Article

Case Studies in Game-Based Complex Learning

Josh Aaron Miller * [ID] and Seth Cooper [ID]

Khoury College of Computer Sciences, Northeastern University, Boston, MA 02115, USA;
se.cooper@northeastern.edu
* Correspondence: miller.josh@northeastern.edu

Abstract: Despite the prevalence of game-based learning (GBL), most applications of GBL focus on teaching routine skills that are easily teachable, drill-able, and testable. Much less work has examined complex cognitive skills such as computational thinking, and even fewer are projects that have demonstrated commercial or critical success with complex learning in game contexts. Yet, recent successes in the games industry have provided examples of success in game-based complex learning. This article represents a series of case studies on those successes. We interviewed game designers Zach Gage and Jack Schlesinger, creators of Good Sudoku, and Zach Barth, creator of Zachtronics games, using reflexive thematic analysis to thematize findings. We additionally conducted a close play of Duolingo following Bizzocchi and Tanenbaum's adaptation of close reading. Several insights result from these case studies, including the practice of game design as instructional design, the use of constructionist environments, the tensions between formal education and informal learning, and the importance of entrepreneurialism. Specific recommendations for GBL designers are provided.

Keywords: case studies; complex learning; game-based learning; gamification

1. Introduction

Although game-based learning (GBL) has become widely popular in recent years, few applications of GBL have considered complex cognitive skills such as computational thinking and language learning. Of these, even fewer have been commercially or critically successful with a wider audience. If the field of GBL is to succeed in teaching more complex skills, we need to understand what aspects of design make them successful.

Our background in game-based learning comes from scholars like Gee (e.g., [1,2]), as well as theories on constructionist and constructivist learning [3], and instructional design for complex learning [4]. Similar work includes the understanding of educational games compared to commercial games [5] and the recent work on resonant games by Klopfer et al. [6], which attempts to present design recommendations for the design and use of educational games in the classroom. From this theoretical background, we conclude that while many design principles are being developed for GBL more generally, there is as of yet scarce information on what makes GBL applications (a) commercially successful, or (b) successful in complex domains such as computational thinking and language learning.

Fortunately, several successful examples of game-based complex learning have recently emerged. Duolingo is a gamified language learning app originally developed in 2011 and continuously developed since then [7]. Good Sudoku is an iOS app about learning and enjoying Sudoku, developed by Zach Gage and Jack Schlesinger in 2020 [8]. Additionally, Zach Barth, designer of Zachtronics, has created several engineering puzzle games that teach computational thinking and general problem-solving, including SpaceChem (2011), SHENZHEN I/O (2016), and Opus Magnum (2017) [9]. These examples were selected for their popularity as exemplars of complex learning in gamified and gameful applications. They are further identified for their educational applications: for example, Duolingo offers "Duolingo for Schools" (https://blog.duolingo.com/duolingo-for-schools/ accessed on 4 October 2021) and Zachtronics offers "Zachademics" (https://www.zachtronics.com/

zachademics/ accessed on 4 October 2021). Despite Good Sudoku's infancy, Gage's earlier games have been critically praised for their ability to turn learning into an enjoyable experience [10,11].

Therefore, in this article we ask the research questions: what makes these games educationally successful? How are the designers conceptualizing their work in order to create these successful outcomes? To answer these questions, we conducted two semi-structured interviews with the designers of Good Sudoku and Zachtronics, respectively, analyzing their responses through a reflexive thematic analysis. We supplement these interviews with a "close play" [12] of Duolingo to understand how a gamified context differs from a gameful context.

We found that in both gamified and gameful approaches to GBL, commercial thinking makes a great impact on GBL success, such as identifying the market value of the application and putting effort into outreach.

In Good Sudoku and Zachtronics' games, the game design of these games *is* the instructional design, as shown through its visual design, emphasis on whole-task practice, and removal of busywork. The instructional design further seeks to establish a constructionist learning environment by providing just-in-time (JIT) information, letting the players drive their own learning and difficulty, and teaching general problem solving. Lastly, the designers noted how the formalities of education and research can obstruct what they saw as the critical learning processes, which included a clear tutorial, experiential learning, and appreciating the domain before going deeper into its nuances. Based on these findings, we make several recommendations in Table 1, including, among others: teach gradually and visibly, iterate, and acknowledge the constraints of education and research.

Table 1. A summary of findings and recommendations from the thematic analysis of case studies in complex GBL.

	Findings and Recommendations
Lessons from Duolingo	**GBL products exist in a real-world market** Adapt to real-world constraints and demands to grow and sustain the product.
	Teach gradually and visibly Be transparent about the learning content. Start simple and gradually increase complexity. Provide ample and thorough feedback.
Holistic game design is instructional design	**Teach through visual design** Use visual design to help players parse and focus on critical information.
	Use varied whole-task practice with emphasis manipulation Focus on one technique at a time in the context of whole, meaningful tasks. Provide varied tasks for learners to practice skills in different contexts.
	Remove busywork Automate routine problem-solving steps that are extrinsic to the skill being practiced, especially perceptual and memory effort.
	Iterate Playtest prototypes early and often with new players; address confusions and UX pain points.

Table 1. *Cont.*

	Findings and Recommendations
Constructionist Environments	**Provide only JIT information** Withhold detail until it is useful in the player's exact situation, and even then provide only as much as is helpful for their situation.
	Let players drive learning Provide information only on player request. Design scenarios where players will discover skills on their own through problem-solving.
	Let players drive difficulty Instead of adjusting the tasks directly (by automatic DDA or difficulty options), provide a range of goals for players to aim for.
	Teach general problem solving Put the model before the problem by providing constraints in an open space instead of specific challenges that appear to afford using superficial strategies.
Do not let education prevent learning	**It's okay to be difficult** Do not be afraid to challenge the player.
	The tutorial should not be difficult Teach new concepts in isolated, structured environments that let the players actively use the new concept without challenge.
	Teach how it's loved, not how it's taught Focus the gameplay first on why the domain is fun or interesting rather than jumping into teaching its nuances.
	Games work as experiences, not content delivery Leverage the interactive, system-exploring nature of games to encourage experiential learning. Minimize content delivery or delegate it to other media. For transfer learning and transformative reflection, include explicit debriefing after the experience.
Success says sell the GBL	**Acknowledge constraints of education and research** Pre-defined learning goals and research questions will likely hinder the product quality unless the project has sustainable funding and ample iteration time.
	If you want buy-in, you have to sell it Work with professional game designers and marketers to design and pitch your game's value. Put effort into outreach and advertisement.

2. Materials and Methods

Two semi-structured interviews were conducted: one with Zach Barth, lasting approximately 1:16 h, the other with Zach Gage and Jack Schlesinger, lasting approximately 1:55 h. Interview questions were tailored to the participants' experiences and creations with respect to the research questions. All methods were approved by the researcher's institutional ethics board, and all participants consented to having their names and quotations be published in this article, which included reviewing this article to ensure that the work is an adequate and accurate representation of their sentiments.

The interviews were analyzed using reflexive thematic analysis [13,14]. We took a deductive approach oriented to both semantic and latent meanings with an emphasis on the discourse [15] used by participants. This could be considered a contextualist or critical realist mindset, sitting between essentialism (the reality of the participants) and constructionism (their experiences as shaped by society).

A close reading, or "close play" was conducted of Duolingo [16] by the first author following the methods described by Bizzocchi and Tanenbaum [12] and later applied by Sullivan and Salter [17]. This qualitative method allows a researcher to deeply examine a work both as a consumer of that media (in this case, a game player) and as a researcher contextualizing the player's experience analytically. The close play focused on Duolingo's Spanish course (one of the original and most thorough courses on Duolingo) and involved playing for 5–20 min each day for over 400 days. This supplementary analysis of a gamified, as opposed to gameful, application provides insight by contrast, allowing us to understand better how each context is functioning in relation to the other.

Games Analyzed

Since this article discusses several games, it is worth briefly describing each in turn. Duolingo is a gamified language learning application playable on mobile devices and in web browsers. Players progress through lessons containing about ten to twenty exercises. Each exercise practices some aspect of language learning, from speaking and listening to writing and comprehending. This core loop is gamified with experience points (XP), leaderboards, and playful animations.

