så ~ at, av, en, et, ik, ul, ur, ål, år, ås
4. ut, ku, åk, åt, få, nå, tå ~ ta, be, fe, te, ni, ti, vi, bo, mo, to, ru, lå, må, rå, så
5. ha, gi, og, do, jo, yr, yt, ly, ry, sy, gå, øl, øs, bø, kø, mø, på ~ ut, ku, åk, åt, få, nå, tå
6. du, by, fy, ny, bæ, hæ, øk, øm, ør, dø ~ ha, gi, og, do, jo, yr, yt, ly, ry, sy, gå, øl, øs, bø, kø, mø, på
7. lam, lar, sal, ler, mer, ser, les, sel, ser, rom, som, mor, ror, los, mos, ros, mot, ris, lim
8. lat, mat, lav, rav, lag, sag, sak, men, ren, sen, let, lek, let, lev, lik, rik
9. sik, lin, liv, rim, riv, siv, rot, lur, rur, mur, sur, lus, mus, sur
10. rop, lyr, myr, syr, lyn, lys, syk, syl, syn, syr, syt, lær, nær, sær, løk, søk
11. røm, søm, mør, rør, sør, røv, møt, søl, søt, måk, råk, mål, lår, mår, rår, sår
12. lås, låt, lån, råd
13. kav, nav, tak, tap, tar, fet, vet, ber, fik, vik, fis, nil, vin, bom, kom, tom
14. vom, bor, kor, for, kos, bok, bot, tok, tom, kul, bur, kur, tur, nut, tut
15. tid, bod, tog, byr, fyr, tyr, fyk, nys, nyt, vær, bær, tær, vær, bøk, tøm, bør
16. før, tør, kåk, våk, bål, kål, nål, tål, får, kår, når, tår, vår, bås, vås, båt
17. det, hat, han, har, den, pen, het, der, her, dis, gir, dor, hos, gul, hul, jul
18. dur, jur, duk, dum, dun, dus, hus, jus, pus, dyr, hær, døm, dør, hør, døv, død
19. hål, går, hår
20. ei, øy, au, ai ~ er, is, om, or, os, la, sa, le, se, li, ri, lo, ro
21. kai, sau, rau, tau, lei, nei, sei, bøy, eng, ung
22. hai, hoi, hei ~ ha, gi, og, do, jo, yr, yt, ly, ry, sy, gå, øl, øs, bø, kø, mø, på

Note: Items following ~ are picked only as distractors.

Word Items (90 dimensions)

1. sale, lese, sele, more, lose, mose, rose, rise, lime, late, mate, rave, lage
2. sage, sene, leve, like, rike, line, rime, rive, rote, lure, mure, sure, rope
3. lyne, lyse, syke, syre, syte, lære, nære, sære, søke, møre, røre, røve, møte
4. søle, søte, måke, måle, såre, låse, låte, låne, råde, kave, tape, fike, vike
5. fise, bore, kore, fore, kose, kule, tute, fyre, fyke, nyse, nyte, være, tære
6. føre, våke, kåre, tåre, våre, våse, hate, hane, hare, pene, gire, gule, dure
7. dyre, døde, kake
8. inn, finn, katt, satt, sett, sitt, bitt, ditt, mitt, mett, pytt, nøtt, natt, titt
9. sopp, sokk, rett, vått, kopp, hopp, topp, mopp, søtt, surr, kott, latt, møkk
10. tørr, tøff, røff, biff, paff, puff, nøff, loff, voff, rigg, pugg, pigg, mygg
11. rygg, vegg, logg, legg, rakk, pakk, pukk, nikk, møkk, vekk, sekk, dykk, dokk
12. bukk, dekk, bekk, hakk, lakk, hekk, voll, null, møll, rull, tull, tall, fyll
13. fall, byll, ball, vinne, penn, tynn, tonn, tinn, tenne, tann, sånn, munn, fonn
14. finne, lønn, rodd, ridd, papp, pupp, napp, lapp, narr, pass, lass, puss, tass
15. buss, boss, bass, gass, ratt, nett, tett, tatt, fått, kutt, kott, matt, nebb
16. mobbe, jobbe, kubbe, rydde, redd, vidde, sydde, gidde, lodd, ledd, loppe, lamme
17. romme, losse, risse, matt, menn, renn, lett, lekk, rikke, rotte, murre, surre
18. sykkel, løkke, søkke, rømme, møtte, søtt, takk, tapp, vinn, bomme, komme, tomme

