



A Systematic Literature Review of Analytics for Adaptivity Within Educational Video Games

Manuel Ninaus^{1,2,3†} and Steve Nebel^{4*†}

¹Department of Clinical Psychology, Institute of Psychology, University of Innsbruck, Innsbruck, Austria, ²LEAD Graduate School and Research Network, University of Tübingen, Tübingen, Germany, ³Leibniz-Institut für Wissensmedien, Tübingen, Germany, ⁴Chair of Psychology of Learning with Digital Media, Faculty of Humanities, Institute of Media Research, Chemnitz University of Technology, Chemnitz, Germany

OPEN ACCESS

Edited by:

Michael Kickmeier-Rust,
University of Teacher Education St.
Gallen, Switzerland

Reviewed by:

Felix Kapp,
Technical University of Berlin,
Germany
Riccardo Berta,
University of Genoa, Italy

*Correspondence:

Steve Nebel
steve.nebel@phil.tu-chemnitz.de

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Digital Education,
a section of the journal
Frontiers in Education

Received: 28 September 2020

Accepted: 21 December 2020

Published: 29 January 2021

Citation:

Ninaus M and Nebel S (2021) A
Systematic Literature Review of
Analytics for Adaptivity Within
Educational Video Games.
Front. Educ. 5:611072.
doi: 10.3389/feduc.2020.611072

Research has shown that serious games, digital game-based learning, and educational video games can be powerful learning instruments. However, experimental and meta-research have revealed that several moderators and variables influence the resulting learning outcomes. Advances in the areas of learning and game analytics potentially allow for controlling and improving the underlying learning processes of games by adapting their mechanics to the individual needs of the learner, to properties of the learning material, and/or to environmental factors. However, the field is young and no clear-cut guidelines are yet available. To shed more light on this topic and to identify common ground for further research, we conducted a systematic and pre-registered analysis of the literature. Particular attention was paid to different modes of adaptivity, different adaptive mechanisms in various learning domains and populations, differing theoretical frameworks, research methods, and measured concepts, as well as divergent underlying measures and analytics. Only 10 relevant papers were identified through the systematic literature search, which confirms that the field is still in its very early phases. The studies on which these papers were based, however, show promise in terms of the efficacy of adaptive educational games. Moreover, we identified an increased interest in the field of adaptive educational games and in the use of analytics. Nevertheless, we also identified a clear lack of common theoretical foundations as well as the application of rather heterogenous methods for investigating the effects of adaptivity. Most problematic was the lack of sufficient information (e.g., descriptions of used games, adaptive mechanisms), which often made it difficult to draw clear conclusions. Future studies should therefore focus on strong theory building and adhere to reporting standards across disciplines. Researchers from different disciplines must act in concert to advance the current state of the field in order to maximize its potential.

Keywords: analytics, educational video games, serious games, game-based learning, adaptivity, learning, systematic review

INTRODUCTION

Digital game-based learning is becoming a powerful tool in education (e.g., Boyle et al., 2016). However, several open issues remain that require further research in order to optimize the use of game-based learning and educational video games. One unique characteristic of digital learning games is the wealth of data they produce, which can be acquired and used for (learning) analytics and adaptive systems. Adaptive learning environments are part of a new generation of computer-supported learning systems that aim to provide personalized learning experiences by capitalizing on the generation and acquisition of knowledge and other types of data regarding learner's cognitive capabilities, knowledge levels, and preferences, among other factors (e.g., Mangaroska and Giannakos, 2019).

Adaptive learning is characterized by an adaptive approach to learner's individual needs and preferences in order to optimize learning outcomes and other learning-related aspects, such as motivation. While the idea of adaptive learning is not new (e.g., mastery learning as discussed by, Bloom, 1968) and has received strong support from researchers in educational psychology (Alexander, 2018), it is surprising how few systematic studies are available on adaptive learning with digital technologies and game-based learning in particular. For instance, a recent review of adaptive learning in digital environments in general found evidence of the effectiveness of adaptivity (Aleven et al., 2016). However, in this review, only one study was identified as having used an educational video game, thus demonstrating the lack of research currently being performed on adaptive learning in the domain of game-based learning.

Most entertainment games are pre-scripted and therefore have static (game) elements such as, content, rules, and narratives (Lopes and Bidarra, 2011). While "fun" is the locus of attention in entertainment games and has been investigated in learning games as well (Nebel et al., 2017c), educational games serve additional purposes, as they need to convey learning content appropriately to learners. According to Schrader et al. (2017) adaptivity in educational games can be defined as "a player-centred approach by adjusting game's mechanics and representational modes to suit game's responsiveness to player characteristics with the purpose of improving in-game behavior, learning processes, and performance" (p. 5). Hence, finding the right balance between the learner's skills and the challenge levels of the games is a critical issue, especially as the perceived difficulty and inferred feedback after facing a task could influence learning outcomes (e.g., Nebel et al., 2017b). Researchers agree that educational video games could utilize adaptivity to optimize knowledge and skills acquisition (e.g., Lopes and Bidarra, 2011; Streicher and Smeddinck, 2016). Potentially, all elements of a game can become adaptive elements (Lopes and Bidarra, 2011). For instance, gameplay mechanics, narrative and scenarios, game content and its objectives, etc., all can contribute to offer and personalized and individualized gaming and learning experience.

It seems natural for learning material to be adapted to individual needs and preferences. In analogue learning

settings, this can be achieved by individualized support from educators, teachers, etc.; in multiplayer games, social processes can trigger similar processes (Nebel et al., 2017a). For single-player games, however, there are several different ways to acquire the data needed to identify user's needs or preferences (for a review see Nebel and Ninaus, 2019) and to change the learning environment accordingly (for a review see Aleven et al., 2016). Numerous studies have demonstrated that data or analytics gathered during play can be used to successfully detect various cognitive (e.g., Witte et al., 2015; Appel et al., 2019), motivational (e.g., Klasen et al., 2012; Berta et al., 2013), and emotional (e.g., Brom et al., 2016; Ninaus et al., 2019a) states of users (for a review see Nebel and Ninaus, 2019). The analytics used in such studies range from simple pre-test measures and self-reports to more complex process measures utilizing (neuro-)physiological sensors (for a review see Ninaus et al., 2014; Nebel and Ninaus, 2019). Consequently, the current systematic review aims to identify if and how such analytics have been used to realize adaptive learning in games. In particular, we wanted to investigate the use of adaptive learning in educational games by utilizing analytics to adapt learning content to the skill level or cognitive capability of the player/learner.

While games theoretically offer many opportunities for adapting their content (e.g., visual presentation, narrative, difficulty), many factors usually need to be considered when implementing adaptivity. These factors include which analytics are used and what as well as how content is actually being adapted by the game or its underlying algorithms. Accordingly, frameworks for adaptive educational games are often guided by two questions (c.f., Shute and Zapata-Rivera, 2012): First, what to adapt: Which analytics and data are utilized to implement adaptivity and which elements are adapted (e.g., feedback, scaffolding, etc.)? Second, how to adapt: Which general methods are used to implement adaptivity? While different frameworks of adaptive (educational) games differ in their granularity and their specific design, they share the common goal of providing a generic approach on how to realize adaptivity (e.g., Yannakakis and Togelius, 2011; Shute and Zapata-Rivera, 2012; Schrader et al., 2017). For instance, Shute and Zapata-Rivera (2012) suggest a four-process adaptive cycle that connects the learner to appropriate educational material through the application of a user model. These generic frameworks are helpful for building adaptive systems, however, they tell very little about their effectiveness. Evaluating adaptive systems with empirical studies is therefore not only informative but absolutely necessary for advancing the field of adaptive educational games. Moreover, as the field of educational psychology aims to improve the theoretical understanding of learning (Mayer, 2018), particular attention should be paid to the theoretical foundations of the adaptation mechanisms being used. This seems to be in contrast to recent trends in learning analytics (e.g., Greller and Drachslers, 2012) and game-learning analytics (e.g., Freire et al., 2016), which employ rather strong data-driven approaches. Thus, in our research, we focused especially on what is adapted, how adaptivity is implemented, and which analytics are utilized to realize adaptivity in game-based learning. In this latter respect, by identifying the analytics used for adaptivity, we

sought to produce a precise overview of successful and less successful approaches in adaptive game-based learning in the interest of identifying practical needs and recommendations. That is, we aimed to provide a systematic overview of the current state of the art of adaptivity in game-based learning by analyzing the ways in which empirical research is currently being conducted in this field of research, which theoretical foundations are being used to realize adaptivity, and what is being targeted by adaptivity.

Description of Research Problem

That learning environments have the potential to act dynamically by gathering user data or pre-test values and responding by altering the learning tasks within a digital environment has been an established fact for decades (e.g., Skinner, 1958; Hartley and Sleeman, 1973; Anderson et al., 1990; Alevan et al., 2009). This approach might be particularly relevant in the field of game-based learning, as games are usually considered to involve highly dynamic environments and adaptive learning seems to be a promising avenue by which learning outcomes in digital learning can be enhanced (for a review see Alevan et al., 2016). However, the extent to which adaptivity has been implemented in empirical studies using game-based learning has thus far not been systematically documented. Consequently, in the current pre-registered systematic literature review (see Ninaus and Nebel, 2020), we pay particular attention to game-based learning environments to uncover the current state of research in this field. An increasing number of studies have demonstrated the use of various analytics to identify different mental states of the users that might be useful for adapting educational games in real time (for a review see Nebel and Ninaus, 2019). However, it remains unknown whether these suggested adaptive approaches have actually been implemented and evaluated in game-based learning. Thus, with this systematic literature review, we aim to address this open question by analyzing the current state of the literature. Instead of motivation or personality-based adaptations (e.g., Orji et al., 2017), we focused on cognition or performance-based adaptations, as learning theories heavily focus on this perspective (e.g., Sweller, 1994; Sweller et al., 1998; Mayer, 2005) and it allows for a more focused analysis of the current literature. Accordingly, we sought to identify successful (learning) analytics for adapting the game-based learning environment to, for instance, the skill level or cognitive capability of the learner. That is, which data about learners or their context can be utilized for understanding and optimizing learning by adapting the learning environment. Doing so might shed light on which approaches are most successful in adaptive game-based learning and thereby advance the field and provide practical recommendations for researchers and educators alike.

