The Role of Feedback Type and Task Performance on Concurrent Emotions and Interest During Game-Based Learning

Elizabeth B. Cloude[1(✉)], Eva Kormann[2], Marco Steiner[2], Antero Lindstedt[1], Kristian Kiili[1], and Manuel Ninaus[2]

[1] Faculty of Education and Culture, Tampere University, 33100 Tampere, Finland
{elizabeth.cloude,antero.lindstedt,kristian.kiili}@tuni.fi
[2] Department of Psychology, University of Graz, Graz, Austria
{e.kormann,marc.steiner,manuel.ninaus}@edu.uni-graz.at

Abstract. Building adaptive game-based learning (GBL) interventions (e.g., immediate feedback) has been a recent effort to maximize learning effectiveness. State-of-the-art algorithms often overlook the concurrent emotions and motivation of individuals and its impact on GBL interventions. This pilot study utilized a 3 (feedback type: results, elaborative, attribution) x 2 (graph type: misleading, non-misleading) within-subjects design with MediaWatch, a GBL environment built to improve critical graph literacy. At an Austrian university, 41 students' concurrent emotions and motivation were measured using validated surveys immediately after different types of feedback on tasks during GBL. Results showed a significant improvement in graph literacy after GBL. Different types of feedback and task performance influenced concurrent emotions and interest, but individual differences accounted for the largest variability explained in emotions and interest. The findings suggest that within-subject variability is crucial for understanding concurrent emotions and motivation to feedback types and task performance during GBL.

Keywords: Emotions · Motivation · Game-based Learning · Feedback

1 Introduction

Game-based learning (GBL) interventions offer promising solutions for improving critical-literacy skills (e.g., graph literacy) [4]. Using inoculation theory [15], GBL environments (GBLEs) expose individuals to weakened doses of misinformation, while simultaneously scaffolding emotional engagement, motivation, and cognitive skills [5]. GBLEs can be further exploited via adaptive scaffolding, where game features, e.g., feedback, can be customized to meet individual learning needs, thereby enhancing learning effectiveness [21]. However, state-of-the-art algorithms driving adaptive scaffolding primarily target the cognitive aspects of learning [8,26], overlooking the influence of emotions and motivation on learning [8,9,16].

F. Bellotti et al. (Eds.): GALA 2024, LNCS 15348, pp. 101–111, 2025.
https://doi.org/10.1007/978-3-031-78269-5_10

Negative emotions can hinder learning and motivation to persist [11,27], while positive emotions can enhance learning and encourage continued engagement [10,11]. A GBL study [17] found that more positive emotions co-occurred with more interest, suggesting emotions and motivation have interdependent relations [9]. Thus, building adaptive GBL interventions that consider emotions and motivational on learning is essential.

1.1 Feedback, Emotions, and Motivation

Feedback on tasks can evoke different emotions and motivation that influence persistence and engagement [12,18]. According to control-value theory (CVT), the attributions an individual makes affects their motivation, emotions, and learning [7]. In response, feedback strategies have been developed to scaffold specific learning processes [6]. In its simplest form, feedback provides task evaluation results: correct vs. incorrect. Studies find that feedback on failed (FOF) tasks is often associated with negative emotions, while feedback on success (FOS) is associated with positive emotions [13]. Whether different emotional responses to FOS or FOF influenced motivation and learning is understudied.

A common feedback strategy is elaborative, providing detailed information about the task evaluation. First, the outcome expectation (standard) and the answer selected is highlighted. Next, hints, examples, and next steps for future improvement are provided [6]. Studies find that when individuals receive elaborative feedback, they demonstrated more learning compared to only feedback of task results [20]. Attribution-based (AB) feedback is another strategy grounded in CVT [7] that strategically directs an individual's attention to the cause of the evaluation. AB highlights internal vs. external attributions that support positive emotions and motivation [14,19]. E.g., if an individual received FOS, the evaluation is attributed to an internal cause, while FOF to an external cause.

[14] found that subjects who received AB feedback outperformed and reported more positive emotions than those receiving results feedback. Another study [19] found that elaborative feedback with AB elements was associated with more learning strategies than only received elaborative feedback. However, both groups used more learning strategies than those who received no feedback. More questions remain about the role of emotions and motivation in response to different types of feedback and its influence on GBL outcomes. In some cases, negative emotions that arise from FOF can motivate individuals to overcome challenges, depending on how they attribute the causes of their performance and whether it is within their control [7,22,23]. Other limitations exist as few studies account for individual variability in emotions and motivation and its influence on GBL outcomes. Prior studies and CVT [7] emphasize that individual characteristics shape the attributions about a task result, which in turn influence the nature, timing, and intensity of emotions and motivation.