Good Sudoku is an iOS application for playing Sudoku, a popular number-placement puzzle. The game consists of a 9×9 grid of cells divided into 3×3 subgrids ("boxes"). Each puzzle is populated with an initial few digits scattered throughout the grid and the goal is to fill in the rest of the grid using deduction and other logic principles such that each row, column, and box contain the digits 1 through 9 with no duplicates. Good Sudoku builds on traditional Sudoku applications by adding novel note-taking, hint systems, tutorials, and game modes that scaffold the player toward better understanding the strategies of Sudoku-solving.

Zachtronics games, including Opus Magnum, SHENZHEN I/O, and SpaceChem, are difficult open-ended logic puzzles. In these games, the player is tasked with engineering a machine from composite units in order to produce some logical output. For example, in SHENZHEN I/O, the player constructs circuits of predefined input and output requirements using a limited array of logic gates and similar components.

3. Results

3.1. Close Play of Duolingo: Gamification of Complex Learning

To borrow the language from rhetorical structure theory [18], the nucleus of Duolingo is its skill tree, a directed acyclic graph of lessons ordered from the easiest introductory vocabulary into the language to the most complex grammatical structures Duolingo teaches. (Notably, Duolingo does not teach language fluency, but typically reaches A2–B1 levels based on the CEFR language education standard, according to their 2019 blog post: https://blog.duolingo.com/how-are-duolingo-courses-evolving/. Accessed October 26, 2021) This is both the core learning mechanic and the core gamification mechanic, since primary play centers around completing lessons within this skill tree to make "permanent" progress on mastering a skill—however, once a skill has been fully mastered, it will occasionally "break", encouraging the user to return and review the skill. Exactly when a break occurs is based on their spaced repetition model [19].

3.1.1. Gamification Mechanics

There are several satellite systems supporting the skill tree. First, users earn XP for all learning exercises. In gamification literature, XP would be considered simple "points" that the user can earn [20]. This is the most fundamental aspect of Duolingo's gamification, since XP earning can belie constant upward progress despite the user's own struggles to master the material. However, XP for me was a meaningless metric, since the most optimal ways to earn it were antithetical to the optimal ways to learn. Practicing new skills or otherwise pushing myself cognitively was the least efficient way to earn XP, and the most

efficient was to mindlessly drill lessons or short stories that I have already completed and mastered beyond need for further practice.

Closely related to XP, the leaderboards show the user's XP earnings relative to a subset of other users to encourage competition. In combination with notifications when one is passed on the leaderboard, this can push the user to engage slightly more than they would otherwise. In practice, this was only effective for me when the competition was close and when there was a reward for coming in first through third place. When users ahead of me had an enormous lead, I made no attempt to catch up. Then, in late Spring of 2021, Duolingo removed rewards for leaderboard rankings, taking away the only reason I was engaged with the competition.

In accordance with classic PBL (points, badges, leaderboards), Duolingo awards a series of badges, or achievements, for certain milestones. These badges elicit the Zeigarnik effect [21], since using the application usually results in at least one achievement always being near completion, prompting the need for closure. While this was effective early on, Duolingo has very few badges to earn beyond the first month of play. I went several months without receiving any badges and thus felt no compulsions to work toward specific milestones.

Recent literature notes that heavy reliance on PBL—as opposed to other possible gamification elements—is likely because they are the easiest features to implement [22]. Yet, points-based rewards can have detrimental effects on motivation in the long-term [23,24]. Scholars conclude that gamification elements have differential, contextual effects, and a careful matching of context to incentive is required for effective gamification [25]. In the case of Duolingo, then, their simple approach to gamification may be effective for recruitment, but not for long-term retention.

Next, Duolingo offers a currency called gems (formerly lingots). Gems can be earned by doing a lesson once per day and completing a lesson segment (typically about four lessons for my Spanish course). To spend gems, there are a few aesthetic items (outfits for Duo the owl) and a couple of optional lessons (e.g., popular idioms). Additionally, once per week the user can wager some gems as a promise that they will practice every day for a week; after seven days of concurrent practice, their wager will be returned doubled. Yet, the most common use for gems in my experience was "testing out" of a skill. Since each skill offers ample lessons for practice, I found myself bored by the additional exercise. Instead, I could spend gems on a test of that skill, and if I passed the test, Duolingo would advance me to the next point of mastery for that skill (skipping three more lessons).

In practice, the high cost of testing felt like pressure to buy into Duolingo's premium services (where tests have no gem cost). This marred my experience with capitalist frustrations and removed me from a more optimal learning and playing experience. On top of this, when I was able to purchase a test, I felt stressed because of the high cost of gems at stake. This sometimes resulted in lower performance because the stress added cognitive load, resulting in me making mistakes that I would have recognized if I were more calm.

To further encourage daily practice, Duolingo makes prominent notice of the user's "streak": how many concurrent days they have met their own learning goal (a minimum of one lesson). Small gem bonuses are rewarded for maintaining this streak. Gems can also be spent to "freeze" the streak, either for a day or for the weekend, helping the user maintain their streak on days they choose to avoid engaging with the app. In practice, I never used the weekend freeze and rarely needed the daily freeze. However, the availability of freezing a streak felt like a quality-of-life feature that I was glad to have present when needed.

The last gamified system Duolingo uses is the heart system. Each mistake a user makes during practice removes one of their hearts. After all five of their hearts are lost, users must regain hearts by either: completing review lessons, paying gems, waiting several hours for their hearts to passively refill (a rate of approximately 5 hearts per day), or paying real money for Duolingo's premium services.

A recent review of persuasive design in mobile games lists several mechanics of player retention and their underlying psychological theories [26]. Based on loss aversion, the he-

donic treadmill, the goal-gradient hypothesis, the endowed progress effect, and two reward schedules (fixed interval and variable ratio), they identify ten retention mechanics: level-up rewards, dynamic experience, daily quests, quest-quests, retroactive quest introduction, interval resource collecting, energy system, random items, rotating shop, welcome gifts, fixed interval schedule login bonus, reward satiation, time-limited rewards, and exhibition. Although a full unpacking of these theories and mechanics is out of scope, we can identify which mechanics Duolingo uses for retention. The heart system is a form of energy system: by restricting the duration of a play session, the system incentivizes returning in order to make use of regenerated energy. The daily gem bonus given for maintaining one's streak is a daily quest. However, Duolingo does not make use of the other eight forms of persuasive mechanics, demonstrating its simplistic approach to persuasion and gamification.

There are likely other subsystems I could describe, such as the motivational messages (which I turned off for being disruptive) and the in-game advertisements; however, the systems I have described adequately capture the major themes of my experience with Duolingo, and further detail would muddy the analysis.

Overall, my experience of the gamification of Duolingo was one of frustration. There were no substantial gameful elements to engage with, having run out of badges and being disinterested in leaderboards for the sake of competition. Yet, what gamified elements remained seemed to exist only to thwart my learning experience. Pushing myself cognitively cost gems, cost hearts, and rewarded little XP. However, I continued to engage with the app for its learning mechanics, despite its gamification. According to its fan wiki, Duolingo has tried many other gameful and gamified techniques, including clubs (groups/guilds) and other social features, duels, and a "full heart" bonus, in addition to several abandoned learning features. One staff comment (https://forum.duolingo.com/comment/31849174?comment_id=31877425 accessed on 26 October 2021) suggests that these decisions are made as a capitalist-focused company considering the costs and benefits rather than taking a philosophy of creating purely the best product. Yet, given the massive success of Duolingo, perhaps this is an unfortunate, uncomfortable truth: that GBL products exist within a real-world market and must adapt to real-world constraints and demands, even if this means sacrificing some aspects of user experience, gamefulness, and learning optimization.

3.1.2. Learning Mechanics

Within the framing of Duolingo's gamification, the application uses several learning mechanics successfully. As with gamification, the nucleus of the application is the skill tree, which forms a hierarchy building from simple vocabulary and grammar to complex linguistic constructions (cf. skill hierarchies and skill chains in instructional design for complex learning [27,28]). In this way, Duolingo scaffolds lessons by ensuring that learners have the prerequisite skills to be prepared for more complex instruction. In practice, I found this to be extremely effective. Very rarely did it feel like there was any spike in difficulty; instead, learning was gradual and smooth, albeit a long, long path to fluency.