Study Objectives

Taken together, this paper systematically reviews the ways in which adaptivity in game-based learning is realized. For this pre-registered systematic literature, we broadly searched for empirical studies that utilized analytics to realize adaptivity in game-based learning scenarios and educational video games. Based on the

previously described research gaps, we were specifically interested in the following three research questions (RQ):

- (RQ1) How is research in the field of analytics for adaptation in educational video games currently conducted? For instance, which learning domains and analytics are most popular as well as most successful in this research field, and which empirical study designs are currently being employed to study the effects of adaptive elements?
- (RQ2) What cognitive/theoretical frameworks within analytics for adaptation in educational video games are currently used? That is, which theoretical underpinnings are currently being used to integrate, justify, and evaluate analytics in adaptive educational video games?
- (RQ3) What types of outcomes are influenced by an adaptive approach? For instance, are the analytics and adaptive mechanisms used for adapting the difficulty of a quiz in a game or is the overall pace of the gameplay altered?

To contribute a precise overview as well as to identify areas for future development in the analytics for adaptation research field, this paper follows the meta-analysis article-reporting standards proposed by the American Psychological Society (2020a), and the PRISMA checklist (Moher et al., 2009) has been used to ensure the inclusion of all relevant information.

MATERIALS AND METHODS

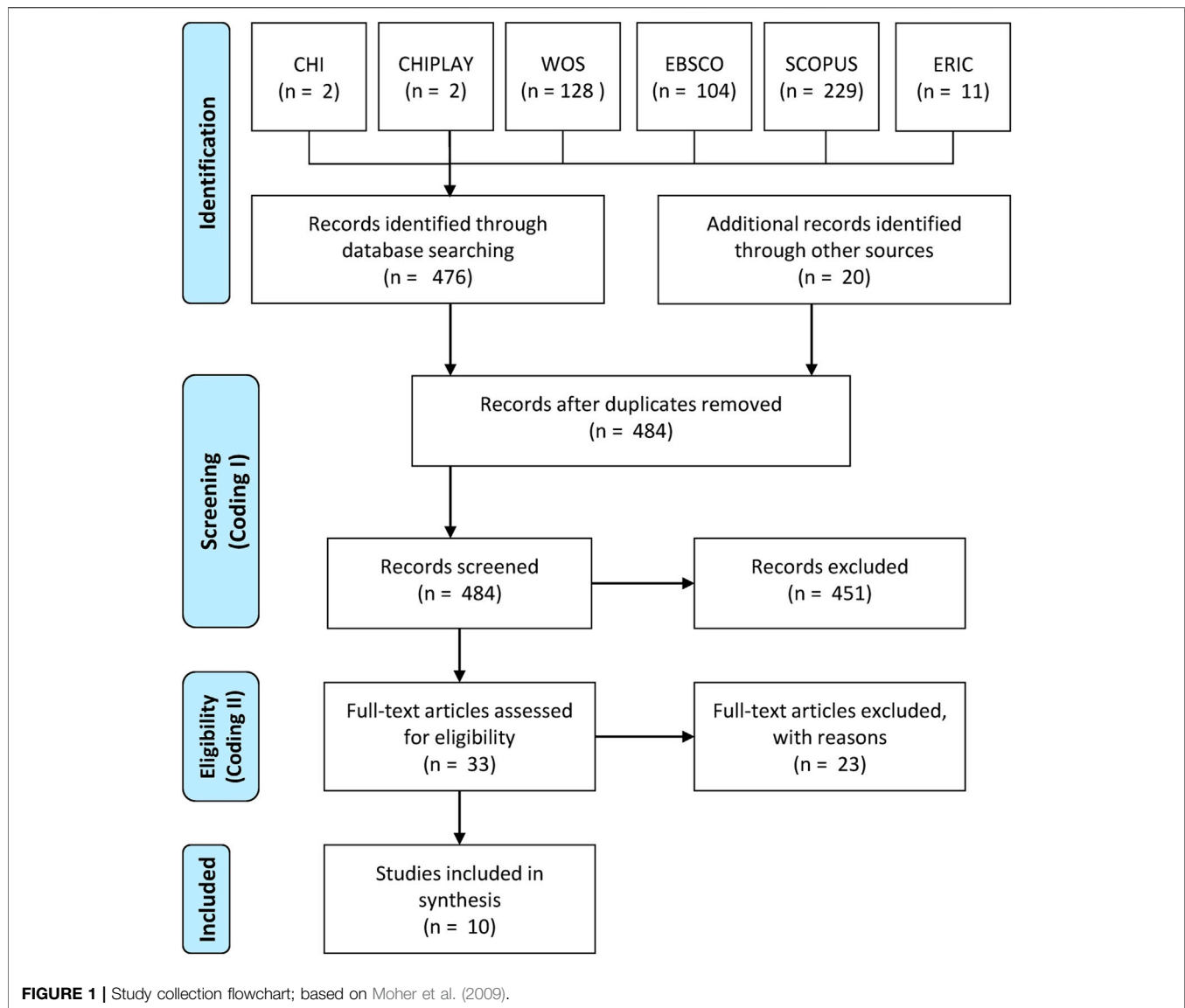
Research Design Overview

This review can be considered systematic (Grant and Booth, 2009) as it includes clarifications of the research questions and a mapping of the literature. Furthermore, the information generated by the review was systematically appraised and synthesized, and a discussion on the types of conclusions that could be drawn within the limits of the review was included (Gough et al., 2017). A systematic approach was chosen to more fully investigate the range of available research in the field and to produce more reliable conclusions with regard to the research questions. Although all literature reviews should be question-led (Booth et al., 2016), rather specific research questions and an overarching synthesis of the findings demarcate this review from similar approaches, such as scoping reviews (Arksey and O'Malley, 2005; Munn et al., 2018). To further enhance methodological integrity, a pre-registration was filed prior to data collection (Ninaus and Nebel, 2020). Finally, the review was conceptualized with a focus on open material by, for instance, providing the developed coding table (Moreau and Gamble, 2020). This, combined with an in-depth description of the approach, should enable future replications of this review, in turn facilitating the systematic identification of additional developments in the field.

Study Data Sources

Researcher Description

Both lead researchers for this study are experts on experimental investigations of learning technologies, such as educational video



games, and actively seek to investigate new approaches for enhancing learning processes. Both researchers have published experimental research in this field and were therefore capable of analyzing and systematizing the sample collected for the current study. However, it should be noted that some papers eligible for inclusion in the analysis were authored by the researchers themselves. Additionally, research assistants with experience working with experimental research and publications supported the coding procedure. The team's affiliation with the field of psychology provided the necessary skills to interpret and evaluate the quality and potential of the measures and frameworks employed as needed for addressing RQ2 and RQ3. That said, given that the overall perspective taken for this work was decidedly psychological in nature, other approaches from relevant fields, such as computer science, received less emphasis.

Study Selection

The following search strategy (**Figure 1**) was pre-registered (Ninaus and Nebel, 2020) and employed for this review. Using the Frontiers research topic "Adaptivity in Serious Games through Cognition-based Analytics" as a basic foundation (Van Oostendorp et al., 2020), articles that addressed some sort of adaptivity, use of a digital educational game, or use of a variation of analytics, and were based on a learning or cognitive framework, were collected. To identify this research, the *Population, Intervention, Comparison, and Outcome* (PICO; Schardt et al., 2007) approach was used, as it results in the largest number of hits compared to other search strategies (Methley et al., 2014). To address *population*, labels describing the desired medium (e.g., educational video games) were used. If necessary, quotation marks were used to indicate the search for a specific term instead of a term's components (e.g., "Serious

Games” to prevent a misleading hit for *serious*). Regarding *intervention*, search terms were used that addressed the topics of our research questions, such as adaptivity and analytics. The *outcome* segment was represented by keywords suitable for capturing the overall cognitive and learning focus of this review (e.g., cognition, learning). Finally, the *comparison* component could not be applied within this review, as a specific empirical procedure was not pre-defined; instead, this component served as a subject of interest for this review. These considerations led to the following search query:

(Adaptivity OR Adaptive OR Adjustment) AND (“Serious Games” OR DGBL OR GBL OR “Educational Videogames” OR “Game Based Learning” OR Simulations) AND (Analytics OR Analytic) AND (Cognitive OR Cognition OR Memory OR Brain OR Learning).

Furthermore, the search engines were adjusted to search within the title, abstract, or keywords of the articles. A quick analysis of the field was conducted to identify the most useful bibliographic databases in line with the psychological and empirical focus the review. As a result, the following databases were used: Association for Computing Machinery—Special Interest Group on Computer-Human Interaction and Special Interest Group on Computer-Human Interaction Play (dl.acm.org/sig/sigchi), Elton Bryson Stephens Company Information Services (search.ebscohost.com), Web of Science (webofknowledge.com), Scopus (scopus.com), and Education Resources Information Center (eric.ed.gov). Other databases with a different focus, such as the technology focused IEEE eXplore, were not used as they might result in no substantial results meeting inclusion- or exclusion criteria presented below (e.g., empirical methodology, inclusion of cognitive aspects, measured outcomes on human participants). In addition, an invitation to recommend articles suitable for the review was sent to colleagues and spread *via* social media. Although the database search was carried out during February 2020, further additions through these additional sources were collected until the end of July 2020.

Altogether, the search returned 496 articles (**Figure 1**; for the full list of coded articles, see: Ninaus and Nebel, 2020). These articles were given an identifier consisting of their database origin and a sequential number (e.g., SCOPUS121). Entries gathered *via* recommendations and other channels were labeled with OTHER. This ID is used throughout this paper when works within the coding table are referenced. The collection process was followed by the first coding of the articles (see **Figure 1**; Coding I) using the pre-registered coding table columns A1 to A18.2 (Ninaus and Nebel, 2020). During this phase, each entry was coded by one coder and verified by a second coder. When disagreement or uncertainty occurred between these two coders, a third coder was consulted and the issue was discussed until the conflict was resolved. For any remaining uncertainty, the rule of thumb was to include rather than exclude the articles in question. The use of at least two independent coders not only increased data quality but also ensured that none of the authors could code their own papers solely by themselves. The focus of this first coding phase was to ensure the eligibility of the search results. More specifically, only papers that presented outcome measures, were published in English, were appropriate in the context of

the research questions, were peer-reviewed, and could be classified as an original research study were included in the review. Furthermore, studies that did not involve digital games, were published prior to 2000, did not document the measures that were used, could be classified as a review or a perspective article, only applied a theoretical or technical framework, or were duplicates from other research results were excluded from the review. In addition, papers for which the full text could not be acquired were excluded as well. The coding procedure was stopped if any of the pre-registered (Ninaus and Nebel, 2020) exclusion criteria or inclusion criteria were met or not met, respectively. For example, the paper WOS32 was published prior to the year 2000 and thus had to be excluded. As a consequence, the columns following the publication year (A6) were not completed.

