What do Log-Files and Learning Outcomes Reveal about Developmental Differences in Self-Regulated Learning with Serious Games?

Saerok Park, University of Central Florida, saerok.park@knights.ucf.edu
Megan Wiedbusch, University of Central Florida, megan.wiedbusch@ucf.edu
Roger Azevedo, University of Central Florida, roger.azevedo@ucf.edu

Abstract: To explore self-regulated learning (SRL) processes with log-files and learning outcome data across developmental levels in game-based learning environments (GBLEs), 26 high school and 26 undergraduates learned microbiology playing Crystal Island as their test scores and log-file data about scientific reasoning activities were collected and analyzed. Results show that undergraduates were more likely to solve a mystery than high school students despite no significant differences in test scores. The frequency of SRL for undergraduates was greater than the frequency for high school students, but the proportional duration of SRL within total duration showed that both groups utilized different strategies. Moreover, developmental levels were related to how likely students were to solve the mystery. Our findings emphasize not only an important theoretical contribution by demonstrating how SRL models should take developmental variations into account, but also educational implications for GBLEs by showing the advantages of providing scaffoldings across developmental levels.

Introduction and theoretical background

Learners often struggle to use effective cognitive, affective, metacognitive, and motivational self-regulated learning (SRL) processes while learning (Winne & Azevedo, 2022). To help learners, educational games have been developed to foster effective SRL (Cloude et al., 2022; Dever et al., 2020; Taub et al., 2020). Beyond understanding the increase of motivation and engagement game-based learning environments (GBLEs) provide, it is critical to investigate our assessment of student learning and the effectiveness of their learning processes as they relate to the contents of GBLEs. Learners may acquire knowledge from GBLEs by interacting with game system elements (e.g., non-player characters, instructional materials). However, these elements may be designed to help learners to enjoy the game, rather than leverage the elements for SRL during GBLEs (Taub et al., 2020). To evaluate the degree of learning in GBLEs, diverse methods can be utilized including game scores, external assessments, and embedded assessments. For example, game scores may reflect if students acquired targets or overcome obstacles during game while external assessments may be post-test scores based on multiple choice questions (Ifenthaler et al., 2012). However, despite efforts to establish best practices for evaluating the effectiveness of GBLEs, there is still a lack of scientific rigor (All et al., 2021). Since assessment after learning in GBLEs which frequently concentrates on outcome and not process may neglect significant changes during the learning process (All et al., 2021), evaluating GBLEs with only post assessment might not be sufficient.

To deal with this issue, trace data can be utilized since trace data can uncover fundamental learner interactions for learning design, identify specific patterns or strategies learners use, and determine key predictors of specific behaviors (Azevedo et al., 2018; Owen & Baker, 2020). Specifically, trace data in GBLEs have the potential to dynamically evaluate learning and provide feedback in ways that are not achievable in traditional learning settings (Nietfeld, 2018). For instance, diverse data could be collected from learner's interactions with some elements in GBLEs such as time spent reading scientific text and collecting evidence, the number of times interacting with non-player characters, the quality of hypotheses generated when collecting and testing data, etc. (Dever et al., 2020). Utilizing various types of data collected with GBLEs can lead to better insight or understanding of learning processes in broader and deeper ways, especially when it comes to identifying the underlying cognitive and metacognitive self-regulatory processes (Taub et al., 2017).

Even though many previous studies have pointed out the effectiveness of GBLEs, there are few studies focusing on learners' SRL from their developmental perspectives (Plass et al., 2020). In addition to this, research related to SRL tends to concentrate primarily on younger children and academic achievement (Bjork et al., 2013), leaving open the question whether different developmental groups would perform tasks with the same SRL processes. Through the previous studies presenting that learning outcomes and learners' metacognitive skills were varied across learners' developmental levels (Mayer, 2019; Veenman et al., 2004), it can be expected that learners' SRL processes and learning outcomes are different based on learners' developmental levels in GBLEs. For example, previous research has shown that undergraduate students typically know and can use more learning strategies and fair better when it comes to the metacognitive monitoring accuracy (Taub et al, 2020), but there has been no comparison across developmental levels and their use of self-regulatory skills with serious games. More studies related to students' developmental levels on SRL with GBLEs are needed to understand differences in self-regulatory behaviors and to subsequently use the data to provide more adaptive scaffolds in order to meet learners' individual needs based on their developmental levels. Therefore, the purpose of this study is to enrich the understanding of SRL based on developmental levels by using log-files and learning outcome data. Our study focuses on whether there are developmental differences in students' learning outcomes with two indicators and SRL processes with log-file data in GBLEs.