1.2 Current Study and Novel Contributions

Emotions and motivation to feedback and its influence on critical-literacy skill development is understudied. The challenges stem from methods for measuring emotions and motivation. Objective, continuous data are common approaches for measuring emotions (e.g., electrodermal [18], neuro-imaging [27]), and while these data provide useful insights, inferences are made about emotions and motivation without input from the individual's subjective experience. This study provides novel insights into concurrent emotions and motivation to feedback during GBL in two ways. First, emotions and motivation were collected using validated surveys to capture the subjective experience following different feedback: AB, elaborative, and results. Second, we studied the role of individual differences on concurrent emotions and motivation after feedback during GBL [9].

2 Methods

2.1 Sample and Materials

Forty-one students ($n = 41$; $M^{\text{Age}} = 23.73$ ($SD = 3.23$); 62% female) enrolled at a University in Austria completed an online study with a GBLE called MediaWatch. Subjects were removed if they did not complete the surveys before and after GBL. One subject was removed for starting the game twice and 2 subjects were removed as they completed the GBL tasks in English, but were instructed to complete them in German. The study lasted roughly 60 - 90 min. After completing the study, subjects could voluntarily enroll in a raffle prize. Pre-/post-surveys were administered before and after GBL to assess graph-literacy skills. Graph-literacy skills were measured using a validated, 13-item instrument [2]. After each GBL task, emotions and motivation were measured using the Emotions & Values (EV) survey [1], a 5-pt Likert scale (*1=Strongly Disagree, 5=Strongly Agree*) capturing anger, boredom, surprise, neutral, confusion, interest, and eureka (e.g., "Right now I feel bored"). Prior work has found a high agreement between facial expressed emotions and EV data (75.6%) [1].

2.2 MediaWatch

MediaWatch is a web-based, GBLE grounded in inoculation theory [15][1]. Subjects are exposed to misleading graphs to build critical graph-literacy skills [5]. On an island called Sahramao, subjects act as a fact-checker to identify misleading graphs marketed by corporations. Subjects must complete 12 graph-reading tasks; each graph-reading task includes a question, a 4-option, multiple-choice answer, and a graph representation. Six of the 12 graphs are misleading, while the rest are non-misleading. Misleading graphs had one of many manipulations, including x- or y-axis scale-/text-spatial conflicts [2,3] (Fig. 1).

[1] https://webpages.tuni.fi/gamelab/2022/mediawatch/.

2.3 Experimental Design

A 2 (graph type: misleading, non-misleading) x 3 (feedback: results, elaborative, AB) block-based, within-subject design was used. Three blocks were used and each block consisted of 4 graphs (2 misleading, 2 non-misleading). Graph type was randomized for each block and subjects received immediate feedback regardless of correctness. The feedback varied for each block: 1) evaluation results, 2) elaborative, and 3) AB feedback (Fig. 1). The results feedback only stated whether their answer was correct or incorrect. For block two, the elaborative message was designed using to the Feedback Strategies Framework [6]. The feedback highlighted (a) the gap between their performance and the standard [6], (b) the answer selected, and (c) a recommendation for how to improve in the future. For all misleading graphs, hints were provided where the misleading graph and its corrected version were presented along with the feedback message (see Fig. 1). In block 3, AB feedback was designed similarly to elaborative but was grounded in CVT [7]. In this case, at least one attribution of the result was highlighted to elicit a positive emotional-motivational response [14]. For FOF, subjects competence was highlighted and the failure was attributed to an external cause, while FOS was attributed to an internal cause. After feedback was provided, EV measures were immediately administered.