The rationale for some of these exclusion and inclusion criteria are evident, such as the exclusion of duplicates. However, six criteria should be clarified further: 1) *inclusion of outcome measures*. For the analysis of RQ3, the papers had to provide detailed insights into the measured outcomes as influenced through adaptive elements. If no outcomes were included (e.g., CHIPLAY1) or could not be interpreted with respect to RQ3 (e.g., WOS87), then this work was excluded; 2) *publication language in English*. In order for the resulting coding table to be interpretable by both the coders and the potentially broad readership, only those works published in the English language were included in the review. Papers whose abstracts were translated into English but whose main text was not were also excluded (e.g., SCOPUS128); 3) *appropriateness for the research question*. As complex research questions were abbreviated to short keywords during the database research, the validity of the search results had to be verified. Doing so was crucial, as some keywords generated for the current study have also been used in different, unrelated fields or have ambiguous meanings when not in context. For example, search items such as “Learning” or “Adaptivity” are also used in the field of algorithm research (e.g., EBSCO80); 4) *original research study*. To avoid redundancies and overrepresentations of specific approaches, only original research studies were included. Additionally, other meta- or review-like publications (e.g., EBSCO40) were excluded, as the current work sought to reach independent conclusions. Because of this, editorial pieces for journals (e.g., EBSCO86) or conferences (e.g., SCOPUS19) were also excluded. As RQ3 required experimentation and/or data collection, theoretical frameworks (e.g., SCOPUS199) or similar publications were omitted as well. Additionally, it was specified to focus on data stemming from human participants, thereby ruling out research using simulations (e.g., WOS42); 5) *digital game use*. As specifically stated within RQ1 and RQ2, this review addresses educational video games. The focus on digital technology was used to shed light on new approaches to adaptivity and assessment not feasible using other games, such as educational board or card games. Clearly, then, papers in which no game at all was included (e.g., OTHER7) were excluded from the review as well. Similar but not identical approaches to video games, such as simulations (e.g., CHI1), were also omitted to retain the focus solely on games; 6) *publication year*. Technology continues to rapidly change and evolve. Thus,

comparing research on specific properties of technology is especially challenging. To face this challenge, and to remain focused on new developments within the field of adaptivity and assessment, papers published before the year 2000 were excluded. Moreover, although the assessment of study quality is part of many systematic review frameworks (e.g., Khan, 2003; Jesson et al., 2011), it is also a much debated issue within review research (Newman and Gough, 2020) and was therefore not used as a selection criterion. Study quality was, however, investigated during the full-text analysis.

In cases where certain criteria could not be conclusively determined based on the information presented in article abstracts and/or titles alone, these works were *not* excluded in the initial step. Overall, 33 articles were deemed appropriate for further full-text review (see **Figure 1**; Coding II) or could not be excluded based solely on title or abstract information. Concerning the remaining, excluded articles, 12 were duplicates of other table entries, 70 did not investigate digital games, 49 did not constitute original research, 318 contained no information relevant to the research questions, and 14 were published prior to the year 2000, resulting in an exclusion rate of 93.35% after the initial coding phase.

Subsequent to the initial coding phase, the authors and assistants coded the remaining articles and completed the pre-registered columns B to G, finalizing the coding table for publication alongside this paper. Similar to the initial phase, the second coding procedure was directed by the same criteria described above, excluding or including published works based on a full-text analysis. Consequently, two papers were identified as duplicates, four did not investigate digital games, 10 did not include original research, and six contained no information relevant to the research questions—thus, these 22 articles were also excluded. Ultimately, then, the second coding procedure resulted in the inclusion and coding of 10 papers (2.01% of the complete sample) for subsequent analysis in this systematic literature review (**Table 1**). One paper (OTHER1 and OTHER2) included multiple experiments and was therefore coded into separate rows in order to investigate the experiments individually. As noted above, three of the sample papers (30.00% of the final sample) were authored or co-authored by the authors of the present review.

Papers Reviewed

In addition to the overview table (**Table 1**), a short summary is presented below, as the final sample was small enough to permit a brief discussion of each paper. The papers will be discussed in no particular order.

Operation ARA: A Computerized Learning Game That Teaches Critical Thinking and Scientific Reasoning (OTHER1 and OTHER2)

The paper by Halpern et al. (2012) includes two separate experiments analyzing the impact of a serious game with respect to scientific reasoning. The authors assessed the student's level of knowledge with scores on multiple-choice tests and as a form of adaptivity, assigning the students based on this classification into three different tutoring conditions.

Within the first experiment, this adaptive approach was accompanied by other adjustments that supposedly support learning and then compared to a control group. The second experiment addressed the tutoring component in more detail. For this, three variations (including the adaptive version) were compared to a control group. Overall, the authors used a pre-post design and a sample size of over 300. As a result, the authors concluded that their game was useful for learning as intended; however, conclusions specifically regarding adaptivity can only be cautiously drawn, as the results were either confounded with other variables or only reached significance for very specific comparisons.

Implementation of an Adaptive Training and Tracking Game in Statistics Teaching (SCOPUS2)

Groeneveld (2014) used a popular approach of difficulty assessment and adaptation (i.e., the Elo-algorithm, e.g., Klinkenberg et al., 2011; Nyamsuren et al., 2017) to match student's skills as well as item difficulty in a statistics learning tool. Groeneveld aimed to reach a 75% success rate of solving tasks among the students. The tool was revealed to be useful in a real-life application that included over 400 students. However, no simultaneous control group was implemented and no specific process data on how the adaptive algorithm influenced learning processes could be gathered.

A Pilot Study on the Feasibility of Dynamic Difficulty Adjustment in Game-Based Learning Using Heart-Rate (OTHER11)

Ninaus et al. (2019b) used physiological measurements (i.e., heart rate) to assess player arousal and defined thresholds to adapt game difficulty according to the Yerkes-Dodson Law (Yerkes and Dodson, 1908). Overall, the authors clearly framed their research as a pilot study, justifying the low count of 15 participants for their main experiment. This also explains the lack of dedicated learning measurements. Nonetheless, their results indicated that the adaptive approach resulted in a more difficult, challenging, and fascinating game experience.

Gamification and Smart Feedback: Experiences With a Primary School Level Math App (SCOPUS1)

Based on the theoretical framework of competence-based knowledge space theory (Doignon, 1994; Albert and Lukas, 1999), Kickmeier-Rust et al. (2014) built a digital agent that provided feedback in a gamified math-learning environment. When pre-defined thresholds of user skill levels were reached, the agent provided adaptive information. In an experiment that included 40 second-grade students, Kickmeier-Rust and colleagues were not able to determine any statistically significant benefits of their method.

Competitive Agents and Adaptive Difficulty Within Educational Video Games (OTHER12)

In the experiment by Nebel and colleagues (2020), two game versions that additively regulated social competition were compared to a non-adaptive game scenario. The authors based their assumptions on, among other theories, cognitive load theory

TABLE 1 | Reviewed studies.

CodingTable ID	Title	Author(s)	Publication year	Disciplinary affiliation of primary author	Geographic location/ language	Method of data collection	Method of analysis	Number of participants	Recruitment method
OTHER 11	A pilot study on the feasibility of dynamic difficulty adjustment in game-based learning using heart-rate	Ninaus, M., Tsarava, K., and Moeller, K.	2019	Psychology	Germany/ German	Questionnaires, in-game measures, physiological signals	Descriptive, inferential	12	Voluntary participation
SCOPUS 2	Implementation of an adaptive training and tracking game in statistics teaching	Groeneveld, C. M.	2014	Psychology	Netherlands/ Dutch	Questionnaires, in-game measures, grades	Descriptive, correlational	~441	Incentivized participation
OTHER 1 and 2	Operation ARA: A computerized learning game that teaches critical thinking and scientific reasoning	Halpern, D., Millis, K., Graesser, A., Butler, H., Forsyth, C., and Cai, Z.	2012	Psychology	United States of Amerika/English	Questionnaires	Descriptive, inferential	136 and 215	–
OTHER 12	Competitive agents and adaptive difficulty within educational video games	Nebel, S., Beege, M., Schneider, S., and Rey, G. D.	2020	Psychology	Germany/ German	Questionnaires, in-game measures	Descriptive, inferential	102	Incentivized participation
OTHER 13	Improving student problem solving in narrative-centered learning environments: a modular reinforcement learning framework	Rowe, L., and Lester J.C.	2015	Computer science	United States of Amerika/English	Questionnaires, in-game measures	Descriptive, inferential	61	–
SCOPUS 1	Gamification and smart feedback: Experiences with a primary school level math app	Kickmeier-Rust, M. D., Hillemann, E. C., and Albert, D.	2014	Education	Austria/German	Questionnaires, in-game measures	Descriptive, inferential	40	–
SCOPUS 8	Predicting learning in a multi-component serious game	Forsyth, C. M., Graesser, A., and Millis, K.	2019	Education	United States of Amerika/English	In-game measures	Descriptive, correlational, inferential	192	Voluntary participation
OTHER 15	Effect of personalized gameful design on student engagement	Mora, A., Tondello, G. F., Nacke, L. E., and Arnedo-Moreno, J.	2018	Computer science	Spain/English	Questionnaires, in-game measures	Descriptive, inferential	81	–
SCOPUS 11	Sensor based adaptive learning - lessons learned	Fortenbacher, A., Ninaus, M., Yun, H., Helbig, R., and Moeller, K.	2019	Computer science	Germany/ German	Physiological signals, interviews	Descriptive	12	–
WOS 57	Logs analysis of adapted pedagogical scenarios generated by a simulation serious game architecture	Callies, S., Gravel, M., Beaudry, E., and Basque, J.	2017	Computer science	Canada/English and French	In-game measures	Descriptive	36	Voluntary participation

(Sweller, 1994; Sweller et al., 2011). Yet, they did not specify how adaptive variations might interact with the implications from their theoretical foundation. Overall, the experiment demonstrated empirical support for each adaptive game version, with the version including an artificial and adaptive opponent exhibiting significant advantages. However, as the three game versions differed in more than one feature, the results were confounded to a certain degree and, as a consequence, associating certain outcomes to specific properties would be challenging.

