Affect-Targeted Interviews for Understanding Student Frustration

Ryan S. Baker[1], Nidhi Nasiar[1], Jaclyn L. Ocumpaugh[1], Stephen Hutt[1], Juliana M.A.L. Andres[1], Stefan Slater[1], Matthew Schofield[1], Allison Moore[2], Luc Paquette[3], Anabil Munshi[2], Gautam Biswas[2]

[1] Graduate School of Education, University of Pennsylvania
[2] Vanderbilt University
[3] University of Illinois at Urbana-Champaign
rybaker@upenn.edu

Abstract. Frustration is a natural part of learning in AIED systems but remains relatively poorly understood. In particular, it remains unclear how students' perceptions about the learning activity drive their experience of frustration and their subsequent choices during learning. In this paper, we adopt a mixed-methods approach, using automated detectors of affect to signal classroom researchers to interview a specific student at a specific time. We hand-code the interviews using grounded theory, then distill particularly common associations between interview codes and affective patterns. We find common patterns involving student perceptions of difficulty, system helpfulness, and strategic behavior, and study them in greater depth. We find, for instance, that the experience of difficulty produces shifts from engaged concentration to frustration that lead students to adopt a variety of problem-solving strategies. We conclude with thoughts on both how this can influence the future design of AIED systems, and the broader potential uses of data mining-driven interviews in AIED research and development.

Keywords: Frustration, Mixed methods, Affect detection, Attitudes, Self-regulated learning.

1 Introduction

Frustration is a natural part of learning, both in the context of AIED systems and more broadly, and yet it remains relatively poorly understood. Some articles have argued that frustration is a negative part of the learning experience, and should be eliminated [1, 2]. Other accounts have argued that frustration is necessary for an appropriate feeling of challenge and retention of knowledge over time (e.g. [3]). Indeed, the relationship between frustration and learning is unclear, with studies finding both negative associations [4, 5], and positive associations [6]. One study's results suggest that it is frustration's duration that matters for learning, not its overall rate of occurrence [7]. Theoretical accounts even disagree about whether frustration is properly understood as a single, discrete affective state, with arguments that there are multiple types of frustration -- some even pleasurable [8] -- or that frustration can be meaningfully split into whether it is germane or extraneous to the learning task [9]. By contrast, other researchers have

argued that confusion and frustration interact with learning in many of the same ways [7]. Hence, it is fair to say that the field of AIED -- and educational psychology more broadly -- is confused about frustration. Many of us even appear to be frustrated about frustration.

In particular, it is poorly understood how frustration interacts with the broader ongoing experience of participating in an AIED learning activity. We know that frustration precedes disengagement and tends to be relieved by disengaged behavior [10]. We know that frustration precedes help-seeking or on-task conversation with other students and can be relieved by those behaviors [11]. We know that frustration varies by learning activity [5] but what we do not know looms large. Researchers have argued that frustration is associated with the experience of difficulty [12, 13], but can we better understand how? How does frustration interact with a student's shifting perspective on whether a learning system is interesting or helpful (cf. [14])? And finally, there is some evidence that frustration is tied to self-regulated learning processes and learning strategy [15], but it is not yet fully clear how.

Although the majority of the past studies on frustration in AIED systems are quantitative, some of the attempts to more deeply understand frustration have leveraged qualitative or even introspective methods [8]. However, it has thus far been challenging to study frustration qualitatively, as out-of-context retrospective descriptions of a frustration experience may no longer have full access to the context or phenomenological experience that accompanies frustration (see review of the memory limitations surrounding retrospective interviews in Huber & Power [16]). Indeed, meta-analyses suggest that naturalistic frustration during learning is not always a particularly frequent or lengthy experience, D'Mello's [17] meta-analysis finds that frustration is rarer than any other commonly-studied affective state except surprise, and other research has shown that a typical occurrence of frustration lasts an average of 8-40 seconds [18, 19]. Thus, a randomly-timed set of interviews would not be expected to capture a particularly large proportion of frustration experiences. Spontaneous self-report in time diaries [20] can capture the context surrounding a specific experience of frustration, but have limited scope for follow-up questions and rely heavily on participant initiative. Artificially-induced frustration [21] may differ from genuine frustration in key fashions – for instance, the stimuli used to create frustration within this methodology may not be representative of the contexts where frustration naturally emerges.