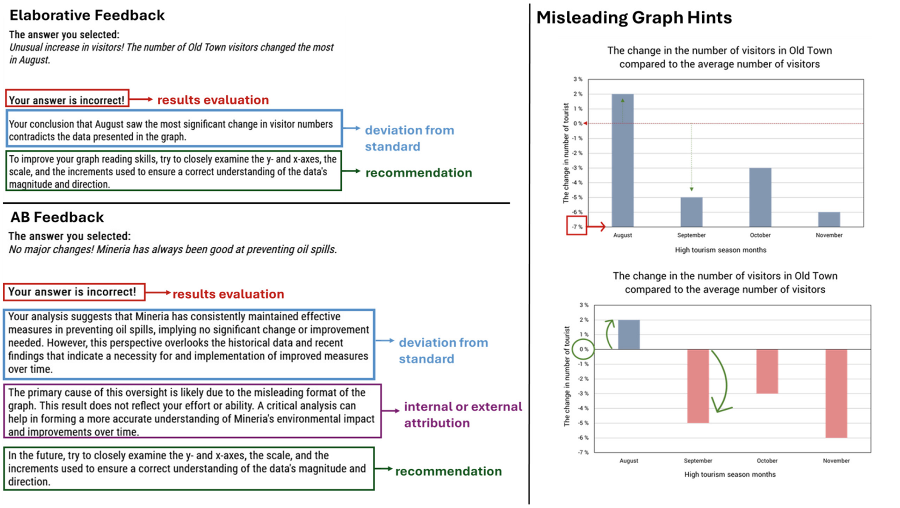

Fig. 1. Elaborative and AB feedback (left); misleading graph hint example for y-axis scale-spatial conflict (right).

2.4 Procedure

Subjects were recruited via email with university listservs. A survey link was shared, which could be accessed using a laptop, where individuals could enroll in the study. Once the link was selected, subjects were directed to the study information. If they decided to participate, they received a random key code. Next, they completed a pre-test survey and were then directed to the game. During GBL, they were presented with brief instructions and then they were asked to complete multiple graph tasks. Once subjects completed the tasks, a post-test survey was immediately administered. After the post-test survey, subjects were thanked and offered enrollment in a raffle.

3 Results

3.1 Are There Differences in Graph-Literacy Skills Before and After GBL with MediaWatch?

A paired Wilcoxon signed rank test was conducted with a one-sided test (we expected improvement in graph-literacy skills after GBL with MediaWatch [5]). Results showed that there was significant improvement in graph-literacy scores from before GBL ($Med = 81.25$, $SD = 7.86$) to after GBL ($Med = 87.5$, $SD = 9.35$), $V = 36.5$, $p = .0006698$, Cohen's $d = .591$. The findings confirmed our hypothesis, where we expected graph-literacy skill improvement after GBL [5].

3.2 Are There Differences in Task Performance During GBL Between Feedback Types (Results, Elaborative, AB) and Graph Types (Misleading, Non-Misleading)?

A standard logistic regression was fitted to assess the relationships between feedback and graph types on task performance during GBL (correct vs. incorrect). The results showed that there were significant, negative relations between task performance and graph type (non-misleading) ($\beta = 1.15$, $p < .01$), $\chi^2(3) = 20.07$, $p < .01$, McFadden $R^2 = .05$. This finding suggests that the odds of an correct answer was approximately .32 times (or 68%) higher when the graph was non-misleading compared to when the graph was misleading. There were no relations with feedback and task performance ($ps < .05$). It is important to note that the model explained 5% of variability in task performance, leaving 95% unexplained by other factors. The findings partially supported our hypothesis, where we expected graph types and feedback to influence task performance [4,5].

3.3 Are There Relations Between Emotions and Motivation, Task Performance (Correct, Incorrect), Feedback Types (Results, Elaborative, AB), and Graph Types (misleading, Non-misleading)?

Cumulative link mixed models (CLMM) make it possible to analyze ordinal responses while allowing the use of random effects [24]. To examine whether feedback, graph type, and task performance were associated with concurrent emotions and motivation, a CLMM with a LaPlace approximation was calculated with a random effects term to partition within-subject variability in dependent variables over time. A separate model was calculated for each dependent variable: anger, boredom, surprise, neutral, confusion, interest, and eureka, using the 'clmm' function with the ordinal package [25] in RStudio [Version 4.3.2] [28]. Results are presented in Table 1 due to page restrictions. Follow-up pairwise Tukey tests were calculated for all significant predictors with a false discovery rate post hoc correction. The post hoc pairwise Tukey results are in Table 2.