Logs Analysis of Adapted Pedagogical Scenarios Generated by a Simulation Serious Game Architecture (WOS57)

Callies et al. (2020) used a Bayesian network to estimate user knowledge and included a planning algorithm to adjust the learning sequence in a real-estate learning simulation with respect to each user. In particular, feedback, challenge, and learning context were adjusted. An evaluation study was conducted, and qualitative considerations supported the feasibility of the chosen approach, although no quantitative information on learning could be evaluated using inferential statistics.

Sensor Based Adaptive Learning—Lessons Learned (SCOPUS11)

This paper reports the findings of ongoing research from the previously discussed paper OTHER11 (Ninaus et al., 2019b). However, more details on potential assessment (e.g., body temperature, CO₂ data) and adaptive mechanisms (e.g., alerts, recommendations) were provided alongside preliminary results, which indicated that participants became aware of the adaptation.

Effect of Personalized Gameful Design on Student Engagement (OTHER15)

For this research, Mora et al. (2018) used the SPARC model, previously reported by the lead author (Mora et al., 2016), to gamify software that teaches statistical computing. Within the experiment, users were categorized according to self-evaluations based on the Hexad User Type scale (Tondello et al., 2016) and assigned to four different implementations of rules and rewards. The inferential statistical analysis could not identify significant deviations induced through this approach. Following our rule of thumb (see *Study Selection*), we retained this study as part of our final sample even though it was not completely clear whether the focal instrument could be considered a game.

Predicting Learning in a Multi-Component Serious Game (SCOPUS8)

This research by Forsyth et al. (2020) used the same software as that employed in previous experiments (OTHER1 and OTHER2; Halpern et al., 2012) with the aim of gaining insights into knowledge formation divided into deep and shallow learning (Marton and Säljö, 1976). For this, students were assigned to different tutoring conditions based on an assessment of prior knowledge with multiple choice tests. The results suggested that some principles, such as generation, might be suitable predictors for learning within the learning environment under study.

Improving Student Problem Solving in Narrative-Centered Learning Environments: A Modular Reinforcement Learning Framework (OTHER13)

Based on theories such as seductive details (Harp and Mayer, 1998) or modular reinforcement learning (Sutton and Barto, 2018), Rowe and Lester (2015) scaffolded adaptive events in microbiology learning and compared them to a non-adaptive version. In summary, the researchers could not identify significant learning improvements with regard to microbiology learning outcomes.

Analysis Data-Analytic Strategies

During the second phase of full article screening, the full text of the articles was reviewed to search for information needed to address RQ1 to RQ3 as well as to generate an overview that could be systematized and presented within this paper. For this, a content analysis approach (Lamnek and Krell, 2016) was followed. More specifically, to answer RQ1 to RQ3, frequency analysis (Lamnek and Krell, 2016) was used and necessary categories, such as type of research or significance of research, were created. Through this, quantifiable information, such as whether a specific model was used more frequently than others, could be inferred. The codes for these categories were built using the prior knowledge of the researchers in the field of experimental research (e.g., study types) or as derived verbatim from the articles (e.g., names of specific theories). Where applicable, existing codes within the literature were used. For example, *game genre* or *subject discipline* were coded using labels from previous meta-analytic and review-like work (Herz, 1997; Connolly et al., 2012). Thus, frequencies could potentially be compared to other reviews in the field. Additionally, other relevant information was narratively systematized and is discussed within the *Narrative Content Analysis* section. For this, no specific qualitative or quantitative approach was used but was instead reliant on an in-depth discussion and interpretation by the authors, one aligned with the overall aim of this review. In line with the systematic approach, differences were discussed and resolved through consensus.

Methodological Integrity

Several aspects of methodological integrity that needed to be discussed have already been addressed within previous sections (e.g., researcher's perspective). Following the Journal Article Reporting Standards by the American Psychological Association (2020b), additional, complementary information was needed. First, to validate the utility of the findings and the general approach to addressing the study problem, a section specifically devoted to this issue is included in the discussion. Second, to firmly base the findings within the evidence (i.e., the papers), the codes should be closely aligned with the sampled literature and sufficient, supportive excerpts should be provided. However, copyright protection of the original articles prevented the inclusion of exhaustive direct quotations. Third, consistency within the coding process was supported by pre-defined entry options that were prepared for several columns within the coding

table. This was especially useful during the first coding phase when the general inclusion or exclusion criteria were verified. For example, game type (column A11) was coded as either *digital*, *non-digital*, or *unknown based on abstract*. Similar, appropriateness for the review (A15.1) was coded as 1 = *should be considered*, 2 = *should NOT be considered*, or 3 = *not sure*. When a paper was coded as 3, the entry was reviewed by a second coder. For the full-text review, such pre-defined entries were less applicable in certain cases, as the codes themselves were of interest in regard to the initial research questions. For example, column B2, *What are the used theoretical frameworks?* needed to be completed during the review process, as the answer to this question was naturally unavailable prior to the review. Overall, the process of pre-registering the coding table, research method, and research questions ensured a high level of methodological integrity throughout the review. Any deviations or extensions of the a priori formulated research plan (which is not unusual for qualitative or mixed method research; Lamnek and Krell, 2016) is clearly indicated throughout the paper. For example, analyses beyond that covering RQ1–3 were included within the *Narrative* subsections of the results section. Finally, the integrity of research like systematic literature reviews is limited by the integrity of the reviewed material. To ensure basic scientific quality, papers that have not yet undergone peer-review were omitted during the first coding phase. In addition, concerns or potential critical issues were identified within the *Papers Reviewed* section.

RESULTS

The findings described below were based on the final sample. Unless stated otherwise, phrases such as “20% of the sample” refer to the final studies reviewed, not to the initial sample after the literature search. In addition, percentages are reported in relation to the final coded sample (11) and not in relation to the ultimately included papers or manuscripts (10).

RQ1—How is Research in the Field of Analytics for Adaptation in Educational Video Games Currently Conducted?

The current systematic literature review identified 10 relevant papers in total. One paper included two studies and was therefore considered as separate entries—that is, as two entries in the coding table. To better understand how research in the field of analytics for adaptation in educational video games is currently conducted, we provide a comprehensive overview of the games and methods currently being used before we report the specifics of the actual implementation of adaptivity.

Games

In the identified 11 studies, eight different games were used. That is, the same games or at least the same game environments were used in multiple studies or papers. The *game genres* used in the studies did not vary considerably, with simulation games (4) and role-playing games (4) being the most popular. The three other studies employed games that did not fit the predefined genres (cf.

Connolly et al., 2012) and were thus classified as “other.” For instance, in OTHER15, a gameful learning experience was designed using Trello boards and the SPARC model (Mora et al., 2016). OTHER12 utilized a game-like quiz, while SCOPUS1 used a game-like calculation app. However, we must note that even for those games that were classified into a predefined genre, the decision to do so was not always clear cut and was at least debatable, as these games were not always sufficiently described.

The studies and games covered different *subject disciplines*, with games covering Science being the most popular (4). Mathematics was the subject discipline for two games. Business and Technology were covered by one study each. Moreover, three studies did not clearly fit into any of the predefined subject disciplines: OTHER12 covered general factual knowledge on animals, whereas OTHER11 and SCOPUS11, which used the same game, covered procedural knowledge for emergency personnel. In this context, the studies mostly targeted higher education content (6) and continuing education (3). Primary and secondary school content was targeted by one study each.

Methods

In the identified sample of papers, the majority utilized quantitative *data* (9). Only one study used qualitative data, while two studies used a combination of both. The overall mean *sample size* of 120.73 could be sufficient to detect differences of medium effect size between two independent samples. For instance, a two-tailed *t*-test with $\alpha = 0.05$, $\text{power} = 0.8$, and $d = 0.5$ requires a sample size of 128 according to *g*power* (Faul et al., 2007). However, the large standard deviation of 120.46 highlights major differences between the individual studies. Furthermore, the studies utilized very different designs, requiring more or less statistical power. To this end, an a priori power analysis could not be identified as a normal procedure within the sample.

Most of the identified studies (8) evaluated their adaptive approach with *participants* from university/higher education (i.e., university students). Although WOS57 did not report specific information on participants, they were recruited with advertisements on a university campus, which suggests a high likelihood that most of the sample consisted of university students as well. Corresponding with the reported target content (see *Games*), one study was performed in primary school and one in secondary school. The majority of studies were conducted in a real-world *setting* (7). Four studies were performed in laboratory settings. The substantial proportion of studies performed in the field indicates the dominance of an applied approach to the field of adaptive educational video games.

In six of the 11 studies, the authors utilized a *control group* or control condition to evaluate the effects of adaptivity. However, we should note that in SCOPUS2, the authors also made an attempt to descriptively compare their investigated student sample with data they had on students from previous years using the same system but without adaptive components. This comparison was accomplished only descriptively and, overall, the comparison was not sufficiently described, which led to it being

classified as not having used a dedicated control group. In any case, the number of studies utilizing a control group was equivalent to the number of studies that did not, underscoring the urgency of increasing empirical standards in this field of research. However, almost all studies used an experimental or quasi-experimental *study design* (9). One study was mostly correlational, while another employed a qualitative study design. Furthermore, seven studies only used one *measurement point* (i.e., post-test measure), while three studies utilized pre- as well as post-test measurements to evaluate their adaptive approach. In SCOPUS1, neither a classic pre- nor post-test was reported, but instead the authors observed primary class children for two learning sessions with or without the adaptive component (i.e., feedback) of the game in question. Only two studies solely used descriptive *statistical analyses*. One study used a combination of descriptive statistics and correlation but without inferential testing. Hence, the large majority (7) ran descriptive as well as inferential statistical analyses to support their conclusions.