To better study frustration, we adopt a novel mixed-methods approach, using affect detection to drive qualitative research. In this approach, a suite of automated affect detectors is integrated into a learning system. When an event of interest occurs – in this case, a transition from a different affective state to frustration, or a student experiencing sustained frustration over a significant period of time – a message is sent to a qualitative researcher present in the classroom, who can conduct an immediate, timely interview. This approach to mixed methods differs from the most common uses of mixed methods in education, which typically involve using both qualitative and more traditional quantitative methods (such as survey instruments or tests) to triangulate a research question, qualitative methods to explain quantitative findings, or using qualitative analysis to

identify behaviors for further quantitative study (e.g. [22]). Instead, we use a quantitative method – automated detection of affect – in support of a qualitative method – field interviews. As such, this method can increase the time-efficiency and cost-efficiency of using qualitative methods to study relatively rare events.

2 Methods

2.1 Betty's Brain

Betty's Brain is an open-ended, computer-based learning system that uses a learning-by-teaching paradigm to teach complex scientific processes [23]. Betty's Brain asks students to teach a virtual agent (Betty) about scientific phenomena (e.g., climate change, ecosystems, thermoregulation) by constructing concept maps that demonstrate the causal relationships involved (see Figure 1.)

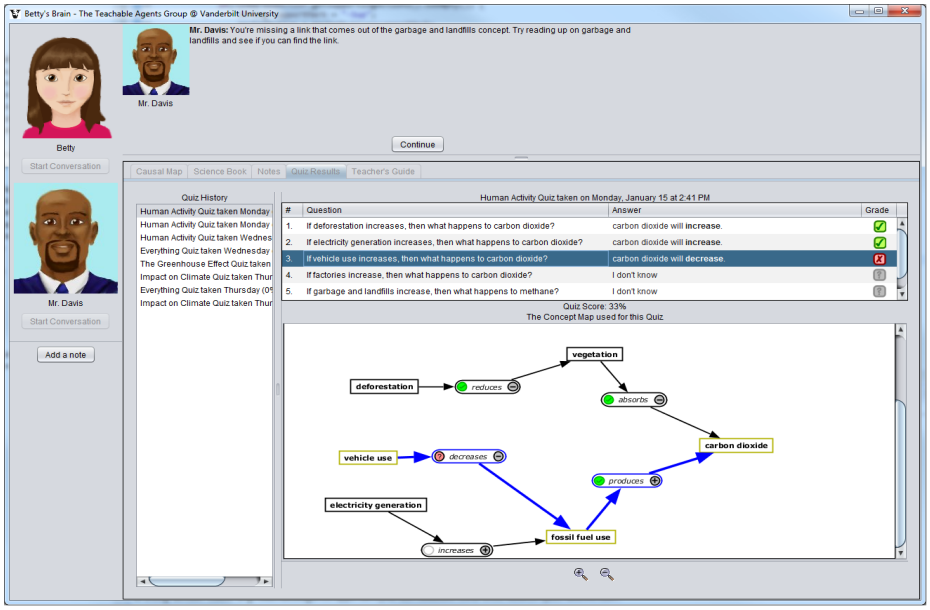

Fig.1. Screenshot of viewing quiz results and checking the chain of links Betty used to answer a quiz question

The learning process required by Betty's Brain necessitates high levels of self-regulation. As students construct their map, they must navigate through multiple hypermedia information sources where they can read about a variety of subthemes. They choose how often to test Betty's knowledge, and they may elect to interact with a virtual mentor agent (an experienced teacher named Mr. Davis) if they are having trouble teaching Betty. Because of these design factors, strong self-regulated learning behaviors are crucial for succeeding within Betty's Brain.

