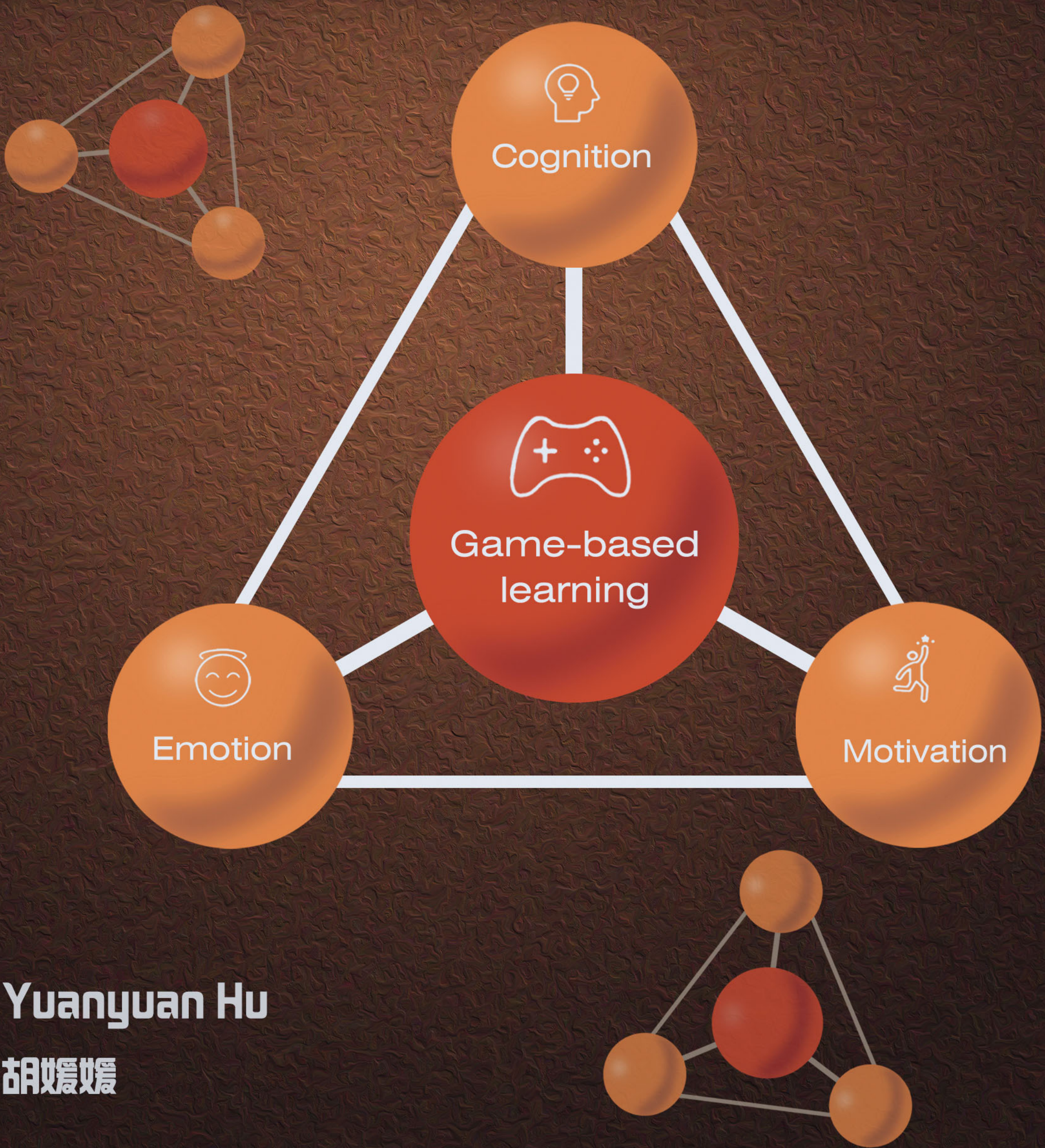
Instructional design of game-based learning in chemistry

- Optimizing cognition,
- motivation,
- and emotion



Yuanyuan Hu

胡媛媛

Instructional design of game-based learning in chemistry

Optimizing cognition, motivation, and emotion

Yuanyuan Hu

DOI: https://doi.org/10.33540/1997

ISBN: 978-94-6483-463-5

Cover design and layout: Fugui Xue, Shufei Liu, and Yuanyuan Hu

Printed by: Ridderprint | www.ridderprint.nl

Instructional design of game-based learning in chemistry

Optimizing cognition, motivation, and emotion

Instructieontwerp van game-based learning in scheikunde:

Optimaliseren van cognitie, motivatie en emotie

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht

op gezag van de rector magnificus,

Prof. Henk Kummeling,

ingevolge het besluit van het college voor promoties in het openbaar te verdedigen

op vrijdag 20 oktober 2023 des ochtends te 10:15 uur

door

Yuanyuan Hu

geboren op 15 juli 1990

te Anhui, China

Promotoren:

Prof. dr. L. Kester

Prof. dr. M.F. van der Schaaf

Copromotor:

Dr. P.J.M. Wouters

Beoordelingscommissie:

Dr. D. Cermak-Sassenrath

Prof. dr. T.A.J.M. van Gog

Prof. dr. T. van Gerven

Prof. dr. J.T. Jeuring

Prof. dr. S.F. te Pas

Dit proefschrift werd (mede) mogelijk gemaakt met financiële steun van het Europese programma Horizon 2020 voor onderzoek en innovatie (No. 812716).

Table of Contents

Preface and acknowledgements

Life is short, but knowledge is long (吾生也有涯,而知也无涯) – Zhuang Zi
Never tire of learning, never tire of teaching (学而不厌,诲人不倦) – Kong Zi
A journey of a thousand miles begins with a single step (千里之行,始于足下) – Lao Zi

Inspired by the three greatest Chinese thinkers, I've always loved things about education. Since 2014, I worked as a chemistry teacher in secondary schools. I found that many students struggled with learning. The heavy teaching load left me no time to research how to make learning effective, engaging, and enjoyable. ‘You should do a PhD in education’, said my husband Fei five years ago. That was the start of my PhD trajectory. With a lot of people, I made my life here. According to attribution theory, I feel happy (due to the outcomes), pride (due to years of effort), and gratitude (due to these people around me).

Met dank aan mijn begeleiders (Yes, de beste die ik kon krijgen). Als ik ergens mee worstel, komen jullie altijd met duizenden oplossingen. Als ik met iets kom, reageren jullie altijd met andere perspectieven. Zelfs als dingen niet goed gaan, lachen we er altijd om. Onze wekelijkse en maandelijkse bijeenkomsten hebben niet alleen dit proefschrift mogelijk gemaakt, maar betekenden ook veel voor mij, vooral toen ik alleen was tijdens COVID-19. Liesbeth, bedankt voor het vertrouwen in mijn capaciteiten, voor je kritische advies en voor het herinneren aan de rode draad in het verhaal. Marieke, bedankt voor je felicitaties bij goed nieuws, voor je troost bij slecht nieuws, en voor je energie en passie voor wat je doet. Pieter, bedankt voor het inbeelden van mijn perspectief, dat je rustig bleef als ik haast had, en dat je me hielp om het simpel te houden. Het is een eer om je eerste en laatste promovendus te zijn (Ja, als het enige kind kreeg ik alle aandacht).

Thanks to the teachers in and out our department (Yes, there are so many people, so I still do not know everyone yet). Lisette, Bert, thank you for helping me as if I was your PhD student. Barbara, Margot, Eva, thank you for the inspiring discussion and accompaniment during the conferences. Vincent, thank you for sharing your funding application experiences and helping our study on emotion. Gesa, thank you for the fruitful discussions in the SIG1 and Pubgroup meetings. Ellen, thank you for demonstrating eye-trackers. Jan, Jeroen, thank you for the job searching tips. Casper, Nies, Kim, Anna, thank you for helping me in Educational Psychology course. Jacqueline, thank you for sharing the responsibilities in Master Thesis course. Brianna, Stephanie, thank you for inviting me to Academic Writing course. Despoina, thank you for inviting me to International Perspectives on Education course. Larike, thank you for sharing Theories in Learning course. Monique, Nienke, Pascale, Ewa, Mayke, thank you for the chats. Andy, thank you for inviting me to visit your group and being the co-author. Reinhard, thank you for the discussion during conferences and being the co-author. Sascha, thank you for the career tips. Min, thank you for the guidance.