Adaptivity

Overall, the general goal of the adaptive mechanism integrated into the games included in the sample was, in most studies, to optimize learning (7). The rest of the studies focused on instigating a change in either a behavioral, motivational, cognitive, or social variable. Hence, there is a clear focus on directly affecting learning in and of itself in the investigated sample of studies. The strategies by which the authors of the studies sought to achieve adaptivity also varied considerably (see also RQ2 on which theoretical frameworks were used). Hence, providing frequencies on the different approaches was not possible. Instead, a few examples are included here to demonstrate the types of approaches used (for more detailed information, see Ninaus and Nebel, 2020). For instance, the adaptive mode used in OTHER11 was aimed at keeping the players in the game loop for as long as possible. In contrast, SCOPUS2 sought to maintain the chance of being correct in the game at 75% by adapting the game's difficulty. Others tried to adapt the feedback provided by the game (e.g., SCOPUS1) or used natural language processing to develop questions that were posed to the players on the basis of their prior knowledge (e.g., SCOPUS8).

To realize adaptivity within a learning environment, different *sources of data* can be used. Although assessment and adaptivity could be potentially realized using only the game system itself, the majority of the sample (80%) utilized additional surveys and questionnaires. Only two papers could be identified with a potentially less intrusive approach of exclusively utilizing in-game measures. In this vein, only two papers used physiological measures instead of behavioral indicators or survey data to realize adaptivity. Using these data, different *adaptive elements* were realized in the sample by either between (8) or within (3) subject designs. The vast majority of studies were aimed at adapting the difficulty of the game (8). Two other studies used adaptive scaffolding to optimize the learning outcomes. One other study (WOS57) investigated pedagogical scenario adaptation (i.e., automatically generated vs. scripted). These elements were, in most of the studies, adapted in real time

(8), followed by between learning sessions (3). However, this differentiation was not always clear as relevant information was in some instances missing. In seven studies, *processing of data* was done using a user model. Two studies processed the data without a user model, while another two used the raw data only.

RQ2—What Cognitive/Theoretical Frameworks Within Analytics for Adaptation in Educational Video Games are Currently Used?

Apart from OTHER1, OTHER2, OTHER11, and SCOPUS11, which were either two experiments within the same paper or articles authored by almost identical authors, each experiment used a unique theoretical approach. Thus, frequency analysis would have been ineffectual. Instead, a few examples can be used to illustrate the encountered theoretical approaches. For instance, cognitive load theory (OTHER12), learner models with Bayesian networks (WOS57), competence-based knowledge space theory (SCOPUS1), or modular reinforcement learning (OTHER13) were applied. This indicates emphasis on data-driven methodology or institutional preferences rather than on slowly evolving and unifying theoretical frameworks.

RQ3—What Kind of Outcomes are Influenced Through the Adaptive Approach?

Every game within our final sample was originally intended as a learning game. Additionally, every application within the sample was intended to increase learning outcomes. Other possible combinations—for example, entertainment-focused commercial games used within an educational context or educational games aimed at improving metacognition or motivation—were not observed. Instead, 60% of the papers reported that the adaptive mechanism was mainly intended to increase learning outcomes. One study aimed to improve user experience (SCOPUS11), while another sought to explore motivational aspects (WOS57); the remaining two studies, on the other hand, indicated mixed goals (OTHER11, OTHER15). In sum, learning improvements can be identified as the main target of adaptive approaches. A different distribution can be observed within the report of statistically significant findings: 50% of the final sample revealed significant findings, 10% reported mixed results, and 40% did not generate statistically significant outcomes. For this frequency analysis, however, it should be noted that statistical significance alone neither necessarily indicates a relevant effect size nor confirms a sufficient methodological approach. In addition, the potential threat of publication bias cannot be ruled out. No study, at least, revealed negative outcomes, and even if not statistically significant or mixed, the majority of outcomes were indicated to be positive.

Narrative Content Analysis Disciplines and Publication

Although the inclusion criteria allowed for the inclusion of research ranging back to the year 2000, the oldest eligible

article was published in 2012, with 50% of the final sample being published from 2018 to 2020. This indicates a substantial increase in research interest in the field. Psychology or computer science were identified as the disciplinary affiliation of the primary author for 80% of the articles, with only 20% originating from the field of education. This might indicate a lack of sufficient support through educational research. The final sample contained only research from North America or Europe, raising questions regarding the availability or visibility of research from other productive regions, such as Asia. Similarly, only 50% of the final sample could be gathered through database research, raising concerns about sufficient visibility or insufficient standardized keywords in the research field. This could potentially be explained by the fact that apart from two papers published in the *International Journal of Game-Based Learning* (Felicia, 2020), all eligible papers were published in different outlets and conferences. This suggests the lack of an overarching community or publication strategy but also indicates that the field addressed by the systematic review remains in its infancy.

Research Strategy and Limitations

Most frequently, the collected articles comprised exploratory research, as only research questions like “Does an online personalized gameful learning experience have a greater impact on student’s engagement than a generic gameful learning experience?” (Mora et al., 2018, p. 1926) or open-ended questions were included. For example, formulations such as “The *primary* research question that the current paper addresses is ...[...].” (Forsyth et al., 2020, p. 254; emphasis added) entailed unspecified additional observations. Rarely, hypothesis testing research was identified: “Hypothesis 4: Learners playing against adaptive competitive elements demonstrate higher retention scores than players competing against human opponents” (Nebel et al., 2020, p. 5). Although exploratory studies are very valuable in early research, their outcomes are subject to more methodological limitations than theoretically and empirically supported hypothesis-testing research. Additional limitations are imposed through frequent applications of quasi-experimental factors using split groups—e.g., “we split the participants into three roughly equal groups based on pretest scores” (Forsyth et al., 2020, p. 268)—or with separations based on other sample properties—e.g., “Students belonged to the CAS or CAT group according to the native language recorded in their academic profile” (Mora et al., 2018, p. 1928).

Occasionally, the authors used limited statistical methods or reported disputable findings if no significant result could be produced: “[...] the descriptive statistics suggest that personalization of gameful design for student engagement in the learning process seems to work better than generic approaches, since the metrics related to behavioral and emotional engagement were higher for the personalized condition in average” (Mora et al., 2018, p. 1932). This was observed even in cases in which the authors were aware of their shortcomings—e.g., “[...] there is a danger of mistaking a correlation for causality [...]” (Groeneveld, 2014, p. 57)—or if critical methodological difficulties, such as alpha-error

inflation, were not considered—“We then computed correlations between all of the measures for the cognitive processes and behaviors (i.e., time-on-task, generation, discrimination, and scaffolding) and the proportional learning gains for the two topics (experimental, sampling) for each of the four groupings” (Forsyth et al., 2020, p. 265). However, rarely, the necessary corrections were applied: “[...] Sidak corrections will be applied to the pairwise comparisons between these groups” (Nebel et al., 2020, p. 8). In contrast, even for process data for which a myriad of potential comparisons could be made, statistical criteria, such as significance levels, were handled incautiously: “A two-tailed *t*-test indicated that students in the Induced Planner condition ($M = 13.7$, $SD = 10.9$) conducted marginally fewer tests than students in the Control Planner condition ($M = 19.5$, $SD = 14.4$), $t(59) = 1.80$, $p < 0.08$ ” (Rowe and Lester, 2015, p. 8). Justifications for such methodological issues relied on *the goal justifies the means* approach: “Although the assumption of independence was violated, the goal of the correlations was to simply serve as a criterion for selecting predictor variables to include in follow-up analyses” (Forsyth et al., 2020, p. 265). As a consequence, the validity of the gathered insights are questionable in light of the methods by which they were generated.

Similar to other emerging fields, yet rarely, the sampled studies employed standardized measurements or comparable indicators: “Second, the high success rate in the final exams is reassuring, but can hardly be considered evidence” (Groeneveld, 2014, p. 57). However, some authors tried to overcome such methodological issues by pre-testing the measures themselves: “There were two versions of our measure of learning [...] Reliability was established using over 200 participants recruited through Amazon Mechanical Turk” (Forsyth et al., 2020, p. 263). However, such pre-testing regarding the effectiveness of the learning mechanism or the suitability of the measures themselves was scarce. Some authors discussed this issue a posteriori: “In hindsight, the lack of a condition effect on learning is unsurprising. A majority of the AESs provided scaffolding for student’s inquiry behaviors, rather than microbiology content exposure, which was the focus of the pre- and post-tests” (Rowe and Lester, 2015, p. 7).

Potential for Improvements

Despite severe methodological challenges, the sampled researchers highlighted various areas of improvement within the field. For instance, in cases where the game was not created by the researchers themselves or not specifically for the addressed research questions, limited insights into the different processes were acknowledged: “[...] we cannot disentangle one theoretical process from each other without restructuring the entire game. For these reasons, we acknowledge that our findings may not be as generalizable as we would hope in regards to the literature of the learning sciences” (Forsyth et al., 2020, p. 274). Another potential area for improvement was the method of adaptation itself. Often, predefined and global thresholds are used to adjust adaptivity, neglecting potential differences between the users: “Feedback is triggered when a certain pre-defined probability threshold is

reached for a skill/skill state” (Kickmeier-Rust et al., 2014, p. 40). Rarely, these thresholds are based on other research or pilot studies: “The 5 bpm threshold was defined based on previous pilot tests with the same game” (Ninaus et al., 2019b, p. 123). In contrast, the thresholds are frequently based on assumptions made by the authors themselves: “Students have, regardless their ability level, a 75% chance of correctly solving a problem, which is motivating and stimulating [...]” (Groeneveld, 2014, p. 54). In addition to the potential for technical improvements, theoretical work could be enhanced as well, especially as exhaustive motivation regarding the use of adaptive features is not presented but their usefulness is rather assumed or briefly mentioned: “This feature also guarantees learner engagement throughout the duration of the game session” (Callies et al., 2020, p. 1196) or “Diverse psychological viewpoints agree that people are not equal, therefore, they cannot be motivated effectively in the same way” (Mora et al., 2018, p. 1925). Rarely, full chapters discussing which processes might be influenced through adaptive elements are included: “[...] adaptive mechanisms [...] offer several benefits in educational settings. For example, [...]” (Nebel et al., 2020, p. 3/4); alternatively, references to methodological approaches are included: “[...] Evidence-Centered Design [...] requires that each hypothetical cognitive process and behavior to be carefully aligned with the measures. For this reason, we needed to identify general processes or actual behaviors with theoretical underpinnings [...]” (Forsyth et al., 2020, p. 259).