These pedagogical agents (Betty and Mr. Davis) provide a social framework for the gradual internalization of effective learning behaviors, and an emphasis on self-regulatory feedback that has been demonstrated to improve these behaviors among students who use Betty's Brain [23]. Prior research [24] has explored the relationships between students' cognitive and affective experiences in Betty's Brain and emphasized how automated affect detector models can be beneficial for providing students with personalized guidance that respond to their affective-cognitive states during learning.

2.2 Study Design

This study uses data from 93 sixth graders who used Betty's Brain during the 2016-2017 school year during their science classes in an urban public school in Tennessee. Data were collected over the course of seven school days. Students and their parents completed a consent form prior to the study. On the first day of the study, students completed a 30-45-minute paper-based pre-test that measured knowledge of scientific concepts and causal relationships. On day 2, students participated in a 30-minute training session that familiarized them with the learning goals and user interface of the software. Following the pre-test and training, students used the Betty's Brain software on days 2 through 6, for approximately 45-50 minutes each session, using concept maps to teach Betty about the causal relationships involved in the process of climate change. On day 7, students completed a post-test that was identical to the pre-test, in order to assess changes in knowledge based on working with Betty's Brain for the week.

As students interacted with Betty's Brain, automatic detectors of educationally relevant affective states [25] and behavioral sequences [24], already embedded in the software, identified key moments in the students' learning processes, either from specific affective patterns or theoretically aligned behavioral sequences. This detection was then used to prompt student interviews. The affect detection used logistic regression or step regression to recognize affect from behavior patterns, and was normed using classroom observations [25].

Interviewers were signaled through a field research app, Quick Red Fox (QRF), which integrates with Betty's Brain events and allows users to record metadata related to each event (in this case, timestamps and which student was being interviewed). A prioritization algorithm was used to select which student should be interviewed in instances where multiple students displayed interesting patterns at roughly the same time. In addition to prioritizing rarer affective sequences (e.g., sustained frustration), prioritization was also given to students who had not yet been interviewed (or who had not been recently interviewed). If interviewers were not comfortable interrupting a student, for any reason, they could skip the prompt within the app, and receive another recommendation from QRF.

Interviewers attempted to take a helpful but non-authoritative role when speaking with students. Interviews were open ended and occurred without a set script; however, they often asked students what their strategies were (if any) for getting through the

system. As new patterns and information emerged in these open-ended interviews, questions designed to elicit information about intrinsic interest (e.g., “What kinds of books do you like to read and why?” or “What do you want to be when you grow up?”) were added. Overall, however, students were encouraged to provide feedback about their experience with the software and talk about their choices as they used the software.

2.3 Interview Coding

A total of 358 interviews were conducted and recorded during this study. Audio files were collated and stored on a secure file management system available only to the research team members. Three members of the research team manually transcribed the interviews, having agreed upon formatting and style. Metadata, including timestamps and recording IDs, were preserved, but student-level information was de-identified (i.e., each student was assigned an alphanumeric identifier, used across data streams).

The code development process followed the recursive, iterative process used in [26] that includes seven stages: conceptualization of codes, generation of codes, refinement of the first coding system, generation of the first codebook, continued revision and feedback, coding implementation, and continued revision of the codes [26]. The conceptualization of codes included a review of related literature to capture meaningful experiences relevant to the study's research questions. Using grounded theory [27], a method that is appropriate for the kind of open-ended interviews where students are being asked to interpret their own experiences, we worked with the lead interviewer (the 3rd author) to identify categories that were (1) relevant to both affective theory (i.e. [28]) and self-regulated learning theory (e.g. [29]) and (2) likely to saliently emerge in the interviews. A draft lexicon and multiple criteria were generated for a coding system to help identify these constructs.