Overall, the findings suggest that concurrent emotions and interest were significantly related to feedback, except for anger. Bored responses were highest after AB (compared to elaborative and results), while surprise, confusion, and interest were lowest after AB feedback (compared to results and elaborative feedback). Eureka responses were highest after elaborative compared to AB feedback. This result did not support our hypothesis where we expected more positive emotions and interest after AB. In contrast, surprise was the only emotion related to graph type. More surprised responses were higher after misleading than non-misleading graphs, which is to be expected since misleading graphs had manipulations. In general, there were positive and significant relationships with anger, surprise, and confused responses with incorrect answers, but neutral responses were higher after correct than incorrect answers. Across the models, the fixed effects (feedback, graph type, and task performance) explained at least 5 - 10% variability in emotional and motivational states. In contrast, the random effects explained 40 - 70% of variability in emotional and motivational states, suggesting within-subject variability played a large role in concurrent emotions and motivation during GBL.

4 Discussion

This study examined concurrent emotions and motivation to different feedback types (AB, elaborative, results) for misleading and non-misleading graphs with MediaWatch and whether graph-literacy skills improved after GBL. Research question one (RQ1) suggested there was significant graph literacy skill improvement after GBL, supporting our hypothesis and prior findings [5]. RQ2 investigated the effect of graph type and feedback on task performance during game-based learning (GBL). The results suggested that non-misleading graphs were associated with a higher likelihood of correct answers compared to misleading graphs. Yet, we did not observe any relations between feedback and task performance, which contradicts prior work [19,20]. A possible explanation for this result could be due to the block-based design of the study. Subjects were exposed to all feedback messages across the GBL tasks, which may have diminished the impact of feedback on task performance over time, unlike a controlled experimental condition. Future studies should leverage experimental designs to investigate the role of different types of feedback on task performance during GBL.

In RQ3, we examined whether task performance, feedback, and graph types were associated with concurrent emotions and motivation during GBL. Results

Table 1. CLMM results.

	Anger			Boredom			Surprise		
Fixed Effects	OR	95% CI	*p*	OR	95% CI	*p*	OR	95% CI	*p*
Feedback (E)	0.67	0.45-0.98	.039	0.52	0.39-0.68	<.001	1.41	1.04-1.91	.028
Feedback (R)	0.62	0.42-0.91	.015	0.34	0.25-0.45	<.001	1.72	1.27-2.32	<.001
Graph type (NM)	0.89	0.64-1.23	.472	1.20	0.96-1.50	.102	0.76	0.59-0.98	.033
TP (INC)	4.02	2.67-6.05	<.001	1.17	0.83-1.65	.358	2.84	2.05-3.94	<.001
Random Effects	σ^2	τ_{00}	ICC	σ^2	τ_{00}	ICC	σ^2	τ_{00}	ICC
Individual	1.00	1.09	.52	1.00	3.17	.76	1.00	0.95	.49
Model Diagnostics									
n	41			41			41		
Observations	492			492			492		
Fixed R^2	0.121			0.047			0.105		
Total R^2	0.580			0.772			0.540		
	Neutral			Confusion			Interest		
Fixed Effects	OR	95% CI	*p*	OR	95% CI	*p*	OR	95% CI	*p*
Feedback (E)	2.08	1.62-2.69	<.001	1.26	0.91-1.75	.162	2.24	1.71-2.95	<.001
Feedback (R)	2.42	1.87-3.13	<.001	1.50	1.09-2.07	.014	3.74	2.82-4.96	<.001
Graph type (NM)	1.22	0.99-1.50	.065	0.81	0.62-1.05	.116	1.06	0.85-1.31	.625
TP (INC)	0.47	0.34-0.64	<.001	5.13	3.63-7.25	<.001	0.80	0.57-1.11	.183
Random Effects	σ^2	τ_{00}	ICC	σ^2	τ_{00}	ICC	σ^2	τ_{00}	ICC
Individual	1.00	2.04	.67	1.00	0.81	.45	1.00	3.14	.76
Model Diagnostics									
n	41			41			41		
Observations	492			492			492		
Fixed R^2	0.071			0.180			0.067		
Total R^2	0.694			0.548			0.775		
	Eureka								
Fixed Effects	OR	95% CI	*p*						
Feedback (E)	1.45	1.07-1.95	.015						
Feedback (R)	1.35	1.00-1.82	.050						
Graph type (NM)	0.89	0.69-1.13	.331						
TP (INC)	1.08	0.75-1.54	.682						
Random Effects	σ^2	τ_{00}	ICC						
Individual	1.00	2.32	.70						
Model Diagnostics									
n	41								
Observations	492								
Fixed R^2	0.009								
Total R^2	0.701								

Note. Feedback(E)/(R) = Elaborative/Results; Graph type (NM) = Non-misleading; TP(INC) = Task Performance Incorrect

Table 2. Significant findings from pairwise Tukey tests for various models.