Thanks to my paranymphs, Tim and Michaela (Yes, we are the poor brothers and sisters of the pandemic generation), for the brainstorm, encouragement, and laughter. We have similar working experience and start more or less at the same time in the same project. It is so great that we share our journey with each other. Tim, thank you for sharing the joys, worries, and complains about life in different countries and the good times with your family. Michaela, thank you for the office pats and hugs, psychologically it is particularly good for my mental health.

Thanks to my roommates from F3.01, E3.13, and E3.36 (Yes, I've moved a lot). David, thank you for being the big brother who is always there to help. Anne, thanks you for sharing the Dutch survival tips. Sophia, thank you for sharing the funny stories. Jael, thank you for caring and making me feel at home from the very beginning. Joris, thank you for the chats. Mei, thank you for helping me settle down. Jonne, thank you for the 'mean' but wise comments. Sophie, thank you for introducing me Jip en Janneke. Marloes, thank you for the Exploring Dutch History adventure. Angela, thank you for the complains about Dutch life. Esther, thank you for the writing tips. My buddy Linda, thank you for the questions and chats. The "newcomers", Nina, thank you for the native Dutch tips; Eleanor, thank you for sharing your travel experiences; Florence, thank you for sharing your China adventure; Melis, thank you for sharing life in Turkey; Jonne, thank for organizing our Thessaloniki adventure; Jane, thank you for sharing the hair troubles; Rowan, thank you for talking about Psychology; Selma, Simone, Adriaan, Anouk, Alex, Christa, Mare, Katrijn, Xiaojing, Dannie, Renée, Monika, Selina, Minke, thank you for the wonderful time at the lunch table and social events.

Thanks to all the people from CHARMING project (Yes, we are charming people). Tom, Ilse, Rabab, Peter, Daniel, Jarka, Fernando, Kasper, Kevin, Thies, Michael, Kristel, Jean Luc, Hans, thank you for the highly successful and well-organized project. This project is the largest I have, which sets a solid foundation for my future career in academia. Michael, Sanne, Nina, Jessica, Serkan, Pedro, Silvia, Chioma, Philippe, Sofia, Ryo, Yusra, thank you for your friendship. I always look forward to the networking events, not only for doing research but also for seeing you in Leuven, Darmstadt, Lyon, and Copenhagen. The dinners and "team-breaking" events remind me that there is more than research in my life.

Met dank aan het ondersteuningsteam. Jos, bedankt voor IT-hulp en de heerlijke noedels. Jolanda, bedankt voor het helpen oefenen met mijn Nederlands. Inge-Liz, onze geldbazin, bedankt voor het beheren van het budget. Caroline, bedankt voor de hulp met ICO. Veronique, bedankt voor het bestellen van boeken. Marianne, bedankt voor het organiseren van de onderzoeksbijeenkomsten, waardoor ik feedback kon verzamelen voor ons onderzoek. Ellie, bedankt dat je me op mijn eerste dag aan iedereen op de afdeling hebt voorgesteld. Andy, bedankt voor het rijden en het boek van je moeder. Jim, Django, Michael, Son, Mark, bedankt voor de IT-hulp. Madelon, bedankt voor de PhD-evenementen.

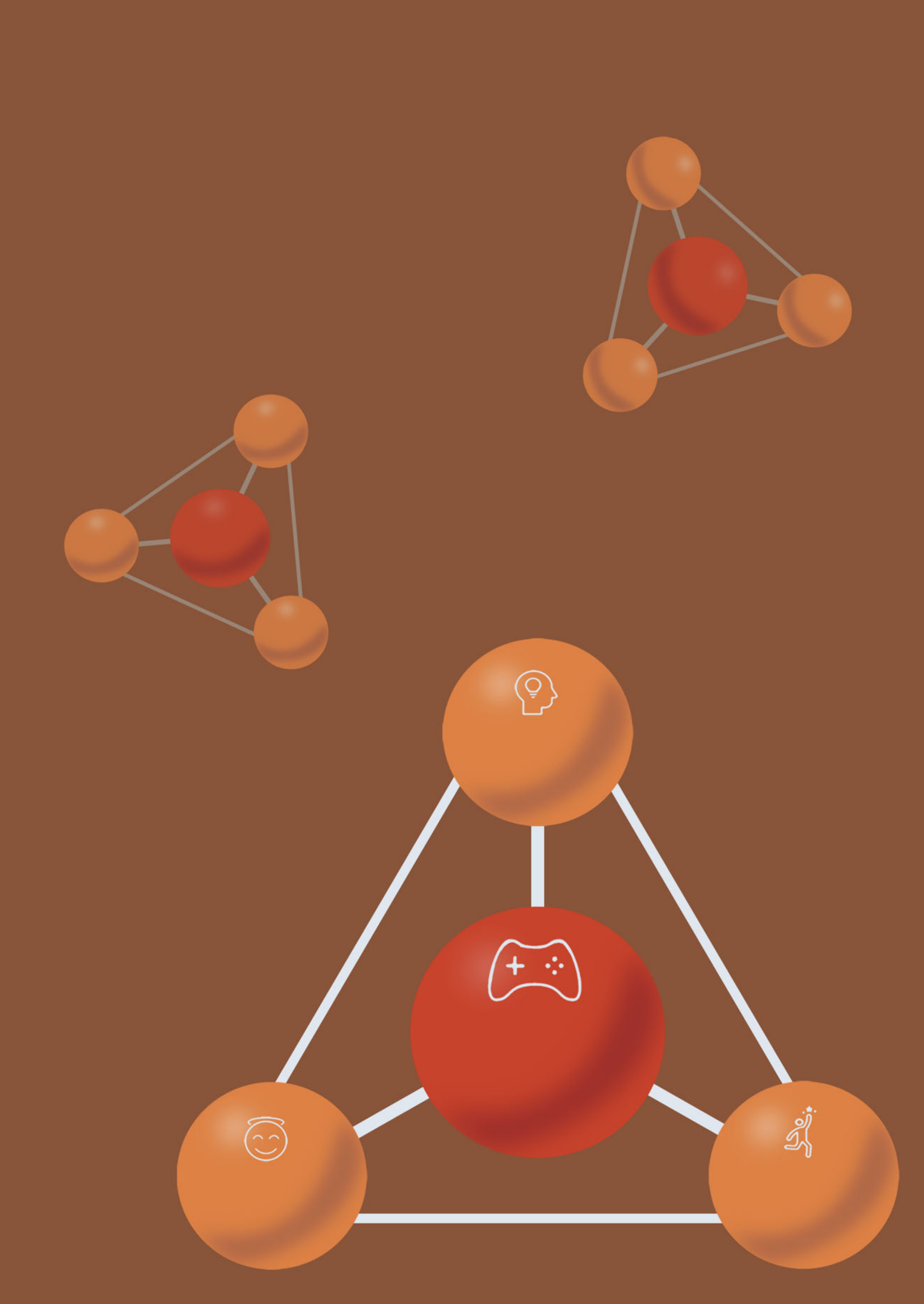
Thanks to the participating teachers and students in our studies. Simone, Reinier, Mirjam, Ferry, Leo, Margarita, Els, Mathijs, Bob, Stephnie, Roy, Janine, Ilse, Theo, Jan, Arnoud, Marieke, Annemiek, Coen, Erik, Dion, Jan-Jaap, Joan, Erwin, Laura, Joost, Wilma, Karin, thank you not only for your interest in our research and helping me collect data, but also for the eye-opening conversations. Because of you, data collection, the most challenging part of research, became very enjoyable.

