To Be or Not to Be Stuck, or Is It a Continuum?: A Systematic Literature Review on the Concept of Being Stuck in Games

TOBIAS DREY, Institute of Media Informatics, Ulm University, Germany
FABIAN FISCHBACH, Institute of Media Informatics, Ulm University, Germany
PASCAL JANSEN, Institute of Media Informatics, Ulm University, Germany
JULIAN FROMMEL, Department of Computer Science, University of Saskatchewan, Canada
MICHAEL RIETZLER, Institute of Media Informatics, Ulm University, Germany
ENRICO RUKZIO, Institute of Media Informatics, Ulm University, Germany

Players can get stuck in video games, which impedes their process to their goal and results in unfavorable outcomes like negative emotions, impediments of flow, and obstacles for learning. Currently, it is not easily possible to assess if a player is stuck, as no widely accepted definition of "being stuck" in games exists. We conducted 13 expert interviews and a systematic literature review with 104 relevant papers selected from 4022 candidates. We present a definition of "being stuck" that conceptualizes the state as a continuum and contextualize it within related concepts. Our stuck continuum can be applied to regulate the player's stuck level. We propose a taxonomy of measures that are useful for the detection of the level of stuckness and discuss the effectiveness of countermeasures. Our stuck concept is crucial for game developers creating an optimal player experience in games.

CCS Concepts: • **General and reference** → **Surveys and overviews**; • **Human-centered computing** → **HCI theory, concepts and models**; • **Software and its engineering** → *Interactive games*; • **Applied computing** → *Computer games.*

Additional Key Words and Phrases: stuck; systematic literature review; survey; continuum; games; taxonomy; interviews

ACM Reference Format:
Tobias Drey, Fabian Fischbach, Pascal Jansen, Julian Frommel, Michael Rietzler, and Enrico Rukzio. 2021. To Be or Not to Be Stuck, or Is It a Continuum?: A Systematic Literature Review on the Concept of Being Stuck in Games. *Proc. ACM Hum.-Comput. Interact.* 5, CHI PLAY, Article 229 (September 2021), 35 pages. https://doi.org/10.1145/3474656

229

1 INTRODUCTION

Video games are increasingly used for leisure as well as educational activities [31, 118]. To achieve a game's purpose to be fun, it is important that players can continuously perform valuable in-game

Authors' addresses: Tobias Drey, Institute of Media Informatics, Ulm University, James-Franck-Ring, Ulm, 89081, Germany, tobias.drey@uni-ulm.de; Fabian Fischbach, Institute of Media Informatics, Ulm University, James-Franck-Ring, Ulm, 89081, Germany, fabian.fischbach@uni-ulm.de; Pascal Jansen, Institute of Media Informatics, Ulm University, James-Franck-Ring, Ulm, 89081, Germany, pascal.jansen@uni-ulm.de; Julian Frommel, Department of Computer Science, University of Saskatchewan, 110 Science Place, Saskatoon, Saskatchewan, S7N 5C9, Canada, julian.frommel@usask.ca; Michael Rietzler, Institute of Media Informatics, Ulm University, James-Franck-Ring, Ulm, 89081, Germany, michael.rietzler@uni-ulm.de; Enrico Rukzio, Institute of Media Informatics, Ulm University, James-Franck-Ring, Ulm, 89081, Germany, enrico.rukzio@uni-ulm.de.

2573-0142/2021/9-ART229 $15.00
https://doi.org/10.1145/3474656

actions to overcome challenges [2, 45, 46, 160], which should be not too easy nor too hard [95], so that all players can reach their personal goal. If not, players can find themselves stuck. Stuck players might stop playing and churn [64, 93], which game developers want to prevent [33]. Therefore, games should be able to recognize situations in which players are in a stuck state and provide assistance to overcome an obstacle without diluting the challenge [62].

Despite its apparent prevalence, it is unclear how to detect if a player is stuck, or even what "being stuck" exactly means, as exemplified by the lack of a widely accepted definition in games. This is surprising, as negative states for players (e.g., player frustration and churn [64, 93]) are frequently associated with stuck players. A variety of concepts of the player experience (PX) appear useful for the conceptualization of being in a stuck state. Flow describes a state of optimal experience that occurs when challenges and skills are optimally balanced between boredom and anxiety [38, 39]. We will show that flow in games is antithetical to the state of "being stuck", as a player who is stuck is very likely not experiencing a flow state, but how it is not its exact opposite. In an affective gaming context, human emotions are considered important for PX [188] and might occur when players experience the stuck state. However, measuring emotions such as frustration has its limitations for assessing if players experience stuck (see types of frustration [64] and the paradox of failure [82, 83], which show that short periods of frustration can be necessary for an overall enjoyable PX). Using only the players' performance is also not sufficient, as personal skills and personal demands such as experience, knowledge, or interest (see Nguyen et al. [122]) influence when a player actually reaches the stuck state. These examples show that there are concepts that describe characteristics of the player's state occurring when entering a stuck state. However, these constructs do not fully capture the nature of this state, as they are conceptualized differently. This suggests the need for a better understanding of the state of "being stuck". To investigate this problem, we investigated the following research questions (RQs):

- *RQ 1:* What are frequently used terms for "being stuck" in games?
- *RQ 2:* How can the state of "being stuck" in games be defined?
- *RQ 3:* How can games detect if players are stuck?
- *RQ 4:* How can games help players when stuck?

To answer these RQs, we conducted expert interviews and a systematic literature review (SLR). With the interviews, we gathered an initial understanding of the state of "being stuck". They resulted in two lists of terms related to the state of "being stuck" and how to measure it. These terms were used to define the search query for the SLR. The SLR was based on the PRISMA guidelines [111, 112]. We selected the three databases ACM Digital Library [9], IEEE Xplore [78], and Science Direct [52], to search for relevant publications, resulting in an initial set of 4022 papers. We defined exclusion and inclusion criteria and reduced the number of relevant publications to 181 in the initial abstract screening phase. Relevant information was then coded in a second phase of full-text screening. Finally, we considered 104 publications as relevant for the synthesis to answer our RQs.

In our SLR results, we show which terms related to the state of "being stuck" were used by the considered publications. To measure the players' stuck level, we provide a taxonomy categorizing measurement techniques stated in the 104 included papers using physiological data, gameplay data, and questionnaires. Further, we summarize existing stuck countermeasures and their effectiveness in helping players who are stuck.

We then discuss how existing concepts such as flow, challenge and difficulty, affective state, engagement, progress and performance, and cognitive load are related to the stuck state. Following this and summing up all the findings of the interviews and the SLR, we provide a definition for the state of "being stuck" as a continuous subjective state that occurs when a player cannot reach the personal goal. Furthermore, we introduce a continuum based on our stuck definition that shows the players' stuck level, i.e., the degree of "being stuck", and propose a game-specific implementation thereof.

The main contributions of our work are:

(1) A list of terms relevant to the state of "being stuck".
(2) A taxonomy of measures useful for the detection of the player's stuck level.
(3) An overview of countermeasures and their effectiveness in reducing the player's stuck level.
(4) A definition of a stuck concept in games including a stuck level continuum.

These contributions are crucial for instructors who use games in education to have a profound way of detecting and intervening when students get stuck, for games themselves to detect if players get stuck (e.g., for automatically adjusting game features), and for game developers to guide the implementation of game mechanics that do not result in states that could negatively affect PX or learning outcomes.

2 RELATED WORK

An initial literature search showed that our stuck research is related to previous work on various concepts of PX, such as flow, affective state, and challenge (see Mekler et al. [109]).

The definition of flow by Csikszentmihalyi [38–40] describes a state of optimal experience between boredom and anxiety. To reach flow, Csikszentmihalyi [39] defined eight conditions such as having the chance to complete the task, having a clear goal, getting immediate feedback, and losing track of time. The flow state is also important for education and the learning experience [91], as it prevents mental over- and underload [61] and provides a positive influence on memorization and comprehension [53]. Further, flow is linked with terms such as concentration, challenge, and skill, which was shown by Sweetser et al. with their GameFlow model [162–164]. Fu et al. [59] introduced an extended version called EGameFlow for e-learning. We conclude that flow as a player state, or more precisely the lack thereof, can be an important measure for detecting if a player is stuck because it is based on goal-orientation and challenge-skill balance.

Affective state is generally considered important for the experience of games, e.g., in its operationalization for game enjoyment [109]. Regardless of conceptualization into dimensions like valence, arousal or dominance [58, 145], or emotional states like frustration [64] and uncertainty [137, 138], it is evident that affective responses are important measures for how players experience games. As such, it makes sense to assume that the experience of "being stuck" in a game results in emotional responses, e.g., in frustration, that might be useful for defining the state of "being stuck".

Challenge is often linked to difficulty and performance. An appropriate level of these results in better PX during gameplay [43, 57], while learning performance and efficiency can benefit, too [121]. This is achieved by avoiding boring or frustrating game situations [57, 178] and by creating a level of challenge that matches the players' skills [43, 44]. Considering that players might feel stuck in situations where challenges are higher than their skills, there is an apparent link between the state of "being stuck" and the degree of challenge, suggesting its value as a useful indicator. The emotional perception of challenge is conceptualized with the concept of perceived challenge [45, 46], describing a PX state.

The concepts flow, affective state, and challenge show that multiple concepts exist to describe the player state or PX. The definition of flow also uses the concepts of challenge and emotions to describe their targeted state [40]. This ambiguous usage of concepts in the context of player states and PX and the lack of a stuck definition highlights the necessity of a holistic definition of "being stuck" and a proper contextualization in light of existing concepts. An overview about related concepts for player state and PX was necessary for this, leading us to conduct a SLR. To prepare it, we started with expert interviews to set the appropriate scope of the SLR.

3 PRELIMINARY EXPERT INTERVIEWS

We used expert interviews (1) to confirm the need to describe the state of "being stuck", (2) to align the direction of this research and our understanding of this state with multiple experts, and (3) to verify the relevance of our RQs.

3.1 Participants

We recruited 13 experts (4 female, 9 male), who were between 27 to 41 years old (M = 33.6, SD = 4.6), through personal and professional contacts choosing experts with human-computer interaction (HCI) background (n = 7), serious games background (n = 3), and psychological and pedagogical background (n = 3). Two had professional experience working in the games industry, and ten had experience with educational games.

3.2 Method

We conducted semi-structured expert interviews in preparation for the SLR in a similar way as Brereton et al. [23]. The interviews had two parts. During the first part, we let our participants rate terms related to the state of "being stuck" on a 7-point Likert scale. We used these terms to define the search query for our SLR, as described later. The second part was a semi-structured interview asking the experts regarding their opinion about the state of "being stuck". We used these findings as additional input for the SLR and the main synthesis.

All interviews were conducted by the lead author. The questions were asked in English, but the experts were free to answer in English or their native language. All interviews were conducted online and audio recorded. Interviews lasted between 15 and 38 minutes (M = 26.8, SD = 8.3). Demographics were collected at the beginning.

3.2.1 Part 1: Rating Terms Related to the State of "Being Stuck". Our goal was to find frequently used terms for states that coincide with the state of "being stuck" and that may have been used in publications to describe a state in which users are stuck without using the specific term *stuck*. The result was a list of terms that can be used to create a search query to find relevant papers for the SLR. First, and not to bias the experts, we asked them to provide relevant terms themselves, which we used to extend our list. Then, using previous works, we presented an initial list of terms that we deemed relevant: *Challenge* [45, 46, 59, 90, 104], *Difficulty* [5, 104], *Emotion* [57, 58, 142], *Engagement* [5, 90, 92, 104, 126, 142, 157], *Enjoyment* [59, 157], *Flow* [57, 59, 90, 92, 104, 157], *Frustration* [104, 126, 142], *Immersion* [3, 59, 92], *Player State* [104, 157], *Presence* [3, 92], *Progress* [58, 126], *Skill Level* [41, 90], and *Stress* [135, 136]. We asked each expert to rate these terms regarding their relevance to "being stuck" on a 7-point Likert scale, from 1 (= "not relevant") to 7 (= "very relevant").

The same approach was used to find related terms for the detection of stuck players. This would be another important part of our search query in answering RQ 3, *"How can games detect if players are stuck?"*. We used the following initial list of terms gained from related work and extended it dynamically as well: *Assess* [58, 126, 142], *Define* [57, 58], *Detect* [58, 142], *Determine* [58, 143], *Measure* [58, 126, 142], and *Recognize* [58, 142].

3.2.2 Part 2: Asking the Experts Regarding Their Opinion about the State of "Being Stuck". We asked our experts questions about when they feel stuck in games, how they would describe such situations, and how they think it is possible to determine this. Additionally, we asked them about their professional background. Those with experience in educational games were asked additionally to answer the questions in relation to educational games and to state similarities and differences between educational/serious games and leisure games in respect to the state of "being stuck". We decided to add additional questions regarding educational games due to their different goals compared to leisure games, as leisure games are played for fun and educational games for learning. Therefore, we assumed that this may affect the stuck state differently and wanted to know our experts' opinion on this. We did not add additional questions for game types or genre, as this has no influence on the overall goals, fun or learning, defining why a game is played and was, therefore, already covered by our questions.

The lead author conducted a reflexive inductive thematic analysis similar to Braun and Clarke [21, 22] to analyze the answers based on the interview notes, which were taken during the interviews and completed afterward using the audio recordings. No coding framework is required for this approach, and it can be conducted by one person [22]. Due to repeating answers and as we had gained a fundamental understanding of our experts' opinions regarding the state of "being stuck" and related terms, we stopped recruiting experts after 13 interviews.

3.3 Terms Rating Results

The following two sections show how the experts rated the terms related to the state of "being stuck" and terms related to the detection of it. These terms were used as base for the search query of the SLR.

3.3.1 Terms Related to the State of "Being Stuck". The following terms were mentioned by our experts and added to the list after verifying with a subsequent literature check that they were actually used in related work in the context of a stuck-related state: *Motivation* [90, 126, 142], *Resignation* [42], *Helpless* [66], *Usability* [191], *Proper Instructions* [90, 92, 154], *Confusion* [194], *Surrender* [42], and *Spatial Orientation* [87, 149].

Figure 1a shows the results of the 7-point Likert scale survey (1 = "not relevant" - 7 = "very relevant") of related terms. All terms were rated by all 13 experts, except those that were suggested by them and then only rated by themselves and the upcoming interviewees.

To tailor our search query for the SLR to the most relevant ones and use not more than ten terms, we chose to exclude all that had a rating below the median (= 5, gray bars in Figure 1a). *Usability* was excluded as well because it is related to a very broad field of user interface (UI) research, not always linked to the state of "being stuck" such as UI design standards [108, 173].

3.3.2 Terms Related to Detecting the State of "Being Stuck". As before, the following terms were named by our experts as related terms regarding stuck detection and added to the list after verifying with a subsequent search that they were actually used in related work in the context of measuring a stuck-related state: *Game Analytics* [126], *Evaluation* [98], *Sensing* [100], and *Discovering* [101].

Figure 1b shows the results of the 7-point Likert scale survey from 1 (= "not relevant") to 7 (= "very relevant") of related terms regarding stuck detection. All terms were rated by all 13 experts, except those that were suggested by them and then only rated by themselves and the upcoming interviewees.

As with the terms related to the state of "being stuck", we wanted to limit the list for the search query to the most relevant ones. Because the list included only ten terms, we excluded all that had a rating below the rank of the first quartile (= 4.7) instead of the median. This list is shown in Figure 1b. *Evaluation* was excluded, as it is a term frequently used in empirical contributions and does not automatically imply a measurement.

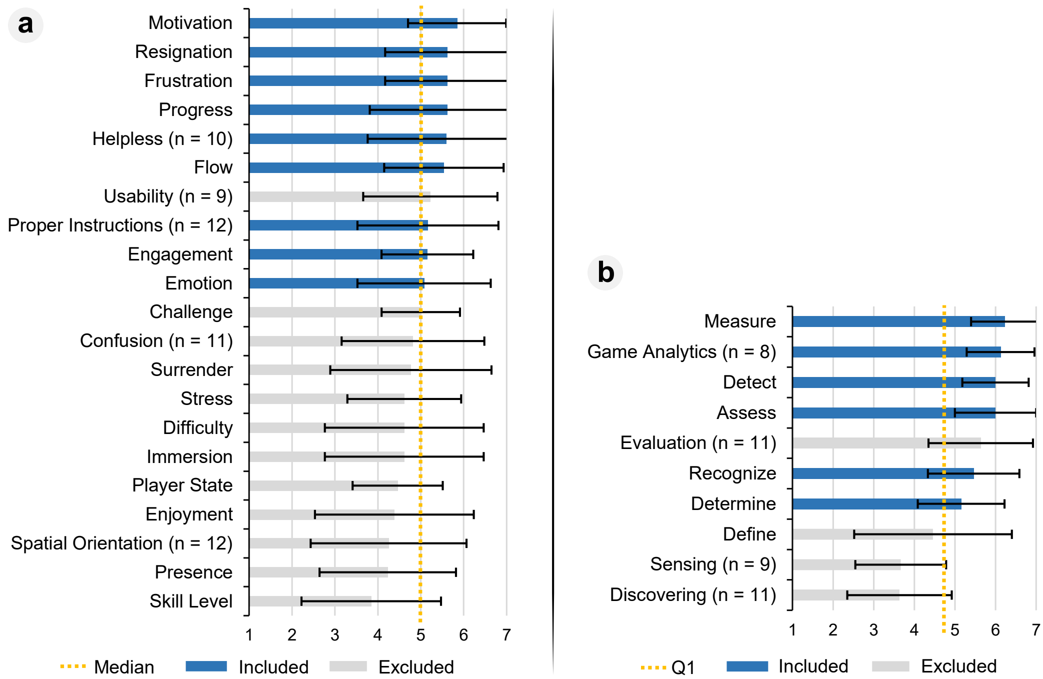

Fig. 1. During the interviews, all 13 experts were asked to rate (a) "being stuck" related and (b) "being stuck" detection related terms on a 7-point Likert scale from 1 (= "not relevant") to 7 (= "very relevant"). Terms that were added based on experts' suggestions between interviews were only rated by themselves and the upcoming interviewees. For those, the number of raters is stated in brackets. Standard deviation is reported as error bars.