Model	Predictor	Comparison	M	p
Anger	Task Performance	Incorrect - Correct	1.83 - 1.18	< .0001
Boredom	Feedback	AB - Elaborative	2.89 - 2.43	< .0001
		AB - Results	2.89 - 2.09	< .001
		Elaborative - Results	2.43 - 2.09	.0061
Surprise	Feedback	Results - AB	1.83 - 1.48	.0011
	Graph Type	Misleading - Non-misleading	1.78 - 1.52	.033
	Task Performance	Incorrect - Correct	2.57 - 1.5	< .0001
Neutral	Task Performance	Correct - Incorrect	3.32 - 2.64	< .0001
	Feedback	Results - AB	3.46 - 2.83	< .0001
Confusion	Feedback	Results - AB	1.67 - 1.41	.0373
	Task Performance	Incorrect - Correct	2.64 - 1.35	< .0001
Interest	Feedback	Results - AB	2.96 - 2.09	< .0001
		Elaborative - AB	2.59 - 2.09	< .0001
		Results - Elaborative	2.96 - 2.59	.0003
Eureka	Feedback	Elaborative - AB	1.87 - 1.66	.0406

showed that the effect of feedback was significantly related to concurrent emotions and interest, with the exception of anger. Boredom was the highest and interest was the lowest after AB compared to elaborative and results feedback, which is contradictory to prior work [7,14,19]. Surprise, neutral, and confused responses were significantly higher after results compared to AB feedback. Eureka was significantly higher after elaborative compared to AB feedback. Our findings partially support prior studies, where there were less positive emotions and motivation after results feedback (possibly explaining the higher confusion). In general, there were less positive emotions and interest after AB compared to results and elaborative feedback. A possible explanation could again be due to an ordering effect caused by the study design. In block 3, individuals reported the most boredom compared to block 2 and 1. This suggests that there was an increase in boredom and decrease in interest across the order of the blocks. In terms of task performance, there were significantly more negative emotions after incorrect compared to correct answers, where there was more anger and confusion after incorrect, but more surprise and neutral responses after correct answers compared to incorrect answers. The findings support our hypothesis, where we expected more confusion and anger after incorrect than correct answers [7,13].

Our study has limitations. First, relying solely on self-report data to measure concurrent emotions and motivation is problematic as individuals may not always be aware of their emotions or motivation. Future studies should utilizing mixed-mulitmodal methods, where both subjective (e.g., think-/emote-alouds, surveys) and objective measures (e.g., facial recognition, electrodermal activity, field observations) are used. Second, the block order design of the study may have influenced emotions and motivational responses. Future studies should consider emotional and motivational changes as well as an control group to account for the possible ordering effect of the feedback blocks on emotions and motivation.

Acknowledgements. This study was funded by the European Union ([EU] ID No. 101105874). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the EU.