Thanks to the assessment and defense committee. Tamara, Johan, Tom, Daniel, Susan, thank you for taking the time and effort to read and review this dissertation. Prof. dr. Harold van Rijen, thank you for making my defense complete.

Thanks to my friends (Yes, I do have a life besides work). Junting and Longfei, thank you for your lifelong friendship and the time in Boston. Junfeng, Xiaoqing, Qianjing, Annabelle, Honghong, Danni, Shanshan, Luolan, Mandy, Wentao, Wenjuan, Zheyan, Min, thank you for hanging out. Nita, Marieke, thank for the fun in and out the Dutch class. Collin, Xira, Leilei, Anne, and Ruud, thank you for the climbing and bouldering time. Lingxiao, thank you for the swim time. Marienke, Mike, John, Eric, Han, thank you for the swing time. Aleza, thank you for the lunch adventures and the comedy time. Natasha, thank you for the jazz time. Annette, Shin-Young, Gwyneth, Katy, thank you for showing Rochester. Yuncai, thank you for being my bodyguard in New York City. Weicheng, Xinyu, thank you for the trip to Rome and

Florence. Oliver, thank you for the walking time in Bern. Meg, Emily, Erika, Tahrim, thank you for the laugh in Dresden. Yihui, thank you for exploring London. Fangzheng, thank you for the brainstorm in Thessaloniki. Feifei, thank you for showing Stockholm. Beibei, Qiaoqiao, Jiangyan, Xingyu, thank you for the encouragement.

Last but not the least, thanks to my family. My moms and dads, thank you for the love and inspiring me to love learning. Xixi, Yue, Lulu, and Miya, thank you for sharing the sorrow and joy. Fei, thank you for everything; You are the one who understands me the most and supports me the best. This Journey to the West would not have been possible without you.



Chapter 1 General introduction

Introduction

From a societal perspective, one of the key challenges for education and professional development is to create effective (e.g., cognition), engaging (e.g., motivation), and enjoyable (e.g., emotion) learning and training experiences (Neelen & Kirschner, 2020). This challenge is urgent in Science, Technology, Engineering, and Mathematics including Computer Science education. Specifically, students have reported low literacy and motivation to study and pursue careers in these subjects (European Commission, 2015a; NRC, 2011b). The continuous professional development of employees in relevant workplace also needs to be improved. This challenge could be addressed by technology-enhanced learning, such as game-based learning (GBL), which is well-known for the motivation appeal (Klopfer & Thompson, 2020; Kärkkäinen & Vincent-Lancrin, 2013; OSTP, 2018).

An example can be found in chemistry and chemical industry. The chemical industry is one of the top five contributors to the manufacturing sales in EU, but the EU is losing its position as the world leader in the global market (Cefic, 2023). In addition, the number of highly educated and trained employees is decreasing in chemical industry. Therefore, the EU urgently needs to improve the training of the current employees and education of the future employees in these workplaces. In particular, the CHARMING project, the European Training Network for Chemical Engineering Immersive Learning (https://charming-etn.eu/), meets this challenge by developing immersive learning technologies, such as games, for children in primary education, for students in secondary and higher education, and for employees in chemical industry in chemistry and chemical engineering. This thesis is part of the CHARMING project, which focuses on *which design features can improve GBL in secondary and higher chemistry education*. Moreover, stakeholders disagree on whether teachers and students should use GBL. For example, students often like games, but parents often do not. Game developers often spend a lot of time and effort making games, but teachers often do not use games in the classroom (Bourgonjon et al., 2017). Entrepreneurs often advocate games, but researchers are often skeptical. Most of these disagreements relate to *whether GBL improves learning compared to non-GBL in secondary and higher chemistry education*.

From a scientific perspective, GBL research, in general, also pursues these two aims. Specifically, *value-added research* focuses on the effectiveness of design features, that is, GBL with a specific design feature versus GBL without this specific design feature (i.e., the first question) and *media comparison research* focuses on the effectiveness of GBL, that is, GBL versus non-GBL (i.e., the second question; Mayer, 2020). For value-added research, previous meta-analyses have found that some instructional design features enhance GBL (Wouters & van Oostendorp, 2013), such as competition (Chen et al., 2020), feedback (Tsai & Tsai, 2020), and scaffolding (Cai et al., 2022). For media comparison research, previous systematic reviews and meta-analyses across multiple subjects have shown mixed outcomes: Some reviews conclude with caution about the effectiveness of GBL and call for more empirical evidence (Boyle et al., 2016; Connolly et al., 2012; Girard et al., 2013; Martinez-Garza et al., 2013; Mayer, 2019, 2020; NRC, 2011a; Young et al., 2012), whereas other meta-analyses support its cognitive benefits (Clark et al., 2016; Karakoç et al., 2020; Lamb et al., 2018; Sitzmann, 2011; Vogel et al., 2006; Wouters et al., 2013), particularly retention (Sitzmann, 2011; Wouters et al., 2013), but not its motivational benefits (Sitzmann, 2011; Vogel et al., 2006; Wouters et al., 2013). Furthermore, it is unclear which instructional and methodological characteristics of the GBL environments moderate effects of GBL. In short, previous research cannot provide concrete answers to these two questions.

Taken together, as part of the CHARMING project, this thesis aims to answer the question: *Which instructional design features improve GBL in secondary and higher chemistry education* *and how?* Practically, this thesis aims to provide more insight into whether teachers and students would do well to use GBL and, if so, how the effectiveness of GBL can be supported by instructional design features. Theoretically, this thesis aims to advance theories of learning and instructional design of GBL. This thesis is innovative in focusing on cognition, motivation, and emotion, considering their interconnections, and using a combination of meta-analytic and experimental methodologies.

Game-Based Learning Defined

GBL is a type of learning pedagogy with game play, accompanied by learning goals, learning outcomes, game goals, and game outcomes, in which a game is the medium for learning (Plass et al., 2020). The games used in GBL are called *serious games*, *educational games*, *learning games*, or *games for learning*. Although there is no agreed definition of games, many researchers endorse the essential features of *games*: play (e.g., games are fun; Homer et al., 2020), goals (e.g., goals for learning outcomes and goals for gameplay; Malone, 1981), rules (e.g., what action are allowed; Garris et al., 2002), interactivity (e.g., the player and the game act upon each other; Vogel et al., 2006), challenges (e.g., games should match the player's skill level; Shute & Ke, 2012), and feedback (e.g., providing timely information about their performance to players; Prensky, 2001), as shown in Figure 1.1. Examples of the games from the CHARMING project are a virtual reality game for chemical lab safety (Chan et al., 2023) or for computational fluid dynamics (Solmaz & Van Gerven, 2022), a mobile game for soap-making (Domínguez Alfaro, Gantois et al, 2022), an augmented reality game for acid−base titration (Domínguez Alfaro, Udeozor et al, 2022), or CHEM game jam (Fornós et al., 2022).

A related but different concept is *gamification*, defined as the addition of specific game features (e.g., mostly narrative and incentives) in existing tedious learning environments in order to make them motivating (Sailer & Homner, 2020). Furthermore, gamification leaves the tedious learning task itself largely unchanged, whereas GBL requires a redesign of the learning task (i.e., the learning task is performed in-game). Another related but different concept is *simulation*, defined as a re-creation of hypothetical or real-world phenomena or situations (Clark et al., 2009). Furthermore, most games are simulations, but not all simulations are games. This thesis focuses on learning with games rather than simulation or gamification.

Theoretical Framework of Game-Based Learning

Theoretically, GBL can impact learning by affecting cognitive processes, motivation to learn, and/or emotion (Plass et al., 2020). Figure 1.1 shows a general theoretical framework of GBL. Design features of GBL include *instructional design features* (e.g., teacher support, task design, and peer interactions) and game design features (e.g., the design of the interface, number of sessions, and challenges) which are further divided into *essential game design features* (e.g., play) and *nonessential game design features* (e.g., narratives). The basic assumption is that well-designed GBL should and can promote cognitive processes, motivation to learn, and positive emotions, all of which contribute to learning (Hu et al., 2021, Plass et al., 2020).