Recommendations

In addition to potential improvements that were more or less explicitly stated or can only be inferred with sufficient knowledge of empirical research, some authors provided clear recommendations. For example, some authors claimed that data extraction and investigation should be intensified: “Thus, we suggest, as demonstrated in this study, that tools [should] be designed to facilitate data extraction and detect learning patterns” (Callies et al., 2020, p. 1196). Furthermore, the complexity of the research field was emphasized: “From our research we learned that quick and easy results are often neither realistic nor meaningful. [...] Using user interaction data for learning analytics is also complex and becomes even more challenging when physiological data are used [...]” (Fortenbacher et al., 2019, p. 197). Within several papers, it was reported that adaptive systems need more time or cases in order to function sufficiently: “[...] for example, students with a very low error rates, with highly unsystematic errors, or students who performed a very small number of tasks, did not received formative feedback because in those cases the system is unable to identify potential problems [...]” (Kickmeier-Rust et al., 2014, p. 45). As a consequence, their full potential could not be assessed within the corresponding studies. As a potential solution, other researchers used pre-test samples to train their algorithms: “[...] we conducted a pair of classroom studies to collect training data for inducing a tutorial planner” (Rowe and Lester, 2015, p. 5). Finally, some authors were aware of methodological limitations and the need for better studies in

the future: “Although the results are encouraging, we recognize that they are not the sort of well-controlled studies that are needed to make strong claims [...]” (Halpern et al., 2012, p. 99) or “However, these results need to be treated with great caution. Future studies with larger sample sizes and more dedicated study designs need to investigate this in more detail” (Ninaus et al., 2019b, p. 126).

Situatedness

Overall, the studies were overwhelmingly conceptualized within an academic context. The researchers worked within various universities, and their samples frequently included students from such institutions. This, consequently, limited their research perspective (e.g., lack of pedagogical input) as well as related applications (e.g., learning impairments, limited technical equipment). In addition, the authors were frequently affiliated with faculty from the same fields, resulting in scarce interdisciplinary discussions within the final sample.

DISCUSSION

Despite the numerous research studies demonstrating the potential of determining user states *via* various analytics relevant for learning (e.g., Klasen et al., 2012; Berta et al., 2013; Brom et al., 2016; Appel et al., 2019; Ninaus et al., 2019a; for a review see; Witte et al., 2015; Nebel and Ninaus, 2019) and supporting their use for adaptation, few studies have actually followed through with this approach as indicated by the current literature review. Consequently, our aim to identify how such analytics have been used to realize adaptive learning in games was compromised by the low number of existing studies in this area. Nevertheless, the current systematic literature provided valuable insights into the nascent field of adaptive educational video games and its use of analytics.

Overall, the existing research on adaptive educational games appears to be somewhat heterogenous in terms of the conceptual approaches applied. However, we did identify clear patterns concerning game genres and subject disciplines. First, there seems to be a clear focus on simulations or simulation games, as well as on role-playing games. Although we can only speculate as to the reasons for this, it would seem that this pattern is completely in line with the overall field of educational games. In a recent review on the effects of serious games and educational games by Boyle et al. (2016), simulation games and role-playing games were also the most popular game genres. Likewise, we would argue that simulations might be easier to design than modeling aspects of learning *via* various game features in such a way that they respond accordingly to the adaptive mechanisms implemented. Moreover, more research is needed to better understand how individual game features affect performance and learning outcomes in general. Importantly, the games in question were not always sufficiently described. That is, it was sometimes unclear which core game loop drove the game and to which genre the game best fit. However, a lack of sufficient details on actual gameplay or of overall information on the games within empirical studies is not unique to the field of adaptive games. As

there is no consensus about how to report and describe educational games in the scientific literature, the whole field of serious games or educational games is impacted. The game attributes taxonomy suggested by Bedwell et al. (2012) might serve as one way to achieve consistent reporting standards.

Second, the majority of studies focused mostly on natural sciences, such as math. So-called “softer” disciplines, such as social sciences, were not found in the final sample. In our opinion, these disciplines might be more difficult to operationalize and evaluate. Consequently, implementing adaptive mechanisms within these disciplines is at once more complex and less reliable.

The target audience of the games varied in the investigated sample, from primary and secondary school pupils to university students and vocational training students. Interestingly, however, most of the studies were performed in real-world settings, thus allowing for high ecological validity. Nevertheless, their overall research designs varied tremendously. While most studies employed experimental or quasi-experimental research designs, only three studies evaluated their adaptive mechanisms with pre- as well as post-tests. It was unfortunate that two studies, which constituted 18% of our total sample, used descriptive statistical analyses alone, further emphasizing how the field remains in its early phases. Moreover, the lack of control groups or specific manipulations of individual elements, as well as the absence of process data, in several studies made clear inferences about the impact of the adaptive systems used especially difficult, if not impossible, to generate. However, it should be noted that creating an appropriate control group for studies of adaptive mechanisms is no trivial endeavor, as learning content between adaptive and non-adaptive learning differs by nature.

In most studies, the general goal of adaptive mechanisms was to optimize or improve learning. While a few studies also investigated adaptive mechanisms on behavioral, motivational, cognitive, or social variables, there was a clear focus on learning or knowledge acquisition. This pattern might have originated from the search query we used, as we intentionally focused on cognitive or learning outcomes. More varied was the pattern in which adaptivity was realized. We could not identify a clear trend with regard to the different mechanisms targeted by the implemented adaptivity. The realization of the adaptive approaches, however, was mostly based on surveys or questionnaires. Only four papers used in-game metrics (SCOPUS8 and WOS57) or physiological signals (SCOPUS11 and OTHER11) directly to adapt the games. Thus, there seems to be room for future improvements, especially minding the various methods of assessing process data within games (for a review see Nebel and Ninaus, 2019).

Overall, there was no clear or coherent pattern of theoretical or cognitive frameworks used within analytics for adaptation in educational video games. That is, almost all studies used unique theoretical approaches to justify their adaptive mechanisms. It seems that the use of general learning theories was mostly neglected in the identified sample. Only OTHER12 and OTHER13 shared similar ideas based on cognitive load theories (e.g., Harp and Mayer, 1998; Mayer, 2005). Some of the presented ideas were technologically impressive but lacked a clear theoretical background. We would suggest that interdisciplinary collaborations might overcome this lack of

theory-driven research and help to advance the field of adaptive educational games, which in turn might also increase the effectiveness of adaptive mechanisms. Researchers from different fields should act in concert to fully utilize current possibilities in adaptive game-based learning from a technological as well as theoretical perspective. Besides new sensor technologies to make data acquisition easier and learning analytics algorithms that permit deeper insights into the learning process and that can potentially identify misconceptions among learners, a strong theoretical foundation is also required—not only of general learning principles (e.g., Mayer, 2005) but also learning domain-specific processes (e.g., embodied learning approaches in mathematics; see Fischer et al., 2011).

As mentioned above, all games or studies were aimed at increasing learning outcomes. However, at the same time, not all of these studies actually evaluated learning outcomes alongside user experiences or motivational outcomes. Only about half of the studies found positive effects of adaptation within their evaluated games. At the same time, no negative effects due to adaptivity were reported. That is, many results did not reach statistical significance, which might be attributed to the varying sample sizes used in individual studies. Hence, this literature review cannot make clear conclusions as to the efficacy of analytics for adaptation in educational video games. However, a recent and more general review on adaptive learning technologies in general reached a more positive verdict on the effectiveness of adaptation (Aleven et al., 2016). We therefore remain cautiously optimistic as to the effectiveness of analytics for adaptation in educational video games.

As the current systematic literature review only identified a rather low number of empirical studies, its results needed to be treated with caution as they might not be representative. However, we need to note that identifying only a rather small number of eligible studies is not completely unusual for the field of serious and educational games, in particular when reviewing a subdiscipline of serious games or focusing on specific constructs [c.f. Lau et al., 2017 (9 studies); Eichenberg and Schott, 2017 (15 studies); Perttula et al., 2017 (19 studies)]. Moreover, other recent and more general systematic literature reviews on adaptive learning systems did also only identify a very small number of empirical studies utilizing games or game-like environments (e.g., Aleven et al., 2016; Martin et al., 2020). That is, our current results seem to be in line with other systematic reviews on adaptive learning systems, suggesting that the field of adaptive educational games and its use of analytics is indeed in its very early phases.

Overall Conclusion

Overall, the presented review contributes a previously lacking overview of and deep exploration into the extant research in the field of analytics for adaptation within educational videogames. Increasing attention to this research area was evident, whereas the overall quantity of relevant experimental research was rather low. In this vein, narrative and frequency analysis could confirm existing opinions about the lack of theory-driven approaches (Van Oostendorp et al., 2020) on a systematic level, although the existence of heterogenous approaches and methodological limitations could prevent further systematization

of the dimensions of adaptive systems. This finding, however, is not entirely surprising, not only because of the different disciplines involved but also because some of the contributing research areas struggle with similar challenges themselves. For instance, Human-Computer-Interaction research that investigates video games often encounters substantial methodological and statistical challenges (Vornhagen et al., 2020), whereas educational psychology faces related issues, such as the replication crisis (Maxwell et al., 2015) or infrequent improvements to theoretical frameworks (Alexander, 2018; Mayer, 2018). Nonetheless, the results of this review can be used to address these critical issues and improve future empirical research in the field. In addition, the empirical evidence, albeit limited, is promising and could encourage future investigations and practical applications of the resulting adaptive systems. Although positive effects were achieved despite the fact that the researchers often worked without pedagogical, instructional, or educational theories or conducted limited exploratory investigations the highly needed quantity of research simply does not exist yet. This was concluded with reasonable certainty after conducting the systematic review. However, the sample supported only a somewhat superficial systematization and assumption of resulting effects. Therefore, it remains to be seen how such conclusions might change as the field matures and improves. This argumentation holds true for subsequent systematic work in the field. Using this article as a starting point, the pre-registered and open-data information can be used to improve the process and gather more fine-grained insights or even yield different conclusions. Furthermore, future systematic reviews following an identical methodical but different theoretical focus could systematize and contrast important literature in adjacent fields (e.g., Bellotti et al., 2009). In addition, different research methods, such as further quantitative or qualitative investigations of the main sample, might enrich the gathered insights. However, in light of the methodological heterogeneity of the current investigation and its small sample size, pursuing such an approach at present is unlikely.