This coding scheme was iteratively refined, allowing us to identify relevant subcategories as they emerged from initial analyses, until the entire research team had reached a shared understanding of the criteria and constructs being examined. Following the production of a coding manual, external coders simultaneously coded with the 5th author to reach acceptable inter-rater reliability before coding all of the transcripts. All codes had Cohen’s kappa > 0.6, and the average Cohen’s kappa across codes was 0.80 -- see Table 1 for details. Throughout the coding, external coders met and clarified any concerns with the codebook authors to avoid misinterpretation or miscoding of the data. A total of 12 interview codes were developed from the interview data; however, we prioritized first coding for experiences involving difficulty (Diff), resource helpfulness (Help), interestingness (Int), and strategic use of resources (Strat) based on the perceived frequency of these experiences and their relevance to the affective experience of frustration. As these qualitative codes are not mutually exclusive, a single interview may be coded under multiple categories.

2.4 Affect Sequence Calculation

Once the interview data from Betty’s Brain was fully labeled, we calculated each

Table 1. The coding scheme used for the interviews.

Code	N	Description, Example
Difficult (Diff)	165	Negative evaluations, confusion, or frustration while interacting with the platform. Ex. "I am reading the science book again but I don't get it." $\kappa = .911$
Helpfulness (Help)	51	Utility of within-game resources in learning, improvement, and positive evaluations of the resources. Ex. "I like how you put in the dictionary all the things that could help you with the – this, 'cause I have no idea." $\kappa = .643$
Interestingness (Int)	11	Interestingness of within-game resources in learning and a continued desire to use the platform. Ex. "Everything I do [in Betty's Brain] interests me, you get one question right or everything right." $\kappa = .726$
Strategic Use (Strat)	205	Indicates a plan for interacting with the platform, notes changes in strategy or interaction with the platform based on experiences. Ex. " I'm just doing one section at a time...one section at a time that I tell Betty to take a test on it...and then I do it in the next section to see if she gets a 100 or if she gets one question wrong I go back and see." $\kappa = .911$

affective pattern's prevalence within each student's log files, looking not just at which patterns triggered a specific interview but all patterns present in the 80 seconds (four affective transitions) immediately before the interview. For each twenty-second period, we labeled it with the most likely (highest probability) affective state. Prior to comparing detector outputs to determine which affective state was most likely, the offset of each detector was mathematically adjusted so that the distribution of the predicted affective states matched the proportions of each affective state within the data originally used to develop the detectors. This step was taken to control for biases potentially introduced through the practice of re-sampling rarer classes, used in the original detector development [25].

In our analyses, we focus on three types of affect patterns that have been previously examined in [30]. Each involved a sequence of either three or four 20-second log-file clips. First, we looked at sequences that mirror the two cycles outlined by D'Mello & Graesser [28] the ENG-CON-DEL-ENG cycle (a student goes from engaged, to confused, to delighted, to engaged again) and the ENG-CON-FRU-BOR cycle (going from engaged, to confusion, to frustration, and boredom). For the purposes of this study, we have limited the analysis to 80 second (four-clip) versions of these cycles.

Next, we considered transitions between two states. For these analyses, we looked for a student having at least two consecutive clips with the same affective state predictions before transitioning to a second state (e.g., ENG-ENG-BOR or CON-CON-FRU). These durations allow us to explore the potential effect that a longer duration (two or more steps) of any given antecedent might have on the subsequent steps in a sequence.

Thus, we are able to explore the possibility that affective states of a longer duration (more than one successive step) might be influencing the results seen for sequences involving multiple transitions without testing all possible durations.

Finally, we consider sustained instances of two affective states that seemed to be driving the other patterns of statistical significance in this study. These are operationalized as 4-clip sequences (BOR-BOR-BOR-BOR and DEL-DEL-DEL-DEL), which we compare to sustained off-task behavior (OFF-OFF-OFF-OFF).