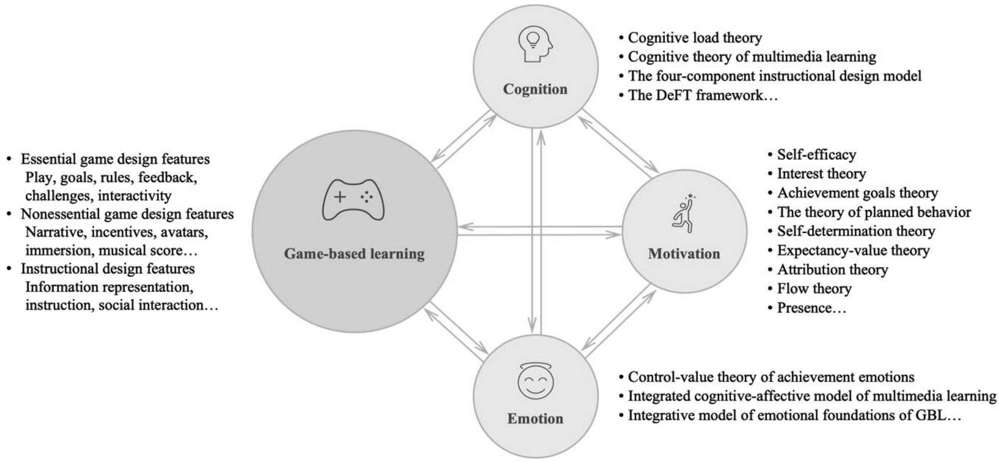
From the cognitive perspective, GBL may affect cognitive processes underlying learning chemistry such as schema construction and schema automation, which is grounded in learning and instructional design theories, such as cognitive load theory (CTL; Sweller et al., 2019), cognitive theory of multimedia learning (CTML; Mayer, 2020), and the four-component instructional design model (4C/ID model; van Merriënboer & Kirschner, 2018). Game design and/or instructional design of GBL aim to optimize cognitive processes and outcomes (Mayer, 2020), such as using feedback (Tsai & Tsai, 2020).

From the motivational perspective, GBL may affect players' values, needs, beliefs, attributions, and goals of learning chemistry (for an overview and comparison, see Cook & Artino, 2016; de Brabander & Martens, 2014; Mayer, 2014b, 2020; Plass et al., 2015), which is grounded in motivation theories, such as self-determination theory (Deci & Ryan, 2000), expectancy-value theory (Wigfield & Eccles, 2000), achievement goal theory (Elliot et al., 2011), self-efficacy (Bandura, 1986), attribution theory (Weiner, 1985), the theory of planned behavior (Ajzen, 1991), flow theory (Csikszentmihalyi, 1990), presence (Cummings & Bailenson, 2016), and interest theory (Hidi & Renninger, 2006). Game design and/or instructional design of GBL also aim to increase motivation to learn (Plass et al., 2020), such as using competition (Chen et al., 2020).

From the emotional perspective, GBL can induce different types of emotions in chemistry learning (e.g., achievement emotions, epistemic emotions, and topic emotions) by shaping their antecedents (e.g., perceived control and perceived value of the learning tasks, cognitive incongruity; for details, see Loderer et al., 2020). These assumptions are grounded in emotion theories, such as the control-value theory of achievement emotions (CVT; Pekrun & Perry, 2014), integrated cognitive-affective model of media (ICALM; Plass & Kaplan, 2015), and the integrative model of emotions in game-based learning (EoGBL; Loderer et al., 2020). Game design and/or instructional design of GBL aim to trigger more positive emotions and less negative emotions (Loderer et al., 2020), such as using emotional design (Plass et al., 2019).

Figure 1.1

The theoretical framework of game-based learning



Note. Single-headed arrows represent causal relations. Cognition, motivation, and emotion can reciprocally affect game-based learning, but the focus of the thesis is on the arrows from game-based learning to cognition, motivation, and emotion.

Instructional Design of Game-Based Learning

An effective, engaging, and enjoyable learning experience requires a well-designed GBL environment. However, most GBL research focuses mainly on cognition, less on motivation, and rarely on emotion, let alone interconnections between them. This thesis focuses on cognition, motivation, and emotion. The three empirical studies in this thesis relate to competence- or achievement-relevant theories, namely, cognitive load theory (Sweller et al., 2019), achievement goals theory (Elliot & Hulleman, 2017), and control-value theory of achievement emotions (Pekrun & Perry, 2014). Figure 1.2 shows an overview of the instructional design features that will be investigated in these three empirical studies (value-added research). In this thesis, *achievement* refers to the attainment of goals (e.g., need for achievement), knowledge, or skills (e.g., academic achievement; see APA Dictionary of Psychology). The center of achievement is *competence*, which is defined as "a condition or quality of effectiveness, ability, sufficiency, or success" and indicates "whether one is doing well or poorly at a task or activity" (Elliot et al., 2017).

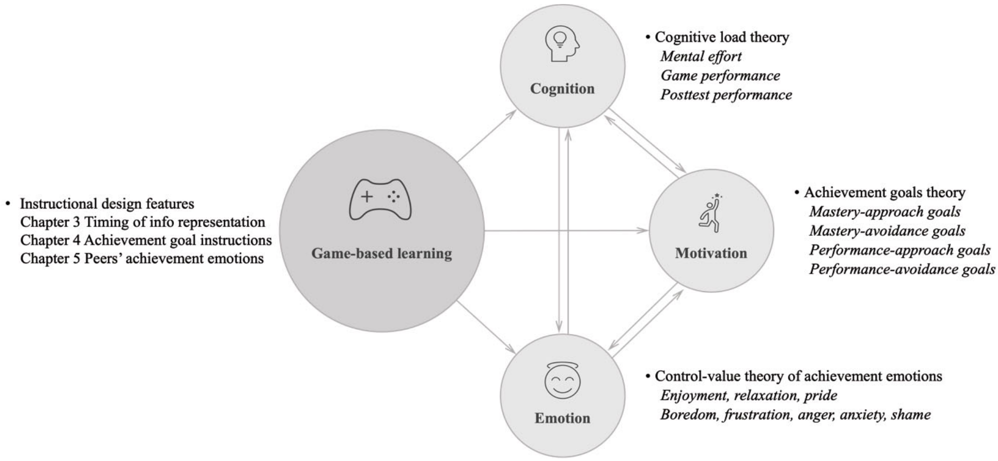
According to cognitive load theory, *cognitive load* is defined as the amount of working memory resources used (Sweller et al., 1988). Cognitive load can be distinguished as *intrinsic load* - load caused by cognitive processes or activities related to learning and performing the task and *extraneous load* - load caused by cognitive processes or activities that are unnecessary for learning and performing the task (Sweller et al., 2019). Because humans have limited working memory capacity, the goal of instructional design is to optimize cognitive load by managing intrinsic load and reducing extraneous load. For example, GBL requires two types of information, namely, *domain-specific information* (i.e., information about the domain), and *game-specific information* (i.e., information about the game), but presenting these two types of information together with gameplay can overload learners. Therefore, the first instructional design feature investigated in this thesis is related to *timing of information presentation* in GBL that aim to optimize cognitive load.

According to achievement goals theory, *achievement goals* is defined as the purpose for engaging in competence-relevant behaviors (Elliot & Hulleman, 2017). Achievement goals can be distinguished as *mastery-approach goals:* striving for task- or self-based competence, such as learning as much as possible, *mastery-avoidance goals*: avoiding task-based or self-based incompetence, such as avoiding learning less than one possibly could, *performance-approach goals*: striving for other-based competence, such as performing better than others, and *performance-avoidance goals*: avoiding other-based incompetence, such as avoiding performing worse than others (Elliot & Hulleman, 2017). Because approach goals are more associated with positive processes and outcomes and avoidance goals are more associated with negative processes and outcomes, the goal of instructional design is to increase mastery-approach goals and performance-approach goals and decrease mastery-avoidance goals and performance-avoidance goals. For example, *achievement goal instructions* - assigning achievement goals before learning - can induce specific achievement goals in learning (e.g., Erhel & Jamet, 2019), but it is unclear which achievement goal instruction is best in GBL. Therefore, the second instructional design feature investigated in this thesis is related to achievement goal instructions in GBL in order to increase approach goals in GBL.