3.4 Thematic Analysis Results

To define the themes, we applied the six phases of thematic analysis suggested by Braun and Clarke [21]. We used the flexibility of the thematic analysis to present our four themes as stories, including multiple references to participants' statements for good quality (see Braun and Clarke [21, 22]). The themes are numerically numbered and provide an overview about the *(1) Experts' Understanding of the State of "Being Stuck"*, if our experts think if there are *(2) Similarities and Differences Between Educational and Leisure Games Regarding the "Being Stuck" State*, and how *(3) Detecting the State of "Being Stuck"* is possible. They further show the *(4) Experts' Definition of the State of "Being Stuck"*. These findings are used as additional input for the main synthesis of the SLR, which is why we describe the themes here, but present their synthesis with the findings from the SLR later in the paper. The themes should not be viewed standalone due to the thematic analysis method's interpretive nature, but they were useful for the preparation and synthesis of the SLR.

3.4.1 Theme 1: Experts' Understanding of the State of "Being Stuck". The experts stated that to be stuck is a feeling *(P1, P2, P5, P10)*. This feeling is caused by reasons such as not reaching your goal (*P5*) or when "*... the actual results of my input do not match with my [expected] results.*" (*P6*). There was a common agreement in the answers of *P1, P4, P5, P8, P10, P11, P13* that "being stuck" describes a state in which players are unable to proceed or find a solution for a problem. The problem itself could be logical or skill-based. Possible results are repetitive player actions or a not goal-oriented *"try and error [approach]"* (*P13*). *P7* stated that it is related to time and progress: *"If you are stuck for a certain period of time ... it is difficult to motivate yourself to continue ... and I find it very frustrating."* *P1* mentioned that this is related to the individual patience and perseverance of a player. *P1*: *"I must admit I am a very impatient player. So I try [to solve a problem] only a few times. When I have the chance, I start asking others questions or search the internet for solutions."* Feedback helps *P1* to limit impatience. *P10* also recommends game feedback as otherwise, players do not know if they are doing things right. This results in a drop of attention (*P10*).

Another reason for getting stuck was seen in missing or unclear instructions so that players do not know what to do (*P2*, *P3*, *P4*, *P9*, *P11*, *P12*). *P9*: *"[I feel stuck] when the goals are unclear, i.e. when I do not know what I am supposed to achieve."* *P12* stated that *"[I am stuck] when game controls are not self-explaining"*. Therefore, a tutorial at the beginning of the game is essential (*P12*). Poor usability can be another reason (*P8*, *P9*, *P11*, *P13*). *P8*: *"[Imagine] a situation when players have to solve a task but do not know which navigation options they have in the game and which actions they can perform."* This was also linked to poor level design (*P3*, *P12*). Players who cannot find the right way or get lost cannot proceed either. This means that it is not only possible to get stuck due to logical or skill-based reasons, it is also possible to feel stuck due to disorientation.

3.4.2 Theme 2: Similarities and Differences Between Educational and Leisure Games Regarding the "Being Stuck" State. All ten experts who had experience with educational games were encouraged to consider similarities and differences between educational and leisure games in their replies to our questions. The experts stated that educational games are often more complex and provide more possibilities to get stuck (*P6*, *P12*), like getting overwhelmed by too much content or being too difficult (*P6*, *P7*, *P13*). It should be considered that the learning unit should always match with the player's prior knowledge so that they can successfully learn and complete it (*P13*).

P1, *P3*, and *P7* considered being stuck mostly similar in educational and leisure games. *P8* and *P9* added that the game itself (e.g., genre or task) has more influence on getting stuck than the context being educational or leisure. According to *P13*, stuck in educational games is more likely caused by a lack of knowledge, while lack of skill might be the reason for getting stuck in leisure games. As our experts identified no major differences between educational and leisure games regarding the stuck state, we do not distinguish between them in the following and generalize our findings.

3.4.3 Theme 3: Detecting the State of "Being Stuck". As game researchers, our experts were familiar with the stuck state, but none of them had explicitly conducted research phrased as stuck research. Nevertheless, most of them (*P1*, *P3*, *P4*, *P5*, *P8*, *P9*, *P10*, *P12*, *P13*) stated that game mechanics could be analyzed to check if milestones were reached in relation to a time interval to detect stuck players. Games could track player behavior for repetitive actions (*P4*, *P5*, *P6*, *P7*, *P8*, *P10*, *P11*, *P13*), count the number of failures (*P7*), measure idle time (*P11*), detect disengagement to the task *"[when a player] looks around questioningly"* (*P10*) or the usage of available help functions (*P11*). Physiological data could be used for the detection too (*P1*, *P4*, *P5*, *P9*), such as skin conductance, heart rate, body posture, eye-tracking, or button pressure level and provide information about the users' emotions. Disoriented, and therefore stuck players, could be detected by the player's movement trajectory (*P3*), observing the player's position during a time period, as done by Kepplinger et al. [87] or Schertler et al. [149]. Time can be a stuck indicator in this case, but this has to be treated with caution, as players could just be looting (*P1*).

3.4.4 Theme 4: Experts' Definition of the State of "Being Stuck". We summarized the answers of the experts and define that "being stuck"

(1) means that players are not able to reach their goal (*P1*, *P3*, *P4*, *P8*, *P11*, *P12*),

(2) is accompanied with negative emotional states (*P5*, *P10*), such as confusion (*P2*) and frustration (*P6*, *P7*), as well as mental overload (*P9*),
(3) and is not a binary state and should be described as a continuum (*P13*).

This definition is applicable to leisure games as well as educational games. We do not distinguish between these types in the following as our experts stated no relevant differences between them regarding the stuck state.

3.5 Discussion

Our thematic analysis confirms that the state of "being stuck" is widely considered important but not well-defined and described somewhat differently. Stuck-related aspects were seen to apply to leisure and educational games in the same way, as only minor differences regarding tasks and topics exist, and the reasons to get stuck are mostly the same. Therefore, we did not explicitly distinguish between these two types of games in our SLR, and will discuss previous work regarding leisure and educational games together. Further important insights were that it is a user and game-specific subjective feeling, which requires user and game-specific methods to detect it. This information helped us to get a common understanding of the stuck state that is in line with our experts' understanding.

The analysis confirms the relevance of our RQs, as they clearly address the relevant topics identified by our themes and the need for a SLR to cover this wide field by analyzing publications systematically and providing an overview. The identified themes guided us through the SLR. Especially important was the rating of the related terms by the experts, which we used to define the search query.

4 SYSTEMATIC LITERATURE REVIEW

The expert interviews provided a basic understanding of the state of "being stuck", but to answer our RQ 1, *"What are frequently used terms for "being stuck" in games?"*, RQ 2, *"How can the state of "being stuck" in games be defined?"*, RQ 3, *"How can games detect if players are stuck?"*, and RQ 4, *"How can games help players when stuck?"*, further information was necessary.

Based on the related terms identified by the experts, our goal was to elaborate a detailed and comprehensive overview of works on the feeling of "being stuck" even if they only indirectly address the respective feeling. Therefore, we needed a structured, objective, and replicable but also broad as well as detailed literature search to cover all relevant works. We decided to apply a SLR, as it matched our requirements. It is considered a methodically rigorous, comprehensive, transparent, and replicable method to analyze literature by simultaneously minimizing subjectivity and bias [111, 155]. It further relies on a search query that can be based on the terms identified by our experts.

4.1 Method

The SLR process of this work is based on the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines proposed by Moher et al. [111, 112], as well as the processes presented by Kitchenham and Brereton [94], and Tsafnat et al. [174]. Their guidelines were iteratively merged and tailored to our research topic. Our process is visualized in Figure 2 and consists of an identification step, a two-step paper screening part, and a conclusive synthesis. The following sections present the process steps in detail so that it could be reproduced by other researchers.

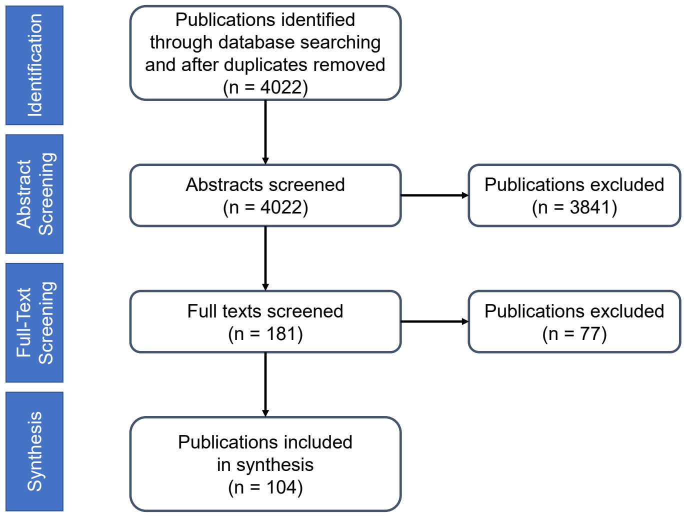

Fig. 2. This flow chart shows our SLR process. It is based on the PRISMA guidelines proposed by Moher et al. [111] and tailored to our needs. It had the purpose to exclude papers so that only the most relevant remained for the synthesis.

4.2 Identification

To identify relevant papers, we created a database query. Following Tsafnat et al. [174], the first step in preparing the database query was the definition of the motivation and the RQs described in our introduction. This step followed the previously described expert interviews, conducted to prepare the SLR. Based on the expert interviews, we created a single search query that combines the identified keywords using Boolean algebra:

(Stuck OR Motivation OR Resignation OR Frustrat* OR Progress OR Helpless* OR Flow OR Instruction OR Engage* OR Emotion*)
AND
(Measure* OR Analy* OR Detect* OR Assess* OR Recogn* OR Determin*)
AND
(Game OR Gaming OR Gamification)

There are three OR blocks connected with AND. The first block ensures that each result at least contained a term of relevance regarding our scope to investigate "being stuck". The second block requires a result to contain some form of measurement or assessment in our defined context. For both these aspects, we used the keywords defined in the interview section (see Figure 1). The last block requires a result to contain game, gaming, or gamification. We added this block to find game-related publications, as our focus included the detection of stuck players. Combinations of terms should be allowed, whereas related keywords are connected with an inclusive OR. Wildcards are used for keywords with a unique root.

We selected the three databases ACM Digital Library [9], IEEE Xplore [78], and Science Direct [52], covering venues such as proceedings from *Conference on Human Factors in Computing Systems (CHI)*, *CHI Play*, *IEEE Transactions on Games*, and *Computers in Human Behavior* and defined three search criteria that were applied directly in the database search: (1) A publication had to be written in English. (2) It had to be a peer-reviewed full paper which excluded extended abstracts and works in progress. (3) Only publication title and abstract were searched similar to Pai et al. [129], Zawacki-Richter et al. [192], and Bargas-Avila et al. [11], as those have to contain an overview of all relevant content of the paper.

The search results were downloaded from the databases with the previously stated query. After removing duplicates with Zotero [37], 4022 papers remained (825 ACM DL (20.5 %), 1,857 IEEE Xplore (46.2 %), 1,340 Science Direct (33.3 %)). To verify that the search results obtained appropriate papers, we checked if they included relevant papers already known to us. This could be verified, as the defined papers for ACM DL [58, 126], IEEE Xplore [157, 186], and Science Direct [92, 147, 154, 183] were found.

4.3 Abstract Screening

Before we started screening the 4022 abstracts, we first defined the inclusion and exclusion criteria that we applied.

4.3.1 Inclusion and Exclusion Criteria. We defined two mandatory inclusion criteria for the abstract screening phase: (1) *"being stuck" related term* and (2) *interactive software application.* If a paper should be included, both mandatory criteria must be *true*. Criteria (1) was *true* when the paper used or described at least one method or concept that we identified as capable to support detecting when a player is stuck. The method or concept had to be used in the context of a game, or a related measurement had to be conducted with some sort of interactive software, which ensures relevance for digital games, to fulfill criteria (2). Further labels were coded but not used in the synthesis. Labels were not set for papers that could be excluded based on their title.

4.3.2 Method. Sysrev [79] was used as an online platform for this screening phase, as it offers tools for collaborative screening of large amounts of publications. The Sysrev project is public[1] to enable future use of the review data in other research projects and to allow researchers to view our work in detail. The abstract screening was conducted by three of the authors.

To create a common understanding of the inclusion and exclusion criteria, the first 20 papers were collaboratively reviewed. A subset of 79 papers was then reviewed by all three to assess agreement. To check inter-rater agreement, p_0 was calculated according to Fleiss [55] (= 93.2 %), which denotes high agreement [30, 54]. The eight conflicts were discussed and resolved, resulting in a refined version of the common understanding. Kappa [35, 36] was not used to assess inter-rater agreement, as according to Feinstein and Cicchetti [30, 54], it is not appropriate for a substantial imbalance in the table's marginal totals. The remaining papers were then rated individually. If a rater was not sure if a paper should be included or excluded, it was flagged for group discussion. This was the case for 32 papers.

4.3.3 Results. In total, 181 papers were included (50 ACM DL (27.6 %), 69 IEEE Xplore (38.1 %), 62 Science Direct (34.3 %)). A spreadsheet was created with one row for each paper, and a code book was defined for the full-text screening, representing our topics of interest for the synthesis. The code book included the contributions according to Wobbrock and Kientz [185], free-text descriptions of contributions, used terms related to the state of "being stuck", measurement and countermeasures of stuck, their impact, the used platform/type of system, the RQs, study details, the education topic (if applicable), and the exclusion criteria if a paper was rated as not relevant after the full-text

[1]https://sysrev.com/u/2468/p/34483 | This public project was used during abstract screening. It provides access to all 4022 screened paper metadata, including the abstract, and shows how we labeled the papers. Statistics can be shown, and filters can be used to show dedicated papers of interest.

screening. We provide the spreadsheet with all 181 included papers as supplementary material so that researchers can access our raw data and use it for future research.

4.4 Full-Text Screening

In preparation of the full-text screening, a rating guideline was defined to get similar results between raters.

4.4.1 Method. The three raters continued in this screening phase. A guideline was created to organize the individual parallel review of the three raters and the code book usage. One initial paper was rated together in the beginning to generate a common understanding. Then, six randomly selected papers (two per database) were rated individually, and inter-rater agreement was qualitatively checked in a discussion similar to Tyack and Mekler [176], and Robinson et al. [140]. This resulted in a refined version of the common understanding and the addition of a platform/type of system column before finalizing the code book. The remaining 174 papers were reviewed individually, whereas 25 of them were marked for a concluding group discussion.

4.4.2 Results. Of the 181 included papers, 104 papers were marked as relevant for the synthesis (23 ACM DL (22.1 %), 43 IEEE Xplore (41.3 %), 38 Science Direct (36.5 %)). 77 papers were excluded after the full-text screening. All exclusion reasons are stated in the supplementary spreadsheet. For example, papers were excluded that are themselves SLRs with no own theoretical contribution regarding stuck-related concepts or used these concepts in situations our experts and we would not consider as a stuck state. Analyzing the publication dates of relevant papers, there is a positive trend for the number of publications that are related to "being stuck" (see Figure 3), showing the increasing relevance of the research domain in recent years. The 104 papers were published in 62 distinct venues. While 45 venues (72.6 %) only appeared once, 17 venues (27.4 %) were represented by multiple publications. The seven most common venues were *Computers & Education* (13 papers), *Computers in Human Behavior* (9 papers), *Entertainment Computing* (5 papers), *Conference on Human Factors in Computing Systems (CHI)* (3 papers), *Frontiers in Education (FIE)* (3 papers), *International Conference on Serious Games and Applications for Health (SeGAH)* (3 papers), and *Conference on Computational Intelligence and Games (CIG)* (3 papers). This distribution confirms that concepts related to "being stuck" are considered mostly in educational and games research.