Theoretical framework

SRL assumes learners monitor and regulate their cognitive, affective, metacognitive, motivational, and social processes (Azevedo et al., 2022; Winne, 2018). Specifically, students can engage in a variety of cognitive, affective, metacognitive, and motivational self-regulatory and reflective processes during GBLEs to ensure they are learning efficiently (Taub et al., 2020). Game-based learning concentrates on the complexities of game design and requires a theoretical framework focused on specific learning processes that are used during SRL (Plass et al., 2020). As such, this study focuses on SRL suggested by Winne and Hadwin's (1998, 2008) model of SRL. According to this model, learning occurs throughout a series of four phases: 1) task understanding, 2) setting goals and planning, 3) engaging in learning strategies, and 4) making adaptations. Each phase which is cyclical and may occur simultaneously allows learners to engage in different self-regulatory processes. While this theory makes assumptions about underlying cognitive and metacognitive processes, it does not make predictions about GBLEs across developmental levels.

Current study

This study aims to further understand developmental differences with self-regulated learning using log-files and learning outcome data during game-based learning. As such, we ask:

RQ 1) Are there differences in learning outcomes (test scores vs. solving mystery) based on learners' developmental levels with Crystal Island? Based on previous studies showing that learning outcomes were different across developmental levels in traditional learning environments (Mayer, 2019; Veenman et al., 2004), we expect that undergraduates will perform better on post-test and solving a mystery during the game compared to high school students.

RQ 2) Are there differences in the frequencies of scientific reasoning activities based on learner's developmental levels during Crystal Island? Considering that learners metacognitive skill increases with age (Veenman et al., 2004) and that undergraduate students can know and utilize more learning strategies (Taub et al., 2020), we expect undergraduate students more often use scientific reasoning activities while learning with Crystal Island.

RQ 3) Are there differences in the proportional duration (compared to total play time) of scientific reasoning activities based on learners' developmental levels during learning with Crystal Island? Considering that learners metacognitive skill increases with age (Veenman et al., 2004) and that undergraduate students can know and utilize more learning strategies (Taub et al., 2020), we expect undergraduate students will spend higher proportion of duration within total game play duration engaging in scientific reasoning activities while learning with Crystal Island.

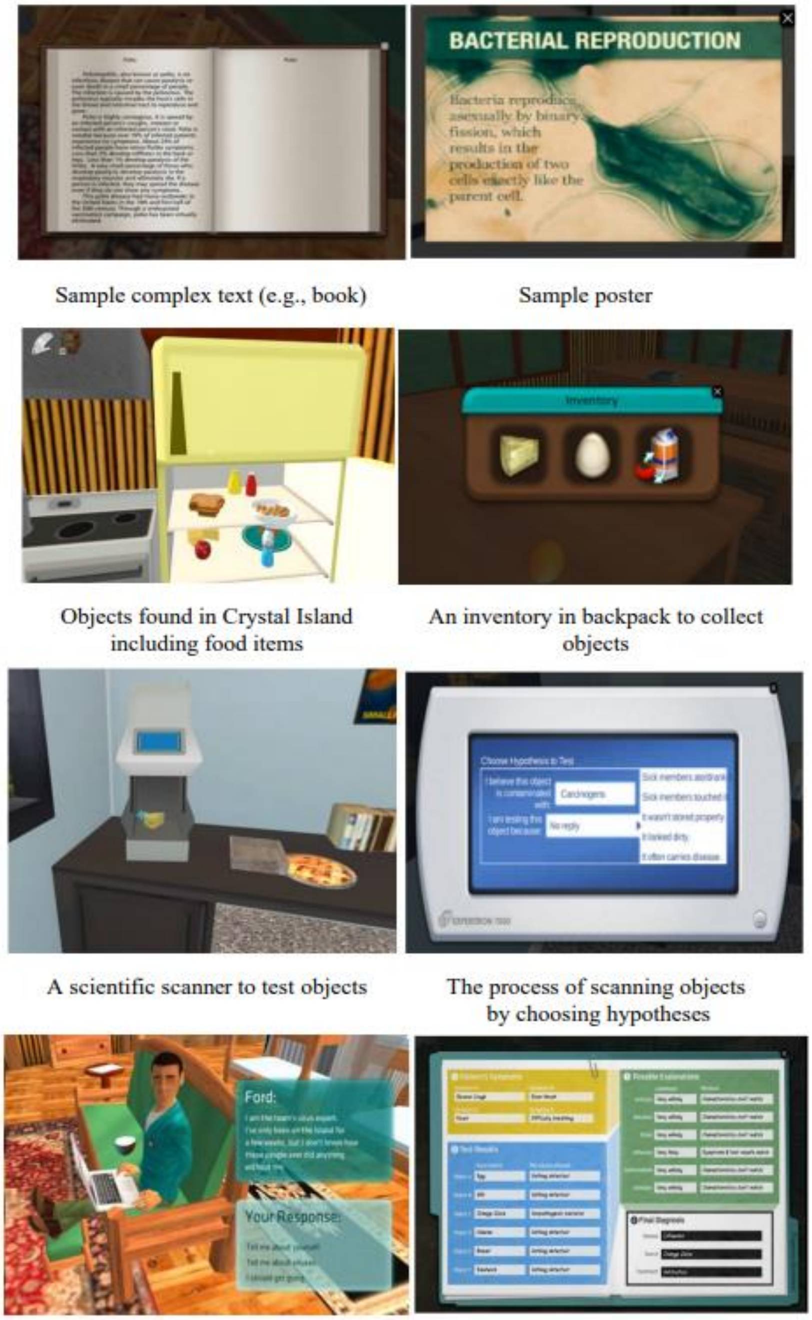
RQ 4) What is the likelihood that a learner's ability to solve a mystery is related to learner's developmental level during learning with Crystal Island? Prior studies show that learning outcomes are varied across developmental levels (Mayer, 2019; Veenman et al., 2004) and that the correlation between students' use of cognitive strategies and academic performance became significantly stronger across developmental levels (Dent & Koenka, 2016). Therefore, we expect that the likelihood of a learner solving a mystery will be significantly related to developmental levels.