According to control-value theory of achievement emotions, *achievement emotions* are emotions related to competence-relevant activities (e.g., attending class) and/or outcomes (i.e., success or failure) in achievement settings (Pekrun, 2006). Depending on the object of emotions, achievement emotions can be distinguished as *activity emotions* related to achievement-relevant activities or tasks (e.g., enjoyment, relaxation, frustration, boredom, and anger) and *outcome emotions* related with the outcomes of these activities (e.g., pride, hope, relief, anxiety, shame, anger, and hopelessness). Furthermore, depending on the valence (positive / negative or pleasant / unpleasant) and activation (physiologically activating / deactivating) of emotions, achievement emotions can be distinguished as *positive activating emotions* (e.g., enjoyment, pride), *positive deactivating emotions* (e.g., relaxation, relief), *negative activating emotions* (e.g., frustration, anxiety), and *negative deactivating emotions* (e.g., boredom, hopelessness; Pekrun & Perry, 2014). Because emotions can help or harm learning, the goal of instructional design is to induce emotions that help learning. For example, teachers' emotions influence students' learning (e.g., Lawson & Mayer, 2022), but it is unclear whether peers' emotions also influence students' learning. Therefore, the third instructional design feature investigated in this thesis is related to peers' achievement emotions in GBL.

Figure 1.2

An overview of instructional design features investigated in this thesis (value-added research)



Note. Single-headed arrows represent causal relations.

Overview of the Chapters

This chapter has introduced a theoretical framework of GBL that describes how GBL affects cognitive processes, motivation to learn, and emotion. Based on this theoretical framework, the focus of this thesis will be: (1) Is the effect of GBL in chemistry education on cognition, motivation, and emotion larger than for non-GBL? (Media comparison research); and (2) Which design features improve the effectiveness of GBL and how? (Value-added research). This thesis systematically reviewed empirical studies on the effects of GBL in comparison to non-GBL (Chapter 2) and investigated the effects of three instructional design features that manipulated students' cognition (i.e., timing of information presentation; Chapter 3), motivation (i.e., achievement goals; Chapter 4), and emotion (i.e., peers' achievement emotions; Chapter 5) on three learning processes and outcomes, namely, cognition (i.e., mental effort and performance), motivation (i.e., achievement goals), and emotion (i.e., achievement emotions) in GBL in chemistry education.

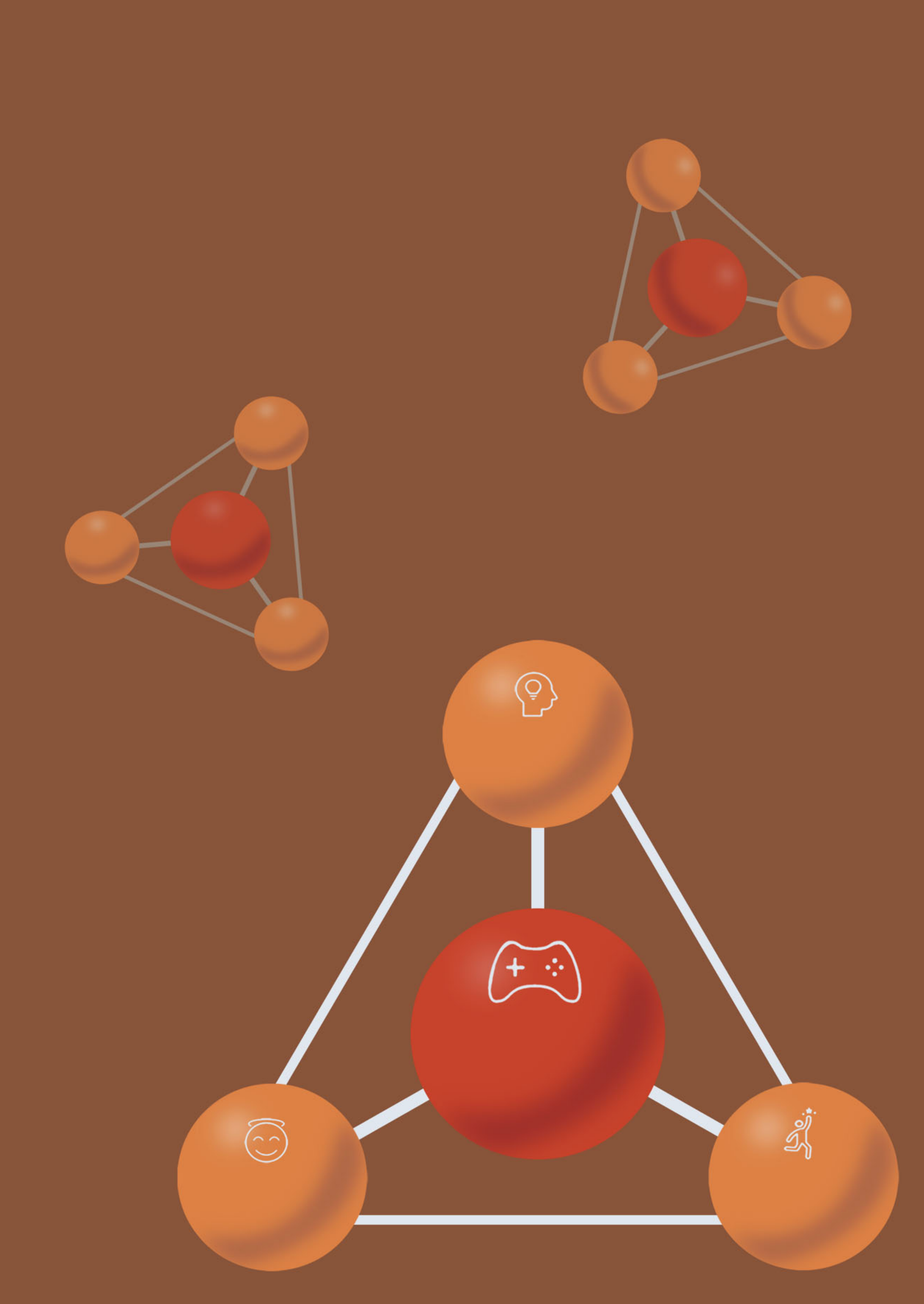
Chapter 2 describes in detail how GBL affects cognition, motivation, and emotion in secondary and higher chemistry education. The specific research questions are: Is the effect of GBL in chemistry education on cognition (including retention), motivation, and emotion larger than for non-GBL (media comparison)? Do instruction characteristics (activity level of control group, additional instruction, user grouping, and number of game sessions) and methodology characteristics (randomization, sample size, publication source, and assessment type) moderate the effect? And which game design or instructional design features improve GBL in chemistry education (value-added research)? We conduct a meta-analysis to answer these questions.

Chapter 3 focuses on an instructional design feature that aims to optimize cognitive load and performance: timing of (domain-specific and game-specific) information presentation. The specific research questions are: How does timing of (domain-specific and game-specific) information presentation affect cognition (i.e., mental effort, posttest performance, and game performance), motivation (i.e., achievement goals), and emotion (i.e., achievement emotions) in GBL in chemistry education? and Do mental effort, perceived competence, perceived control, and perceived value mediate the effects? We compare four conditions in an experimental study to investigate this: Timing of domain-specific information presentation (before/during gameplay) and timing of game-specific information presentation (before/during gameplay).