Utility of the Findings and Approach in Responding to the Initial RQs

The approach and its findings can be considered to be successful and insightful with regard to major aspects of the initial research questions. A clear picture of current research was gathered (RQ1),

and crucial gaps and heterogenous approaches were clearly identified (RQ2). In this latter respect, the systematic approaches also increased the validity of the conclusions, thereby supporting previous considerations within the field. However, some aspects could not be completely assessed by the pre-registered categories (RQ3), consequently reducing the systematized information collected by this review and its subsequent conclusions. In order to compensate for this limitation, which was discovered after pre-registration, an additional, non-systematic narrative analysis was conducted. Taken together, the approach can be considered fruitful, even though the coding table could be further optimized. In addition, the findings are capable of addressing the initial research questions, even though the answers sometimes contained less information than first assumed during the conceptualization of the review.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repository and accession number(s) can be found below: The datasets (i.e., complete coding tables) for this study can be found in the OSF: <https://osf.io/dvzsa/>.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

The publication of this article was funded by Chemnitz University of Technology.

ACKNOWLEDGMENTS

We want to sincerely thank our student assistants Stefanie Arnold, Gina Becker, Tina Heinrich, Felix Krieglstein, Selina Meyer, and Fangjia Zhai for supporting the initial screening of articles for this review.

REFERENCES

- D. Albert and J. Lukas (Editors) (1999). *Knowledge spaces: theories, empirical research, and applications*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Aleven, V., McLaren, B. M., Sewall, J., and Koedinger, K. R. (2009). A new paradigm for intelligent tutoring systems: example-tracing tutors. *Int. J. Artif. Intell. Educ.* 19 (2), 105–154.
- Aleven, V., McLaughlin, E. A., Glenn, R. A., and Koedinger, K. R. (2016). “Instruction based on adaptive learning technologies,” in *Handbook of research on learning and instruction*. 2nd Edn, Editors R. E. Mayer and P. Alexander (New York, NY: Routledge), 552–560.
- Alexander, P. A. (2018). Past as prologue: educational psychology’s legacy and progeny. *J. Educ. Psychol.* 110 (2), 147–162. doi:10.1037/edu0000200
- American Psychological Association (2020a). Jars—QUAL table 2 qualitative meta-analysis article reporting standards. Available at: <https://apastyle.apa.org/jars/qual-table-2.pdf> (Accessed September 15, 2020).
- American Psychological Association (2020b). JARS—Qual | table 1 information recommended for inclusion in manuscripts that report primary qualitative research. Available at: <https://apastyle.apa.org/jars/qual-table-1.pdf> (Accessed September 15, 2020).
- Anderson, J. R., Boyle, C. F., Corbett, A. T., and Lewis, M. W. (1990). Cognitive modeling and intelligent tutoring. *Artif. Intell.* 42 (1), 7–49. doi:10.1016/0004-3702(90)90093-f

- Appel, T., Sevchenko, N., Wortha, F., Tsarava, K., Moeller, K., Ninaus, M., et al. (2019). Predicting cognitive load in an emergency simulation based on behavioral and physiological measures. 2019 International Conference on Multimodal Interaction (ICMI '19), Tübingen, Germany. (Tübingen, Germany: University of Tübingen), 154–163. doi:10.1145/3340555.3353735
- Arksey, H., and O'Malley, L. (2005). Scoping studies: towards a methodological framework. *Int. J. Soc. Res. Methodol.* 8 (1), 19–32. doi:10.1080/1364557032000119616
- Bedwell, W. L., Pavlas, D., Heyne, K., Lazzara, E. H., and Salas, E. (2012). Toward a taxonomy linking game attributes to learning. *Simulat. Gaming* 43 (6), 729–760. doi:10.1177/1046878112439444
- Bellotti, F., Berta, R., De Gloria, A., and Primavera, L. (2009). Adaptive experience engine for serious games. *IEEE Trans. Comput. Intell. AI Games* 1 (4), 264–280. doi:10.1109/tciaig.2009.2035923
- Berta, R., Bellotti, F., De Gloria, A., Pranantha, D., and Schatten, C. (2013). Electroencephalogram and physiological signal analysis for assessing flow in games. *IEEE Trans. Comput. Intell. AI Games* 5 (2), 164–175. doi:10.1109/tciaig.2013.2260340
- Bloom, B. S. (1968). Learning for mastery. instruction and curriculum. regional education laboratory for the Carolinas and Virginia, topical papers and reprints, number 1. *Eval. Comment* 1 (2).
- Booth, A., Sutton, A., and Papaioannou, D. (2016). *Systematic approaches to a successful literature review*. 2nd Edn. Thousand Oaks, CA: SAGE.
- Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., et al. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Comput. Educ.* 94, 178–192. doi:10.1016/j.compedu.2015.11.003
- Brom, C., Šisler, V., Slussareff, M., Selmbacherová, T., and Hlávka, Z. (2016). You like it, you learn it: affectivity and learning in competitive social role play gaming. *Intern. J. Comput.-Support. Collab. Learn* 11 (3), 313–348. doi:10.1007/s11412-016-9237-3
- Callies, S., Gravel, M., Beaudry, E., and Basque, J. (2020). “Logs analysis of adapted pedagogical scenarios generated by a simulation serious game architecture,” in *Natural language processing: concepts, methodologies, tools, and Applications* (Hershey, PA: IGI Global 2019), 1178–1198.
- Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., and Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Comput. Educ.* 59 (2), 661–686. doi:10.1016/j.compedu.2012.03.004
- Doignon, J.-P. (1994). “Knowledge spaces and skill assignments,” in *Contributions to mathematical psychology, psychometrics, and methodology*. Editors G. H. Fischer and D. Laming (New York, NY: Springer), 111–121.
- Eichenberg, C., and Schott, M. (2017). Serious games for psychotherapy: a systematic review. *Game. Health J.* 6 (3), 127–135. doi:10.1089/g4h.2016.0068
- Faul, F., Erdfelder, E., Lang, A. G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi:10.3758/bf03193146
- Felicia, P. (2020). International journal of game-based learning (IJGBL). Available at: www.igi-global.com/journal/international-journal-game-based-learning/41019 (Accessed September 15, 2020).
- Fischer, U., Moeller, K., Bientzle, M., Cress, U., and Nuerk, H. C. (2011). Sensorimotor spatial training of number magnitude representation. *Psychon. Bull. Rev.* 18, 177–183. doi:10.3758/s13423-010-0031-3
- Forsyth, C. M., Graesser, A., and Millis, K. (2020). Predicting learning in a multi-component serious game. *Technol. Knowl. Learn.* 25 (2), 251–277. doi:10.1007/s10758-019-09421-w
- Fortenbacher, A., Ninaus, M., Yun, H., Helbig, R., and Moeller, K. (2019). “Sensor based adaptive learning—lessons learned,” in *Die 17. Fachtagung bildungstechnologien, lecture notes in informatics (LNI)*. Editors N. Pinkwart and J. Konert (Bonn, Germany: Gesellschaft für Informatik), 193–198.
- Freire, M., Serrano-Laguna, Á., Iglesias, B. M., Martínez-Ortiz, I., Moreno-Ger, P., and Fernández-Manjón, B. (2016). “Game learning analytics: learning analytics for serious games,” in *Learning, design, and technology: an international compendium of theory, research, practice, and policy*. Editors M. J. Spector, B. B. Lockee, and M. D. Childress (Cham, Switzerland: Springer International Publishing), 1–29.
- Gough, D., Oliver, S., and Thomas, J. (2017). “Introducing systematic reviews,” in *An introduction to systematic reviews*. 2nd Edn. Editors D. Gough, S. Oliver, and J. Thomas (Thousand Oaks, CA: SAGE Publications), 1–17.
- Grant, M. J., and Booth, A. (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Inf. Libr. J.* 26 (2), 91–108. doi:10.1111/j.1471-1842.2009.00848.x
- Greller, W., and Drachslers, H. (2012). Translating learning into numbers: a generic framework for learning analytics. *J. Educ. Technol. Soc.* 15 (3), 42–57.
- Groeneveld, C. M. (2014). “Implementation of an adaptive training and tracking game in statistics teaching,” in *Computer assisted assessment. Research into E-assessment*. Editors M. Kalz and E. Ras (Cham, Switzerland: Springer International Publishing), 53–58.
- Halpern, D. F., Millis, K., Graesser, A. C., Butler, H., Forsyth, C., and Cai, Z. (2012). Operation ARA: a computerized learning game that teaches critical thinking and scientific reasoning. *Think. Skills Creativ.* 7 (2), 93–100. doi:10.1016/j.tsc.2012.03.006
- Harp, S. F., and Mayer, R. E. (1998). How seductive details do their damage: a theory of cognitive interest in science learning. *J. Educ. Psychol.* 90, 414–434. doi:10.1037/0022-0663.90.3.414
- Hartley, J. R., and Sleeman, D. H. (1973). Towards more intelligent teaching systems. *Int. J. Man Mach. Stud.* 5 (2), 215–236. doi:10.1016/s0020-7373(73)80033-1
- Herz, J. C. (1997). *Joystick nation*. Boston, MA: Little Brown and Company.
- Jesson, J., Matheson, L., and Lacey, F. M. (2011). *Doing your literature review: traditional and systematic techniques*. Thousand Oaks, CA: SAGE Publications.
- Khan, K. S., Kunz, R., Kleijnen, J., and Antes, G. (2003). Five steps to conducting a systematic review. *J. R. Soc. Med.* 96, 118. doi:10.1177/014107680309600304
- Kickmeier-Rust, M. D., Hillemann, E.-C., and Albert, D. (2014). Gamification and smart feedback. *Int. J. Game Base. Learn.* 4 (3), 35–46. doi:10.4018/ijgbl.2014070104
- Klasen, M., Weber, R., Kircher, T. T., Mathiak, K. A., and Mathiak, K. (2012). Neural contributions to flow experience during video game playing. *Soc. Cognit. Affect Neurosci.* 7 (4), 485–495. doi:10.1093/scan/nsr021
- Klinkenberg, S., Straatemeier, M., and van der Maas, H. L. J. (2011). Computer adaptive practice of Maths ability using a new item response model for the fly ability and difficulty estimation. *Comput. Educ.* 57 (2), 1813–1824. doi:10.1016/j.compedu.2011.02.003
- Lamnek, S., and Krell, C. (2016). *Qualitative sozialforschung*. Weinheim, Germany: Beltz Verlagsgruppe.
- Lau, H. M., Smit, J. H., Fleming, T. M., and Riper, H. (2017). Serious games for mental health: are they accessible, feasible, and effective? A systematic review and meta-analysis. *Front. Psychiatr.* 7, 209. doi:10.3389/fpsy.2016.00209
- Lopes, R., and Bidarra, R. (2011). Adaptivity challenges in games and simulations: a survey. *IEEE Trans. Comput. Intell. AI Games* 3 (2), 85–99. doi:10.1109/tciaig.2011.2152841
- Mangaroska, K., and Giannakos, M. (2019). Learning analytics for learning design: a systematic literature review of analytics-driven design to enhance learning. *IEEE Trans. Learn. Technol.* 12 (4), 516–534. doi:10.1109/ltl.2018.2868673
- Martin, F., Chen, Y., Moore, R. L., and Westine, C. D. (2020). Systematic review of adaptive learning research designs, context, strategies, and technologies from 2009 to 2018. *Educ. Technol. Res. Dev.* 68 (4), 1903–1929. doi:10.1007/s11423-020-09793-2
- Marton, F., and Säljö, R. (1976). On qualitative differences in learning: I—outcome and process*. *Br. J. Educ. Psychol.* 46 (1), 4–11. doi:10.1111/j.2044-8279.1976.tb02980.x
- Maxwell, S. E., Lau, M. Y., and Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *Am. Psychol.* 70 (6), 487–498. doi:10.1037/a0039400
- Mayer, R. E. (2005). “Cognitive theory of multimedia learning,” in *The Cambridge handbook of multimedia learning*. Editor R. E. Mayer (Cambridge, United Kingdom: Cambridge University Press), 31–48.
- Mayer, R. E. (2018). Educational psychology’s past and future contributions to the science of learning, science of instruction, and science of assessment. *J. Educ. Psychol.* 110 (2), 174–179. doi:10.1037/edu0000195
- Methley, A. M., Campbell, S., Chew-Graham, C., McNally, R., and Cheraghi-Sohi, S. (2014). PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. *BMC Health Serv. Res.* 14, 579. doi:10.1186/s12913-014-0579-0
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and Group, T. P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 6 (7), e1000097. doi:10.1371/journal.pmed.1000097