Before every lesson, Duolingo offers an optional text briefing on the contents of the lesson. This can be seen as cognitive priming or, using Gagné's nine events of instruction, the preparation phase: informing the learning objectives, stimulating recall of prior learning, and presenting the new content [3,29]. I found these guides helpful but insufficient. Often, I would return to them after making a mistake to try to understand better what linguistic nuance I was missing, only to re-discover that the guide did not mention that detail at all. Usually, the per-lesson forums would provide an answer, but it felt frustrating both to enter the lesson unprepared and to practice a concept without a formal explanation for it, relying only on previous experiences with the concept as decision-making cues.

Next, there is an abundance of varied task types. Task variety is typically recommended for both better engagement and more robust mental models [3,4,30,31]. In my Spanish course, I experienced short stories with comprehension questions, word matching, sentence completion, grammar drills, naming objects, listening exercises, speaking

exercises, and fully written translation exercises. If we were to extend the analysis to the wider Discourse of Duolingo, this list would also include the podcasts they produce and social practice activities organized by users in their forums and other affinity spaces such as their subreddit. Although I did not take part in social activities and rarely listened to the podcasts, I felt sufficiently engaged with the diversity of task types. The only major drawback I experienced was the lack of variety in vocabulary within lessons. For any given lesson, the number of sentences in that lesson was quite limited, and often I would repeat the same sentence multiple times in each task type (e.g., speak a sentence, then hear it and translate it, then see it and translate it). Despite the varied task types building on each other and reinforcing the material in multiple ways, the lack of content variety meant that I was often using rote memorization rather than actual language skills to perform the tasks. If I had just heard a sentence, I do not need to think about how to translate it, I only need to recall my aural short-term memory. The instructional designers themselves recognize this, since they recommend "hovering" around lessons, (https://blog.duolingo.com/whats-the-best-way-to-learn-with-duolingo/, accessed on 26 October 2021), which would provide the learner with more short-term variety across tasks. However, Duolingo has not integrated this technique into the application itself, placing the burden of lesson sequencing on the learner.

Finally, Duolingo offers sufficient corrective, but not cognitive, feedback (cf. [32]). After each exercise, Duolingo responds whether I was correct or incorrect: if I am correct, they may offer additional insights such as typos I made, accents I missed, or what the sentence translates to (if, for example, the goal was only to speak the sentence without parsing it). If I am incorrect, I am told what the correct answer is. On occasion, Duolingo attempts to identify where I made a mistake and provide further guidance on that concept. In practice, however, this worked only for the most basic concepts. Beyond the basic grammar of gendered articles and having adjectives agree with nouns in gender and number, there were rarely instances where Duolingo's tip identified the issue I was struggling with. Instead, I would either know immediately what I did wrong (ah, that noun is masculine not feminine, no I do not need a reminder about gendered nouns) or I would remain unsure (why is this preposition required here?). To Duolingo's defense, computationally identifying useful cognitive feedback remains a challenging problem [33]. However, perhaps a better approach would be simply answering the question: "What concepts that are new to this lesson are used in this exercise, and how and why?" As we will see in the analysis of Good Sudoku and Zachtronics games, this just-in-time approach can be used to great effect.

Holistically, after using Duolingo every day for more than a year, my understanding of Duolingo is that it has a solid nucleus of learning and gamification and struggles with its satellite scaffolding. By centering around a detailed, thorough, well-defined skill tree, progress was always clear, gradual, and rewarding, albeit slow. It is this core, I believe, that supports Duolingo's success as a gamified learning application. Where there is room for improvement, on the other hand, are its supporting systems. The other gamification features detract from the learning experience, and the other learning features are sparse and insufficient. If I were actually depending on Duolingo to learn another language, I would likely need a more formal curriculum, using Duolingo only as supplementary rather than primary didactic material.

I include this close reading for contrast. As we shift into the analysis of Good Sudoku and Zachtronics games, trying to understand their gameful approaches to learning and by what means they operate, I will use this reflection on Duolingo as a comparison. What makes gamification different from gameful? What makes mass-market different from indie? As we will see, both contexts focus on marketing the game for its value, and both contexts teach gradually (though Duolingo focuses on content delivery, while Good Sudoku and Zachtronics focus on experiences), and varied practice. Clear feedback, too, is a key component of successful GBL in any context. Furthermore, perhaps most importantly, the actual learning in these games and gamified apps is inherent to the task structure, not to the support around it. However, as will be described below, there are some

aspects which are unique to gameful GBL: whole-task practice, teaching general problem solving, and teaching with an emphasis first on appreciating the domain before exploring its complexities.

3.2. Thematic Analysis of Good Sudoku and Zachtronics Games

The reflexive thematic analysis described in Materials and Methods resulted in four themes summarizing the approaches that Zach Gage, Jack Schlesinger, and Zach Barth take in their design of complex learning games. Their descriptions suggest that *Holistic game design* is *instructional design*, i.e., the scaffolding techniques they use to onboard players in aggregate sums to good instructional design. Second, they use *constructionist environments* to allow players to explore, fail, learn, and generalize at their own pace and in their own ways. Third, they *do not let education prevent learning*. These designers value difficulty while simultaneously respecting the learning needs of their players. Moreover, they take creative liberties to teach in a manner they find more effective, rather than teaching in what might be a more traditional approach. In practice, this means designing for the enjoyment of the intrinsic material rather than with the goal of teaching that material. In doing so, their work is inspired by pedagogical theory, but not driven by it: they make their own interpretations of research and education and explicitly separate themselves from formal learning. Finally, they offer practical insights, such as the need for outreach and entrepreneurialism, an oft-ignored aspect of GBL research.

3.2.1. Holistic Game Design Is Instructional Design

Both Zach Gage and Jack Schlesinger have been teachers in the traditional sense. Schlesinger further has designed curricula, and in reflecting on these experiences they argue that game design *is* instructional design. "I have had ... hours-long discussion[s] ... over... the choice between two words in a rulebook..." says Schlesinger. It is these small considerations, they argue, which make up the whole of an instructional design. Furthermore, together, their many micro moves have a sort of Gestalt effect. Gage describes it as "one hundred different things and every single thing is like a feedback loop that's pointing to another learning tool."

Many of these design decisions are about the visual design of the game. For example, the note-taking system uses vibrant red and blue to identify which cells can be which digits—the player is drawn (sometimes unconsciously) to identify color contradictions as indicators of what moves can be made. Additionally, when a player selects a cell, the game highlights that cell's row, column, and box. When selecting a digit, other instances of that digit are highlighted throughout the grid. In these ways and others, the low-level decisions about the app's visual design contributes to reducing the player's perceptual and memory loads so that they can focus on critical information.

Moreover, Gage and Schlesinger describe their design as not only what they included, but what they consciously excluded. Other Sudoku apps, for example, highlight wrong inputs in red. However, this incentivizes guessing behavior and disincentivizes applying intelligent techniques. Instead, Good Sudoku allows the player to get into an unsolvable state, but detects when this occurs and allows them to backtrack to the last solvable step.

Not only do they use visual design to reduce perceptual load, they also teach perceptual skills as part of the complex task. Focus Mode, for example, demonstrates and guides players through the strategy of grid-scanning.

This approach to instructional design is reminiscent of common techniques in level design, which takes root in the field of architecture, using light, color, and space (among others) for aesthetic appeal and direction [34,35]. Although an unpacking of this connection is out of scope, instructional designers can look to architectural design and level design (see especially [35]) for ways to guide visual attention.

Next, Good Sudoku teaches via whole-task practice with emphasis manipulation [4]. That is, from the beginning, players engage with the whole Sudoku challenge and all of its mechanics. This is accomplished by several scaffolding mechanisms.

First, the difficulty levels of Good Sudoku are segmented and structured by the techniques they require, allowing each difficulty to provide emphasis on increasingly complex techniques. Second, by being explicit about this structure, the game primes the player for what challenges they can expect in their next puzzle. Third, when the game detects that a technique is needed, but the player has not used or learned about it yet, it recommends the tutorial they need. Each tutorial demonstrates a specific technique in situations that focus on the value of the technique and how it can transfer to other contexts. Then the tutorial lets you practice a specific technique repeatedly in a sandbox that generates practice examples in diverse contexts.

Finally, once the player is in the puzzle, an AI assistant provides hints on request. The AI is able to do this because the linear nature of Sudoku—combined with the note-taking feature—identifies exactly what step in the problem-solving process the player is at. With all of these design elements in combination, Gage and Schlesinger create a system loop: anticipate what the learner is trying to accomplish or practice, then provide opportunities for them to experience that challenge in a variety of contexts. Improve Mode builds on this loop by training note-taking and specific techniques through emphasis manipulation.