Method

Research context: Crystal island

Crystal Island is a game-based learning environment designed to teach microbiology and scientific inquiry skills. During the game, participants play the role of a medical detective investigating a mysterious infectious disease outbreak on a remote island. To solve this mystery, participants are required to identify a type of disease (e.g., influenza or salmonellosis), a type of infection (e.g., viral or bacterial), a cause of the infectious disease (e.g., egg, bread, etc.), and a solution treatment for the disease (e.g., rest or vaccination). To do so (see Figure 1), participants read complex texts (e.g., articles and books) and posters and then summarize the texts, as they gather important clues. These clues can be collected by finding objects to put into their backpacks while exploring the island, by scanning the objects with hypotheses that they create, and by communicating with non-player characters (NPCs) (e.g., a nurse, patients, a virus expert, etc.) through dialog selection. Based on this evidence, they complete a diagnosis worksheet to solve a mystery. Thus, SRL is fundamental in Crystal Island to foster content knowledge in microbiology and scientific reasoning abilities (Dever et al., 2020).

Figure 1
Activities in Crystal Island to support scientific reasoning and learning about microbiology



Sample complex text (e.g., book)

Sample poster

Objects found in Crystal Island including food items

An inventory in backpack to collect objects

A scientific scanner to test objects

The process of scanning objects by choosing hypotheses

Communication with non-player characters (NPCs) through dialog selection

Diagnosis worksheet to be used throughout the learning session to record data, findings, hypotheses, etc.

Data collection and analysis

Data were collected from 26 high school students (*M*_age=16.04; *SD*_age=0.34; 65% female) and 26 undergraduate students (*M*_age=19.73; *SD*_age=1.49; 69% female). The data for the high school students were collected in a classroom setting while the data on the undergraduates were collected from individual students in a university research lab. On average, high school students spent 48 minutes (*SD*=6.96) and undergraduate students spent 87 minutes (*SD*=22.50) with the game.

This study focuses on two ways to assess learning outcomes in GBLEs (game scoring and external assessment) distinguished by Ifenthaler et al. (2012). We operationalize game scoring by determining if participants solved a mystery while playing Crystal Island. External assessment was operationalized as performance on the post-test about knowledge of microbiology while accounting for pre-test knowledge. Pre-test scores and post-test scores were calculated as normalized change score to alleviate pre-test score biases (Marx & Cummings, 2007).