2.5 Identifying and Studying Relationships

Calculating the relationships between affect sequences and interview codes would ideally involve statistical significance testing but doing so is infeasible for two reasons. First, the number of affect sequences and interview codes being studied is sufficiently large that studying their combination would require a much larger data set than is feasible for interview data, for even a very liberal false discovery rate post-hoc control. This could be controlled for by selecting a much smaller number of affective sequences in advance. However, doing so would miss the opportunity to explore the space of affective sequences, a still incompletely-understood area. A second limitation, even stronger, is that many of the interview codes and affective sequences are rare within the data set, requiring even more data to be able to capture the relationships between them.

Instead, we look for the largest magnitudes of relationship, looking at the relative differences in frequency of an affective sequence when an interview code is present or absent. This provides a set of potentially interesting relationships to investigate in further detail. Having found the largest-magnitude relationships, we examine the transcripts of the interviews to understand the relationships better, presented below. Pseudonyms were assigned to participants using http://random-name-generator.info/ which generates names based on the frequencies within all U.S. census data, ignoring local community or subgroup variation, and ignoring the actual gender or age of the student.

3 Results

The top five strongest associations between affective sequences and specific interview codes were [helpful, sustained FRU], [helpful, BOR-->FRU], [difficult, ENG-->FRU], [interesting, BOR-->ENG], [strategic, ENG-->FRU]. Table 2 shows the magnitude of the relationships between these affective sequences and interview codes. In examining the interviews in detail, we were able to better understand many of these relationships.

Strategic: ENG->FRU and *Difficulty: ENG->FRU*

Many of the same interviews that immediately followed ENG-FRU affect transitions involved reports of both difficulty and strategic behavior.

Several students seemed to transition from engaged to frustrated when they

Table 2. The five strongest associations between affective sequences and interview codes

Interview Code	Affective Sequence	Pct Code when Affect	Pct Code when ~Affect	Pct Affect when Code	Pct Affect when ~Code	Relative Diff (by Code)
Helpful	SusFRU	66.7%	13.4%	4.5%	0.4%	12.4x
Helpful	BOR->FRU	50.0%	13.7%	2.3%	0.4%	6.2x
Difficult	ENG->FRU	83.3%	44.8%	3.5%	0.6%	6.0x
Interesting	BOR->ENG	12.5%	2.3%	12.5%	2.3%	5.5x
Strategic	ENG->FRU	83.3%	53.9%	2.9%	0.7%	4.2x

experienced difficulty and did not understand the system's feedback. For example, Gretchen said, "I change one thing and then I go back to the clues on those things and then I'm yeah. My head hurts… He keeps giving me zeros even though I on yesterday I got a percent when I know this there… I don't even know what correct is."

Gretchen responded to her frustration by trying different approaches, such as taking notes -- "I know it's very good to take his clues… Sometimes I had trouble with sea ice so I would go to sea ice now change and thing that was put in the ice and then you see that's not the thing and then you would have to come in for what he's...So whatever I do like I go back to…", to which the interviewer responded "Yeah that's that's one way to keep a note."

Willard adopted a different strategy: "So I'm focusing on one subject and taking a quiz. Sometimes I think I'm not going…", to which the interviewer responded "But now you've got that. Yeah but things that want to test it. OK. So, you'll work from this quiz." and Willard responded, "Yeah it's [inaudible], so whatever he doesn't you know that's wrong."

Another student, Shawn, also adopted a different strategy. Shawn expressed his frustration -- "I keep getting them wrong and I don't really know what I'm doing [inaudible] trying to learn a bit more about it" -- but adopted a strategic response to the situation: "I've been trying to fix my links."

Helpful: Sustained FRU

Two of the three cases where sustained frustration occurred had student reports of the system being helpful. These students experienced sustained frustration, but then experienced breakthroughs -- even eureka moments -- after that sustained frustration. The system's features were helpful, but the students' own strategies also played a role.