Chapter 4 focuses on an instructional design feature that aims to increase approach goals: achievement goal instructions (mastery-approach goals and performance-approach goals). The research questions are: How do achievement (mastery-approach and performance-approach) goal instructions affect motivation (i.e., achievement goals), cognition (i.e., mental effort and posttest performance), and emotion (i.e., achievement emotions) in GBL in chemistry education? Do prior achievement goals moderate the effects? And do achievement goals mediate the effects? We compare four conditions to investigate the research questions: Mastery-approach goal instructions (yes/no) and performance-approach goal instructions (yes/no).

Chapter 5 focuses on an instructional design feature that aims to induce positive achievement emotions: emotional contagion of peers' achievement emotions (enjoyment, frustration, and neutral state). The research questions are: How do peers' achievement emotions (enjoyment/frustration/neutral state) affect students' emotion (i.e., achievement emotions), motivation (i.e., achievement goals), and cognition (i.e., mental effort, posttest performance, and game performance) in GBL in chemistry education? And do students' achievement emotions mediate the effects? We compare three conditions to investigate these research questions: Peers' enjoyment, peers' frustration, and peers' neutral state.

Chapter 6 reviews the results presented in Chapter 2, 3, 4, and 5, discusses the implications for instruction design and theories of learning (e.g., cognitive load theory, achievement goals theory, and control-value theory of achievement emotions), reflects on the practical implications for educators, learners, game designers, and researchers, presents the limitations and suggestions for future research, and provides a general conclusion.



Chapter 2 Game-based learning has good chemistry with chemistry education: A meta-analysis

This chapter is based on:

Hu, Y., Gallagher, T., Wouters, P., van der Schaaf, M., & Kester, L. (2021). Game-based learning has good chemistry with chemistry education: A three-level meta-analysis. *Journal of Research in Science Teaching*, 1–45. https://doi.org/10.1002/tea.21765

Acknowledgement of author contributions

All authors designed the study. Yuanyuan Hu collected and analyzed the data and drafted the manuscript. All authors contributed to critical revision of the manuscript. Pieter Wouters, Marieke van der Schaaf and Liesbeth Kester supervised the study.

Acknowledgment

The authors would like to thank Caspar van Lissa, Wolfgang Viechtbauer, David van Alten, Renée Jansen, Guido Schwarzer, Michael Borenstein, Mathias Harrer, Pim Cuijpers for sharing their statistical expertise; Judith M. Conijn, Belén Fernández-Castilla, Cheryl Bodnar, Kim Chwee Daniel Tan, Kian Seh Low, Kaoru Sumi, Kermin Joel Martínez-Hernández, Jason M. Harley, José Nunes da Silva Júnior, Mike Coffey, and Su Cai for providing additional information on their studies.

Abstract

Game-based learning (GBL) may address the unique characteristics of a single subject, such as chemistry. Previous systematic reviews on the effects of GBL have yielded contradictory results concerning cognition and motivation. This meta-analysis aims to: (a) estimate the overall effect size of GBL in chemistry education on cognition (including retention), motivation, and emotion compared with non-GBL (i.e., media comparison); (b) examine possible moderators of the effects; and (c) identify the more effective game design and instructional design features (i.e., value-added comparison). We screened 842 articles and included 34 studies. This study is the first GBL meta-analysis that employed a three-level random-effects model for the overall effects. Moderator analysis used a mixed-effects meta-regression model. Results from the media comparison suggest GBL in chemistry education was more effective for cognition ($g = .70$, $k = 30$, $N = 4155$), retention ($g = .59$, $k = 20$, $N = 2860$), and motivation ($g = .35$, $k = 7$, $N = 974$) than non-GBL and the substantial heterogeneity ($I^2 = 86\%$) for cognition. No study reported emotions, and studies considering value-added comparisons of GBL with versus without specific design features ($k = 3$) were too few to perform a meta-analysis. Moderator analyses implied that except for publication source and sample size, no other moderator was related to effect sizes. There may be the small-study effects, particularly publication bias. Although we conclude that GBL enhances chemistry learning more than non-GBL, the results also make clear that additional high-quality value-added research is needed to identify design features that may further improve GBL in chemistry education. More GBL meta-analyses on subjects other than chemistry are also needed. As the first GBL meta-analysis that emphasizes emotion, we call for more research on emotion and on relationships between cognition, motivation, and emotion in GBL.

Keyword: Chemistry; Game-based learning; Meta-analysis; Cognition; Motivation

Introduction

Every time a new medium (e.g., radio, television, computer) emerges, stakeholders (e.g., teachers, students, parents, researchers, entrepreneurs, developers, administrators, policy makers etc.) expect new opportunities for education (Kirschner & Hendrick, 2020). This also applies to game-based learning (GBL) – learning with games. This expectation of GBL happens for a reason. For example, games are fun (Plass et al., 2022). Also, they are very popular: In 2020, people spent approximately 6.33 hours per week playing games worldwide (Limelight Networks, 2020). Moreover, GBL may promote diversity, equality, and inclusion in education (Stewart et al., 2013; OET, 2017).

As shown in Chapter 1, the evidence for media comparison research in GBL is inconsistent. This inconsistency may be due to the differences among the empirical studies (NRC, 2011a; Vogel et al., 2006; Wouters et al., 2013; Young et al., 2012). Most important, the effectiveness of GBL may depend on the subjects or nature of the content (Acquah & Katz, 2020; Hung et al., 2018; Rutten et al., 2012; Wouters et al., 2013; Young et al., 2012). That is, GBL should address the unique characteristics of each subject. In line with discipline-based educational research (NRC, 2012a; Rahman & Lewis, 2020), meta-analyses examining a particular subject (see Table 2.1), such as math (Byun & Joung, 2018; Tokac et al., 2019), second language (e.g., English; Chen, Tseng, & Hsiao, 2018; Thompson & von Gillern, 2020), and science (Riopel et al., 2020; Setiawan & Phillipson, 2019; Tsai & Tsai, 2020) will add value to our existing knowledge about GBL.

Perhaps due to the unique characteristics of physics, chemistry, and biology, this inconsistency is even more obvious in science GBL (Cheng et al., 2015; Klopfer & Thompson, 2020; Li & Tsai, 2013; Mayer, 2014b, 2020; NRC, 2011a; Riopel et al., 2020; Setiawan & Phillipson, 2019; Tsai & Tsai, 2020; Wouters et al., 2013; Young et al., 2012). Physics is mainly characterized by highly abstract and idealized mathematical expressions (Docktor & Mestre, 2014; Duit et al., 2007; Opfermann et al., 2017), and biology is mainly characterized by multiple and hierarchical levels of organization in living organisms (NRC, 2009; Tsui & Treagust, 2013; Wandersee et al., 2000). Although multilevel thinking plays a role in many STEM subjects such as physics and biology, chemistry is mainly characterized by multilevel thinking (American Chemical Society, 2018; de Jong & Taber, 2007; Gilbert & Treagust, 2009; NRC, 2009; NRC, 2012a). The triple nature of chemistry is difficult to learn mostly because students struggle to coordinate thinking within three unique levels of chemical knowledge: (1) macro—tangible and visible phenomena, such as chemical reactions; (2) submicro—invisible atoms, ions, molecules, or structures; and (3) symbolic—representational symbols, formulas, or equations (called Johnstone's triangle or chemistry triplet; de Jong & Taber, 2007; Gilbert & Treagust, 2009; Johnstone, 1991, 2000; Sirhan, 2007; Taber, 2009, 2013; Talanquer, 2011; Towns & Kraft, 2011).