Log-file data collected during game play was extracted and several activities were analyzed to compare across development levels. Through these data, we identified how many times students interacted with scientific reasoning activities in the GBLE (e.g., opening articles, books, posters, scanning objects, opening backpacks to gather clues, communicating non-player characters (NPCs) through dialog selection, and opening a diagnosis worksheet). Also, we analyzed the proportion of the duration that students interacted with the scientific reasoning within their total amount of the game play.

Results

Learning outcomes

A *t*-test using participants' normalized change scores did not show any significant differences (t=-1.92, p=0.061) between high school students (*M*=0.12, *SD*=0.23) and undergraduate students (*M*=0.23, *SD*=0.22) (see Table 1). Despite no significant difference in pre-test scores, there was a mean difference in normalized changes scores for each group, but the huge variabilities in both groups lead to no statistically significant difference between developmental levels. That is, both developmental levels showed evidence of learning about the same with Crystal Island.

When it comes to a difference in solving the mystery based on developmental levels, a 2X2 chi-square test revealed a significant difference in the distribution of students who solved the mystery correctly across development levels ($\chi 2$=16.65, p<.01) (see Table 2). More specifically, 24 undergraduates (92%) solved the mystery compared to only 10 high school students (38%). This suggests that outcomes or performance-based goals were different, but microbiology content knowledge acquisition was not.

Table 1
Pre-test and normalized change scores across developmental levels

	High school students *M(SD)*	Undergraduates *M(SD)*	*t*	*p*
Prior knowledge (pre-test scores)	10.15 (48.33%) (2.09 (9.95%))	11.15 (53.10%) (2.44 (11.62%))	-1.59	.119
Normalized change score (pre-test and post-test scores)	0.12 (0.23)	0.23 (0.22)	-1.92	.061

Note. *p<0.05

Table 2
Frequency of participants who solved the mystery

	High school students (*N*=26)	Undergraduates (*N*=26)
Mystery solved	10 (38.46%)	24 (92.30%)
Mystery unsolved	16 (61.54%)	2 (7.7%)

Frequency of scientific reasoning activities

A *t*-test was calculated using a Bonferroni correction for multiple tests ($p<0.05/6 = 0.008$) and showed that undergraduates had significantly greater frequencies of opening complex scientific texts (*M*=23.00, *SD*=6.79; *t*=-4.90, *p*<.008), scientific posters (*M*=13.88, *SD*=5.48; *t*=-4.79, *p*<.008), scanning evidence with a scientific tool (*M*=26.85, *SD*=15.62; *t*=-5.81, *p*<.008), opening backpack with objects (*M*=100.31, *SD*=38.03; *t*=-6.33, *p*<.008), communicating with NPCs through dialog selection (*M*=59.65, *SD*=18.87; *t*=-3.52, *p*<.008), and opening diagnosis worksheet (*M*=29.81, *SD*=15.92; *t*=-6.29, *p*<.008) compared to high school students (see Table 3). Overall, frequently using these game features related to scientific reasoning by undergraduates is indicative of their high use of SRL strategies compared to high school students.

Table 3
Summary statistics in frequency of scientific reasoning activities

Activities	High school students *M(SD)*	Undergraduates *M(SD)*	*t*	*p*
Complex text (Books and articles)	15.12 (4.59)	23.00 (6.79)	-4.90	.000*
Poster	7.46 (4.10)	13.88 (5.48)	-4.79	.000*
Scanning	7.88 (5.74)	26.85 (15.62)	-5.81	.000*
Backpack open	47.69 (18.79)	100.31 (38.03)	-6.33	.000*
Dialog selection	44.46 (11.31)	59.65 (18.87)	-3.52	.001*
Diagnosis worksheet	9.00 (5.59)	29.81 (15.92)	-6.29	.000*

Note. *$p<0.008$ after Bonferroni correction for multiple post-hoc comparisons

Duration of scientific reasoning activities proportionate to total game play time

High school students played the game for an average of 48 minutes (*SD*=6.96) while undergraduate students played for an average of 87 minutes (*SD*=22.50). A t-test was calculated using a Bonferroni correction for multiple tests ($p<0.05/4 = 0.012$) in order to compare the two groups' time spent on each scientific reasoning activity within their total game play time. Considering each proportion of activities within total game play duration (see Table 4), high school students spent a higher percentage of their time reading complex text such as articles and books (*M*=43.77, *SD*=11.91; *t*=2.88, *p*<.012) during the game compared to undergraduates. This can help explain our earlier finding such that it appears high school students were focused more on the concrete knowledge acquisition through reading and information gathering while undergraduates were focused more on solving the mystery and potentially using scientific reasoning skills. There was no significant difference between high school students and undergraduates in the proportions of poster (*p*=.165) and diagnosis worksheet (*p*=.020). However, undergraduates had a significantly greater proportion of durations on scanning objects (*M*=2.33, *SD*=1.10; *t*=-3.70, *p*<.012) compared to high school students while they played. These results show that undergraduate students spent a considerable amount of time within the total time on gathering clues by scanning objects to test hypotheses for solving the mystery rather than reading complex text or poster while playing games while they learned in GBLE.