This learning loop—identify what the learner is trying to solve and help them solve it—is a particular affordance of the problem-solving of Sudoku. Whereas, in other tasks with more open problem-solving, either: (a) the student and teacher need to have a collaborative dialogue to establish the context, or (b) the teacher needs to provide an excess of information in case some information is useful to the situation, or (c) the teacher has to fabricate a specific and inauthentic context for providing the information, which can feel "super hand-holdy" (Gage).

Not only are learners engaging with whole tasks, they are offered a variety of means of practice. Interleaving practice in multiple contexts increases the contextual interference, which in turn builds more robust mental models for transfer learning [4]. Gage further notes that these different game modes (and Good Sudoku's three kinds of daily puzzles) also appeal to different kinds of learners (cf. learning styles [36–38]), especially since different pressures—such as time pressure or analysis paralysis—can add stress to different learners.

In order to support engagement with the critical task, Gage and Schlesinger advocate for removing "busywork." For example, the auto-note feature allows players to identify where digits can be legally placed, rather than forcing them through the tedium of counting every cell. Moreover, Good Sudoku's smart auto-completion feature selectively fills in cells that have only one legal move while making sure not to "step in front of somebody's thought process" (Gage) by performing any problem-solving steps for them. As Gage summarizes, "Not having to do that work frees your brain up from, like, being distracted by all this menial stuff to actually pay attention to the cooler, deeper structures of the thing."

Looking to instructional design theory, the removal of busywork is in line with Cognitive Load Theory, which states that cognitive processing extrinsic to the learning task is extraneous load that is best to eliminate [39]. By focusing the learner's attention, this technique provides a kind of scaffolding through simplification, allowing the learner to focus on the more intrinsic, structural features of the problem [4].

However, Gage and Schlesinger did not arrive at this design on their first try. They note that their many micro moves of mentorship required a lot of small iterations, such as finding the optimal way to phrase a hint. Similarly, Barth is constantly working toward a smooth onboarding, achieved via iteration. "The learnability of systems is always on our mind all the time," he explains, "...Because everything we design and present to the user, they have to be able to learn it. Furthermore, it's not really enough to just be like, 'oh yeah, there is a wiki, you can read about it all'—no, it has to be something that people can pick up smoothly and it has to make sense and it has to all fit together and it has to be intuitive to some degree."

To design a game is to design the instruction for that game, and these designers take a holistic approach consisting of dozens of different micro-features that play into the resulting smooth experience. Each scaffolding technique ties into the rest of the assemblage.

This design work includes: visual design guiding perception, whole tasks teaching in a variety of contexts, removing busywork to free cognitive resources, and iterating on the whole experience. The result is a game that teaches holistically and that is fun to learn from. As Barth says, quoting online comments about their games, "the whole game is the tutorial"—because, as Barth argues, "what makes games fun is the learning aspect of it."

3.2.2. Constructionist Environments

Although they do not explicitly describe it as such, these complex learning games bear striking resemblance to the design of constructionist learning environments (cf. [2,40]). Rather than focusing on content delivery, these designs try to help the players discover the intrinsic value of the games' subjects while letting them set their own direction and pace. Moreover, Gage, Schlesinger, and Barth all attempt to teach a generalized problem-solving mindset rather than teach how to solve specific problems, thus encouraging the learner to construct their own problem-solving strategies.

Gage nearly definitionally describes using just-in-time (JIT) information: "Only give people information when they ask for it and only give them the exact information that they're looking for." Gage adds that in some cases, Good Sudoku will withhold nuanced details on the obscure uses for techniques for as many as thirty-five instances of it, providing these details only when the user is ready and the context is right. He acknowledges that trying to give someone information when they are not ready is futile. By giving learners the space to grapple with a problem before learning more about it, Gage is providing what Schwartz and Bransford call "time for telling" [41]. In their implementation of this, Gage adds a book icon next to nearly every concept in Good Sudoku—clicking on this icon provides on-demand information for that concept. According to Gage's philosophy, the only time it's acceptable to provide information without "explicit approval" (Gage) is when a player first opens the game for the first time, since this action marks an explicit curiosity about what the application is.

Barth similarly describes learning that he "can not dump information on people" and cannot offer a help screen expecting players to navigate to it voluntarily. Instead, Barth and Gage both lean toward a model of self-paced discovery. According to Barth, this philosophy was strongly prevalent in designs from 2011: "you build a little obstacle course, you put a sign on the wall so they can just see 'X is jump' and then people will get through it.... you set up little learning labs for them where they were being guided towards being successful without having to just force them and tell them what to do."

This example is the prototype of the constructionist learning environment that Deterding describes of games [40]. In this interpretation of games, games are systems that players can explore, learning by experimenting with interactions and observing results. Gee breaks this model down further, giving us additional vocabulary: fish tanks are simplified ecosystems that highlight critical variables and their interactions, helping players learn what to pay attention to before more complexity is introduced; sandboxes, as Gee defines, are safe havens that look and feel like authentic environments, but with greatly mitigated risk, enabling performance before competence and providing "horizontal learning" (or time to practice and explore) for those who need it [2,42,43]. Gee breaks this down further into supervised and unsupervised sandboxes, with the former providing explicit scaffolding and the latter being a "hidden tutorial," blending tutorial and gameplay [2].

In fact, Gage explicitly describes constructivist sandboxes in their games during their 2014 talk at PRACTICE, noting that this design tool is the inverse of modern achievement systems: whereas achievements encourage players to explore the irrelevant corners of the game, sandboxes encourage players to explore the nuances of the core mechanics and build expert mental models [44]. Both fish tanks and sandboxes encourage system-thinking and mental model construction [1,42], which is why the Game Approachability Principles list sandboxes as one of the design tools for onboarding, among others like JIT information, scaffolding, knowledge transfer, and self-efficacy [45].

Similar to sandboxes, games by Gage and Barth are about engaging with an interesting set of tools. However, unlike sandbox games, the player is working toward producing a specific output from a specific input, especially for Zachtronics games. This creates the sought-after dynamic of goal-oriented constructivist learning.

To summarize the shared design philosophy, these designers are constructing sandboxes wherein players solve their own problems while still being oriented toward a guiding goal provided by the game. Furthermore, in this way, players' self-chosen goals acts as its own dynamic difficulty adjustment (DDA).

DDA is traditionally considered the automatic refinement of game parameters, such as adjusting enemy health or reaction speed based on player performance, though DDA has recently extended beyond pure system control and to other aspects of game adjustment, such as level design [46–48]. However, DDA has been criticized for causing player frustration and dissatisfaction when players become aware that their difficulty is being adjusted without their informed consent [49,50].

Barth recounts that several of their design principles for Zachtronics games were formed as a reaction against DDA and gamification mechanics, two components of the trend of 'RPGification' [51]. Rather than gamifying the task or reducing the challenge, Barth widened the possibility space between passing a level and excelling at it: their games allow you to proceed with a passable solution to the challenge rather than demanding great solutions. For example, he describes, their earlier game SHENZHEN I/O had tight spatial requirements for solutions, which made levels more difficult because there was less room for mediocre solutions. Opus Magnum, on the other hand, removed this space requirement: players now had unrestricted time and space to find a solution that solves the problem. For Barth, this is not DDA because the player is still solving the whole problem, whereas traditional DDA techniques modify parameters intrinsic to the challenge.

In Good Sudoku, Gage and Schlesinger designed their levels very intentionally with respect to how techniques are grouped and ranked and which techniques are used in which difficulty levels. By doing this, players could make informed decisions about the level of difficulty they want to experience—without affecting the intrinsic challenge of the task itself.

In addition to self-driven learning, both Barth and Gage also mentioned trying to teach general problem-solving skills. For Gage, their design philosophy for problem solving is based on their mother's research on math education, Kilpatrick [52], and other resources on constructionist teaching [44]. Through this lens, building problem-solving skills is about forming mental models about the structural relationships between elements, rather than the problems' surface features (see also [53,54]). "If you provide novice problem-solvers with a problem," Gage describes, "they'll attempt to solve it using superficial strategies, comparing it to routine problems that they already understand... However, if you provide novice problem-solvers with—instead of a problem—a set of constraints, and then ask them to form and solve their own complex problems, something amazing happens—they solve these problems with expert-level strategies" [44].