To date, the biggest challenge in education, such as chemistry, is how to create effective, efficient, motivating, and enjoyable learning experiences (Neelen & Kirschner, 2020), particularly how to increase chemistry literacy, motivate learners to learn chemistry, and/or pursue chemistry-relevant advanced degrees and careers (European Commission, 2015b; NRC, 2011a, 2011b, 2012b, 2012c, 2014). Theoretically, this challenge could be addressed by instructional methods such as GBL (Cooper & Stowe, 2018; Klopfer & Thompson, 2020). First, interactivity in combination with multiple representations in GBL requires learners to connect all three levels of chemistry knowledge and switch from one level to another, which may help overcome learning difficulties in Johnstone's triangle (de Jong & Taber, 2007). For example, to figure out how suspects make fake coins, players must watch an animation of zinc, water, and chloride (submicro), write its chemical equation (symbolic), and conduct a gold rush experiment in virtual labs (macro; Hodges et al., 2018). In this process, GBL can also demonstrate the chemical phenomena, visualize the underlying submicroscopic processes, and show symbolic representations. Second, real-time feedback in GBL enables learners to identify chemistry content that they may be struggling with. Third, as the essential activity of GBL (Sicart, 2014), play is critical for cognitive and emotional development (Homer et al., 2020). For example, play allow learners to retain multiple representations of the same subject (Plass et al., 2015), which may also help multilevel thinking. Fourth, challenges in GBL that are neither too easy nor too difficult ask players to master certain chemistry content before moving to next level and playing games (e.g., Sokobond) is usually fun, which may help support a zone of proximal development (Vygotsky, 1978), motivate, and enjoy learning chemistry (Homer et al., 2020; Malone, 1981). Fifth, through GBL, learners can enjoy experiences free of real-life constraints and practice repeatedly. For example, although lab works play a central role in secondary (ACS, 2018) and higher education (ACS, 2015a, 2015b), sometimes they are rarely implemented due to limited curriculum time or costly infrastructures such as nuclear magnetic resonance (NRC, 2011a, 2014). GBL can create virtual laboratories or scenes to conduct scientific inquiry (e.g., HoloLAB Champions), particularly in dangerous experiments and environments that are physically inaccessible (Parker et al., 2008), which may help develop chemical practices (NRC, 2011a, 2012b). Thus, GBL has great potential to boost chemistry education.

Empirically, little attention has been paid to chemistry (Cheng et al., 2015)—the central science that connects physics and biology (Brown et al., 2018). In previous systematic reviews, most primary research in science GBL was conducted in physics or biology (Cheng et al., 2015; Li & Tsai, 2013), with only five studies in chemistry. This may be due to limitations such as examining media comparison research without value-added research (e.g., Riopel et al., 2020; Setiawan & Phillipson, 2019), a single learning outcome (e.g., achievement; Riopel et al., 2020; Setiawan & Phillipson, 2019; Tsai & Tsai, 2020; Young et al., 2012), a narrow publication source (e.g., peer-reviewed articles; Setiawan & Phillipson, 2019), and/or narrow range of grades (e.g., K-8; Setiawan & Phillipson, 2019; K-12, Young et al., 2012).

Although GBL has been emerging in chemistry education over the past 20 years, little is known about its effectiveness. Therefore, systematic knowledge is needed about whether GBL makes a difference in chemistry education and how to support GBL chemistry (Bellou et al., 2018). Hence, this meta-analysis investigates the effects of GBL in chemistry education (media comparison) and design features (value-added comparison) on cognition, motivation, and emotion (i.e., cognitive, motivational, and emotional processes and outcomes); that is, to estimate the effect size, indicate whether the effect size is consistent across empirical studies, and/or to identify more sources of diversity (Borenstein et al., 2009). To include exhaustive studies, we broadened the learning processes and outcomes, age groups, and publication sources.

Table 2.1

Overview of previous meta-analyses in game-based learning

Study	Year range	Target	Subject	Comparison	K1	Independent variable	Dependent variable	K2	ES
Byun & Joung (2018)	2000-2014	K-12	Math	Media	17	Digital game-based learning	Achievement	25	$d = .37^{na}$
Chen et al. (2020)	2008-2019	K-16	All	Value	25	Competition	Cognitive outcomes	25	$g = .37^{*}$
							Non-cognitive outcomes	82	$g = .40^{*}$
M.H. Chen et al. (2018)	2003-2014	All	English	Media	10	Digital game-based learning	Vocabulary acquisition	10	$d = 1.03^{*}$
Clark et al. (2016)	2000-2012	K-16	All	Media	69	Digital games	Cognitive learning outcomes	173	$g = .35^{*}$
							Intrapersonal learning outcomes	35	$g = .35^{*}$
				Value		Enhanced design	Learning outcomes	40	$g = .34^{*}$
Karakoç et al. (2020)	2000-2018	K-16	All	Media	35	Game-based learning	Achievement	38	$g = 1.70^{*}$
Lamb et al. (2018)	2002-2015	K-14	All	Media	28	Serious educational games	Cognition	na	$d = .67^{na}$
						Serious games	Affect	na	$d = .51^{na}$
						Simulations	Behavior	na	$d = .04^{na}$
Riopel et al. (2020)	-2020	All	Science	Media	79	Serious games incl. simulations	Declarative knowledge	65	$d = .34^{*}$
							Retention	8	$d = .31^{*}$
							Procedural knowledge	7	$d = .41^{*}$
Setiawan & Phillipson (2019)	2010-2017	K-8	Science	Media	12	Digital games	Cognitive outcomes	12	$g = .67^{*}$
Sitzmann (2011)	1976-2009	> 18 years old	All	Media	65	Computer-based simulation games	Declarative knowledge	39	$d = .28^{na}$
							Retention	8	$d = .22^{na}$
							Procedural knowledge	22	$d = .37^{na}$
							Self-efficacy	8	$d = .52^{na}$
Thompson & Gillern (2020)	-2017	K-16	English	Media	19	Video game-based learning	Vocabulary acquisition	20	$d = .70^{na}$
Tokac et al. (2019)	2000-2012	K-12	Math	Media	24	Game-based learning	Achievement	39	$d = .13^{*}$
Tsai & Tsai (2020)	2000-2018	K-16	Science	Media	26	Digital games incl. simulations	Knowledge acquisition	14	$g = .65^{*}$
				Value				13	$g = .41^{*}$
Vogel et al. (2006)	1986-2003	All	All	Media	32	Games	Cognitive gains	na	$z = 6.05^{*}$
						Interactive simulations	Attitude	na	$z = 13.74^{*}$
Wouters et al. (2013)	1990-2012	All	All	Media	39	Serious games	Knowledge	25	$d = .27^{*}$
							Skills	52	$d = .29^{*}$
							Retention	16	$d = .36^{*}$
							Motivation	31	$d = .26$
Wouters & Van Oostendorp (2013)	1990-2012	All	All	Value	29	Instructional support	Knowledge	36	$d = .33^{*}$
							Skill	32	$d = .62^{*}$
							In-game performance	38	$d = .19^{*}$

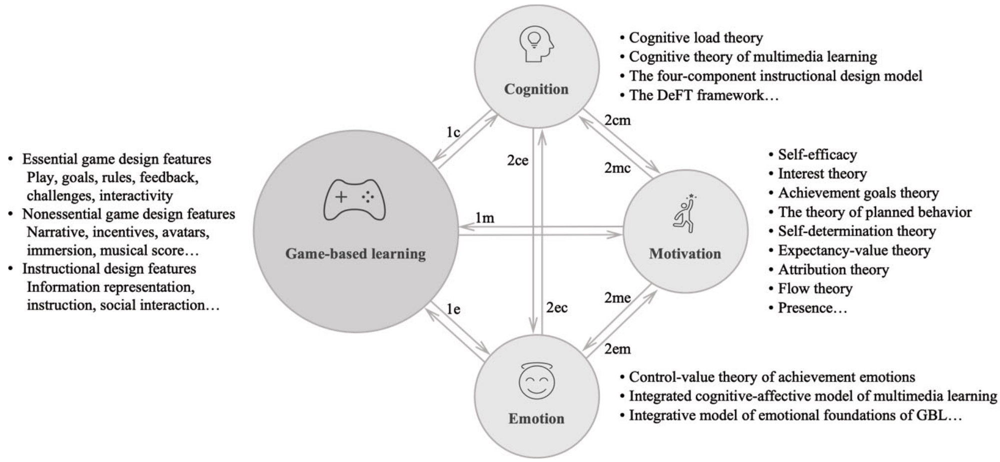
Note. K1 = Number of primary studies included in the meta-analysis; K2 = Number of pairwise comparisons in each category; ES = effect size; na = not available; * $p < .05$.