Table 4
Summary statistics in proportion of activities within total game play duration

Activities	High school students *M(SD)*	Undergraduates *M(SD)*	*t*	*p*
Complex text (Books and articles)	43.77 (11.91)	34.95 (10.05)	2.88	.006*
Poster	1.06 (0.63)	1.29 (0.54)	-1.41	.165
Scanning	1.30 (0.92)	2.33 (1.10)	-3.70	.001*
Diagnosis worksheet	6.45 (3.29)	8.90 (4.02)	-2.41	.020

Note. *$p<0.012$ after Bonferroni correction for multiple post-hoc comparisons

Predicting solving the mystery based on learners' developmental level

A binomial logistic regression analysis was conducted to determine the likelihood that a learner solves the mystery illness given the learners' developmental levels with Crystal Island (see Table 5). The likelihood of a learner solving the mystery was significantly related to learners' developmental levels (p<0.025). In other words, the higher learner's developmental level was, the more likely the learner was to solve the mystery. Specifically, the odds of a learner solving the mystery were 0.05 times greater if the learner was an undergraduate student.

Table 5
Logistic regression analysis of solving mystery by developmental levels

Predictor	β	*SE*	*Wald*	*df*	*p*	e^{β}
Constant	2.49	0.74	11.40	1	0.001	12.00
Developmental levels	-2.96*	0.84	12.40	1	0.000	0.05

Note. e^{β}= exponentiated beta or odds ratio; *SE*=standard error; **p*<0.025

Discussion and conclusion

Despite the panacea of using GBLEs to solve STEM learning issues, we argue that these environments do not always lead to significant learning outcomes because students have different self-regulatory skills. Secondly, GBLEs designed on the principles of agency or without instructional scaffoldings need to be examined especially when it comes to understanding how students of different ages use and learn from them. Results found that undergraduate students were more likely to solve a mystery than high school students with Crystal Island although there was no significant difference in learning gains measured by test scores. The fact that both age groups did equally well on the post-test learning outcomes could reflect the fact that high school students are not that much younger than the undergraduates. High school students are currently learning about biology while it may have been several years since undergraduates learned about biology. Furthermore, high school students may have spent just enough time while undergraduates spent significantly more time (to activate their prior knowledge) while learning with Crystal Island. The significant difference in solving the mystery in Crystal Island can be understood through investigation of learner's SRL processes with previous studies (Nietfeld et al, 2014; Sabourin et al, 2013) showing that students' use of SRL was different based on their learning gains or performances. For instance, Nietfeld et al. (2014) found that the use of SRL strategies was positively related to performance in Crystal Island by revealing that learners using more SRL strategies had higher game-score performance.

Furthermore, there were significant differences in the use of SRL strategies depending on developmental levels. In terms of frequency, undergraduates not only opened books, articles, posters, or diagnosis worksheets more often than high school students, but also scanned food items by opening backpack more often than high school students. Also, the undergraduates spent more time communicating with NPCs through dialog selection than high school students. Considering the higher frequency of dialog selection compared to the frequencies of scanning and opening backpack for both groups, we can assume that high school students did not have huge difficulties communicating with non-player characters (NPCs) while they played compared to university students. This can be explained by the fact that NPCs may be easier to find than other clues such as food items which students need to explore carefully to find on the island. For example, some specific items (e.g., cheese) which give students important clues for the mystery might be hard to detect for the students since they only can find certain food items when they open a refrigerator during the game. They then must develop a hypothesis about that food item (usually through talking with NPCs) and then scan the items that may be important. This difficulty to utilize game elements for high school students is also supported by the frequencies of scanning and backpack open. Compared to university students, high school students showed significantly lower frequency in scanning and opening backpack. This shows that high school students might have difficulties finding clues while playing or did not make it as much of a priority as reading materials about microbiology. To scan food items to formulate hypotheses in Crystal Island, students need to find and put food items into their backpack or change the items, which means that the more they find the food items, the more they can open backpack and scan. These features led to a low possibility to solve the mystery, presenting that learners' use of SRL strategies with GBLEs is varied based on their developmental levels.