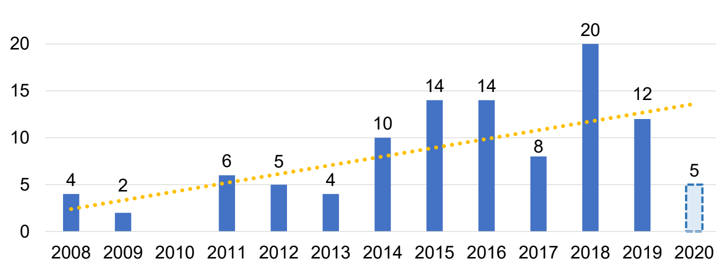

Fig. 3. All 104 publications included in the synthesis are shown here by their publication year. The yellow trend line shows a linear increase, which indicates that our research topic has a growing interest during the last years. As this work was done in 2020, the number for 2020 is not final.

5 RESULTS

In this section, we present the results of the synthesis based on the 104 papers that were selected as relevant to answer our RQs. We start with an overview of the included papers by providing a list of all stuck-related terms and a description of how they were used. As they often measure the used state, we continue by presenting a taxonomy of stuck-related measurement techniques. As some papers developed approaches to prevent the unwanted stuck-related states, we conclude by providing an overview of countermeasures.

5.1 Terms Related to the State of "Being Stuck"

The first part of the synthesis consisted of coding publications regarding the use of terms related to the state of "being stuck". For example, we added a code for *performance* only if there was a clear connection to a related state or concept such as a player's in-game performance, and not just a description of, e.g., a software's technical performance. Another goal of the coding was to discover whether the terms identified in the expert interviews reappeared in the SLR, how frequently they were used, and if there were additional terms not stated by our experts.

In total, we found 75 related terms. While 39 terms (52.0 %) appeared only once, we coded 36 terms (48.0 %) at least twice[2]. We created a list of the 23 most relevant terms and a side-by-side comparison with the expert interview terms (see Figure 4). All terms with a relative frequency greater than 1.0 % were selected. These were 19 in total. We further selected four additional terms for the figure, as they were mentioned as relevant by the experts to allow side-by-side comparison: *immersion*, *confusion*, *presence*, and *spatial orientation*. As we will show in the discussion, we based our "being stuck" state definition on many of these terms and, therefore, show how they were used in the included papers.

5.1.1 Flow Theory. The most frequently used concept with 12.6 % was *flow* (see Figure 4) initially introduced by Csikszentmihalyi in 1975 [38] and extended by him and others in the following years [39, 40, 124]. It was used by educational games to optimize learning. Liu et al. [103] compared a game and a traditional lecture for teaching computational problem solving and showed that flow is higher when using the game, had an influence on the used learning strategy, and increased motivation. Flow is further in line with the players' acceptance of a game, as shown by Hou and Li [73]. Therefore, Kiili et al. [92] created a flow framework to analyze the quality of educational games and their learning experience.

Flow itself can be increased by allowing players to be self-directed, as shown by Hsu [75] and Chen and Sun [29]. Sanjamsai and Phukao [147] stated that flow in games can have the dimensions cognitive flow and emotional flow. Chanel et al. [28] used boredom, engagement, and anxiety as indicators and adjusted the game difficulty accordingly to maintain flow. This is in line with Csikszentmihalyi [38–40] who also identified multiple influencing factors. This shows that flow theory is linked with further concepts found with our SLR.

5.1.2 Challenge and Difficulty. Challenge (7.2 %) and difficulty (7.5 %) are the next two most frequently found concepts. They are often linked and influence the experience of a player (see [4, 6–8, 12, 14, 19, 27, 41, 67, 80, 96, 127, 133, 148]). Hung et al. [76] have shown that an adequate difficulty in educational games results in the best learning performance. This is in line with Sampayo-Vargas et al. [146], who showed that learning is best with a dynamic difficulty adjustment (DDA) game. The DDA approach is similar to the approaches of Chanel et al. [28] previously shown in the flow theory section, who tried to keep the flow state by providing a perfect relation between challenge and emotions such as boredom or anxiety. The link of challenge and emotions is described with the perceived challenge (see Denisova et al. [45, 46]) and used by Hardy et al. [67], who investigated what is perceived as a challenge in games and how this is related to the difficulty level. This shows

[2]See the column ""being stuck" related terms" in the supplementary spreadsheet.

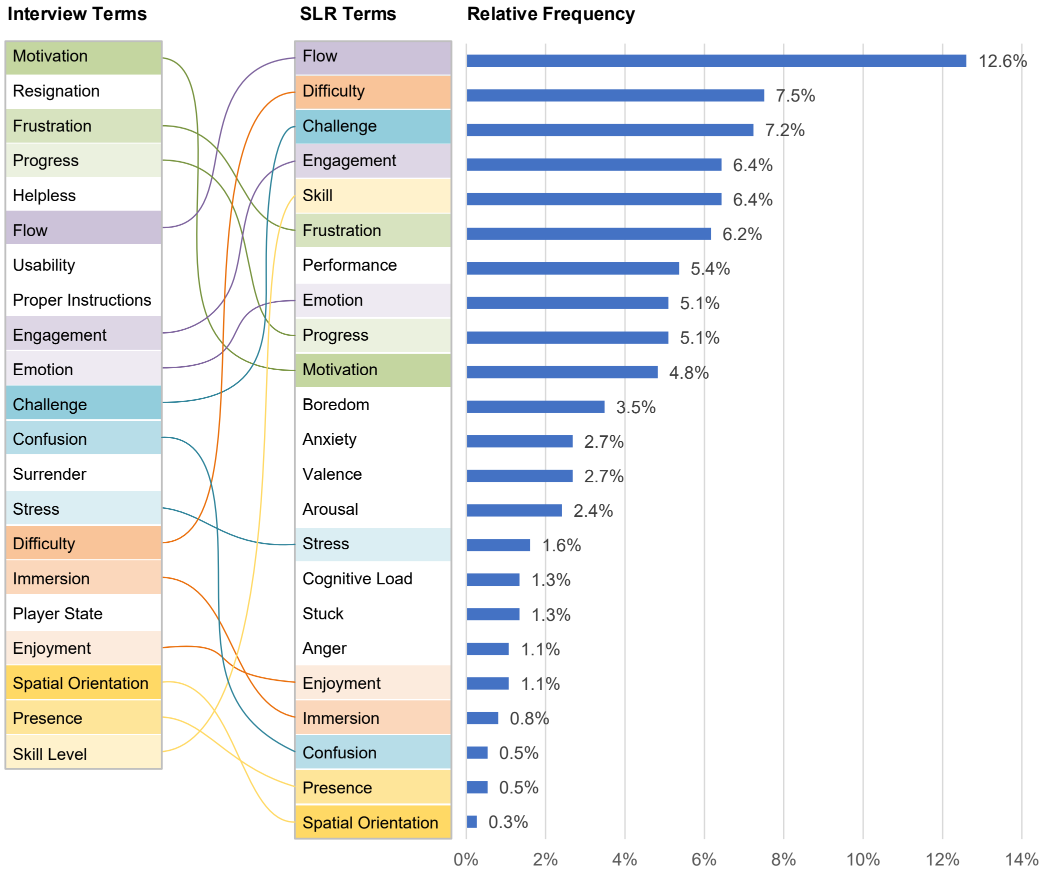

Fig. 4. Comparison of expert interview terms (left) and the most frequently used terms related to the state of "being stuck" in the SLR (right). The relative frequencies of the SLR terms are shown in descending order. The order of the interview terms reflects their relevance stated by the experts also in descending order. Each term from the expert interviews that reoccurred in the SLR is highlighted with a color and connected with its counterpart.

how challenge and flow are in relation to each other, which also can be seen by the definition of Csikszentmihalyi [40].

In the context of challenge and difficulty, we also found the works of Anderson et al. [7], Yang et al. [187], and Sun et al. [161], who used the word stuck when challenge or difficulty was imbalanced but did not define the stuck state.

5.1.3 Affective State. Affective state, which is common in different ways like with specific emotions (5.1 %), or concepts like mood, and core affect [50], was frequently used, with different terms comprising several emotions (see Figure 4): frustration (6.2 %), boredom (3.5 %), anxiety (2.7 %), anger (1.1 %), or enjoyment (1.1 %). Emotion recognition was often used for DDA with automatic as well as self-reported approaches [19, 28]. Frommel et al. [57] used self-reported emotions for DDA and showed that this increases PX compared to constant or increasing difficulty approaches. Frommel et al. [58] further reported automatic approaches considering input and performance features to detect the emotional state. Further automatic approaches were shown by Mostefai et al. [114] by considering the personality type and the playing style as well as game events. As previously stated, Chanel et al. [28] showed that there is a link between emotions, challenge, and flow theory.

5.1.4 Engagement. Engagement (6.4 %; see Figure 4) can be split according to Abbasi et al. [1] into cognitive, affective, and behavioral engagement. Psaltis et al. [139] extracted features of cognitive and behavioral engagement based on players' interaction with the game. Engagement was also used for DDA (see Chanel et al. [28]), which shows that engagement is often used together with other concepts such as affective state [27, 28], challenge [141, 167], or flow [28]. The versatility of engagement with multiple possible measurement techniques can also be seen in Figure 5 that we will describe in the next section.

5.1.5 Progress and Performance. Progress (5.1 %) and performance (5.4 %) can be used to detect the stuck state (see Figure 4) and were used as a measurement to detect the players' success. Tao et al. [172] assessed the learning performance in a business simulation game and set it in correlation with the game's difficulty. Schadenberg et al. [148] investigated how children could be kept motivated by assessing the skill level and performance by using DDA. Similar DDA approaches were investigated by Silva et al. [156] or Orvis et al. [127]. Progress and performance are related to challenge and difficulty (e.g., Hung et al. [76], Tao et al. [172]) and can be used as a measurement technique for several other values (see Figure 5).

In this context, we found the works of Johnson et al. [81] and Tang et al. [170], who used the word stuck to describe players who deviated from their path to their goal.

5.1.6 Cognitive Load. Our experts mentioned mental overload as a possible result of being stuck, which we also found with the SLR (1.3 %; see Figure 4). The basic cognitive load theory was described by Sweller in 1988 [165] and extended by him and other researchers in the following years [34, 166]. The players' mental effort is linked to their individual cognitive load [128, 158].

Yang et al. [187] investigated how cognitive complexity influences learning performance and had the result that an appropriate amount of cognitive load improves the learning results, especially of low-achieving students, but also has a negative interaction with their anxiety. Hsu [75] showed that cognitive load depends on a students' learning style and that it is lower for a serial approach compared to a global one. They further evaluated the effects on flow (see previous flow theory section). These works show that cognitive load is also linked to other concepts such as performance, affective state, and flow.

5.2 Taxonomy of Measures for Stuck-Related States

During our synthesis, we found a variety of techniques for measuring stuck-related states, and we created a taxonomy of useful measures for detecting them. After performing a morphological analysis, we obtained a multidimensional matrix (see Figure 5), also known as Zwicky-Box [195], which is a well-established tool for the creation of design spaces or taxonomies (e.g., [47, 71]). The two axes that span the taxonomy are the dimensions **measurement value (D1)** and **measurement technique (D2)**. Each dimension has several subcategories, which define the rows and columns of the matrix. The combination of row and column subcategories defines a cell that summarizes specific measurement methods used in the analyzed papers.

D1 represents a selection of feeling-based stuck-related measures based on the most frequently used SLR terms shown in Figure 4, which are *anxiety*, *arousal*, *boredom*, *cognitive load*, *engagement*, *flow*, *frustration*, *motivation*, *stress*, and *valence*. **D2** represents measurement techniques, which are grouped similarly as by Schrader et al. [151] into the subcategories *physiological data*, *gameplay data*, and *questionnaire*. As part of the synthesis, we analyzed 104 papers to determine which techniques were used for measuring stuck-related concepts[3]. We only included papers in the taxonomy that applied measurement techniques and no summaries of different methods such as Foutsitzi et al. [56].

5.2.1 Physiological Data. Physiological data are used in automated data collection approaches based on the human body, which often require dedicated sensors (e.g., cameras [139, 190] or electrodes [51, 74]). Following Bian et al. [15], we divide physiological measures into *first-level indicators* consisting of *second-level indicators* that measure different states of the same body function and use their classification. For example, the two *second-level indicators*, *respiratory rate* [15] and *chest cavity expansion* [27] measure human respiration and are, therefore, part of the *first-level indicator respiratory activity*. In our synthesis, we found many *second-level indicators* that only appeared once. Therefore, we categorized them into the *first-level indicators*, *cardiovascular activity (CA)* and *electrodermal activity (EDA)*, to provide a more compact overview. *CA* includes, e.g., measures for *heart rate* [14], *heart interbeat interval* [68], and *blood volume pulse* [27], and has been mostly used to detect *arousal* [96], *emotions* (*boredom* [157], *frustration* [97]), and to assess player *engagement* [27]. *EDA* includes, e.g., measures for *skin conductance* [116, 186] and was used by Katmada et al. to detect *arousal* and *anxiety* [85], and *engagement* [86]. Other *first-level indicators* are *cortisol concentration* [189], *electroencephalography (EEG)* [74], *electromyography (EMG)* [15], and *skin temperature* [86]. In addition, researchers used body movement-related measures, such as *facial expressions* [159], *eye-tracking data* [97], and *head posture* [19]. We found that most physiological measurement techniques can provide information of multiple stuck-related states but were seldom or never used for *cognitive load* or *motivation* in our papers sample.

5.2.2 Gameplay Data. Gameplay data measurement techniques are automated online evaluation approaches embedded into the game or employ offline evaluations via data collections. The goal is to infer players' state by evaluating information of what the player does in-game [151]. We identified two dominant groups according to their characteristics, which allowed us to cluster many of the measurement techniques. The first group contains *player performance* measures (e.g., [7, 69, 141]), which measure the current progress of the player within the game, such as completed levels [7], gameplay time [88], or number of points [120]. Such data can then be used to derive the players' state using online or offline methods. For example, Frommel et al. [58] applied a machine learning approach using player performance (in combination with input features) to predict affect. While the majority of measures from the *player performance* group were applied online during gameplay, Fu et al. [60] and O'Rourke et al. [126] used *activity logs* to log all player data to analyze it offline. The second group contains *player behavior* measures (e.g., [89, 161, 167]), which describe the behavior of the player during the execution of a task [167] and provides, for example, information about player movement [181]. In addition to these two groups, we found other gameplay measures worth mentioning individually, such as *controller sensors* [58, 170, 181], *keystroke analysis* [157, 170], and *touch pressure* [58, 110].

5.2.3 Questionnaires. Questionnaires examine the player's subjective perception of stuck-related states. We found at least one questionnaire for each measurement value that was *self-developed* by the respective authors and therefore had no name (e.g., [14, 18, 180]). Established questionnaires were also frequently used, which contain subscales for assessing e.g., *flow* or *emotions* such as *frustration*, *boredom*, and *engagement*. Some of these questionnaires are, e.g., *Game Engagement Questionnaire (GengQ)* [69], *Game Experience Questionnaire (GexpQ)* [113], *Immersive Experience Questionnaire (IEQ)* [114], *Self-Assessment Manikin (SAM)* [148], *Simulator Sickness Questionnaire (SSQ)* [113], *4-Alternative-Forced-Choice (4-AFC)* [181], *Game Flow Inventory (GFI)* [157], *Modified Differential Emotions Scale (M-DES)* [96], *Visual Analogue Scale (VAS)* [184], *Engagement Sample*

[3]See the column "Stuck measurement concept" in the supplementary spreadsheet.