Similarly, Barth tries to impart general problem solving skills in Zachtronics games. Citing Papert's movement toward teaching computational thinking [55], Barth tries to encourage skill transfer. Rather than enforcing a skill chain (see [56]), Barth employs a bubble sort [57] to order levels in an intuitive, artistic process. "We do not make levels to try to teach a specific thing. We just make levels that introduce a specific new wrinkle. Furthermore, that really is not even specific. It's just like make a bunch of stuff that's interesting and then sort them in a way that scaffolds the introduction of new things we are asking you to do one at a time and something that feels like the complexity increases instead of being all over the place" (Barth). Interestingly, Barth has experimented with offering branching choices of levels, but "even when we offer them a bunch of levels in a set, almost everybody plays them in the order that we give them" (Barth).

These designers aim to teach their players general critical thinking skills that can transfer from the game into their daily lives. They achieve this by creating constructionist

and constructivist learning environments that provide JIT information and enable self-driven learning through player-made goals.

3.2.3. Do Not Let Education Prevent Learning

Puzzle games have a strange tension, says Barth. As puzzles, their goal is to challenge players. However, as products, they do not want their users to feel frustrated. Contrary to flow theory [58], Barth argues that it's okay to be difficult. "Vegetables do not have flow, you know, going to school does not have flow... I think a lot of the stuff in life that's like, real and worth doing, flow is not part of it."

Contrast this with educational games, which have been measured to be easier to learn, less complex, shorter, less challenging, and using fewer forms of fun than their commercial counterparts [5]. Rather, entertainment games rely on "hard fun" by challenging the players and encouraging a sense of mastery through difficulty [59,60]. This dynamic is enabled, in part, because of the designers' aforementioned difficulty scaling—or lack thereof. By letting the players choose their own goals, as long as the list of available goals covers a wide range of difficulties, players will find an appropriate difficulty for them.

However, the tutorial should not be difficult. Barth describes that although all of the levels in a Zachtronics game are hidden tutorials—in that they teach the player something—there is an initial set of levels that their team internally call tutorials; these levels do not try to challenge the player, instead they focus only on introducing core mechanics and controls.

Barth recalls that in Ironclad Tactics, he tried making the tutorials "little puzzles", but soon realized that learning the mechanics was puzzling enough for players. He summarizes: "So that was that was where we made the rule for ourselves, which is that if you have something that is like *a tutorial*, there should be nothing puzzling in it whatsoever." Barth believes only controls and basic concepts can be taught in tutorials (and should be, he says: "if people can not figure out your controls, they're going to not play your game."); higher-order concepts like strategies must be learned through experience.

Moreover, Barth found that forcing or hand-holding players through a set of actions was not an effective tutorial, nor was SpaceChem's approach of presenting "walls of text with pictures" (Barth). Instead, as elaborated in the previous section, Barth found it more effective to set up small, designed (constructionist) environments where players would be drawn to the correct solution by it being the natural, intuitive way to interact with the environment. In this way, they drive their own learning and encounter no difficulties when exploring a new concept for the first time.

For Gage and Schlesinger, the tutorials were the one aspect of design they strongly playtested. This was to ensure that anyone could access Good Sudoku and eventually reach its higher-level instructional design. Their focus on the tutorials emphasizes the need for usability, especially in the first-time user experience.

In summary, these designers tried to remove all challenge and frustration from the tutorials, with respect to usability, playability, and difficulty. By ensuring that players could access the tutorials, the designers could be assured that players were prepared for the rest of the game's challenges.

Next, the designers took creative liberties with their subject matter in order to highlight the intrinsically enjoyable aspects of the domain. In doing so, they developed their own language for the topic—sometimes quite literally in the case of Zachtronics' programming games. Barth intentionally places their games in fictional domains in order to "level the playing field" (Barth), ensuring that anyone coming in will have (as much as possible) the same experience.

In developing Good Sudoku, Gage and Schlesinger encountered a different problem: they were working in a real domain that already has terminology for specific strategies. Furthermore, to their dismay, the existing jargon was unintuitive. On the other hand, because the jargon was already deep-seated in the community, renaming them was not an option. Instead, they extended the terminology to tease out important distinctions, such

as by dividing techniques into their single-house versions (i.e., contained within a row, column, or box) and their "split" counterparts, noting that the perceptual contexts used in each case makes the technique conceptually different, even if the heuristic is the same. With expert techniques, however, Gage and Schlesinger were able to rename them because the community was small enough—and opinions diverse enough—that no standard had already been established.

Another aspect which significantly aided the development of Good Sudoku, Gage and Schlesinger claim, was that neither of them were Sudoku experts when they began. If they were, they explained, then Good Sudoku would have ended up teaching Sudoku the way it has already been taught—and interesting only the people who were already interested.

Although he does not use this terminology, Gage was describing a functional fixedness [61] to trying to teach something one already knows. Making Good Sudoku and Really Bad Chess (2016; a chess game by Zach Gage that randomizes all starting pieces except the king, based on the player's skill rating) required a paradigm shift away from how these games are typically taught and toward the novice view of seeing them for what is inherently interesting about the domain. This shift evokes the criticisms against modern academic writing, including issues with metadiscourse, the curse of knowledge, and functional fixedness, and suggests a return to the "classic style" of helping the audience see something in the world that the author has noticed, but they have not [62]. Moreover, as Gage notes, by teaching about something the designer themself does not understand, the designer gets to go on that journey of self-discovery and appreciation alongside their audience. In this way, Gage and Schlesinger sought to create the tools they wanted to learn from. They ask, "what are people experiencing when they're playing Sudoku that we are not experiencing?" (Schlesinger). This philosophy is summarized by Gage:

> "I think fundamentally, the way that you teach someone to do something is the same way you teach them to play a video game, which is teach them how to have fun, like focus on the thing that's fun and teach them how to do that... Really Bad Chess is not trying to teach you how to play chess, it's trying to teach you how to enjoy chess."

This point is where Gage differs from educational design. Rather than teach people to be good at something, Gage aims to show their audience what's fun about a subject and encourage them to appreciate it on their own, giving them the agency and ability to seek out more about it if they so desire. Similarly, by choosing a fictional programming context, Barth avoids any real-life complexities that would arise with using a realistic setting. Instead, he gains full control over the content, which allows them to focus solely on the interesting, intrinsic challenges and what is inherently fun about programming and logic.

In separating their work from educational design, the designers emphasized that games work as experiences, not as content delivery mechanisms, as is common (but not effective) with educational games [63]. Gage and Schlesinger speculated that "edutainment" games are produced more akin to software, taking a top-down approach to building a game around some desired learning outcomes.

In contrast, Gage did not intend to teach Sudoku with Good Sudoku. Instead, as described above, he aimed for teaching a love for Sudoku and a repertoire of self-driven learning and critical thinking skills. This is the difference between Gage's games and edutainment. Rather than fact-based learning, these successful commercial games teach by experience. Schlesinger calls out Duolingo as fact-based learning, instead wishing that it was able to simulate experiential contexts such as living in another country and being immersed in the language. "The thing that is powerful about games is that games give you experiences," Gage says, "and experiences are a completely different kind of learning than fact-based learning."

However, speaking to the educational potential of games, Gage notes that the experience alone is not enough for transfer learning. Although he does not use these words, Gage describes the potential for (exo-)transformative reflection in games [64–68], and

emphasizes the importance of feedback and debriefing in transferring experiential learning to transformative reflections. Regarding their work with NYSCI's *Connected Worlds* (https://nysci.org/home/exhibits/connected-worlds/ accessed October 13, 2021), Gage commented that "the only way to drive that [learning goal] home is to then have teachers have essentially like group meetings afterwards." For transformative reflection to have specific outcomes or transfer specific skills, there needs to be explicit debriefing: "You can teach a lot more than you would expect to be able to teach in a game if you remove the burden of processing the game and you put that onto like an educator or something like that, it would just work to have the game provide an experience for players that they will then interrogate later" (Gage).