The difference in frequency of activities might not be surprising since undergraduates spent more overall time playing Crystal Island than high school students. This is also true that undergraduates spent greater durations on all elements related to scientific reasoning while they played Crystal Island compared to high school students. In addition to this, the fact that there was no significant difference between normalized change scores showed that high school students performed just as well as on the post-test with less time compared to undergraduates. However, considering the proportion of each activity within total game play duration, the results could provide a different new perspective in the students' SRL abilities. Although high school students showed a greater proportion of durations on reading articles and books within their total game play duration compared to university students, they were less likely to solve the mystery. Unlike high school students, college students spent a significant proportion of time gathering clues and formulating their hypotheses through scanning objects, which led to a higher possibility to solve the mystery. This shows that high school students might lack the same level of SRL abilities such as planning or monitoring the difficulties of utilizing game elements while learning in GBLEs. This is consistent with the previous study showing that learner's metacognitive skillfulness increases with age (Veenman et al., 2004). However, we also acknowledge that the main goal of these two groups could be driving their action differences. That is, if high school students were more focused on learning as much as they could about microbiology (knowing there is a test at the end of the game), they would not be as focused on the game goal of solving the mystery. In comparison, the undergraduates appear to be more focused on solving the mystery as shown by the larger duration time spent doing other activities (e.g., scanning objects with a scientific tool to test hypotheses). Both goals were presented to the learners, however, our results suggest that each developmental group focused on a different aspect of these two goals. Task-wise learning about microbiology is simpler and more straightforward while developing non-explicit scientific reasoning skills through exploratory learning is more abstract and requires more metacognitive skills and abilities.

According to the binomial logistic analysis, older learners (i.e., undergraduates) are more likely to solve the mystery than high school students. This is supported by the results of the previous studies indicating that learning outcomes were different for different age groups (Mayer, 2014; Veenman et al., 2004). Specifically, Mayer (2014) presented that GBLEs influenced learning outcomes differentially across learner's developmental levels with different effect sizes; the effect size for college students was higher than secondary school students.

An important theoretical implication of our work is a general suggestion for models to incorporate the role of developmental differences in the assumptions of monitoring and strategy use (e.g., Winne's [2018] COPES and SMART schemas should have different assumptions based on learner's developmental levels). Our work provides empirical evidence to further support the idea that even when close in age, SRL processes may have developmental differences that impact one's efficiency of learning with GBLEs. Moreover, our findings have implications for designing various types of scaffolds in GBLEs, depending on developmental levels. Since the effects of learning at the age groups can be different through theoretical background and instructional strategies (Dignath & Büttner, 2008), utilizing differentiated scaffolds based on developmental levels in GBLEs can lead to more effective learning. For instance, high school students may benefit from procedural scaffolding to help them to fully explore learning elements in GBLEs or to foster increases in the frequency and duration of specific activities which might be useful for learning and scientific reasoning. In contrast, undergraduates may benefit from more conceptual and metacognitive scaffolds given their higher frequency and duration of activities (Lim et al., 2023).

Despite these results worth noting related to developmental differences of SRL in GBLEs, this study still has several limitations for future studies. This study focused on two different age groups related to developmental differences with the small sample size. In order to understand specific SRL patterns from various age groups with developmental differences, more studies comparing the diverse groups with developmental differences are needed. Furthermore, while log-file data were effective in measuring learning processes in this study, a deeper understanding of developmental differences of SRL in GBLEs can be strengthened with the consideration of more diverse trace data (e.g., eye-tracking, facial expressions of emotions, think-aloud protocols, etc.). Along with the rich trace data, a mixed methods approach and semi-structured interviews can be supplemented to lead to a richer and better-informed explanation of learners' SRL and decision making within GBLEs. Future research should compare multimodal multichannel data to provide a comprehensive understanding of the temporally-based developmental differences across age groups and design intelligent, adaptive scaffolding in GBLEs for STEM.

References

All, A., Castellar, E. N. P., & Van Looy, J. (2021). Digital Game-Based Learning effectiveness assessment: Reflections on study design. *Computers & Education, 167*, 104160.