- Mora, A., Planas, E., and Arnedo-Moreno, J. (2016). "Designing game-like activities to engage adult learners in higher education," in Proceedings of the fourth international conference on technological ecosystems for enhancing multiculturalism—TEEM'16, Salamanca, Spain, November 2–4, 2016 (ACM), 755–762.
- Mora, A., Tondello, G. F., Nacke, L. E., and Arnedo-Moreno, J. (2018). Effect of personalized gameful design on student engagement. *IEEE Global Eng. Educ. Conf. (EDUCON)*, 1925–1933. doi:10.1109/EDUCON.2018.8363471
- Moreau, D., and Gamble, B. (Forthcoming 2020). Conducting a meta-analysis in the age of open science: tools, tips, and practical recommendations. *Psychol. Methods* doi:10.31234/osf.io/t5dwg
- Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., and Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med. Res. Methodol.* 18 (1), 143. doi:10.1186/s12874-018-0611-x
- Nebel, S., Beege, M., Schneider, S., and Rey, G. D. (2020). Competitive agents and adaptive difficulty within educational video games. *Front. Educ.* 5, 129. doi:10.3389/educ.2020.00129
- Nebel, S., and Ninaus, M. (2019). "New perspectives on game-based assessment with process data and physiological signals," in *Game-based assessment revisited*. Editors D. Ifenthaler and Y. Jeon (Cham, Switzerland: Springer), 141–161.
- Nebel, S., Schneider, S., Beege, M., Kolda, F., Mackiewicz, V., and Rey, G. D. (2017a). You cannot do this alone! increasing task interdependence in cooperative educational videogames to encourage collaboration. *Educ. Technol. Res. Dev.* 65 (4), 993–1014. doi:10.1007/s11423-017-9511-8
- Nebel, S., Schneider, S., Beege, M., and Rey, G. D. (2017b). Leaderboards within educational videogames: the impact of difficulty, effort and gameplay. *Comput. Educ.* 113, 28–41. doi:10.1016/j.compedu.2017.05.011
- Nebel, S., Schneider, S., Schledjewski, J., and Rey, G. D. (2017c). Goal-setting in educational video games. *Simulat. Gaming* 48 (1), 98–130. doi:10.1177/1046878116680869
- Newman, M., and Gough, D. (2020). "Systematic reviews in educational research," in *Methodology, perspectives and application*. Editors O. Zawacki-Richter, M. Kerres, S. Bedenlier, M. Bond, and K. Buntins (Wiesbaden, Germany: Springer VS), 3–22.
- Ninaus, M., Greipl, S., Kiili, K., Lindstedt, A., Huber, S., Klein, E., et al. (2019a). Increased emotional engagement in game-based learning—a machine learning approach on facial emotion detection data. *Comput. Educ.* 142, 103641. doi:10.1016/j.compedu.2019.103641
- Ninaus, M., Tsarava, K., and Moeller, K. (2019b). "A pilot study on the feasibility of dynamic difficulty adjustment in game-based learning using heart-rate," in *Games and learning alliance*. Editors A. Liapis, G. N. Yannakakis, M. Gentile, and M. Ninaus (Cham, Switzerland: Springer International Publishing), 117–128.
- Ninaus, M., Kober, S. E., Friedrich, E. V. C., Dunwell, I., Freitas, S. D., Arnab, S., et al. (2014). Neurophysiological methods for monitoring brain activity in serious games and virtual environments: a review. *Ijtel* 6 (1), 78–103. doi:10.1504/ijtel.2014.060022
- Ninaus, M., and Nebel, S. (2020). A systematic literature review of analytics for adaptation within educational videogames. Available at: <https://osf.io/dvzsa/> (Accessed September 15, 2020).
- Nyamsuren, E., van der Vegt, W., and Westera, W. (2017). "Automated adaptation and assessment in serious games: a portable tool for supporting learning," in *Advances in computer games*. Editors M. H. M. Winands, H. J. van den Herik, and W. A. Kusters (Cham, Switzerland: Springer International Publishing), Vol. 10664, 201–212.
- Orji, R., Nacke, L. E., and Di Marco, C. (2017). "Towards personality-driven persuasive health games and gamified systems," in Proceedings of the 2017 CHI conference on human factors in computing systems, Denver, CO, May 6–11, 2017 (ACM), 1015–1027.
- Perttula, A., Kiili, K., Lindstedt, A., and Tuomi, P. (2017). Flow experience in game based learning—a systematic literature review. *Int. J. Serious Games* 4 (1), 57–72. doi:10.17083/ijsg.v4i1.151
- Rowe, J. P., and Lester, J. C. (2015). "Improving student problem solving in narrative-centered learning environments: a modular reinforcement learning framework," in *Artificial intelligence in education*. Editors C. Conati, N. Heffernan, A. Mitrovic, and M. F. Verdejo (Cham, Switzerland: Springer International Publishing), Vol. 9112, 419–428.
- Schardt, C., Adams, M. B., Owens, T., Keitz, S., and Fontelo, P. (2007). Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Med. Inform. Decis. Mak.* 7 (1), 16. doi:10.1186/1472-6947-7-16
- Schrader, C., Brich, J., Frommel, J., Riemer, V., and Rogers, K. (2017). "Rising to the challenge: an emotion-driven approach toward adaptive serious games," in *Serious games and edutainment applications*. Cham, Switzerland: Springer, 3–28. doi:10.1007/978-3-319-51645-5_1
- Shute, V. J., and Zapata-Rivera, D. (2012). "Adaptive educational systems," in *Adaptive technologies for training and education*. Editors P. J. Durlach and A. M. Lesgold (New York, NY: Cambridge University Press), 7–27.
- Skinner, B. F. (1958). Teaching machines: from the experimental study of learning come devices which arrange optimal conditions for self instruction. *Science* 128, 969–977. doi:10.1126/science.128.3330.969
- Streicher, A., and Smeddinck, J. D. (2016). "Personalized and adaptive serious games," in *Entertainment computing and serious games*. Editors R. Dörner, S. Göbel, M. Kickmeier-Rust, M. Masuch, and K. Zweig (Cham, Switzerland: Springer International Publishing), Vol. 9970, 332–377.
- Sutton, R. S., and Barto, A. G. (2018). *Reinforcement learning an introduction*. 2nd Edn. Cambridge, MA: MIT Press.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learn. Instruct.* 4 (4), 295–312. doi:10.1016/0959-4752(94)90003-5
- Sweller, J., Ayres, P., and Kalyuga, S. (2011). *Cognitive load theory*. New York, NY: Springer.
- Sweller, J., van Merriënboer, J. J. G., and Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educ. Psychol. Rev.* 10, 251–296. doi:10.1023/a:1022193728205
- Tondello, G. F., Wehbe, R. R., Diamond, L., Busch, M., Marczewski, A., and Nacke, L. E. (2016). "The gamification user types hexad scale," in Proceedings of the 2016 annual symposium on computer-human interaction in play, Austin, TX, October 18, 2016 (ACM), 229–243.
- Van Oostendorp, H., Bakkes, S. C. J., Jay, T., Habgood, J., Brom, C., and Kickmeier-Rust, M. (2020). Adaptivity in serious games through cognition-based analytics | Frontiers research topic. Available at: <https://www.frontiersin.org/research-topics/11739/adaptivity-in-serious-games-through-cognition-based-analytics> (Accessed September 15, 2020).
- Vornhagen, J. B., Tyack, A., and Mekler, E. D. (2020). *Statistical significance testing at CHI PLAY: challenges and opportunities for more Transparency CHIPLAY'20*. Canada: ACM.
- Witte, M., Ninaus, M., Kober, S. E., Neuper, C., and Wood, G. (2015). Neuronal correlates of cognitive control during gaming revealed by near-infrared spectroscopy. *PLoS One* 10 (8), e0134816. doi:10.1371/journal.pone.0134816
- Yannakakis, G. N., and Togelius, J. (2011). Experience-driven procedural content generation. *IEEE Trans. Affective Comput.* 2 (3), 147–161. doi:10.1109/t-affc.2011.6
- Yerkes, R. M., and Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *J. Comp. Neurol. Psychol.* 18, 459–482. doi:10.1002/cne.920180503

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Ninaus and Nebel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.