In summary, the designers note that several of the ways in which educational games are currently produced are counter to successful strategies. Rather than being afraid to challenge the player, provide multiple difficulties that let them challenge themselves. However, do not feel obligated to make every moment of the game challenging—the tutorial, especially, should be as clear and simple as possible, with gradual increases in difficulty after the player demonstrates a basic understanding. Moreover, instead of starting with the concept itself and all of its nuances, start from a place of curiosity: why is this topic interesting? Why should the learner care about or appreciate this topic? By designing from this angle, the gameplay will more likely reflect something inherently fun about the material. Furthermore, this will enhance the game's ability to teach as an experience instead of as content delivery, which contributes to a deeper understanding of the material and better leverages the affordances of the medium.

3.2.4. Success Says Sell the GBL

Perhaps this will come as an uncomfortable truth to some readers, but the success stories of these industry professionals suggest that the constraints of education and research weaken the final product. Instead, taking an entrepreneurial and open mindset to creating an enjoyable experience seems to be the most effective strategy for a well-received game. Barth, for example, explicitly separates their work from academia. He argues that games with "competent game design" is an "exercise for academics," because for the purpose of selling games, competent design is overshadowed by other marketing factors. The more powerful marketing factors include sociocultural perceptions, which is partly why Gage and Schlesinger question the "educational" label being applied to their games. "There's stigma," says Gage.

Although their games interact with education, these developers do not approach design from an academic perspective; however, they do look to academia for inspiration. For example, Barth took inspiration from a paper on tutorials [69], but drew their own conclusions of "the thing that felt true to us." The high-level points he took away were that you "can not dump information on people" and you "can not give them a help screen to look at on their own volition." However, then he constructs his own narrative, rather than actually building on the publication. Similarly, Gage draws inspiration from constructivist learning, such as Kilpatrick's work [52]. Like Barth, Gage takes the aspects that resonate with him and uses a few key ideas as the foundation of his design. In doing so, Gage focuses on putting the model before the problem via self-driven goal-setting, as discussed earlier.

Yet, Schlesinger notes that his design process would never work with a starting constraint on the learning goal, as is the common approach when developing educational or serious games. Barth shares this sentiment and expands on it:

> "Normally when we make games, we get to make up whatever we want. When you make educational games, you do not get to make up whatever you want. Furthermore, so a lot of the things that people say about like 'oh game design, good game design, and good games are like so magical and compelling.' Furthermore, it's like, that is assuming that you can do whatever you want."

Making games that are both fun and educational, Barth found, was "actually really hard because it really does constrain your design space." This is why, instead of building directly from academic research, these developers draw ideas from research and create their own interpretations. Doing so allows them to be more fluid with their designs and find learning opportunities that fit the gameplay they can create, rather than trying to force gameplay to fit a learning goal. Furthermore, by creating a more natural fit of gameplay and learning, they are better able to sell their games.

To this point, Gage focuses strongly on outreach when making their games. He has collected a list of press contacts and works with other industry professionals to develop pitches, reach out to games journalists, and create targeted advertisements for specific markets. Gage put a great deal of effort into the "messages" of their game and how pitches were delivered, including taglines and the perception of the core gameplay at a glance.

In this way, the commercial success of Barth, Gage, and Schlesinger is due, in part, to actually treating their products commercially. They value *value*—what their products contribute to the market. Barth also noted reusing "conceptual tools" across their games to save time and effort, explaining why several Zachtronics games have similar experiences. Through this, Barth builds an audience for players who appreciate this style of play, which helps grow the Zachtronics brand for further success.

This finding agrees with an ongoing discussion of entrepreneurialism in educational technologies [6]. The Radix Endeavor [70], for example, failed to maintain or further develop as a product due to a lack of sustainable funding (personal communication, Louisa Rosenheck, at the 2021 Connected Learning Summit's Hall of Failure). Scholars argue that the GBL academic community has been taking the wrong approach to creating effective game-based learning environments and testing the success of GBL [63]. The most recent addition to this conversation is by Ike et al. [71], who argue that "educational game ideas should be practical and commercially viable to develop," achieved—among other means—through feasibility studies during pre-production to identify a target market and the target audience's motivation. Although Barth and Gage did not speak to such formal approaches, it seems they have naturally developed a target market and appealed to their motivations through consistent design and branding.

In short, historical approaches to GBL have been weighed down by narrowly constrained goals and strict academic approaches, relying on the assumption that students will prefer any game to traditional learning. Motivation is not magically achieved by having a game. If you want buy-in, you have to sell it.

4. Discussion

In this article, we explored the successes of commercial game-based complex learning in Duolingo, Good Sudoku, and Zachtronics games via a close reading of Duolingo and interviews with the designers of Good Sudoku and Zachtronics games.

Despite the popular praise of Duolingo as the best gamified language learning app available, we found the gamification mechanics sparse and antagonistic to the learning goals (cf. Morschheuser et al. [25], gamification mechanics should match complexity). However, given Duolingo's commercial success, we draw the conclusion that this end-product is the result of years of testing and iteration, and GBL products must operate in a real-world market with real user data and financial constraints, even if the final outcome is not an ideal user experience.

The learning mechanics, on the other hand, were more cohesive. The skill tree, forming the nucleus of the learning system, as well as satellite systems of practice and feedback, provided smooth and gradual progression through increasingly complex content. Through transparency and visibility, the learning was made clear and simple despite the complexity of the material.

Although all of these GBL elements have been suggested in prior literature (PBL [72,73], retention mechanics [26], skill trees [27,28], feedback [32], task variety [3,4,30,31], etc.), this close reading contributes an exploration of what has been practically most effective for

successful GBL in a complex, real-world use case. Furthermore, what was practically most effective was a foundation of strong instructional design: gradual, well-paced, experiential learning with varied tasks and thorough feedback. We found similar results in the gameful case studies of Good Sudoku and Zachtronics games.

Through interviewing designers Barth, Gage, and Schlesinger and following with reflexive thematic analysis, we generated four themes. **Holistic game design *is* instructional design**: between visual design guiding players toward critical information, using varied whole-task practice, removing busywork, and iterating on prototypes, the designers are able to achieve effective instructional design through a focus on the game experience itself. This instructional design takes the form of **constructionist learning environments**, including JIT information, player-driven learning and difficulty, and ultimately teaching general problem solving.

These industry professionals also emphasized that they **do not let education prevent learning**: by letting the game be difficult—but not the tutorial—they enable learning in both the early and late stages of gameplay. Moreover, they focus on teaching the topic how it is loved, rather than how it is taught. Gage offers clear advice on how to transfer this approach:

> "Make something you do not understand. Do not do not try to teach people something that you already know how to play really well, because you're too far away from the experience of somebody who is just starting to even understand it."

Furthermore, Barth, Gage, and Schlesinger highlight the affordances of games as experiences rather than a medium for content delivery. This echoes criticisms of GBL research that focus heavily on putting games in the classroom without considering the contexts in which games are most effective for learning [63].

Finally, the industry professionals are successful because they acknowledge their products for commercial value and sell them accordingly—that is, they **sell the GBL**. Predefined learning goals and research questions hinder the flexibility of the game during development, making it harder to design for appeal to a particular audience. Furthermore, to the point of audiences, these designers have one in mind. They construct specific pitches for those audiences and put concerted efforts into reaching them.

Our results echo common themes of GBL. Many of the game design points discussed here agree with Gee's argument of game learning as instructional design [1,2,15,42], as well as other arguments for games as constructionist learning environments [40] and constructivism in games [38,74]. Previous work has also identified the affordances of games as experiences rather than content delivery [42,75,76]. Like the close reading, the novel contributions of these case studies are the emphasis on what was practically most useful for commercial and critical success in complex GBL.

In these case studies, two major points stand out: first, teaching how the topic is loved rather than how it is traditionally taught—this concept is rare in GBL literature, though it has ties to existing conversations such as using authentic problems to demonstrate meaning and value [6]. Second, a failure to treat the game as a commercial product that demands advertisement, market analysis, and so on, is an issue which has caused the downfall of many educational and research games. Although a discussion of the systemic financial problems with educational technology is out of scope for this article, GBL practitioners would benefit from planning their funding from prototype to long-term sustainability early in the project design.