In one example, Jason tells the interviewer “I haven't done too well but it helped me know a lot of things I have wrong and I still have a few areas to [inaudible]. I think I'm making progress…”

Jason expresses his frustration -- matching the detector assessment -- “I feel like I kind of skimmed a little on the science book. And reading it again and it's like wait, I don't remember reading that and then I add it to the concept map… We have like I have something that like connects something that's not supposed to even though overall it should do that. Then it’s like messing me up… I suspect I'm still missing something, and I need to add something on.”

But later, Jason explains the strategy he is using, and how it helps the system help him: “I was having trouble finding what was wrong so then I tried to make more specific quizzes and... it's helped me understand more. Because I'm pretty sure I fixed this.” When prompted, by the interviewer, “Yeah? Are you using them more frequently now or?” Jason responds, “Yes.”

Helpful: BOR->FRU

Looking at this relationship shows the limitation of this method. In this case, the fairly large difference in relative magnitude came down to the very limited number of cases of BOR->FRU, only 2, one of which coincided with helpful. The resulting interview demonstrates boredom, frustration, and helpfulness, but it appears they may coincide due to the interviewer’s choice of questions rather than a genuine interrelationship.

Early in the interview, Shirley indicates frustration: “Because when I do shortcuts I get it wrong when I’m pretty sure it was right so I’m trying to fix this shortcut…” -- later she discusses her lack of interest in the topic -- in response to the interviewer’s question “Is it the sort of thing that you like to do generally?” the student responds “After school I usually read… I like to read fanfiction… Anime. A lot of anime.”

She discusses with the interviewer when Mr. Davis is helpful within the system, in response to a specific question on what is helpful. First, the interviewer asks, “So tutorial this morning help?”, to which the student responds “Yeah.” Then the interviewer asks “Yeah okay, is there anything else you figured out the last day or so help?”, and Shirley responds “I figured out that with my quiz result, if I get something wrong I let [Mr. Davis] try to figure out why I got it.” -- but these questions and responses are not connected to the immediate context of the interview.

4 Discussion and Conclusions

In this paper, we have used a novel multi-method approach to better understand the student perceptions surrounding the experience of frustration while learning from Betty’s Brain. In this approach, automated detectors were used to identify affective

transitions involving frustration (and sustained frustration) while using Betty's Brain, in real-time, and then field researchers conducted in-the-moment interviews with students experiencing those affective patterns. The interviews were coded for experiences of difficulty, perceptions of helpfulness, perceptions of interestingness, and use of strategic behaviors. We then distilled the top five sequences of affect that were most associated with a difference in the frequency of specific interview codes and analyzed cases where these affective sequences and interview codes co-occurred.

Through this analysis process, we found patterns that provided insights on the "why" and "what next" of frustration. Students who went from experiencing engaged concentration to frustration often reported both experiencing difficulty and using strategic behavior to resolve it. It may be possible to leverage this pattern, a productive response to experiencing difficulty, to better support students. These findings suggest that if a student goes from engaged to frustrated when encountering difficulty, but does not adopt a strategic behavior (which can be automatically detected as well [24, 31]), it may be appropriate for the learning system to offer recommendations of learning strategies. However, the best strategy may vary from case to case. Gretchen, Willard, and Shawn all adopted different learning strategies in response to the combination of frustration and difficulty. Jason's experience shows that the right system support can help resolve frustration -- even sustained frustration. Therefore, a learning system such as Betty's Brain may be able to use an approach such as reinforcement learning [32] to identify which strategy to recommend to which student, using the qualitative findings presented here to drive the design of learning strategy supports.

At the same time, Shirley's example -- where an affective state and interview code coincided due to the interviewer's choice of questions rather than a more useful overlap -- shows that there are still limitations to our method to be worked out. Another limitation is seen in our method's speed. Our method successfully focused interviewer time on key events of interest and facilitated the collection of interviews involving relatively rare events. However, the coding required afterwards was time-intensive, and is still ongoing for additional interview codes. It may be possible to improve the method -- to address both these limitations – by following interviews with an immediate end-of-day round of interview data coding, while the interview experience is still fresh in the field researchers' minds. This would also support the possibility of using this method not just for research, but for fast-paced iterative design.