D1: Measurement Value	D2: Measurement Technique: Physiological Data	D2: Measurement Technique: Gameplay Data	D2: Measurement Technique: Questionnaire
Anxiety	**CA** [4, 27, 85, 86, 157], **EDA** [4, 27, 85, 86], **EEG** [4], **Facial Expressions** [169], **Respiratory Activity** [4, 27], **Skin Temperature** [27, 85, 86]	**Keystroke Analysis** [157]	**Flow State Scale** [29], **Language Anxiety Scale** [75, 187], **Learning Experience Survey** [103], *Self-Developed* [18, 72]
Arousal	**CA** [15, 96], **EDA** [27, 51, 85, 96, 186], **EEG** [15, 74, 113], **EMG** [15], **Facial Expressions** [26], **Respiratory Activity** [15]	**Controller Sensors** [58], **Player Performance** [58], **Touch Pressure** [58, 110]	**Affect Grid** [68], **GexpQ** [113], **IEQ** [113], **SAM** [27, 58, 148], **SSQ** [113], *Self-Developed* [1, 27, 51, 74]
Boredom	**CA** [4, 14, 27, 86, 97, 157], **EDA** [4, 27, 86, 157], **EEG** [4, 97, 157], **Eye Tracking Data** [97], **Facial Expressions** [14, 26, 169], **Respiratory Activity** [4, 27], **Skin Temperature** [27, 86]	**Controller Sensors** [181], **Keystroke Analysis** [157], **Player Behaviour** [181], **Player Performance** [57]	**4-AFC** [181], **Dialogue-Based Self-Reports** [57], **Flow State Scale** [29], **GFI** [157], **Learning Experience Survey** [103], **M-DES** [96], **VAS** [184], *Self-Developed* [14, 18, 97]
Cognitive Load	**EEG** [74]	**Working Memory Task** [189]	**Cognitive Load Scale** [75, 187], **Subjective Mental Effort Questionnaire** [67], *Self-Developed* [74]
Engagement	**Body Movement** [139], **CA** [27, 86, 97], **EDA** [27, 86], **EEG** [97, 113], **Eye Tracking Data** [97], **Facial Expressions** [19, 139], **Head Posture** [19], **Respiratory Activity** [27], **Skin Temperature** [27, 86]	**Player Behaviour** [167], **Player Performance** [7, 60, 69, 88, 99, 120, 126, 134, 139, 141]	**Consumer Video Game Engagement** [1], **Engagement Survey** [146], **ESQ** [150], **GengQ** [69, 88, 139], **GexpQ** [113], **IEQ** [113, 114], **SSQ** [113], **VAS** [184], *Self-Developed* [32, 70, 97, 190]
Flow	**CA** [4, 15, 68], **Respiratory Activity** [4, 15, 68], **EDA** [4, 157], **EEG** [4, 15, 157], **EMG** [15], **Facial Expressions** [26, 169]	**Keystroke Analysis** [157], **Player Performance** [106, 193]	**EGameFlow** [153], **Flow Scale** [73, 75, 92], **FSS-2** [29, 68], **Flow Short-Scale** [15, 53, 193], **GengQ** [4, 139], **GFI** [157], **Learning Experience Survey** [103], *Self-Developed* [18, 72, 147, 180]
Frustration	**CA** [97], **EDA** [116, 186], **EEG** [74, 97], **Eye Tracking Data** [97], **Facial Expressions** [14, 26, 116, 159], **Speech** [159]	**Controller Sensors** [170, 181], **Keystroke Analysis** [170], **Player Behaviour** [161, 181], **Player Performance** [57], **Touch Pressure** [110]	**4-AFC** [19, 181], **Dialogue-Based Self-Reports** [57], *Self-Developed* [24, 97, 105, 110, 159]
Motivation		**Player Behaviour** [89], **Player Performance** [69, 167]	**AMS** [179], **Engagement Survey** [146], **QCM** [67], **Free-Choice Method** [148], **IMI** [10, 69, 119], **Learning Motivation Measure** [187], **MSLQ** [103], *Self-Developed* [18, 105, 127, 141, 172]
Stress	**CA** [14, 186], **Cortisol Concentration** [189], **EDA** [186], **EEG** [74], **Facial Expressions** [14], **Skin Temperature** [190]		**Rating Scale Mental Effort** [186], *Self-Developed* [14, 74]
Valence	**CA** [15], **EEG** [15, 113, 74], **EMG** [15], **Facial Expressions** [26], **Respiratory Activity** [15]	**Controller Sensors** [58], **Player Performance** [58], **Touch Pressure** [58]	**Affect Grid** [68], **GexpQ** [113], **I3E** [189], **IEQ** [113], **SAM** [27, 58], **SSQ** [113], *Self-Developed* [27, 51, 74]

Fig. 5. A taxonomy of measures used for the detection of stuck-related states obtained by a morphological analysis. They are categorized into the taxonomy's cells by measurement value (D1) and the used measurement technique (D2). Abbreviations are explained in the text.

Questionnaire (ESQ) [150], *Flow State Scale 2 (FSS-2)* [68], *Academic Motivation Scale (AMS)* [179], *Questionnaire on Current Motivation (QCM)* [67], *Intrinsic Motivation Inventory (IMI)* [10], *Motivated Strategies for Learning Questionnaire (MSLQ)* [103], and *Inventory of Three-dimensional Emotions (I3E)* [189].

5.3 Countermeasures for Stuck-Related States

In our synthesis, we identified 59 papers (56.7 %) that contained countermeasure techniques to prevent or exit stuck-related states. 34 papers contained an evaluation of the countermeasure effectiveness, from which 25 papers reported a positive effect in the form of increased player performance and/or PX, while eight papers reported at least partially or neutral countermeasure effects[4]. We grouped all found countermeasure techniques into three types, according to their aim: adjusting the difficulty of the game content (56.3 %), supporting the players' problem-solving process (38.0 %), and positively regulating the players' behavior (5.6 %) (see Figure 6).

			Results			
			No Evaluation	**Positive Results**	**Neutral/Mixed Results**	**Negative Results**
Type	**Game Difficulty Adjustment**	Dynamic Difficulty Adjustment	[12, 13, 26, 28, 56, 58, 80, 85, 86, 97, 102, 115, 116, 130, 144, 168, 169]	[4, 19, 20, 57, 69, 107, 114, 119, 127, 133, 156, 175, 190]	[32, 88, 125, 134, 146, 148]	
		Self-adjustment		[119, 152]	[24, 75]	
	Problem-Solving Support	Performance Feedback	[49, 56, 58, 63, 65, 73, 193]	[8, 29, 70, 120, 177]	[148]	
		Hints	[56, 63, 81, 85, 171]	[76, 89, 99, 161, 170]		[126]
		Supporting Tools	[84]	[177, 187]		
	Positive Behavior Regulation		[63]	[8, 29, 161]		

Fig. 6. An overview of the countermeasures to prevent or exit stuck-related states that we found in our synthesis. They are categorized into the cells according to their type and the reported results.

5.3.1 *Game Difficulty Adjustment.*

Game difficulty adjustment techniques are based on the principle that it is easier to reach a goal in a game when the difficulty is appropriate for the players. This prevents them from perceiving negative stuck-related states by altering in-game elements. 65.2 % of the papers that used and evaluated these countermeasures could show a positive effect on player performance and/or enjoyment.

With 36 papers, the most common adjustment technique in our SLR is DDA, which automatically changes gameplay parameters. Frommel et al. [57] showed that DDA based on players' self-reported emotions boredom and frustration led to a decreased number of fails and a level of challenge that was perceived as more suitable in comparison to constantly and increasing difficulty in a platformer online game. Similar performance increases, as well as motivational ones, were stated by Orvis et al. [127]. But DDA does not always mean to decrease the difficulty. Masanobu et al. [107] showed that a difficulty increase when players feel pressure followed by a decrease just exactly before the point when a player actually fails can be enjoyable.

Instead of automatically adapting the game difficulty, another approach is to give the control to the player. Nagle et al. [119] used this concept of self-adjustment in a memory training game

[4]See the columns "Stuck countermeasures", "Countermeasure effectiveness measured", and "Results/effectiveness of countermeasure" in the supplementary spreadsheet.

and compared it with a performance-based DDA approach. Player performance was higher in the DDA variant, but PX was rated higher in the self-adjustment variant. A positive effect of self-adjustment was also reported by Schrader and Nett [152], leading to a higher perception of control and enjoyment as well as a lower perception of frustration and anger. Self-adjustment is further applicable to educational games. Ting-Chia [75] let participants select their vocabulary learning sequence, which resulted in higher flow but also higher cognitive load. This trade-off between enjoyment and performance has to be considered when choosing the appropriate difficulty adjustment technique.

5.3.2 Problem-Solving Support. Problem-solving process support techniques facilitate players to make progress and achieve goals by supporting them while in-game content and difficulty stay the same. 85.7 % of the papers that used and evaluated these countermeasures could show a positive effect on player performance and/or enjoyment.

13 of the papers that used a problem-solving process support technique offered some kind of performance feedback to the player. This helps players to assess themselves by presenting game scores, wrong answer correction, or gameplay progress to them [63]. Chen and Sun [29] and Hicks et al. [70] showed that performance feedback can have a positive effect on player performance and flow.

Giving hints in the form of short instructions or clues according to the in-game progress is more advanced, and eleven papers used them to support the player. The most were textual, as shown by Kickmeier-Rust et al. [89], who gave guidance and support in a geographical knowledge game leading to better player performance. Lee et al. [99] came to the same result by displaying textual suggestions in a game to learn programming concepts. Hung et al. [76] used different visual hints in a puzzle game, which improved learner involvement and decreased anxiety. Sun et al. [161] placed visual hints in a Sudoku game, which resulted in a higher sense of challenge, control, and desire for achievement. In contrast, O'Rourke et al. [126] are the only ones who demonstrated negative effects of hints. They compared four different textual hint systems in a browser game and noticed worse player performance compared to a version without hints, but their hints did not always address the players' current confusion due to a missing adaption to the players.

Besides offering feedback and hints, supporting tools can be used to decrease the cognitive load of players. Kang et al. [84] provided statistics of the use of ten different tools to share cognitive load, support cognitive processes, support cognitive activities, and support hypothesis generation in a serious game. As shown by Yang et al. [187], balancing cognitive load can lead to better learning performance.

5.3.3 Positive Behavior Regulation. Positive behavior regulation techniques have the goal to positively influence the behavior and motivation of the player while the game content stays the same and no additional support is provided. The idea is to achieve better performance by raising the players' motivation to search for solutions and thereby increase the probability of achieving personal goals. Techniques from this countermeasure type are quite rare, and only four papers used techniques from this type. While Sun et al. [161] and Gavriushenko et al. [63] presented the concept of a reward system, Apostolakis et al. [8] made use of positive reinforcement, and Chen and Sun [29] relied on self-regulation of individuals by promoting players to constantly evaluate and analyze their ongoing situations, in order to make better investments in activities. Three concepts (see Figure 6) led to positive effects on player performance or PX in combination with other countermeasure types but were not evaluated independently.

6 DISCUSSION

In this section, we will discuss the previously presented results of our SLR and combine them with the results of our expert interviews to answer our RQs. The first section will answer RQ 1, *"What are frequently used terms for "being stuck" in games?"*, by discussing the found related terms. The subsequent three sections will answer RQ 2, *"How can the state of "being stuck" in games be defined?"*, by first discussing major stuck-related concepts in preparation for presenting our stuck state definition and finally describing this definition with a continuum. The following section will answer RQ 3, *"How can games detect if players are stuck?"*, by discussing the taxonomy of the found stuck measurements, before the last section will answer RQ 4, *"How can games help players when stuck?"*, by discussing the found stuck countermeasures.

6.1 Terms Related to the State of "Being Stuck"

Since 15 of our expert terms have reappeared in the SLR, we conclude that the experts' opinions are generally consistent with existing research (see Figure 4). Some terms (e.g., *skill* and *difficulty*) were not rated highly relevant by the experts but appeared more frequently in the SLR. In contrast, more relevant rated terms did not appear in the SLR as often as expected (e.g., *motivation* and *confusion*). Six interview terms the experts considered relevant did not reappear in the SLR. These findings indicate that existing research does not cover the whole state of "being stuck". This is in line with the result that the term stuck itself was only found five times. Since the expert interviews showed that "being stuck" is a relevant state, the reason could be that publications avoid using this term, as it is not yet properly defined. We assume that publications, therefore, often rely on well-defined approaches to describe stuck-related states (e.g., *flow*, *difficulty*, and *challenge*; see Figure 4). This underlines the relevance of our investigation of the state of "being stuck" in this work so that researchers in the future can rely on our definition when using this term and therefore use it more often.

We conclude that Figure 4 provides a hierarchical list of frequently used terms related to the state of "being stuck" that will help researchers to find related works more easily. This list answers our RQ 1, *"What are frequently used terms for "being stuck" in games?"*, and is furthermore a basis for our following "being stuck" state definition.

6.2 Major Stuck-Related Concepts

Figure 4 shows that we found several terms related to the state of "being stuck". As we have shown in our results section most of them are related among themselves. We will now discuss how they are related to the state of "being stuck" and why they are not similar.

6.2.1 *Flow Theory.*

Flow is a concept that is different from "being stuck" but closely related because it describes a state between boredom and anxiety and is related to skill and challenge [38]. Further characteristics stated by our experts match with those shown by Csikszentmihalyi [39]. These are the chance to reach a clear goal, having positive emotions like enjoyment, and to define it as a continuum.

In a flow state, the player is unlikely to be stuck. Therefore, the flow level can be used as an inverse measure to detect the stuck state. Whenever players leave perfect flow, this means that they possibly could have become stuck. Nevertheless, flow is not the opposite of stuck. Inverting the flow theory and calling it stuck theory will not work. There are situations when players are not in flow but are not stuck either. Repetitive tasks could lead to boredom, which will end the flow state without players ever getting stuck. There are also situations when players are in flow but stuck. Players could experience flow and high levels of enjoyment when they are solving side quests in role-playing games (RPGs). However, they are maybe only engaging with these side quests because they are stuck on the main story without even noticing it. This may be only a minor issue for leisure games, but considering educational games, it is important that players who are stuck with their main progress are directed back to their main topic of interest. This shows that flow theory can be used as an indicator for stuck detection, but it cannot completely describe the state of "being stuck".

6.2.2 Challenge and Difficulty. The concept of perceived challenge [45, 46], i.e., challenge conceptualized as a player experience state, likely is associated with the feeling of "being stuck", in that a player who feels stuck probably perceives a challenge that is too high. However, it is not so easy to just conceptualize the experience of "being stuck" as another form of perceived challenge or difficulty. For instance, a player might feel stuck when not knowing what to do next without engaging in, and as a result perceiving, any challenge at all. Nevertheless, the association with the feeling of "being stuck" makes it possible to use challenge and difficulty as indicators for the stuck state.

6.2.3 Affective State. According to the expert interviews, negative emotions such as frustration are directly correlated with the feeling of "being stuck", whereas positive emotions such as enjoyment are inversely correlated. Similar to flow, emotions can be indicators for stuck but cannot predict it in every situation. An example is anxiety that can be high when players face new challenges. This can be caused by actual consequences when not fulfilling a task but also just by the fear that their own skills are not sufficient but never becoming stuck.

6.2.4 Engagement. Affective engagement describes a complex emotional state and is further linked to flow [20]. Flow and engagement can perfectly correlate, as can be seen by the flow associated questions of the Game Engagement Questionnaire (GengQ) [25]. Engagement is, therefore, an inverse measure for detecting stuck. Emotions can also influence cognitive and behavioral engagement [131], which shows that there are interdependencies.

Engagement is a perfect example why we propose to use multiple measures for describing a player's stuck state. Concepts described in literature often use multiple basic aspects like several emotions or define conditions and group them together to new concepts like flow or engagement. Although these concepts overlap, their basic principles and characteristics are different. In a similar way, multiple aspects can and should be considered to describe the stuck state in a reliable way.

6.2.5 Progress and Performance. Whenever progress is not possible for a defined period of time, this results in low performance, when defined as progress per time interval. In such situations, it is, according to our experts, very likely that players feel stuck. However, as a feeling is subjective, the lack of progress and performance does not necessarily instantly cause the stuck state. It will take a different subjective amount of time until different players actually feel it. Even with bad performance or low progress, some players might not feel stuck immediately. This has to be considered when using progress or performance as a measure, and maybe another measure such as flow or affective state should be used as well.

6.2.6 Cognitive Load. Cognitive load theory is related to the capacity of the working memory [34]. When a high cognitive load is needed to master a task, further cognitive capabilities are not available for other processes such as learning [165] and reaching one's goal. Due to this correlation, it can be used as a measure for the stuck state. An example is a player who is overextended by the game controls, resulting in not being able to learn the content presented in the virtual environment. Negative effects of high cognitive load are not only limited to educational settings. In leisure games, it could be possible that players are not able to solve a puzzle or perform a complex combination of inputs, which results in them entering the stuck state.

6.3 Stuck Definition

Considering RQ 2, *"How can the state of "being stuck" in games be defined?"*, our findings show that none of the related and already existing concepts is able to fully describe the state of "being stuck". Based on the expert interviews, we present a preliminary definition involving that (1) it means that players cannot reach their goal, (2) it is associated with negative emotions and mental overload and therefore subjective, and (3) is a non-binary continuous state. This definition can be grounded and specified by the data of the SLR.

The SLR showed that in 88.5 % of the synthesized 104 papers, participants had to solve a task to reach a goal such as achieving a high score in Tetris [14, 27, 28, 68, 97, 107, 157], solving an escape-room game [73, 171, 189], or learning a language [58, 72, 75, 146, 159, 175, 187][5]. This highlights the relevance and relationship of a task with a goal in every research setup that did measure stuck-related concepts. Considering these results and the statements of our experts, we conclude that this relationship also applies to the "being stuck" state.

Figure 5 shows how the synthesized papers measured concepts related to the state of "being stuck". All measurement techniques and measurement values target emotions, cognitive load, or other highly individual or personal, and therefore subjective, states. This is in line with our experts, and we conclude that this also applies to the "being stuck" state.