Comparing our close reading and case studies, what are the similarities and differences between gamified and gameful approaches to GBL? Both Duolingo and the games in the case studies are commercial products and are treated accordingly, with developer intent toward advertising and market value as well as iterative design. Moreover, all games studied offer gradual learning, JIT information, and player-set difficulty. However, Good Sudoku and Zachtronics games take a much more holistic, whole-task approach to their puzzles and experiential learning that focuses on teaching general problem-solving skills,

while Duolingo situates itself in part-task practice and content delivery. Although this is likely due in part to the nature of the domains, it is true that the motivational power of gamification is better at encouraging practice and "drilling," while gameful applications are more suited to holistic experiences. Each of these approaches, then, should be considered as two separate sets of tools in the toolkit or repertoire of GBL practitioners.

In summary, successful game-based complex learning is about leveraging game design techniques and the power of games as a medium, constructionist and constructivist learning, and strong marketing.

If you take one point away from these case studies, take this: design from a place of curiosity and authenticity and create the learning tools you want to see and play with yourself—the rest will follow.

Author Contributions: Conceptualization, J.A.M.; methodology, J.A.M.; formal analysis, J.A.M.; investigation, J.A.M., writing—original draft, J.A.M.; writing—review and editing, J.A.M., S.C.; supervision, S.C. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The study was approved by the Institutional Review Board of Northeastern University (#21-01-23, approved 22 January 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the participants to publish this paper.

Data Availability Statement: For access to the transcripts and/or the analysis audit trail, please contact the first author. Some data may not be available to protect participant privacy.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DDA	Dynamic Difficulty Adjustment
GBL	Game-Based Learning
JIT	Just-In-Time
PBL	Points, Badges, Leaderboards

References

1. Gee, J.P. Good video games and good learning. *Phi Kappa Phi Forum. Honor. Soc. Phi Kappa Phi* **2005**, *85*, 33.
2. Gee, J.P. *Good Video Games+ Good Learning: Collected Essays on Video Games, Learning, and Literacy*; Peter Lang: New York, NY, USA, 2007.
3. Schunk, D.H. *Learning Theories an Educational Perspective*, 6th ed.; Pearson: Boston, MA, USA, 2012.
4. Van Merriënboer, J.J.; Kirschner, P.A. *Ten Steps to Complex Learning: A Systematic Approach to Four-Component Instructional Design*, 3rd ed.; Routledge: New York, NY, USA, 2018.
5. Heeter, C.; Chu, C.; Maniar, A.; Winn, B.; Mishra, P.; Egidio, R.; Portwood-Stacer, L. Comparing 14 plus 2 forms of fun (and learning and gender issues) in commercial versus educational space exploration digital games. In Proceedings of the International Digital Games Research Conference, Utrecht, The Netherlands, 4–6 November 2003; pp. 4–6.
6. Klopfer, E.; Haas, J.; Osterweil, S.; Rosenheck, L. *Resonant Games: Design Principles for Learning Games That Connect Hearts, Minds, and the Everyday*; MIT Press: Cambridge, MA, USA, 2018.
7. Von Ahn, L.; Hacker, S. Duolingo. 2011. Available online: https://www.duolingo.com (accessed on 4 October 2021).
8. Gage, Z.; Schlesinger, J. Good Sudoku. 2020. Available online: https://www.playgoodsudoku.com/ (accessed on 12 September 2021).
9. Barth, Z. Zachtronics. Available online: https://www.zachtronics.com/ (accessed on 23 September 2021).
10. Funnell, R. 'Really Bad Chess' Review—Rook No Further. 2016. Available online: https://toucharcade.com/2016/10/14/really-bad-chess-review/ (accessed on 13 October 2021).
11. Kidwell, E. Really Bad Chess is Bad in All the Right Ways. 2016. Available online: https://killscreen.com/previously/articles/really-bad-chess-bad-right-ways/ (accessed on 13 October 2021).
12. Bizzocchi, J.; Tanenbaum, T. Well read: Applying close reading techniques to gameplay experiences. In *Well Played 3.0: Video Games, Value and Meaning*; ETC Press: Pittsburgh, PA, USA, 2011; Volume 3.

13. Braun, V.; Clarke, V. Using thematic analysis in psychology. *Qual. Res. Psychol.* **2006**, *3*, 77–101. [CrossRef]

14. Braun, V.; Clarke, V. Reflecting on reflexive thematic analysis. *Qual. Res. Sport. Exerc. Health* **2019**, *11*, 589–597. [CrossRef]

15. Gee, J.P. *An Introduction to Discourse Analysis: Theory and Method*; Routledge: London, UK, 2004.

16. Duolingo. Duolingo. Available online: https://www.duolingo.com/ (accessed on 4 October 2021).

17. Sullivan, A.; Salter, A. A taxonomy of narrative-centric board and card games. In Proceedings of the 12th International Conference on the Foundations of Digital Games, Association for Computing Machinery, Hyannis, MA, USA, 14–17 August 2017; pp. 1–10.

18. Thompson, S.A.; Mann, W.C. Rhetorical structure theory: A framework for the analysis of texts. *IPRA Pap. Pragmat.* **1987**, *1*, 79–105. [CrossRef]

19. Settles, B.; Meeder, B. A trainable spaced repetition model for language learning. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics: Berlin, Germany, 2016; pp. 1848–1858.

20. Toda, A.M.; Valle, P.H.; Isotani, S. The dark side of gamification: An overview of negative effects of gamification in education. In *Researcher Links Workshop: Higher Education for All*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 143–156.

21. van Roy, R.; Deterding, S.; Zaman, B. Uses and gratifications of initiating use of gamified learning platforms. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*; 2018; Association for Computing Machinery: New York, NY, USA; pp. 1–6.

22. Tondello, G.F.; Mora, A.; Nacke, L.E. Elements of gameful design emerging from user preferences. In Proceedings of the Annual Symposium on Computer-Human Interaction in Play, Association for Computing Machinery: New York, NY, USA, 2017; pp. 129–142.

23. Eveleigh, A.; Jennett, C.; Blandford, A.; Brohan, P.; Cox, A.L. Designing for dabblers and deterring drop-outs in citizen science. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Association for Computing Machinery: New York, NY, USA, 2014; pp. 2985–2994.

24. Tang, J.; Prestopnik, N.R. *Toward an Understanding of the Influences of Meaningful Framing on User Participation in a Gamified Information System*; In In Proceedings of the 24th European Conference on Information Systems, Istanbul, Turkey, 12–15 June 2016.

25. Morschheuser, B.; Hamari, J.; Koivisto, J.; Maedche, A. Gamified crowdsourcing: Conceptualization, literature review, and future agenda. *Int. J. -Hum.-Comput. Stud.* **2017**, *106*, 26–43. [CrossRef]

26. Legner, L.; Eghtebas, C.; Klinker, G. Persuasive mobile game mechanics for user retention. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*; 2019; Association for Computing Machinery: New York, NY, USA; pp. 493–500.

27. Van Merriënboer, J.J.; Clark, R.E.; De Croock, M.B. Blueprints for complex learning: The 4C/ID-model. *Educ. Technol. Res. Dev.* **2002**, *50*, 39–61. [CrossRef]

28. Horn, B.; Cooper, S.; Deterding, S. Adapting cognitive task analysis to elicit the skill chain of a game. In Proceedings of the Annual Symposium on Computer-Human Interaction in Play, Amsterdam, The Netherlands, 15–18 October 2017; pp. 277–289.

29. Kumari, S.; Deterding, S.; Kuhn, G. Why game designers should study magic. In Proceedings of the 13th International Conference on the Foundations of Digital Games, Association for Computing Machinery, New York, NY, USA, 7–10 August 2018; pp. 1–8.

30. Dickey, M.D. Engaging by design: How engagement strategies in popular computer and video games can inform instructional design. *Educ. Technol. Res. Dev.* **2005**, *53*, 67–83. [CrossRef]

31. Bond, M.; Beale, R. What makes a good game? Using reviews to inform design. In Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology, Cambridge, UK, 1–5 September 2009; pp. 418–422.

32. Shannon, A.; Boyce, A.; Gadwal, C.; Barnes, T. *Effective Practices in Game Tutorial Systems*; In Proceedings of the 8th International Conference on the Foundations of Digital Games, Chania, Crete, Greece, 14–17 May 2013; pp. 338–345.

33. Keuning, H.; Jeuring, J.; Heeren, B. A systematic literature review of automated feedback generation for programming exercises. *ACM Trans. Comput. Educ. (TOCE)* **2018**, *19*, 1–43. [CrossRef]

34. Bacher, D. Design Patterns in Level Design: Common Practices in Simulated Environment Construction. Master's Thesis, Iowa State University, Ames, IA, USA, 2008.