Azevedo, R., Bouchet, F., Duffy, M., Harley, J., Taub, M., Trevors, G., Cloude, E., Dever, D., Wiedbusch, M., Wortha, F., & Cerezo, R. (2022). Lessons learned and future directions of metatutor: leveraging multichannel data to scaffold self-regulated learning with an intelligent tutoring system. *Frontiers in Psychology, 13*. 813632. https://doi.org/10.3389/fpsyg.2022.813632

Azevedo, R., Taub, M., & Mudrick, N.V. (2018). Using multi-channel trace data to infer and foster self-regulated learning between humans and advanced learning technologies. In D. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 254-270). Routledge.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417-444. https://doi.org/10.1146/annurev-psych-113011-143823

Cloude, E., Dever, D., Hahs-Vaughn, D., Emerson, A., Azevedo, R., & Lester, J. (2022). Affective dynamics and cognition during game-based learning. *IEEE Transactions on Affective Computing*, *13*(4), 1705-1717. https://doi.org/10.1109/TAFFC.2022.3210755

Dent, A. L., & Koenka, A. C. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis. *Educational Psychology Review, 28*, 425-474. https://doi.org/10.1007/s10648-015-9320-8

Dever, D. A., Azevedo, R., Cloude, E. B., & Wiedbusch, M. (2020). The impact of autonomy and types of informational text presentations in game-based environments on learning: Converging multi-channel processes data and learning outcomes. *International Journal of Artificial Intelligence in Education*, *30*, 581-615. https://doi.org/10.1007/s40593-020-00215-1

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, *3*, 231-264. https://doi.org/10.1007/s11409-008-9029-x

Ifenthaler, D., Eseryel, D., & Ge, X. (2012). Assessment for game-based learning. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in game-based learning: Foundations, innovations, and perspectives* (pp. 3-10). Springer.

Lim, L., Bannert, M., van der Graaf, J., Singh, S., Fan, Y., Surendrannair, S., ... & Gašević, D. (2023). Effects of real-time analytics-based personalized scaffolds on students' self-regulated learning. *Computers in Human Behavior, 139*, 107547. https://doi.org/10.1016/j.chb.2022.107547

Mayer, R. E. (2019). Computer games in education. *Annual review of psychology, 70*, 531-549. https://doi.org/10.1146/annurev-psych-010418-102744

Marx, J. D., & Cummings, K. (2007). Normalized change. *American Journal of Physics, 75*, 87–91. https://doi.org/10.1119/1.2372468

Nietfeld, J. L., Shores, L. R., & Hoffmann, K. F. (2014). Self-regulation and gender within a game-based learning environment. *Journal of Educational Psychology, 106*, 961–973. https://doi.org/10.1037/a0037116

Nietfeld, J.L. (2018). The role of self-regulated learning in digital games. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 271-284). Routledge.

Owen, V.E., & Baker, R.S. (2020). Learning analytics for games. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 513-536). MIT Press.

Plass, J. L., Mayer, R. E., & Homer, B. D. (2020) (Eds.). *Handbook of game-based learning*. MIT Press.

Sabourin, J. L., Shores, L. R., Mott, B. W., & Lester, J. C. (2013). Understanding and predicting student self-regulated learning strategies in game-based learning environments. *International Journal of Artificial Intelligence in Education, 23*, 94–114. https://doi.org/10.1007/s40593-013-0004-6

Taub, M., Azevedo, R., Bradbury, A. E., & Mudrick, N. V. (2020). Self-regulation and reflection during game-based learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 239-262). MIT Press.

Taub, M., Mudrick, N. V., Azevedo, R., Millar, G. C., Rowe, J., & Lester, J. (2017). Using multi-channel data with multi-level modeling to assess in-game performance during gameplay with Crystal Island. *Computers in Human Behavior, 76*, 641-655. http://dx.doi.org/10.1016/j.chb.2017.01.038

Veenman, M. V., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and instruction*, *14*, 89-109.

Winne, P. H. (2018). Cognition and metacognition within self-regulated learning. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 36-48). Routledge.

Winne, P.H., & Azevedo, R. (2022). Metacognition. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences (*pp. 93-113*)*. Cambridge University Press.

Winne, P., & Hadwin, A. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 227–304). Routledge.

Winne, P., & Hadwin, A. (2008). The weave of motivation and self-regulated learning. In D. Schunk & B. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297–314). Taylor & Francis.