All discussed concepts have in common that they describe states e.g., flow or cognitive load, as continuous. This is in line with our synthesis, which shows that 82.7% of the 104 papers defined their stuck-related states as non-binary, which means that these states could increase and decrease their value on a continuum[6]. Even those papers that used binary states often only applied thresholds to continuous concepts (e.g., see Frommel et al. [58] or Bitrián et al. [18]). Considering these results and the statements of our experts, we conclude that a continuous definition should also be used for the "being stuck" state, as this allows a dynamically changing stuck level representing the degree to which a player feels stuck.

Based on these findings, we define "being stuck" as follows:

Being stuck is a continuous subjective state that occurs when a player cannot reach the personal goal.

To describe this definition further and make it applicable for researchers and game developers, we created a stuck continuum based on this definition (see Figure 7).

6.4 Stuck Continuum

Our definition states that stuck is a continuous state and as such defined by a continuum as shown in Figure 7. The personal stuck level could have any value inside the *stuck continuum*, representing players' subjective feelings of how much they feel stuck. The stuck level can increase or decrease, depending on specific causes, while measurements can determine its value. We will refer to causes that lead to an increase of the personal stuck level as *triggers* and to approaches that reduce it as *countermeasures*. We use the term *indicators* for measures that are useful to determine the stuck level and detect changes. We visualize this relationship between *triggers*, *indicators*, and *countermeasures* as a functional chain in our continuum (see Figure 7).

An *indicator* can be directly measured (e.g., stress level [189]), and it can also be part of a broader concept (e.g., flow with anxiety and boredom [4, 29], or affective state with arousal and valence [68]). *Indicators* measure a subjective feeling of players that is directly linked to the feeling of "being stuck"

[5]See the column "Participants' goal + task" in the supplementary spreadsheet.
[6]See the column "Are the "being stuck" related terms defined as binary state (stuck or not)" in the supplementary spreadsheet.

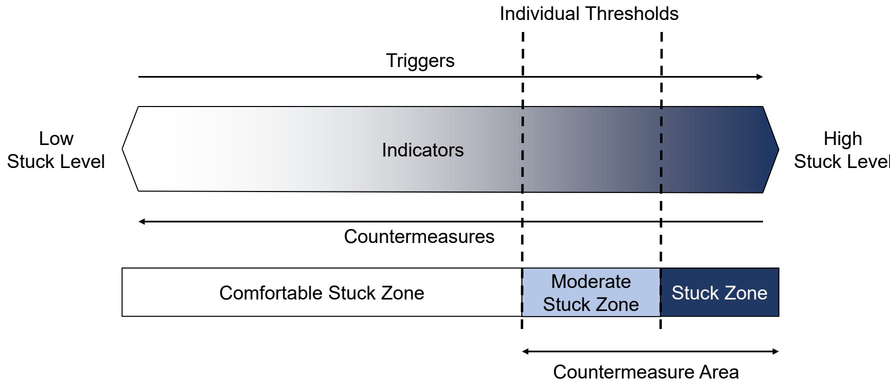

Fig. 7. Stuck Continuum: This continuum represents the player's subjective stuck level, which could have any value between the maximum and the minimum and has no scale to be generalizable. It shows the triggers, indicators, and countermeasures functional chain to manipulate and measure the player's personal stuck level. Two thresholds are defined to create stuck level zones and an area to apply countermeasures whenever the stuck level leaves the comfortable zone.

and their stuck level. As previously discussed, several of our explained stuck-related concepts can be used as an *indicator*, but a single arbitrary *indicator* is no exclusive or absolute stuck measurement, as their concepts are related but not equal to the state of "being stuck". Therefore, the measurement accuracy to determine the player's stuck level increases with the number of *indicators* that can be validated and interpreted. Further possible *indicators* are shown in our stuck measurement taxonomy (see Figure 5) introduced in the previous results section.

Triggers are the reason why a stuck level increases, such as low performance [127, 146] or too high challenge [6, 178], which are mostly present for a longer period of time. *Triggers* and *indicators* are both quantifiable. An occurrence of a *trigger* does not necessarily instantly affect the stuck level as they represent an objective measure and no subjective state. They only affect the stuck level when they are perceived by the player and therefore influence *indicators*, e.g., when they increase frustration. Therefore, *triggers* are only indirectly correlated to the stuck level. Multiple *triggers* may influence the stuck level at the same time, while a single *trigger* may influence multiple *indicators* simultaneously.

Countermeasures are approaches to decrease the stuck level. They try to support the player by modifying game difficulty [57, 107, 119, 127] or providing hints [76, 89, 99]. Several examples and their effectiveness were shown in the previous results section (see Figure 6). Similar to *triggers*, *countermeasures* have only an indirect correlation to the stuck level, as they represent no subjective states either, but may positively manipulate stuck *indicators* and therefore may decrease a player's stuck level. As with *triggers*, multiple *countermeasures* may influence the stuck level at the same time, while a single *countermeasure* may influence multiple *indicators* simultaneously.

Measures exist that can be interpreted as *triggers*, *countermeasures*, or *indicators*. Challenge, for example, can be objectively adjusted, like the selected difficulty, which is a characteristic of *triggers* if increased or of *countermeasures* when decreased. But challenge also has a subjective interpretation, which can be referred to as *perceived challenge* (see Denisova et al. [45, 46]). In this case, challenge is related to a player's feeling, which is a characteristic of *indicators*. We will clarify this in this work whenever we use an ambiguous type of measure and advice to consider this when using our continuum.

Our stuck definition states that a personal stuck threshold exists, which divides the continuum into an area of an acceptable and unacceptable stuck state. According to our definition, this means that the stuck level is acceptable for players until they have the subjective feeling that the personal goal cannot be reached. When this occurs, meaning that the personal stuck threshold is exceeded, players enter the unacceptable stuck state. Since *countermeasures* should be taken before the player exceeds this unacceptable stuck threshold, we have integrated a second moderate stuck threshold, which spans the range from which *countermeasures* should be taken. These two individual thresholds span three distinct zones, which are the individual *comfortable stuck zone*, *moderate stuck zone*, and *stuck zone*, which can be used to define concrete *countermeasures*. Whereas no actions are necessary if players are within the *comfortable stuck zone*, *countermeasures* should be applied when the *moderate stuck zone* is reached, which in the best case prevent players from ever reaching the unacceptable stuck threshold and getting *stuck*. Nevertheless, *countermeasures* should not stop when a player is stuck, which is why the *countermeasure area* covers both zones.

We summarize that we created a stuck continuum based on our definition of the state of "being stuck", which helps to actively detect and prevent this unintended state in the future. We propose that games should implement our stuck continuum in their game mechanics to improve PX and prevent players from churning.

6.5 Taxonomy of Measures for Stuck-Related States

Figure 5 shows that the total number of *gameplay data measures* was lower than the number of *physiological data measures* and *questionnaire measures*. This highlights that researchers determined indicators reflective of stuck-related states mostly via game-independent measurements, which shows promising potential for future work. Only gameplay data-based measures are capable of online stuck state detection compared to interrupting or delayed questionnaires (e.g., [57, 58, 67, 68, 96, 113, 157]) and require no external sensors as most physiological measures (e.g., [51, 74, 139, 190]). New emerging hardware such as virtual reality (VR) offers further possibilities and built-in sensors for measurement as proposed by Drey et al. [48], and learning analytics can detect patterns as proposed by Müller and Fritz [117], which shows potential for future work. As data-based gameplay measures work with standard hardware and standard games, they should be investigated more in the future to enable online stuck detection for the mass market.

Our taxonomy does not claim completeness, as it is based on our SLR's papers set. Therefore, cells in the matrix may be empty (e.g., measuring *motivation* with *physiological data*), but this does not mean that no measurement technique exists. It just means that no papers in our synthesis used such measures. Nevertheless, we argue that the population of our taxonomy is representative, as our synthesis paper set was gained through a SLR following the PRISMA guidelines [111, 112]. With our taxonomy, we want to support researchers and game developers in selecting appropriate indicators for our stuck definition and continuum by providing a structured overview of measurement techniques and answer therewith RQ 3, *"How can games detect if players are stuck?"*.

6.6 Countermeasures for Stuck-Related States

Our countermeasures overview shows that *game difficulty adjustment*, *problem-solving support*, and *positive behavior regulation* can all positively influence e.g., challenge, performance, motivation, enjoyment, flow or anxiety and, therefore, the stuck level according to our definition and continuum. The results of Ting-Chia [75] for self-adjustment difficulty and O'Rourke et al. [126] for supportive hints both endorse our stuck definition in being a subjective individual state, which also requires individualized countermeasures. Ting-Chia [75] further states that there is a trade-off between enjoyment and performance regarding the appropriate difficulty. This highlights that the multiple possible triggers and indicators of our continuum can sometimes be contradictory. This is not an issue, on the contrary, it supports our stuck continuum in including multiple triggers and indicators to describe the perceived stuck level of a player.

Our definition and the continuum define stuck as an undesirable state with multiple negative effects on the PX. Countermeasures to reduce the stuck level are equally important as triggers or indicators (see Figure 7). Therefore, it is problematic that only a few papers investigated countermeasures and evaluated their effectiveness. This offers opportunities for future research, and we encourage researchers to close this gap.

While this summary of countermeasures is not exhaustive, we provide a representative overview (see Figure 6), as it is based on a SLR. This overview should help researchers and game developers in selecting appropriate countermeasures for our stuck definition and continuum and answers therewith RQ 4, *"How can games help players when stuck?"*.

7 LIMITATIONS

While there was scarce research explicitly investigating "being stuck" in games, we had assumptions about what "being stuck" in games could mean, which influenced how we defined our RQs. To avoid selection of existing research biased by our assumptions, we widened our expectations through expert interviews and conducted a SLR based on the PRISMA guidelines [111, 112]. This way, we might have missed concepts that are potentially relevant for the state of "being stuck", but we achieved a more objective and holistic definition.

We limited the terms of the SLR database search query to a subset of those mentioned in the expert interviews. We did this to keep the SLR focused and the resulting set of papers manageable. Even without explicitly searching for some terms, our results show that relevant concepts were part of the results, despite not being part of the query (see Figure 4). This can be explained by the fact that 79.8 % of our 104 analyzed papers used more than one related term, which means that a paper found by one term delivered further terms that were not part of the query. Our taxonomy of stuck-related measures (see Figure 5) emphasizes this as well, as it is quite uniformly filled, which implies that our 104 paper SLR set is covering a wide variety of important research.

Our findings are limited by our search context (particularly considering games), with the results potentially excluding research in non-game contexts. We assume our findings might be useful for understanding "being stuck" in non-game contexts as well, but this requires further research, e.g., considering that lack of progress might be more detrimental in less playful contexts.

As previously stated, the results of the SLR show that there is scarce existing research that explicitly investigates "being stuck" (see Figure 4). Therefore, we used related concepts, such as flow, challenge, affective state, or cognitive load, to triangulate what constitutes the nature of "being stuck" in games. As a result, our definition of "being stuck" is affected by those states and does overlap their definitions in some regard. Nevertheless, "being stuck" is not fully covered by any existing concept. This suggests the value of our findings, while future research is necessary to improve the theoretical separation to other concepts. It is further necessary to validate our definition as well as the continuum with quantified measures including questionnaires. This will also show if our definition can cover all possible types of "being stuck".

8 FUTURE WORK

We presented the continuum of "being stuck" (see Figure 7) based on findings from expert interviews and a SLR, but we have not yet implemented and evaluated it in practice. There is much room for such future research involving research questions, such as which triggers, indicators, or countermeasures to choose. Further, we have yet to define how to quantify the stuck level, e.g., in terms of measurement units or how much individual concepts such as flow contribute to it. We require more research to define specific thresholds and the sizes of the zones, which depend on the players' individual characteristics, such as frustration tolerance. Measurement methods should be created that are able to determine them based on appropriate indicators.

To do so, we would advise researchers first to select a subset of indicators, such as arousal, boredom, cognitive load, flow, or frustration including appropriate measures (see Figure 5). They should be measurable in a game that also has stuck triggers, e.g., increasing difficulty. In a user study, the indicators can then be continuously measured while the game difficulty is constantly increased. Participants should be instructed to report when they feel a moderate stuck level (first threshold of the continuum; see Figure 7) as well as when they feel stuck (second threshold of the continuum). The measurement values of the indicators can then be correlated with the participants' stuck reports to create an automatic stuck level detection for games. As a next step, we would advise researchers to combine the automatic stuck level detection with automatic countermeasures, such as to provide hints or decrease the difficulty. In a user study, it should be evaluated if this leads to a decreased stuck level as our continuum predicts it. When this is confirmed, an approach for automatic stuck level detection as well as automatic reduction was created that could be implemented in games for stuck prevention.

We only presented and discussed the results of our SLR synthesis that were necessary to answer our RQs. Our supplementary spreadsheet shows that we have not evaluated information about the contributions according to Wobbrock and Kientz [185], the free-text descriptions of contributions, the used platform/type of system, the used RQs, study details, and the education topic. These data are useful for further insights about the included papers and might be used to find other underexplored research fields related to the state of "being stuck" for future work.

9 CONCLUSION

To define the state of "being stuck" in games, we conducted 13 expert interviews and a SLR with 104 relevant papers selected from 4022 candidates. We found that "being stuck" in games is related to multiple existing concepts but is still a unique state that we defined as continuous and subjective, which occurs when a player's personal goal cannot be reached. The SLR results showed that existing literature mainly covers established related concepts, but explicit stuck-related conceptual research is rarely present, a gap we targeted with our RQs. In conclusion, we have created a structured concept for a previously undefined but commonly used term by defining the state of "being stuck" in games that complements existing ones. Further, we provided a continuum based on our definition, presented a taxonomy of stuck-related measurements, and discussed countermeasures and their effectiveness. Our concept is a profound base for the implementation of game mechanics, preventing players from getting stuck.

ACKNOWLEDGMENTS

This work was conducted within the project "AuCity 2 - AR in University Teaching Using the Example of Civil Engineering", funded by the Federal Ministry of Education and Research (BMBF) and within the German Research Foundation (DFG) project "Empirical Assessment of Presence and Immersion in Augmented and Virtual Realities" (RU 1605/4-1).

REFERENCES

References marked with • are in the set of the 104 papers included in the synthesis of the SLR (see supplementary spreadsheet).

[1] • Amir Zaib Abbasi, Ding Hooi Ting, Helmut Hlavacs, Liliana Vale Costa, and Ana Isabel Veloso. 2019. An empirical validation of consumer video game engagement: A playful-consumption experience approach. *Entertainment Computing* 29 (2019), 43 – 55. https://doi.org/10.1016/j.entcom.2018.12.002

[2] Ernest Adams. 2014. *Fundamentals of Game Design.* Pearson Education.

[3] Amy L Alexander, Tad Brunyé, Jason Sidman, and Shawn A Weil. 2005. From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in pc-based simulations and games. *DARWARS Training Impact Group* 5 (2005), 1–14.

[4] • T. Alves, S. Gama, and F. S. Melo. 2018. Flow adaptation in serious games for health. In *2018 IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH)*. 1–8. https://doi.org/10.1109/SeGAH.2018.8401382

[5] Erik Andersen. 2012. Optimizing Adaptivity in Educational Games. In *Proceedings of the International Conference on the Foundations of Digital Games* (Raleigh, North Carolina) *(FDG '12)*. Association for Computing Machinery, New York, NY, USA, 279–281. https://doi.org/10.1145/2282338.2282398

[6] • Craig G. Anderson, Kathryn Campbell, and Constance Steinkuehler. 2019. Building Persistence through Failure: The Role of Challenge in Video Games. In *Proceedings of the 14th International Conference on the Foundations of Digital Games* (San Luis Obispo, California, USA) *(FDG '19)*. Association for Computing Machinery, New York, NY, USA, Article 34, 6 pages. https://doi.org/10.1145/3337722.3337741

[7] • Craig G. Anderson, Jen Dalsen, Vishesh Kumar, Matthew Berland, and Constance Steinkuehler. 2018. Failing up: How failure in a game environment promotes learning through discourse. *Thinking Skills and Creativity* 30 (2018), 135–144. https://doi.org/10.1016/j.tsc.2018.03.002 The Role of Failure in Promoting Thinking Skills and Creativity.

[8] • K. C. Apostolakis, K. Stefanidis, A. Psaltis, K. Dimitropoulos, and P. Daras. 2017. Offline and online adaptation in prosocial games. In *2017 9th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*. 201–208. https://doi.org/10.1109/VS-GAMES.2017.8056602

[9] Association for Computing Machinery. 2020. *ACM Digital Library*. https://dl.acm.org/ (Accessed on 08/23/2020).