35. Totten, C.W. *Architectural Approach to Level Design*; CRC Press: Boca Raton, FL, USA, 2019.

36. Kolb, D.A. *Experiential Learning: Experience as the Source of Learning and Development*; Prentice Hall, Inc: Upper Saddle River, NJ, USA, 2014.

37. Felder, R.M.; Silverman, L.K. Learning and teaching styles in engineering education. *Eng. Educ.* **1988**, *78*, 674–681.

38. Becker, K. *The Invention of Good Games: Understanding Learning Design in Commercial Video Games*; Doctoral Thesis, University of Calgary, Calgary, Alberta, 2008.

39. Sweller, J. Cognitive Load Theory. In *Psychology of Learning and Motivation*; Elsevier: Amsterdam, The Netherlands, 2011; Volume 55, pp. 37–76.

40. Deterding, S. The ambiguity of games: Histories and discourses of a gameful world. In *SP Walz, & S. Deterding, The Gameful World: Approaches, Issues, Applications* 2014; MIT Press: Cambridge, MA, USA; pp. 23–64.

41. Schwartz, D.L.; Bransford, J.D. A time for telling. *Cogn. Instr.* **1998**, *16*, 475–5223. [CrossRef]

42. Gee, J.P. Learning by design: Good video games as learning machines. *E-Learn. Digit. Media* **2005**, *2*, 5–16. [CrossRef]

43. Goto, S.T. Basic writing and policy reform: Why we keep talking past each other. *J. Basic Writ.* **2002**, *21*, 4–20. [CrossRef]

44. Gage, Z. Designing for Problem-Solvers. Presented at PRACTICE. 2014. Available online: https://www.youtube.com/watch?v=yyrRVQnnBFs (accessed on 13 October 2021).

45. Desurvire, H.; Wiberg, C. User experience design for inexperienced gamers: GAP—Game Approachability Principles. In *Game User Experience Evaluation*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 169–186.

46. Hunicke, R. The case for dynamic difficulty adjustment in games. In Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, Valencia, Spain, 15–17 June 2005; pp. 429–433.

47. Ang, D.; Mitchell, A. Representation and frequency of player choice in player-oriented dynamic difficulty adjustment systems. In Proceedings of the Annual Symposium on Computer-Human Interaction in Play, Association for Computing Machinery, New York, NY, USA, 22–25 October 2019; pp. 589–600.

48. Jennings-Teats, M.; Smith, G.; Wardrip-Fruin, N. Polymorph: dynamic difficulty adjustment through level generation. In Proceedings of the 2010 Workshop on Procedural Content Generation in Games, Association for Computing Machinery, New York, NY, USA, 18 June 2010; pp. 1–4.

49. Gerling, K.M.; Miller, M.; Mandryk, R.L.; Birk, M.V.; Smeddinck, J.D. Effects of balancing for physical abilities on player performance, experience and self-esteem in exergames. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA, 26 April–1 May 2014; pp. 2201–2210.

50. Liu, C.; Agrawal, P.; Sarkar, N.; Chen, S. Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback. *Int. J. -Hum.-Comput. Interact.* **2009**, *25*, 506–529. [CrossRef]

51. Gallagher, R. Case Study# 4: Exaptation, Interpretation, PlayStation. Available online: http://median.newmediacaucus.org/case-study-4-exaptation-interpretation-playstation/ (accessed on 26 October 2021).

52. Kilpatrick, J.; Swafford, J.; Findell, B.; National Research Council. *Adding It Up: Helping Children Learn Mathematics*; Citeseer: Princeton, NJ, USA, 2001.

53. Chase, W.G.; Simon, H.A. Perception in chess. *Cogn. Psychol.* **1973**, *4*, 55–81. [CrossRef]

54. Chi, M.T.; Feltovich, P.J.; Glaser, R. Categorization and representation of physics problems by experts and novices. *Cogn. Sci.* **1981**, *5*, 121–152. [CrossRef]

55. Papert, S.A. *Mindstorms: Children, Computers, and Powerful Ideas*; Basic Books: New York, NY, USA, 1980.

56. Cook, D. The Chemistry of Game Design. Game Developer. Available online: http://www.gamasutra.com/view/feature/1524/the_chemistry_of_game_design.php (accessed on 27 October 2021).

57. Astrachan, O. Bubble sort: An archaeological algorithmic analysis. *ACM Sigcse Bull.* **2003**, *35*, 1–5. [CrossRef]

58. Csikszentmihalyi, M.; Csikszentmihaly, M. *Flow: The Psychology of Optimal Experience*; Harper & Row: New York, NY, USA, 1990; Volume 1990.

59. Lazzaro, N. The 4 Keys 2 Fun. Nicole Lazzaro. Available online: http://www.nicolelazzaro.com/the4-keys-to-fun (accessed on 28 October 2021).

60. Lazzaro, N. Why we Play: Affect and the fun of games. *Hum.-Comput. Interact. Des. Divers. Users Domains* **2009**, *155*, 679–700.

61. Duncker, K.; Lees, L.S. On problem-solving. *Psychol. Monogr.* **1945**, *58*, i-113. [CrossRef]

62. Pinker, S. Why academics stink at writing. *Chron. High. Educ.* **2014**, *61*, 148989.

63. Young, M.F.; Slota, S.; Cutter, A.B.; Jalette, G.; Mullin, G.; Lai, B.; Simeoni, Z.; Tran, M.; Yukhymenko, M. Our princess is in another castle: A review of trends in serious gaming for education. *Rev. Educ. Res.* **2012**, *82*, 61–89. [CrossRef]

64. Fleck, R.; Fitzpatrick, G. Reflecting on reflection: framing a design landscape. In Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction, Brisbane, Australia, 22–26 November 2010; pp. 216–223.

65. Moon, J.A. *Reflection in Learning and Professional Development: Theory and Practice*; Routledge: London, UK, 1999.

66. Baumer, E.P. Reflective informatics: conceptual dimensions for designing technologies of reflection. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA, 18–23 April 2015; pp. 585–594.

67. Khaled, R. Questions over answers: Reflective game design. In *Playful Disruption of Digital Media*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–27.

68. Whitby, M.A.; Deterding, S.; Iacovides, I. "One of the Baddies All along": Moments That Challenge a Player's Perspective. In Proceedings of the Annual Symposium on Computer-Human Interaction in Play; Association for Computing Machinery, CHI PLAY'19, New York, NY, USA, 22–25 October 2019; pp. 339–350. doi: 10.1145/3311350.3347192. [CrossRef]

69. Andersen, E.; O'Rourke, E.; Liu, Y.E.; Snider, R.; Lowdermilk, J.; Truong, D.; Cooper, S.; Popovic, Z. The impact of tutorials on games of varying complexity. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, TX, USA, 5–10 May 2012; pp. 59–68.

70. Clarke-Midura, J.; Rosenheck, L.; Haas, J.; Klopfer, E. *The Radix Endeavor: Designing a Massively Multiplayer Online Game around Collaborative Problem Solving in STEM*; In Proceedings of the Computer Supported Collaborative learning Conference, Madison, WI, USA, 15–19 June 2013.

71. Ike, T.C.; Hoe, T.W.; Yatim, M.H.M. Designing Elements for Immersive User Experience in Educational Games Using the Entertainment Game Development Approach. *Rev. Int. Geogr. Educ. Online* **2021**, *11*, 738–747.

72. Mekler, E.D.; Brühlmann, F.; Tuch, A.N.; Opwis, K. Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Comput. Hum. Behav.* **2017**, *71*, 525–534. [CrossRef]

73. Orji, R.; Nacke, L.E.; Di Marco, C. Towards personality-driven persuasive health games and gamified systems. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; pp. 1015–1027.
74. Catalano, C.E.; Luccini, A.M.; Mortara, M. Guidelines for an Effective Design and Evaluation of Serious Games. *Int. J. Serious Games* **2014**, *1*. [CrossRef]
75. Iacovides, I.; McAndrew, P.; Scanlon, E.; Aczel, J. The gaming involvement and informal learning framework. *Simul. Gaming* **2014**, *45*, 611–626. [CrossRef]
76. Becker, K. *Design Paradox: Instructional Games*; Presented at the University of Calgary: Calgary, AB, Canada, 2006.