Our next steps, therefore, are to use these findings to refine Betty's Brain. In that work, we will study the potential of Quick Red Fox -- with some procedural adjustments -- to enhance our process for rapid iterative design. At the same time, we will continue to study the rich data set we have obtained for further insights on student affect and perceptions. Overall, we believe these results demonstrate the potential of integrating data mining and qualitative research in new ways, facilitating the process of better understanding learners and improving learning experiences.

References

1. Grawemeyer, B., Wollenschlaeger, A., Santos, S. G., Holmes, W., Mavrikis, M., & Poulovassilis, A. Using Graph-based Modelling to explore changes in students' affective states during exploratory learning tasks. In Proceedings of the International Conference on Educational Data Mining, pp. 382-383. (2017).
2. DeFalco, J. A., Rowe, J. P., Paquette, L., Georgoulas-Sherry, V., Brawner, K., Mott, B. W., ... & Lester, J. C. Detecting and addressing frustration in a serious game for military training. International Journal of Artificial Intelligence in Education, 28(2), 152-193 (2018).
3. Sottilare, R., & Goldberg, B. Designing adaptive computer-based tutoring systems to accelerate learning and facilitate retention. Cognitive Technology, 17(1), 19-33 (2012).
4. Forbes-Riley, K., Rotaru, M., & Litman, D. The relative impact of student affect on performance models in a spoken dialogue tutoring system. User Modeling and User-Adapted Interaction, 18, 11-43 (2007).
5. D'Mello, S. K., Lehman, B., & Person, N. Monitoring affect states during effortful problem solving activities. International Journal of Artificial Intelligence in Education, 20(4), 361-389 (2010).
6. Pardos, Z.A., Baker, R.S., San Pedro, M.O.C.Z., Gowda, S.M., Gowda, S.M. Affective states and state tests: Investigating how affect and engagement during the school year predict end of year learning outcomes. Journal of Learning Analytics, 1 (1), 107-128 (2014).
7. Liu, Z., Pataranutaporn, V., Ocumpaugh, J., & Baker, R. Sequences of frustration and confusion, and learning. In Proceedings of the International Conference on Educational Data Mining (2013).
8. Gee, J.P. Good video games+ good learning: Collected essays on video games, learning, and literacy. Peter Lang Pub Incorporated, Bern, Switzerland (2007).
9. Richey, J. E., Andres-Bray, J. M. L., Mogessie, M., Scruggs, R., Andres, J. M., Star, J. R., ... & McLaren, B. M. More confusion and frustration, better learning: The impact of erroneous examples. Computers & Education, 139, 173-190 (2019).
10. Sabourin, J., Rowe, J. P., Mott, B. W., & Lester, J. C.. When off-task is on-task: The affective role of off-task behavior in narrative-centered learning environments. In Proceedings of the International Conference on Artificial Intelligence in Education (pp. 534-536). Springer, Berlin, Heidelberg (2011).
11. Baker, R. S., Moore, G. R., Wagner, A. Z., Kalka, J., Salvi, A., Karabinos, M., ... & Yaron, D. The dynamics between student affect and behavior occurring outside of educational software. In Proceedings of the International Conference on Affective Computing and Intelligent Interaction (pp. 14-24). (2011).
12. Valitutti, A. Action decomposition and frustration regulation in the assisted execution of difficult tasks. In Proceedings of the AIED 2009 Workshops, Brighton, UK (2009).
13. Miller, M. K., & Mandryk, R. L. Differentiating in-game frustration from at-game frustration using touch pressure. In Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces, pp. 225-234 (2016).
14. McCuaig, J., Pearlstein, M., & Judd, A. Detecting learner frustration: towards mainstream use cases. In Proceedings of the International Conference on Intelligent Tutoring Systems (pp. 21-30). Springer, Berlin, Heidelberg (2010).
15. Buono, S., Zdravkovic, A., Lazic, M., & Woodruff, E. The effect of emotions on self-regulated-learning (SRL) and story comprehension in emerging readers. In Frontiers in Education (Vol. 5, p. 218). (2020).
16. Huber, G. P., & Power, D. J. Retrospective reports of strategic-level managers: Guidelines for increasing their accuracy. Strategic management journal, 6(2), 171-180 (1985).