19. nytt, dette, hatt, hann, denne, penn, gull, dukke, dumme, pusse, dømme
20. ul, ull,
21. ak, akk
22. tet, tett
23. tut, tutt
24. tyn, tynn
25. tør, tørr
26. os, oss
27. mur, murr
28. fin, finn
29. søt, søtt
30. sur, surr
31. tak, takk
32. pen, penn
33. pir, pirr
34. ven, venn
35. bake, bakke
36. bøte, bøtte
37. duke, dukke
38. halen, hallen
39. ire, irre
40. kake, kakke
41. knipe, knippe
42. kul, kull
43. lake, lakke
44. leke, lekke
45. lese, lesse
46. lose, losse
47. luke, lukke
48. mase, masse
49. møte, møtte
50. nepe, neppe
51. nise, nisse
52. nyte, nytte
53. nåde, nådde
54. pine, pinne
55. pute, putte
56. rape, rappe
57. rate, ratte
58. rede, redde
59. reke, rekke
60. ripe, rippe
61. same, samme
62. sipe, sippe
63. slipe, slippe
64. søke, søkke
65. tele, telle
66. vane, vanne
67. vase, vasse
68. vipe, vippe
69. gran, grand
70. gren, grend
71. grin, grind
72. hån, hånd
73. mil, mild
74. lin, lind
75. lun, lund
76. mel, meld
77. ran, rand
78. sen, send
79. sil, sild
80. syn, synd
81. tin, tind
82. ven, vend
83. vin, vind
84. von, vond
85. laus, saus, leir, leik, leit, meis, taus, feil, neie, bøye, føye, køye, nøye
86. tøye, eng, ung, bang, fang, lang, sang, tang, seng, ring, ting, ving, tung, haik
87. heil, heis, peile, peis, hoie, gøye, høye, naust, gang, pang, heng, pung, syng
88. kjas, kjøl, kjøp, sjø, sju, sjuk, sjal, sjef, sjakk, skje, kjapp, sjokk ~ kav, kø, kår, sør, sur, sang, seng, sol, sal, ser
89. hjelp, hjelm, hjørne, gjær, gjesp, gjedde, skjul, skjenn, skjedd ~ her, hør, jul, går, sur, ser, seng
90. sjakt, sjark, sjau, skjørt, skjold, skjerm, skjorte, skjerf ~ saft, sau, sot, ser

Note. Dimensions 20–84 are minimal pairs. Items following ~ are picked only as distractors.

APPENDIX B: PROGRESSION TABLES

The following tables show the number of items known at each of the five measuring points (M1 … M5) for all items combined and items for each content type. The tables show both the item count and the percentage of total items. Further, these data are divided between the regular group and the at-risk group, and include the 25, 50 (median) and 75 percentiles, as well as the minimum and maximum values.

Table B1. Overall User Progression at Each Measuring Point.

Group		M1		M2		M3		M4		M5	
		count	%	count	%	count	%	count	%	count	%
Regular	Minimum	13	1.7	22	3.0	23	3.1	40	5.4	47	6.4
	25 percentile	70	9.5	124	16.9	198	27.1	249	34.1	280	38.3
	50 percentile/Mdn	114	15.6	193	26.4	282	38.6	320	43.8	363	49.7
	75 percentile	189	25.8	292	40.0	375	51.3	489	66.9	507	69.4
	Maximum	332	45.4	500	68.4	663	90.8	703	96.3	706	96.7
At Risk	Minimum	5	0.6	9	1.2	4	0.5	8	1	13	1.7
	25 percentile	23	3.1	39	5.3	58	7.9	80	10.9	87	11.9
	50 percentile/Md	42	5.7	60	8.2	107	14.6	131	17.9	166	22.7
	75 percentile	53	7.2	104	14.2	139	19.0	163	22.3	222	30.4
	Maximum	77	10.5	150	20.5	228	31.2	255	34.9	297	40.6

Table B2. Letter–Content Progression at Each Measuring Point.

Group		M1		M2		M3		M4		M5	
		count	%	count	%	count	%	count	%	count	%
Regular	Minimum	12	50.0	14	58.3	13	54.2	15	62.5	18	75.0
	25 percentile	24	100.0	24	100.0	24	100.0	24	100.0	24	100.0
	50 percentile/Mdn	24	100.0	24	100.0	24	100.0	24	100.0	24	100.0
	75 percentile	24	100.0	24	100.0	24	100.0	24	100.0	24	100.0
	Maximum	24	100.0	24	100.0	24	100.0	24	100.0	24	100.0
At Risk	Minimum	5	20.8	9	37.5	4	16.6	8	33.3	13	54.2
	25 percentile	15	62.5	18	75.0	24	100.0	21	87.5	23	95.8
	50 percentile/Mdn	20	83.3	23	95.8	24	100.0	24	100.0	24	100.0
	75 percentile	24	100.0	24	100.0	24	100.0	24	100.0	24	100.0
	Maximum	24	100.0	24	100.0	24	100.0	24	100.0	24	100.0

Table B3. Syllable–Content Progression at Each Measuring Point.

Group		M1		M2		M3		M4		M5	
		count	%	count	%	count	%	count	%	count	%
Regular	Minimum	0	0.0	8	2.9	10	3.7	24	8.8	29	10.7
	25 percentile	46	16.9	100	36.8	174	64	224	82.4	250	91.9
	50 percentile/Mdn	90	33.1	169	62.1	251	92.3	263	96.7	266	97.8
	75 percentile	165	60.7	259	95.2	266	97.8	270	99.3	270	99.3
	Maximum	271	99.6	271	99.6	272	100	272	100	272	100
At Risk	Minimum	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
	25 percentile	9	3.3	19	7.0	34	12.5	56	20.6	67	24.6
	50 percentile/Mdn	23	8.5	37	13.6	83	30.5	111	40.8	142	52.2
	75 percentile	30	11	80	29.4	115	42.3	140	51.5	198	72.8
	Maximum	53	19.5	126	46.3	203	74.6	232	85.3	266	97.8