References

1. Harley, J., Bouchet, F., Hussain, M., Azevedo, R., Calvo, R.: A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. Comp. Hum. Behav. **48**, 615–625 (2015)
2. Galesic, M., Garcia-Retamero, R.: Graph literacy: a cross-cultural comparison. Med. Decis. Making **31**(3), 444–457 (2011)
3. Okan, Y., Garcia-Retamero, R., Galesic, M., Cokely, E.: When higher bars are not larger quantities: on individual differences in the use of spatial information in graph comprehension. Spat. Cogn. Comput. **12**(2–3), 195–218 (2012)
4. Kiili, K. Siuko, J. Ninaus, M.: Tackling misinformation with games: a systematic literature review. Interact. Learn. Environ. 1–16 (2024)
5. Siuko, J. Cloude, E. Kiili, K.: Improving critical graph reading skills: the potential might lie in game-based learning. In: DSouza, P., Bujić, M. Xi, N., Hamari, J. (eds.) Proceedings of 8th International GamiFIN Conference, vol. 3669, pp. 79–87 (2024)
6. Narciss, S., et al.: Exploring feedback and student characteristics relevant for personalizing feedback strategies. Comput. Educ. **71**, 56–76 (2014)
7. Pekrun, R.: The control-value theory of achievement emotions: assumptions, corollaries, and implications for educational research and practice. Edu. Psych. Rev. **18**, 315–341 (2006)
8. Aleven, V., McLaughlin, E., Glenn, R.A., Koedinger, K.: Instruction based on adaptive learning technologies. Handb. Res. Learn. Instr. **2**, 522–560 (2016)
9. Pekrun, R., Marsh, H.: Research on situated motivation and emotion: progress and open problems. Learn. Instr. **81**, 101664 (2022)
10. Goetz, T. Hall, N.: Emotion and achievement in the classroom. In: Hattie, J., Anderman, E.M. (eds.) International Guide to Student Achievement, pp. 192–195. Routledge (2013)
11. Loderer, K., Pekrun, R., Lester, J.: Beyond cold technology: a systematic review and meta-analysis on emotions in technology-based learning environments. Learn. Instr. **70**, 101162 (2020)
12. Peterson, E., Brown, G., Jun, M.: Achievement emotions in higher education: exploring emotions across an assessment event. Cont. Edu. Psych. **42**, 82–96 (2015)
13. Lipnevich, A., Murano, D., Krannich, M., Goetz, T.: Should I grade or should I comment: links among feedback, emotions, and performance. Learn. Individ. Differ. **89**, 102020 (2021)
14. Schrader, C. Grassinger, R.: Tell me that I can do it better: the effect of attributional feedback from a learning technology on achievement emotions and performance and the moderating role of individual adaptive reactions to errors. Comp. Edu. **161**, 104028 (2021)

15. Ecker, U.K.H., et al.: The psychological drivers of misinformation belief and its resistance to correction. Nat. Rev. Psychol. **1**(1), 13–29 (2022)
16. Cloude, E.B., Taub, M., Lester, J., Azevedo, R.: The role of achievement goal orientation on metacognitive process use in game-based learning. In: Isotani, S., Millán, E., Ogan, A., Hastings, P., McLaren, B., Luckin, R. (eds.) AIED 2019. LNCS (LNAI), vol. 11626, pp. 36–40. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-23207-8_7
17. Kiili, K., Juho, S., Cloude, E., Dindar, M.: Demystifying the relations of motivation and emotions in game-based learning: insights from co-occurrence network analysis. Int. J. Serious Games **10**(4), 93–112 (2023)
18. Horvers, A., Molenaar, I.V., Der West, H., Bosse, T., Lazonder, A.: Multimodal measurements enhance insights into emotional responses to immediate feedback. Front. Psych. **14**, 1294386 (2024)
19. De Sixte, R. Mañá, A. Ávila, V. Sánchez, E.: Warm elaborated feedback: exploring its benefits on post-feedback behaviour. Edu. Psychol. **40**(9), 1094–1112 (2020)
20. Van der Kleij, F., Feskens, R., Eggen, T.: Effects of feedback in a computer-based learning environment on students' learning outcomes: a meta-analysis. Rev. Educ. Res. **85**(4), 475–511 (2015)
21. Koskinen, A., McMullen, J., Hannula-Sormunen, M., Ninaus, M., Kiili, K.: The strength and direction of the difficulty adaptation affect situational interest in game-based learning. Comp. Edu. **194**, 104694 (2023)
22. Vogl, E. Pekrun, R.: Emotions that matter to achievement: student feelings about assessment. In: Handbook of Human and Social Conditions in Assessment, pp. 111–128. Routledge (2016)
23. Lim, L.-A., et al.: Students' perceptions of, and emotional responses to, personalised learning analytics-based feedback: an exploratory study of four courses. Assess Eval. High Edu. **46**(3), 339–359 (2021)
24. Agresti, A.: Categorical Data Analysis, vol. 792. Wiley (2012)
25. Christensen, R.: Ordinal-regression models for ordinal data. R Package Vers. **10**(2019), 54 (2019)
26. Chen, C., Law, V., Huang, K.: Adaptive scaffolding and engagement in digital game-based learning. Educ. Tech. Res. Dev. **71**(4), 1785–1798 (2023)
27. Lei, H., Wang, C., Chiu, M., Chen, S.: Do educational games affect students' achievement emotions? Evidence from a meta-analysis. J. Comp. Assist. Learn. **38**(4), 946–959 (2022)
28. R Core Team: R: a language and environment for statistical computing. R Foundation for Statistical Computing (2023)