17. D'Mello, S. A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. Journal of Educational Psychology, 105(4), 1082 (2013).
18. D'Mello, S., & Graesser, A. The half-life of cognitive-affective states during complex learning. Cognition & Emotion, 25(7), 1299-1308 (2011).
19. Botelho, A.F., Baker, R., Ocumpaugh, J., Heffernan, N. Studying Affect Dynamics and Chronometry Using Sensor-Free Detectors. Proceedings of the 11th International Conference on Educational Data Mining, 157-166 (2018).
20. Lazar, J., Bessiere, K., Ceaparu, I., Robinson, J., & Shneiderman, B. Help! I'm lost: User frustration in web navigation. IT & Society, 1(3), 18-26 (2003).
21. Taylor, B., Dey, A., Siewiorek, D., & Smailagic, A. Using physiological sensors to detect levels of user frustration induced by system delays. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 517-528 (2015).
22. Canossa, A., Drachen, A., & Sørensen, J. R. M. Arrrgghh!!! blending quantitative and qualitative methods to detect player frustration. In Proceedings of the 6th international conference on foundations of digital games, pp. 61-68 (2011).
23. Leelawong, K., & Biswas, G. Designing learning by teaching agents: The Betty's Brain system. International Journal of Artificial Intelligence in Education, 18(3), 181-208 (2008).
24. Munshi, A., Rajendran, R., Ocumpaugh, J., Biswas, G., Baker, R., Paquette, L. Modeling Learners' Cognitive and Affective States to Scaffold SRL in Open-Ended Learning Environments. Proceedings of the 25th Conference on User Modeling, Adaptation, and Personalization, 131-138 (2018).
25. Jiang, Y., Bosch, N., Baker, R. S., Paquette, L., Ocumpaugh, J., Andres, J. M. A. L., ... & Biswas, G. Expert feature-engineering vs. deep neural networks: Which is better for sensor-free affect detection?. In Proceedings of the International Conference on Artificial Intelligence in Education (pp. 198-211). Springer, Cham (2018).
26. Weston, C., Gandell, T., Beauchamp, J., McAlpine, L., Wiseman, C., & Beauchamp, C. Analyzing interview data: The development and evolution of a coding system. Qualitative sociology, 24(3), 381-400 (2001).
27. Charmaz, K. The grounded theory method: An explication and interpretation. Contemporary field research, 109-126 (1983).
28. D'Mello, S., & Graesser, A. Dynamics of affective states during complex learning. Learning and Instruction, 22(2), 145-157 (2012).
29. Winne P. H., Hadwin A. F. Studying as self-regulated engagement in learning. Metacognition in Educational Theory and Practice. Hacker D., Dunlosky J., Graesser A. Hillsdale (Eds.), NJ: Erlbaum, 277–3048 (1998).
30. Andres, J.M.A.L., Ocumpaugh, J., Baker, R., Slater, S., Paquette, S., Jiang, Y., Bosch, N., Munshi, A., Moore, A., Biswas, G. Affect Sequences and Learning in Betty's Brain. Proceedings of the 9th International Learning Analytics and Knowledge Conference, 383-390 (2019).
31. Azevedo, R., & Gašević, D. Analyzing multimodal multichannel data about self-regulated learning with advanced learning technologies: Issues and challenges. Computers in Human Behavior, 96, 207-210 (2019).
32. Chi, M., VanLehn, K., Litman, D., & Jordan, P. Empirically evaluating the application of reinforcement learning to the induction of effective and adaptive pedagogical strategies. User Modeling and User-Adapted Interaction, 21(1), 137-180 (2011).