[10] • Youngkyun Baek and Achraf Touati. 2017. Exploring how individual traits influence enjoyment in a mobile learning game. *Computers in Human Behavior* 69 (2017), 347–357. https://doi.org/10.1016/j.chb.2016.12.053

[11] Javier A. Bargas-Avila and Kasper Hornbæk. 2011. Old Wine in New Bottles or Novel Challenges: A Critical Analysis of Empirical Studies of User Experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 2689–2698. https://doi.org/10.1145/1978942.1979336

[12] • Glen Berseth, M. Brandon Haworth, Mubbasir Kapadia, and Petros Faloutsos. 2014. Characterizing and Optimizing Game Level Difficulty. In *Proceedings of the Seventh International Conference on Motion in Games* (Playa Vista, California) *(MIG '14)*. Association for Computing Machinery, New York, NY, USA, 153–160. https://doi.org/10.1145/2668064.2668100

[13] • N. Beume, H. Danielsiek, C. Eichhorn, B. Naujoks, M. Preuss, K. Stiller, and S. Wessing. 2008. Measuring flow as concept for detecting game fun in the Pac-Man game. In *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*. 3448–3455. https://doi.org/10.1109/CEC.2008.4631264

[14] • Fernando Bevilacqua, Henrik Engström, and Per Backlund. 2018. Changes in heart rate and facial actions during a gaming session with provoked boredom and stress. *Entertainment Computing* 24 (2018), 10–20. https://doi.org/10.1016/j.entcom.2017.10.004

[15] • Y. Bian, C. Yang, F. Gao, H. Li, X. Sun, X. Meng, and Y. Wang. 2015. A Physiological Evaluation Model for Flow-Experience in VR Games: Construction and Preliminary Test. In *2015 International Conference on Identification, Information, and Knowledge in the Internet of Things (IIKI)*. 244–249. https://doi.org/10.1109/IIKI.2015.59

[16] • Francisco Bicho and Carlos Martinho. 2018. *Multi-Dimensional Player Skill Progression Modelling for Procedural Content Generation*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3235765.3235774

[17] • G. Biswas, R. Rajendran, N. Mohammed, B. S. Goldberg, R. A. Sottilare, K. Brawner, and M. Hoffman. 2020. Multilevel Learner Modeling in Training Environments for Complex Decision Making. *IEEE Transactions on Learning Technologies* 13, 1 (2020), 172–185. https://doi.org/10.1109/TLT.2019.2923352

[18] • Paula Bitrián, Isabel Buil, and Sara Catalán. 2020. Flow and business simulation games: A typology of students. *The International Journal of Management Education* 18, 1 (2020), 100365. https://doi.org/10.1016/j.ijme.2020.100365

[19] • Paris Mavromoustakos Blom, Sander Bakkes, and Pieter Spronck. 2019. Modeling and adjusting in-game difficulty based on facial expression analysis. *Entertainment Computing* 31 (2019), 100307. https://doi.org/10.1016/j.entcom.2019.100307

[20] • Boyan Bontchev and Dessislava Vassileva. 2016. Assessing Engagement in an Emotionally-Adaptive Applied Game. In *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality* (Salamanca, Spain) *(TEEM '16)*. Association for Computing Machinery, New York, NY, USA, 747–754. https://doi.org/10.1145/3012430.3012602

[21] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa arXiv:https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp063oa

[22] Virginia Braun and Victoria Clarke. 2020. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology* 0, 0 (2020), 1–25. https://doi.org/10.1080/14780887.2020.1769238 arXiv:https://doi.org/10.1080/14780887.2020.1769238

[23] Pearl Brereton, Barbara A. Kitchenham, David Budgen, Mark Turner, and Mohamed Khalil. 2007. Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software* 80, 4 (2007), 571 – 583. https://doi.org/10.1016/j.jss.2006.07.009 Software Performance.

[24] • J. Brich, K. Rogers, J. Frommel, M. Weidhaas, A. Bruckner, S. Mirabile, T. Dorn, V. Riemer, C. Schrader, and M. Weber. 2015. LiverDefense: Using a Tower Defense Game as a Customisable Research Tool. In *2015 7th International Conference on Games and Virtual Worlds for Serious Applications (VS-Games)*. 1–8. https://doi.org/10.1109/VS-GAMES.2015.7295779

[25] Jeanne H. Brockmyer, Christine M. Fox, Kathleen A. Curtiss, Evan McBroom, Kimberly M. Burkhart, and Jacquelyn N. Pidruzny. 2009. The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology* 45, 4 (2009), 624 – 634. https://doi.org/10.1016/j.jesp.2009.02.016

[26] • A. Burns and J. Tulip. 2017. Detecting flow in games using facial expressions. In *2017 IEEE Conference on Computational Intelligence and Games (CIG)*. 45–52. https://doi.org/10.1109/CIG.2017.8080414

[27] • G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun. 2011. Emotion Assessment From Physiological Signals for Adaptation of Game Difficulty. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 41, 6 (2011), 1052–1063. https://doi.org/10.1109/TSMCA.2011.2116000

[28] • Guillaume Chanel, Cyril Rebetez, Mireille Bétrancourt, and Thierry Pun. 2008. Boredom, Engagement and Anxiety as Indicators for Adaptation to Difficulty in Games. In *Proceedings of the 12th International Conference on Entertainment and Media in the Ubiquitous Era* (Tampere, Finland) *(MindTrek '08)*. Association for Computing Machinery, New York, NY, USA, 13–17. https://doi.org/10.1145/1457199.1457203

[29] • Li-Xian Chen and Chuen-Tsai Sun. 2016. Self-Regulation Influence on Game Play Flow State. *Comput. Hum. Behav.* 54, C (Jan. 2016), 341–350. https://doi.org/10.1016/j.chb.2015.08.020

[30] Domenic V Cicchetti and Alvan R Feinstein. 1990. High agreement but low kappa: II. Resolving the paradoxes. *Journal of clinical epidemiology* 43, 6 (1990), 551–558.

[31] Serdar Çiftci. 2018. Trends of serious games research from 2007 to 2017: A bibliometric analysis. *Journal of Education and Training Studies* 6, 2 (2018), 18–27.

[32] • Douglas B. Clark, Satyugjit S. Virk, Jackie Barnes, and Deanne M. Adams. 2016. Self-explanation and digital games: Adaptively increasing abstraction. *Computers & Education* 103 (2016), 28–43. https://doi.org/10.1016/j.compedu.2016.09.010

[33] Oscar Clark. 2014. *Games as a service: How free to play design can make better games.* CRC Press.

[34] Ruth Colvin Clark, Frank Nguyen, John Sweller, and Melissa Baddeley. 2006. Efficiency in learning: Evidence-based guidelines to manage cognitive load. *Performance Improvement* 45, 9 (2006), 46–47.

[35] Jacob Cohen. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20, 1 (1960), 37–46. https://doi.org/10.1177/001316446002000104 arXiv:https://doi.org/10.1177/001316446002000104

[36] Jacob Cohen. 1968. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin* 70, 4 (1968), 213.

[37] Corporation for Digital Scholarship. 2020. *Zotero*. https://www.zotero.org/ (Accessed on 08/23/2020).

[38] Mihaly Csikszentmihalyi. 1975. Beyond boredom and anxiety: Experiencing flow in work and play. *San Francisco/Washington/London* (1975).

[39] Mihaly Csikszentmihalyi. 1990. *Flow: The psychology of optimal experience.* Vol. 1990. Harper & Row New York.

[40] Mihaly Csikszentmihalyi. 1997. Finding flow: The psychology of engagement with everyday life. (1997).

[41] • Simon Demediuk, Marco Tamassia, William L. Raffe, Fabio Zambetta, Florian "Floyd" Mueller, and Xiaodong Li. 2018. Measuring Player Skill Using Dynamic Difficulty Adjustment. In *Proceedings of the Australasian Computer Science Week Multiconference* (Brisband, Queensland, Australia) *(ACSW '18)*. Association for Computing Machinery, New York, NY, USA, Article 41, 7 pages. https://doi.org/10.1145/3167918.3167939

[42] Guillaume Denis and Pierre Jouvelot. 2005. Motivation-Driven Educational Game Design: Applying Best Practices to Music Education. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology* (Valencia, Spain) *(ACE '05)*. Association for Computing Machinery, New York, NY, USA, 462–465. https://doi.org/10.1145/1178477.1178581

[43] Alena Denisova and Paul Cairns. 2015. Adaptation in Digital Games: The Effect of Challenge Adjustment on Player Performance and Experience. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play* (London, United Kingdom) *(CHI PLAY '15)*. Association for Computing Machinery, New York, NY, USA, 97–101. https://doi.org/10.1145/2793107.2793141

[44] Alena Denisova and Paul Cairns. 2019. Player experience and deceptive expectations of difficulty adaptation in digital games. *Entertainment Computing* 29 (2019), 56 – 68. https://doi.org/10.1016/j.entcom.2018.12.001

[45] Alena Denisova, Paul Cairns, Christian Guckelsberger, and David Zendle. 2020. Measuring perceived challenge in digital games: Development & validation of the challenge originating from recent gameplay interaction scale (CORGIS). *International Journal of Human-Computer Studies* 137 (2020), 102383. https://doi.org/10.1016/j.ijhcs.2019.102383

[46] Alena Denisova, Christian Guckelsberger, and David Zendle. 2017. Challenge in Digital Games: Towards Developing a Measurement Tool. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 2511–2519. https://doi.org/10.1145/3027063.3053209

[47] Tobias Drey, Jan Gugenheimer, Julian Karlbauer, Maximilian Milo, and Enrico Rukzio. 2020. VRSketchIn: Exploring the Design Space of Pen and Tablet Interaction for 3D Sketching in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376628

[48] Tobias Drey, Pascal Jansen, Fabian Fischbach, Julian Frommel, and Enrico Rukzio. 2020. Towards Progress Assessment for Adaptive Hints in Educational Virtual Reality Games. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3334480.3382789

[49] • I. Dunwell, P. Petridis, M. Hendrix, S. Arnab, M. AL-Smadi, and C. Guetl. 2012. Guiding Intuitive Learning in Serious Games: An Achievement-Based Approach to Externalized Feedback and Assessment. In *2012 Sixth International Conference on Complex, Intelligent, and Software Intensive Systems*. 911–916. https://doi.org/10.1109/CISIS.2012.205

[50] Panteleimon Ekkekakis. 2012. Affect, mood, and emotion. *Measurement in sport and exercise psychology* 321 (2012).

[51] • Lucca Eloy, Angela E.B. Stewart, Mary Jean Amon, Caroline Reinhardt, Amanda Michaels, Chen Sun, Valerie Shute, Nicholas D. Duran, and Sidney D'Mello. 2019. Modeling Team-Level Multimodal Dynamics during Multiparty Collaboration. In *2019 International Conference on Multimodal Interaction* (Suzhou, China) *(ICMI '19)*. Association for Computing Machinery, New York, NY, USA, 244–258. https://doi.org/10.1145/3340555.3353748

[52] Elsevier. 2020. *ScienceDirect.com | Science, health and medical journals, full text articles and books.* https://www.sciencedirect.com/ (Accessed on 08/23/2020).

[53] • Séverine Erhel and Eric Jamet. 2019. Improving instructions in educational computer games: Exploring the relations between goal specificity, flow experience and learning outcomes. *Computers in Human Behavior* 91 (2019), 106 – 114. https://doi.org/10.1016/j.chb.2018.09.020

[54] Alvan R Feinstein and Domenic V Cicchetti. 1990. High agreement but low kappa: I. The problems of two paradoxes. *Journal of clinical epidemiology* 43, 6 (1990), 543–549.

[55] Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin* 76, 5 (1971), 378.

[56] • S. Foutsitzi, S. Asteriadis, and G. Caridakis. 2019. An overview of Affective Models and ICT in Education. In *2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA)*. 1–8. https://doi.org/10.1109/IISA.2019.8900783

[57] • Julian Frommel, Fabian Fischbach, Katja Rogers, and Michael Weber. 2018. Emotion-Based Dynamic Difficulty Adjustment Using Parameterized Difficulty and Self-Reports of Emotion. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play* (Melbourne, VIC, Australia) *(CHI PLAY '18)*. Association for Computing Machinery, New York, NY, USA, 163–171. https://doi.org/10.1145/3242671.3242682

[58] • Julian Frommel, Claudia Schrader, and Michael Weber. 2018. Towards Emotion-Based Adaptive Games: Emotion Recognition Via Input and Performance Features. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play* (Melbourne, VIC, Australia) *(CHI PLAY '18)*. Association for Computing Machinery, New York, NY, USA, 173–185. https://doi.org/10.1145/3242671.3242672

[59] Fong-Ling Fu, Rong-Chang Su, and Sheng-Chin Yu. 2009. EGameFlow: A scale to measure learners' enjoyment of e-learning games. *Computers & Education* 52, 1 (2009), 101 – 112. https://doi.org/10.1016/j.compedu.2008.07.004

[60] • Xin Fu, Xi Chen, Yu-Tong Shi, Indranil Bose, and Shun Cai. 2017. User segmentation for retention management in online social games. *Decision Support Systems* 101 (2017), 51–68. https://doi.org/10.1016/j.dss.2017.05.015

[61] Clive J Fullagar, Patrick A Knight, and Heather S Sovern. 2013. Challenge/skill balance, flow, and performance anxiety. *Applied Psychology* 62, 2 (2013), 236–259.

[62] Tracy Fullerton. 2014. *Game design workshop: a playcentric approach to creating innovative games.* CRC press.

[63] • M. Gavriushenko, L. Karilainen, and M. Kankaanranta. 2015. Adaptive systems as enablers of feedback in English language learning game-based environments. In *2015 IEEE Frontiers in Education Conference (FIE)*. 1–8. https://doi.org/10.1109/FIE.2015.7344107

[64] Kiel M Gilleade and Alan Dix. 2004. Using Frustration in the Design of Adaptive Videogames. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology* (Singapore) *(ACE '04)*. Association for Computing Machinery, New York, NY, USA, 228–232. https://doi.org/10.1145/1067343.1067372

[65] • Benjamin Goldberg and Janis Cannon-Bowers. 2015. Feedback source modality effects on training outcomes in a serious game: Pedagogical agents make a difference. *Computers in Human Behavior* 52 (2015), 1–11. https:

//doi.org/10.1016/j.chb.2015.05.008
[66] Fateme Hajiarbabi, Hassan Ahadi, Ali Delavar, and Hasan Asadzadeh. 2012. Computer Games and Its Effect on Attribution of Learned Helplessness Students. *Life Science Journal* 9, 3 (2012).
[67] • Sandro Hardy, Angelika Kern, Tim Dutz, Christoph Weber, Stefan Göbel, and Ralf Steinmetz. 2014. What Makes Games Challenging? Considerations on How to Determine the "Challenge" Posed by an Exergame for Balance Training. In *Proceedings of the 2014 ACM International Workshop on Serious Games* (Orlando, Florida, USA) *(SeriousGames '14)*. Association for Computing Machinery, New York, NY, USA, 57–62. https://doi.org/10.1145/2656719.2656728
[68] • László Harmat, Örjan de Manzano, Töres Theorell, Lennart Högman, Håkan Fischer, and Fredrik Ullén. 2015. Physiological correlates of the flow experience during computer game playing. *International Journal of Psychophysiology* 97, 1 (2015), 1–7. https://doi.org/10.1016/j.ijpsycho.2015.05.001
[69] • B. Harrison and D. L. Roberts. 2015. An Analytic and Psychometric Evaluation of Dynamic Game Adaption for Increasing Session-Level Retention in Casual Games. *IEEE Transactions on Computational Intelligence and AI in Games* 7, 3 (2015), 207–219. https://doi.org/10.1109/TCIAIG.2015.2410757
[70] • Drew Hicks, Michael Eagle, Elizabeth Rowe, Jodi Asbell-Clarke, Teon Edwards, and Tiffany Barnes. 2016. Using Game Analytics to Evaluate Puzzle Design and Level Progression in a Serious Game. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge* (Edinburgh, United Kingdom) *(LAK '16)*. Association for Computing Machinery, New York, NY, USA, 440–448. https://doi.org/10.1145/2883851.2883953
[71] Teresa Hirzle, Jan Gugenheimer, Florian Geiselhart, Andreas Bulling, and Enrico Rukzio. 2019. A Design Space for Gaze Interaction on Head-Mounted Displays. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300855
[72] • Jon-Chao Hong, Kai-Hsin Tai, Ming-Yueh Hwang, and Yen-Chun Kuo. 2016. Internet cognitive failure affects learning progress as mediated by cognitive anxiety and flow while playing a Chinese antonym synonym game with interacting verbal–analytical and motor-control. *Computers & Education* 100 (2016), 32–44. https://doi.org/10.1016/j.compedu.2016.04.009
[73] • Huei-Tse Hou and Ming-Chaun Li. 2014. Evaluating multiple aspects of a digital educational problem-solving-based adventure game. *Computers in Human Behavior* 30 (2014), 29–38. https://doi.org/10.1016/j.chb.2013.07.052
[74] • X. Hou, Y. Liu, O. Sourina, and W. Mueller-Wittig. 2015. CogniMeter: EEG-based Emotion, Mental Workload and Stress Visual Monitoring. In *2015 International Conference on Cyberworlds (CW)*. 153–160. https://doi.org/10.1109/CW.2015.58
[75] • Ting-Chia Hsu. 2017. Learning English with Augmented Reality: Do learning styles matter? *Computers & Education* 106 (2017), 137–149. https://doi.org/10.1016/j.compedu.2016.12.007
[76] • C. Hung, F. Kuo, J. Chih-Yuan Sun, and P. Yu. 2014. An Interactive Game Approach for Improving Students' Learning Performance in Multi-Touch Game-Based Learning. *IEEE Transactions on Learning Technologies* 7, 1 (Jan 2014), 31–37. https://doi.org/10.1109/TLT.2013.2294806
[77] • Ming-Yueh Hwang, Jon-Chao Hong, Jian-Hong Ye, Yu-Feng Wu, Kai-Hsin Tai, and Ming-Chi Kiu. 2019. Practicing abductive reasoning: The correlations between cognitive factors and learning effects. *Computers & Education* 138 (2019), 33–45. https://doi.org/10.1016/j.compedu.2019.04.014
[78] IEEE. 2020. *IEEE Xplore*. https://ieeexplore.ieee.org/Xplore/home.jsp (Accessed on 08/23/2020).
[79] Insilica LLC. 2020. *Sysrev*. https://sysrev.com/ (Accessed on 08/23/2020).
[80] • A. Isaksen, D. Gopstein, J. Togelius, and A. Nealen. 2018. Exploring Game Space of Minimal Action Games via Parameter Tuning and Survival Analysis. *IEEE Transactions on Games* 10, 2 (2018), 182–194. https://doi.org/10.1109/TCIAIG.2017.2750181
[81] • A. Johnson, Y. Tang, and C. Franzwa. 2014. kNN-based adaptive virtual reality game system. In *Proceedings of the 11th IEEE International Conference on Networking, Sensing and Control*. 457–462. https://doi.org/10.1109/ICNSC.2014.6819669
[82] Jesper Juul. 2009. Fear of Failing? The Many Meanings of Difficulty in Video Games. *The Video Game Theory Reader* 2, 237-252 (2009).
[83] Jesper Juul. 2013. *The Art of Failure: An Essay on the Pain of Playing Video Games*. MIT Press.
[84] • Jina Kang, Min Liu, and Wen Qu. 2017. Using gameplay data to examine learning behavior patterns in a serious game. *Computers in Human Behavior* 72 (2017), 757–770.
[85] • A. Katmada, M. Chatzakis, H. Apostolidis, A. Mavridis, and S. Panagiotis. 2015. An adaptive serious neuro-game using a mobile version of a bio-feedback device. In *2015 International Conference on Interactive Mobile Communication Technologies and Learning (IMCL)*. 416–420. https://doi.org/10.1109/IMCTL.2015.7359633
[86] • A. Katmada, A. Mavridis, H. Apostolidis, and T. Tsiatsos. 2015. Developing an adaptive serious game based on students' bio-feedback. In *2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA)*. 1–6. https://doi.org/10.1109/IISA.2015.7387975

[87] Daniel Kepplinger, Günter Wallner, Simone Kriglstein, and Michael Lankes. 2020. See, Feel, Move: Player Behaviour Analysis through Combined Visualization of Gaze, Emotions, and Movement. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376401

[88] • Mohammad M. Khajah, Brett D. Roads, Robert V. Lindsey, Yun-En Liu, and Michael C. Mozer. 2016. *Designing Engaging Games Using Bayesian Optimization.* Association for Computing Machinery, New York, NY, USA, 5571–5582. https://doi.org/10.1145/2858036.2858253

[89] • M. D. Kickmeier-Rust, C. M. Steiner, and D. Albert. 2009. Non-invasive Assessment and Adaptive Interventions in Learning Games. In *2009 International Conference on Intelligent Networking and Collaborative Systems*. 301–305. https://doi.org/10.1109/INCOS.2009.30

[90] Kristian Kiili. 2005. Digital game-based learning: Towards an experiential gaming model. *The Internet and Higher Education* 8, 1 (2005), 13 – 24. https://doi.org/10.1016/j.iheduc.2004.12.001

[91] Kristian Kiili, Sara de Freitas, Sylvester Arnab, and Timo Lainema. 2012. The Design Principles for Flow Experience in Educational Games. *Procedia Computer Science* 15 (2012), 78 – 91. https://doi.org/10.1016/j.procs.2012.10.060 4th International Conference on Games and Virtual Worlds for Serious Applications(VS-GAMES'12).

[92] • Kristian Kiili, Timo Lainema, Sara de Freitas, and Sylvester Arnab. 2014. Flow framework for analyzing the quality of educational games. *Entertainment Computing* 5, 4 (2014), 367 – 377. https://doi.org/10.1016/j.entcom.2014.08.002

[93] Seungwook Kim, Daeyoung Choi, Eunjung Lee, and Wonjong Rhee. 2017. Churn prediction of mobile and online casual games using play log data. *PloS one* 12, 7 (2017).

[94] Barbara Kitchenham and Pearl Brereton. 2013. A systematic review of systematic review process research in software engineering. *Information and Software Technology* 55, 12 (2013), 2049 – 2075. https://doi.org/10.1016/j.infsof.2013.07.010

[95] Madison Klarkowski, Daniel Johnson, Peta Wyeth, Mitchell McEwan, Cody Phillips, and Simon Smith. 2016. Operationalising and Evaluating Sub-Optimal and Optimal Play Experiences through Challenge-Skill Manipulation. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5583–5594. https://doi.org/10.1145/2858036.2858563

[96] • Julia Kneer, Malte Elson, and Florian Knapp. 2016. Fight fire with rainbows: The effects of displayed violence, difficulty, and performance in digital games on affect, aggression, and physiological arousal. *Computers in Human Behavior* 54 (2016), 142–148. https://doi.org/10.1016/j.chb.2015.07.034

[97] • M. Köles, L. Szegletes, and B. Forstner. 2015. Towards a physiology based difficulty control system for serious games. In *2015 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. 323–328. https://doi.org/10.1109/CogInfoCom.2015.7390612

[98] Effie Lai-Chong Law and Xu Sun. 2012. Evaluating user experience of adaptive digital educational games with Activity Theory. *International Journal of Human-Computer Studies* 70, 7 (2012), 478 – 497. https://doi.org/10.1016/j.ijhcs.2012.01.007 Special Issue on User Experience (UX) in Virtual Learning Environments.

[99] • Michael J. Lee, Andrew J. Ko, and Irwin Kwan. 2013. In-Game Assessments Increase Novice Programmers' Engagement and Level Completion Speed. In *Proceedings of the Ninth Annual International ACM Conference on International Computing Education Research* (San Diego, San California, USA) *(ICER '13)*. Association for Computing Machinery, New York, NY, USA, 153–160. https://doi.org/10.1145/2493394.2493410

[100] Serena Lee-Cultura, Kshitij Sharma, Sofia Papavlasopoulou, Symeon Retalis, and Michail Giannakos. 2020. Using Sensing Technologies to Explain Children's Self-Representation in Motion-Based Educational Games. In *Proceedings of the Interaction Design and Children Conference* (London, United Kingdom) *(IDC '20)*. Association for Computing Machinery, New York, NY, USA, 541–555. https://doi.org/10.1145/3392063.3394419

[101] Hua Li, Hector Munoz-Avila, Lei Ke, Carl Symborski, and Rafael Alonso. 2013. Discovery of player strategies in a serious game. In *First AAAI Conference on Human Computation and Crowdsourcing*. Citeseer.

[102] • Y. Li and A. S. Elmaghraby. 2014. A framework for using games for behavioral analysis of autistic children. In *2014 Computer Games: AI, Animation, Mobile, Multimedia, Educational and Serious Games (CGAMES)*. 1–4. https://doi.org/10.1109/CGames.2014.6934157

[103] • Chen-Chung Liu, Yuan-Bang Cheng, and Chia-Wen Huang. 2011. The effect of simulation games on the learning of computational problem solving. *Computers & Education* 57, 3 (2011), 1907–1918. https://doi.org/10.1016/j.compedu.2011.04.002

[104] Pramila Rani Nilanjan Sarkar Changchun Liu. 2005. Maintaining optimal challenge in computer games through real-time physiological feedback. *Foundations of Augmented Cognition* (2005), 184.

[105] • Christian E. Lopez and Conrad S. Tucker. 2019. The effects of player type on performance: A gamification case study. *Computers in Human Behavior* 91 (2019), 333–345. https://doi.org/10.1016/j.chb.2018.10.005

[106] • L. B. Martins Kummer, J. Cesar Nievola, and E. C. Paraiso. 2018. Applying Commitment to Churn and Remaining Players Lifetime Prediction. In *2018 IEEE Conference on Computational Intelligence and Games (CIG)*. 1–8. https:

//doi.org/10.1109/CIG.2018.8490443
[107] • E. Masanobu, H. D. F. B., and K. Mikami. 2017. Dynamic Pressure Cycle Control: Dynamic Diffculty Adjustment beyond the Flow Zone. In *2017 Nicograph International (NicoInt)*. 9–14. https://doi.org/10.1109/NICOInt.2017.12
[108] Deborah J. Mayhew. 1999. The Usability Engineering Lifecycle. In *CHI '99 Extended Abstracts on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania) *(CHI EA '99)*. Association for Computing Machinery, New York, NY, USA, 147–148. https://doi.org/10.1145/632716.632805
[109] • Elisa D. Mekler, Julia Ayumi Bopp, Alexandre N. Tuch, and Klaus Opwis. 2014. A Systematic Review of Quantitative Studies on the Enjoyment of Digital Entertainment Games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 927–936. https://doi.org/10.1145/2556288.2557078
[110] • Matthew K. Miller and Regan L. Mandryk. 2016. Differentiating In-Game Frustration from at-Game Frustration Using Touch Pressure. In *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces* (Niagara Falls, Ontario, Canada) *(ISS '16)*. Association for Computing Machinery, New York, NY, USA, 225–234. https://doi.org/10.1145/2992154.2992185
[111] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G Altman, Prisma Group, et al. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS med* 6, 7 (2009), e1000097.
[112] David Moher, Larissa Shamseer, Mike Clarke, Davina Ghersi, Alessandro Liberati, Mark Petticrew, Paul Shekelle, Lesley A Stewart, et al. 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic reviews* 4, 1 (2015), 1.
[113] • D. Monteiro, H. Liang, A. Abel, N. Bahaei, and R. d. C. Monteiro. 2018. Evaluating Engagement of Virtual Reality Games Based on First and Third Person Perspective Using EEG and Subjective Metrics. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. 53–60. https://doi.org/10.1109/AIVR.2018.00015
[114] • Belkacem Mostefai, Amar Balla, and Philippe Trigano. 2019. A generic and efficient emotion-driven approach toward personalized assessment and adaptation in serious games. *Cognitive Systems Research* 56 (2019), 82–106. https://doi.org/10.1016/j.cogsys.2019.03.006
[115] • J. E. Muñoz, S. Cao, and J. Boger. 2019. Kinematically Adaptive Exergames: Personalizing Exercise Therapy Through Closed-Loop Systems. In *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. 118–1187. https://doi.org/10.1109/AIVR46125.2019.00026
[116] • L. Müller, A. Bernin, A. Kamenz, S. Ghose, K. von Luck, C. Grecos, Q. Wang, and F. Vogt. 2017. Emotional journey for an emotion provoking cycling exergame. In *2017 IEEE 4th International Conference on Soft Computing Machine Intelligence (ISCMI)*. 104–108. https://doi.org/10.1109/ISCMI.2017.8279607
[117] S. C. Müller and T. Fritz. 2015. Stuck and Frustrated or in Flow and Happy: Sensing Developers' Emotions and Progress. In *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, Vol. 1. 688–699.
[118] Lennart E. Nacke, Anna Cox, Regan L. Mandryk, and Paul Cairns. 2016. SIGCHI Games: The Scope of Games and Play Research at CHI. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 1088–1091. https://doi.org/10.1145/2851581.2886438
[119] • A. Nagle, D. Novak, P. Wolf, and R. Riener. 2014. The effect of different difficulty adaptation strategies on enjoyment and performance in a serious game for memory training. In *2014 IEEE 3nd International Conference on Serious Games and Applications for Health (SeGAH)*. 1–8. https://doi.org/10.1109/SeGAH.2014.7067088
[120] • V. Naik and V. Kamat. 2015. Adaptive and Gamified Learning Environment (AGLE). In *2015 IEEE Seventh International Conference on Technology for Education (T4E)*. 7–14. https://doi.org/10.1109/T4E.2015.23
[121] Steve Nebel, Maik Beege, Sascha Schneider, and Günter Daniel Rey. 2020. Competitive Agents and Adaptive Difficulty Within Educational Video Games. *Frontiers in Education* 5 (2020), 129. https://doi.org/10.3389/feduc.2020.00129
[122] Loc Nguyen and Phung Do. 2008. Learner model in adaptive learning. *World Academy of Science, Engineering and Technology* 45, 70 (2008), 395–400.
[123] • Ogechi Nnadi, Ute Fischer, Michael Boyce, and Michael Nitsche. 2008. Effect of Dynamic Camera Control on Spatial Reasoning in 3D Spaces. In *Proceedings of the 2008 ACM SIGGRAPH Symposium on Video Games* (Los Angeles, California) *(Sandbox '08)*. Association for Computing Machinery, New York, NY, USA, 157–162. https://doi.org/10.1145/1401843.1401873
[124] Thomas P Novak and Donna L Hoffman. 1997. Measuring the flow experience among web users. (1997).
[125] • B. Omella Mainieri, V. L. Azevedo, P. H. Cacique Braga, and N. Omar. 2018. Development and Assessment of an Adaptive Difficulty Arithmetic Game Based Learning Object. In *2018 XIII Latin American Conference on Learning Technologies (LACLO)*. 232–239. https://doi.org/10.1109/LACLO.2018.00051
[126] • Eleanor O'Rourke, Christy Ballweber, and Zoran Popovií. 2014. Hint Systems May Negatively Impact Performance in Educational Games. In *Proceedings of the First ACM Conference on Learning @ Scale Conference* (Atlanta, Georgia, USA) *(L@S '14)*. Association for Computing Machinery, New York, NY, USA, 51–60. https://doi.org/10.1145/2556325.2566248

[127] • Karin A. Orvis, Daniel B. Horn, and James Belanich. 2008. The roles of task difficulty and prior videogame experience on performance and motivation in instructional videogames. *Computers in Human Behavior* 24, 5 (2008), 2415–2433. https://doi.org/10.1016/j.chb.2008.02.016 Including the Special Issue: Internet Empowerment.

[128] Fred G. W. C. Paas and Jeroen J. G. Van Merriënboer. 1993. The Efficiency of Instructional Conditions: An Approach to Combine Mental Effort and Performance Measures. *Human Factors* 35, 4 (1993), 737–743. https://doi.org/10.1177/001872089303500412 arXiv:https://doi.org/10.1177/001872089303500412

[129] Madhukar Pai, Michael McCulloch, Jennifer D Gorman, Nitika Pai, Wayne Enanoria, Gail Kennedy, Prathap Tharyan, and John M Colford Jr. 2004. Systematic reviews and meta-analyses: an illustrated, step-by-step guide. *The National medical journal of India* 17, 2 (2004), 86–95.

[130] • K. Park, T. Kihl, S. Park, M. Kim, and J. Chang. 2016. Narratives and sensor driven cognitive behavior training game platform. In *2016 IEEE 14th International Conference on Software Engineering Research, Management and Applications (SERA)*. 125–131. https://doi.org/10.1109/SERA.2016.7516137

[131] Reinhard Pekrun and Lisa Linnenbrink-Garcia. 2012. *Academic Emotions and Student Engagement*. Springer US, Boston, MA, 259–282. https://doi.org/10.1007/978-1-4614-2018-7_12

[132] • H. A. Pereira, A. F. D. Souza, and C. S. D. Menezes. 2018. Obtaining evidence of learning in digital games through a deep learning neural network to classify facial expressions of the players. In *2018 IEEE Frontiers in Education Conference (FIE)*. 1–8. https://doi.org/10.1109/FIE.2018.8659216

[133] • M. Pirovano, R. Mainetti, G. Baud-Bovy, P. L. Lanzi, and N. A. Borghese. 2012. Self-adaptive games for rehabilitation at home. In *2012 IEEE Conference on Computational Intelligence and Games (CIG)*. 179–186. https://doi.org/10.1109/CIG.2012.6374154

[134] • J.L Plass, B.D. Homer, S. Pawar, C. Brenner, and A.P. MacNamara. 2019. The effect of adaptive difficulty adjustment on the effectiveness of a game to develop executive function skills for learners of different ages. *Cognitive Development* 49 (2019), 56–67. https://doi.org/10.1016/j.cogdev.2018.11.006

[135] Michela Ponticorvo, Angelo Rega, Andrea Di Ferdinando, Davide Marocco, and Orazio Miglino. 2017. Approaches to Embed Bio-inspired Computational Algorithms in Educational and Serious Games.. In *CAID@ IJCAI*. 8–14.

[136] Anne Marie Porter and Paula Goolkasian. 2019. Video Games and Stress: How Stress Appraisals and Game Content Affect Cardiovascular and Emotion Outcomes. *Frontiers in Psychology* 10 (2019), 967. https://doi.org/10.3389/fpsyg.2019.00967

[137] Christopher Power, Paul Cairns, Alena Denisova, Themis Papaioannou, and Ruth Gultom. 2019. Lost at the Edge of Uncertainty: Measuring Player Uncertainty in Digital Games. *International Journal of Human–Computer Interaction* 35, 12 (2019), 1033–1045. https://doi.org/10.1080/10447318.2018.1507161 arXiv:https://doi.org/10.1080/10447318.2018.1507161

[138] Christopher Power, Alena Denisova, Themis Papaioannou, and Paul Cairns. 2017. Measuring Uncertainty in Games: Design and Preliminary Validation. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI EA '17)*. Association for Computing Machinery, New York, NY, USA, 2839–2845. https://doi.org/10.1145/3027063.3053215

[139] • A. Psaltis, K. C. Apostolakis, K. Dimitropoulos, and P. Daras. 2018. Multimodal Student Engagement Recognition in Prosocial Games. *IEEE Transactions on Games* 10, 3 (2018), 292–303. https://doi.org/10.1109/TCIAIG.2017.2743341

[140] Raquel Robinson, Katelyn Wiley, Amir Rezaeivahdati, Madison Klarkowski, and Regan L. Mandryk. 2020. *"Let's Get Physiological, Physiological!": A Systematic Review of Affective Gaming*. Association for Computing Machinery, New York, NY, USA, 132–147. https://doi.org/10.1145/3410404.3414227

[141] • Miia Ronimus, Janne Kujala, Asko Tolvanen, and Heikki Lyytinen. 2014. Children's engagement during digital game-based learning of reading: The effects of time, rewards, and challenge. *Computers & Education* 71 (2014), 237–246. https://doi.org/10.1016/j.compedu.2013.10.008

[142] Shaghayegh Roohi, Elisa D. Mekler, Mikke Tavast, Tatu Blomqvist, and Perttu Hämäläinen. 2019. Recognizing Emotional Expression in Game Streams. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Barcelona, Spain) *(CHI PLAY '19)*. Association for Computing Machinery, New York, NY, USA, 301–311. https://doi.org/10.1145/3311350.3347197

[143] Ira J Roseman. 1984. Cognitive determinants of emotion: A structural theory. *Review of personality & social psychology* (1984).

[144] • F. Rozi, Y. Rosmansyah, and B. Dabarsyah. 2019. A Systematic Literature Review on Adaptive Gamification: Components, Methods, and Frameworks. In *2019 International Conference on Electrical Engineering and Informatics (ICEEI)*. 187–190. https://doi.org/10.1109/ICEEI47359.2019.8988857

[145] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.

[146] • Sandra Sampayo-Vargas, Chris J. Cope, Zhen He, and Graeme J. Byrne. 2013. The effectiveness of adaptive difficulty adjustments on students' motivation and learning in an educational computer game. *Computers & Education* 69 (2013), 452–462. https://doi.org/10.1016/j.compedu.2013.07.004

[147] • Supat Sanjamsai and Darunee Phukao. 2018. Flow experience in computer game playing among Thai university students. *Kasetsart Journal of Social Sciences* 39, 2 (2018), 175 – 182. https://doi.org/10.1016/j.kjss.2018.03.003

[148] • B.R. Schadenberg, M.A. Neerincx, F. Cnossen, and R. Looije. 2017. Personalising game difficulty to keep children motivated to play with a social robot: A Bayesian approach. *Cognitive Systems Research* 43 (2017), 222–231. https://doi.org/10.1016/j.cogsys.2016.08.003

[149] Ruben Schertler, Simone Kriglstein, and Günter Wallner. 2019. User Guided Movement Analysis in Games Using Semantic Trajectories. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Barcelona, Spain) *(CHI PLAY '19)*. Association for Computing Machinery, New York, NY, USA, 613–623. https://doi.org/10.1145/3311350.3347156

[150] • Henrik Schoenau-Fog, Alexander Birke, and Lars Reng. 2012. Evaluation of Continuation Desire as an Iterative Game Development Method. In *Proceeding of the 16th International Academic MindTrek Conference* (Tampere, Finland) *(MindTrek '12)*. Association for Computing Machinery, New York, NY, USA, 241–243. https://doi.org/10.1145/2393132.2393185

[151] Claudia Schrader, Julia Brich, Julian Frommel, Valentin Riemer, and Katja Rogers. 2017. *Rising to the Challenge: An Emotion-Driven Approach Toward Adaptive Serious Games*. Springer International Publishing, Cham, 3–28. https://doi.org/10.1007/978-3-319-51645-5_1

[152] • Claudia Schrader and Ulrike Nett. 2018. The perception of control as a predictor of emotional trends during gameplay. *Learning and Instruction* 54 (2018), 62–72.

[153] • Chen Shu-Hui, Wu Wann-Yih, and Jason Dennison. 2018. Validation of EGameFlow: A Self-Report Scale for Measuring User Experience in Video Game Play. *Comput. Entertain.* 16, 3, Article 6 (Sept. 2018), 15 pages. https://doi.org/10.1145/3238249

[154] Valerie J. Shute, Sidney D'Mello, Ryan Baker, Kyunghwa Cho, Nigel Bosch, Jaclyn Ocumpaugh, Matthew Ventura, and Victoria Almeda. 2015. Modeling how incoming knowledge, persistence, affective states, and in-game progress influence student learning from an educational game. *Computers & Education* 86 (2015), 224 – 235. https://doi.org/10.1016/j.compedu.2015.08.001

[155] Andy P. Siddaway, Alex M. Wood, and Larry V. Hedges. 2019. How to Do a Systematic Review: A Best Practice Guide for Conducting and Reporting Narrative Reviews, Meta-Analyses, and Meta-Syntheses. *Annual Review of Psychology* 70, 1 (2019), 747–770. https://doi.org/10.1146/annurev-psych-010418-102803 arXiv:https://doi.org/10.1146/annurev-psych-010418-102803 PMID: 30089228.

[156] • Mirna Paula Silva, Victor do Nascimento Silva, and Luiz Chaimowicz. 2017. Dynamic difficulty adjustment on MOBA games. *Entertainment Computing* 18 (2017), 103–123. https://doi.org/10.1016/j.entcom.2016.10.002

[157] • A. Sinha, R. Gavas, D. Chatterjee, R. Das, and A. Sinharay. 2015. Dynamic assessment of learners' mental state for an improved learning experience. In *2015 IEEE Frontiers in Education Conference (FIE)*. 1–9.

[158] Alexander Skulmowski and Günter Daniel Rey. 2017. Measuring cognitive load in embodied learning settings. *Frontiers in Psychology* 8 (2017), 1191.

[159] • M. Song, Z. Yang, A. Baird, E. Parada-Cabaleiro, Z. Zhang, Z. Zhao, and B. Schuller. 2019. Audiovisual Analysis for Recognising Frustration during Game-Play: Introducing the Multimodal Game Frustration Database. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. 517–523. https://doi.org/10.1109/ACII.2019.8925464

[160] Jaakko Stenros. 2017. The Game Definition Game: A Review. *Games and culture* 12, 6 (2017), 499–520.

[161] • Chuen-Tsai Sun, Li-Xian Chen, and Hsiu-Mei Chu. 2018. Associations among scaffold presentation, reward mechanisms and problem-solving behaviors in game play. *Computers & Education* 119 (2018), 95–111. https://doi.org/10.1016/j.compedu.2018.01.001

[162] Penny Sweetser, Daniel Johnson, and Peta Wyeth. 2012. Revisiting the GameFlow model with detailed heuristics. *Journal of Creative Technologies* 2012, 3 (2012), 1–16. https://eprints.qut.edu.au/58216/

[163] Penelope Sweetser, Daniel Johnson, Peta Wyeth, Aiman Anwar, Yan Meng, and Anne Ozdowska. 2017. GameFlow in Different Game Genres and Platforms. *Comput. Entertain.* 15, 3, Article 1 (April 2017), 24 pages. https://doi.org/10.1145/3034780

[164] Penelope Sweetser and Peta Wyeth. 2005. GameFlow: A Model for Evaluating Player Enjoyment in Games. *Comput. Entertain.* 3, 3 (July 2005), 3. https://doi.org/10.1145/1077246.1077253

[165] John Sweller. 1988. Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science* 12, 2 (1988), 257–285. https://doi.org/10.1207/s15516709cog1202_4 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog1202_4

[166] John Sweller, Jeroen JG Van Merrienboer, and Fred GWC Paas. 1998. Cognitive architecture and instructional design. *Educational psychology review* 10, 3 (1998), 251–296.

[167] • Samira Syal and John L. Nietfeld. 2020. The impact of trace data and motivational self-reports in a game-based learning environment. *Computers & Education* 157 (2020), 103978. https://doi.org/10.1016/j.compedu.2020.103978

[168] • M. A. Syufagi, P. Mauridhi Hery, and M. Hariadi. 2011. Modeling serious games based on Cognitive Skill classification using Learning Vector Quantization with Petri net. In *2011 International Conference on Advanced Computer Science and Information Systems*. 159–164.

[169] • R. Tadayon, A. Amresh, T. McDaniel, and S. Panchanathan. 2018. Real-time stealth intervention for motor learning using player flow-state. In *2018 IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH)*. 1–8. https://doi.org/10.1109/SeGAH.2018.8401360

[170] • Y. Tang, J. Liang, R. Hare, and F. Wang. 2020. A Personalized Learning System for Parallel Intelligent Education. *IEEE Transactions on Computational Social Systems* 7, 2 (2020), 352–361. https://doi.org/10.1109/TCSS.2020.2965198

[171] • Y. Tang and S. Shetty. 2011. Adaptive virtual reality game system for personalized problem-based learning. In *2011 International Conference on Networking, Sensing and Control*. 1–6. https://doi.org/10.1109/ICNSC.2011.5874957

[172] • Yu-Hui Tao, C. Rosa Yeh, and Kung Chin Hung. 2012. Effects of the heterogeneity of game complexity and user population in learning performance of business simulation games. *Computers & Education* 59, 4 (2012), 1350–1360. https://doi.org/10.1016/j.compedu.2012.06.003

[173] Henrik Thovtrup and Jakob Nielsen. 1991. Assessing the usability of a user interface standard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 335–341.

[174] Guy Tsafnat, Paul Glasziou, Miew Keen Choong, Adam Dunn, Filippo Galgani, and Enrico Coiera. 2014. Systematic review automation technologies. *Systematic reviews* 3, 1 (2014), 74.

[175] • Tzu-Wei Tsai, Hsiao Yu Lo, and Kai-Shao Chen. 2012. An Affective Computing Approach to Develop the Game-Based Adaptive Learning Material for the Elementary Students. In *Proceedings of the 2012 Joint International Conference on Human-Centered Computer Environments* (Aizu-Wakamatsu, Japan) *(HCCE '12)*. Association for Computing Machinery, New York, NY, USA, 8–13. https://doi.org/10.1145/2160749.2160752

[176] April Tyack and Elisa D. Mekler. 2020. Self-Determination Theory in HCI Games Research: Current Uses and Open Questions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–22. https://doi.org/10.1145/3313831.3376723

[177] • Andika Y. Utomo and Harry B. Santoso. 2015. Development of Gamification-Enriched Pedagogical Agent for e-Learning System Based on Community of Inquiry. In *Proceedings of the International HCI and UX Conference in Indonesia* (Bandung, Indonesia) *(CHIuXiD '15)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/2742032.2742033

[178] Giel van Lankveld, Pieter Spronck, H. Jaap van den Herik, and Matthias Rauterberg. 2010. Incongruity-Based Adaptive Game Balancing. In *Advances in Computer Games*, H. Jaap van den Herik and Pieter Spronck (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 208–220.

[179] • Rob van Roy and Bieke Zaman. 2018. Need-supporting gamification in education: An assessment of motivational effects over time. *Computers & Education* 127 (2018), 283–297. https://doi.org/10.1016/j.compedu.2018.08.018

[180] • Matti Vuorre and Janet Metcalfe. 2016. The relation between the sense of agency and the experience of flow. *Consciousness and Cognition* 43 (2016), 133–142. https://doi.org/10.1016/j.concog.2016.06.001

[181] • V. R. Warriar, J. R. Woodward, and L. Tokarchuk. 2019. Modelling Player Preferences in AR Mobile Games. In *2019 IEEE Conference on Games (CoG)*. 1–8. https://doi.org/10.1109/CIG.2019.8848082

[182] • R. M. Weerakoon, P. Chundi, and M. Subramaniam. 2011. Dynamically Adapting Training Systems Based on User Interactions. In *Proceedings of the 2011 Workshop on Knowledge Discovery, Modeling and Simulation* (San Diego, California, USA) *(KDMS '11)*. Association for Computing Machinery, New York, NY, USA, 31–38. https://doi.org/10.1145/2023568.2023578

[183] Russell B. Williams. 2018. Conceptual models and mental models in operation: Frustration, performance and flow with two different video game controllers. *Entertainment Computing* 28 (2018), 2 – 10. https://doi.org/10.1016/j.entcom.2018.07.004

[184] • Harry J. Witchel, Carlos P. Santos, James K. Ackah, Julian Tee, Nachiappan Chockalingam, and Carina E.I. Westling. 2016. The Complex Relationship Between Empathy, Engagement and Boredom. In *Proceedings of the European Conference on Cognitive Ergonomics* (Nottingham, United Kingdom) *(ECCE '16)*. Association for Computing Machinery, New York, NY, USA, Article 4, 4 pages. https://doi.org/10.1145/2970930.2970935

[185] Jacob O. Wobbrock and Julie A. Kientz. 2016. Research Contributions in Human-Computer Interaction. *Interactions* 23, 3 (April 2016), 38–44. https://doi.org/10.1145/2907069

[186] • S. Wu and T. Lin. 2011. Exploring the use of physiology in adaptive game design. In *2011 International Conference on Consumer Electronics, Communications and Networks (CECNet)*. 1280–1283.

[187] • Qi-Fan Yang, Shao-Chen Chang, Gwo-Jen Hwang, and Di Zou. 2020. Balancing cognitive complexity and gaming level: Effects of a cognitive complexity-based competition game on EFL students' English vocabulary learning performance, anxiety and behaviors. *Computers & Education* 148 (2020), 103808. https://doi.org/10.1016/j.compedu.2020.103808

[188] W. Yang, M. Rifqi, C. Marsala, and A. Pinna. 2018. Physiological-Based Emotion Detection and Recognition in a Video Game Context. In *2018 International Joint Conference on Neural Networks (IJCNN)*. 1–8.

[189] • Yu-chu Yeh, Guey-Jen Lai, Chun Fu Lin, Chung-Wei Lin, and Hua-Chun Sun. 2015. How stress influences creativity in game-based situations: Analysis of stress hormones, negative emotions, and working memory. *Computers & Education* 81 (2015), 143–153. https://doi.org/10.1016/j.compedu.2014.09.011

[190] • Chang Yun, Dvijesh Shastri, Ioannis Pavlidis, and Zhigang Deng. 2009. O' Game, Can You Feel My Frustration? Improving User's Gaming Experience via Stresscam. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 2195–2204. https://doi.org/10.1145/1518701.1519036

[191] Panagiotis Zaharias. 2009. Usability in the context of e-learning: A framework augmenting 'traditional'usability constructs with instructional design and motivation to learn. *International Journal of Technology and Human Interaction (IJTHI)* 5, 4 (2009), 37–59.

[192] Olaf Zawacki-Richter, Victoria I Marín, Melissa Bond, and Franziska Gouverneur. 2019. Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education* 16, 1 (2019), 39.

[193] • J. Zhao and Y. Li. 2020. Experimental Study on Gamification Teaching of Mathematics Classroom Practice Based on Flow Experience. In *2020 IEEE 2nd International Conference on Computer Science and Educational Informatization (CSEI)*. 36–40. https://doi.org/10.1109/CSEI50228.2020.9142485

[194] Yun Zhou, Tao Xu, Shaoqi Li, and Ruifeng Shi. 2019. Beyond engagement: an EEG-based methodology for assessing user's confusion in an educational game. *Universal Access in the Information Society* 18, 3 (2019), 551–563.

[195] F. Zwicky. 1967. The Morphological Approach to Discovery, Invention, Research and Construction. In *New Methods of Thought and Procedure*, Fritz Zwicky and Albert G. Wilson (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 273–297.

Received February 2021; revised June 2021; accepted July 2021