Learning Processes and Outcomes in Chemistry Game-Based Learning

Chemistry learning involves not only scientific practices (e.g., ask questions; develop and use models; plan and carry out investigations), crosscutting concepts that bridge across other disciplines (e.g., patterns; cause and effect; scale, proportion, and quantity), and chemistry core ideas (e.g., matter and its interactions; energy) but also motivation (e.g., attitude; interest) and feelings toward chemistry (e.g., emotions; ACS, 2018; European Commission, 2015b; Forsthuber et al., 2011; NRC, 2012a, 2012b, 2014, 2016; Schola Europaea, 2019). Theoretically, GBL can impact chemistry learning by affecting cognitive processes, motivation to learn, and/or emotion (NRC, 2011a; Plass et al., 2015; Plass et al., 2020), as shown in Figure 2.1. To solve the aforementioned challenges regarding low motivation to learn chemistry, we focus on motivation to learn a subject (e.g., chemistry) instead of motivation to play the game.

Figure 2.1

The theoretical framework of game-based learning



Note. Single-headed arrows represent causal relations; All numbers are explained in the text.

Cognition in Chemistry Game-Based Learning

From the cognitive perspective, the goal of chemistry education involves scientific practices related to chemistry core ideas and crosscutting concepts (ACS, 2018; NRC, 2012a, 2012b, 2014; Schola Europaea, 2019). Generally, GBL may affect cognitive processes underlying learning chemistry such as schema construction and schema automation, as displayed in Figure 2.1 line 1c. Take the previous gold rush game for example. GBL can provide multiple representations: Learners learn different representations, mentally relate representations to one another, and integrate them into coherent mental models (multimedia principle; Mayer, 2014a), which may foster the aforementioned multilevel thinking (Chiu & Wu, 2009; Wu & Shah, 2004). According to the DeFT framework (Ainsworth, 2006), multilevel thinking requires not only multiple representations but also dynamic linking between these representations (multiple representation principle; Ainsworth, 2014), which can be facilitated by interactivity in GBL. Furthermore, learning chemistry such as multilevel thinking may pose considerable cognitive demands on learners and GBL may affect three demands: essential processing aiming at mentally representing the essential material, generative processing aiming at making sense of materials, and extraneous processing that does not contribute to learning Mayer, 2014b, 2020).

For example, in HoloLAB Champions, narrative by the virtual host provides a relevant and meaningful context for scientific practices (situational learning; Plass et al., 2015; Prensky, 2001), which may facilitate essential processing; Game interactivity allows learners to learn chemistry lab skills by doing, which may facilitate generative processing (e.g., Moreno & Mayer, 2005); and Ongoing feedback assesses learners' performance and directs their attention to relevant information, which may reduce extraneous processing (Johnson et al., 2017). Given that players' cognitive capacity is limited, complex GBL, particularly when multiple representations are involved, is demanding.

Motivation in Chemistry Game-Based Learning

From the motivational perspective, the goal of chemistry education is to increase motivation to learn, complete degrees, or pursue careers in chemistry (ACS, 2018; European Commission, 2015b; NRC, 2012a, 2012b, 2014; Schola Europaea, 2019). Generally, GBL may affect players' values, needs, beliefs, attributions, and goals of learning chemistry (for an overview and comparison, see Cook & Artino, 2016; de Brabander & Martens, 2014; Mayer, 2014b, 2020; Plass et al., 2015), as displayed in Figure 2.1 line 1m. For example, according to the player experience of the need satisfaction model, features of GBL environments such as HoloLAB Champions can support basic psychological needs for autonomy (e.g., choices regarding the level of challenge, strategies, or tools), competence (e.g., experience growth and leveling up by optimal challenge), and relatedness (e.g., opportunities to contribute, communicate, and cooperate with the virtual host), resulting in intrinsic motivation and cognition (Ryan & Rigby, 2020).

Emotion in Chemistry Game-Based Learning

Emotions are also involved in chemistry learning (Jaber & Hammer, 2016; King et al., 2017; Maria et al., 2003; NRC, 2012a; Raker et al., 2019; Sinatra et al., 2014). Generally, GBL can induce different types of emotions in chemistry learning (e.g., achievement emotions, epistemic emotions, and topic emotions) by shaping their antecedents (e.g., perceived control and perceived value of the learning tasks, cognitive incongruity; for details, see Loderer et al., 2020; Plass et al., 2019), as displayed in Figure 2.1 line 1c. For example, according to CVT, the optimal level of challenges and scaffolding in HoloLAB Champions may promote a higher perceived control and value of learning chemistry, and, consequently, induce more positive achievement emotions (e.g., enjoyment) and less negative achievement emotions (e.g., boredom).

Furthermore, chemistry game designers and researchers must consider which design features facilitate cognitive process, motivation, and emotions in players because they may reciprocally affect each other (e.g., Pekrun & Linnenbrink-Garcia, 2014; Robbins et al., 2004; Talsma et al., 2018; Valentine et al., 2004), as showed in Figure 2.1 lines 2. Such influence may apply to GBL in chemistry education contexts. First, past performance in GBL (e.g., success or failure) may be the sources of motivation (e.g., self-efficacy; Bandura, 1986) and (achievement) emotions (e.g., enjoyment; Pekrun & Perry, 2014). For example, a successful performance on one chemistry game level is likely to promote higher motivation and more positive emotions on the following level. Second, GBL motivates players to invest sustained effort and time to engage in selecting, organizing, and integrating information, improving learning and emotion (Mayer, 2014b, 2019). For example, a higher motivation on a chemistry game is likely to promote higher performance and more positive emotions. Third, positive emotions in GBL induce intrinsic motivation to invest effort (Loderer et al., 2018), reduce cognitive load (Plass & Kaplan, 2015), sustain attention on relevant information (Park et al., 2015), lead to flexible and creative learning strategies (Fiedler & Beier, 2014), facilitate self-regulated learning

(Artino & Jones, 2012; Pekrun & Perry, 2014), and, consequently, increase performance (Loderer et al., 2020; Sabourin & Lester, 2014). For example, more positive emotions on a chemistry game are likely to promote higher motivation and performance. Unfortunately, data from included studies in this meta-analysis do not allow us to formulate a research question on relations between cognition, motivation, and emotion in GBL in chemistry education. Further research on GBL in chemistry education is needed to confirm this assumption.

The Present Study

Although GBL and chemistry education align, overview research regarding the learning effects and the determining factors is limited. Despite focusing on chemistry education, this meta-analysis builds on and considers some limitations from previous meta-analyses. First, we focus on unresolved issues: whether GBL is more motivating (Sitzmann, 2011; Vogel et al., 2006; Wouters et al., 2013) than non-GBL (i.e., any type of learning activities without using games, such as learning with lectures) and which design features enhance GBL (Chen et al., 2020; Clark et al., 2016; Tsai & Tsai, 2020; Wouters & van Oostendorp, 2013). Second, to avoid deviating definitions of key concepts including cognitive gains (Vogel et al., 2006), motivation (Clark et al., 2016; Lamb et al., 2018), or simulation games (Riopel et al., 2020; Setiawan & Phillipson, 2019; Sitzmann, 2011), we define GBL narrowly by the aforementioned essential game design features such as play (i.e., exclude pure simulations) and classify learning processes and outcomes into cognition, motivation, and emotion. Third, as learning content varies, game genres vary and, consequently, game effectiveness may also vary (Wouters et al., 2013). Given that game genre is critical in game design and depends on to-be-learned knowledge, we examine it and follow Chen et al. (2020) and Ke's (2016) classification of game genre such as role-playing (see Table 2.2). Fourth, some studies only included attitude (Vogel et al., 2006), self-efficacy (Sitzmann, 2011), interest, and engagement (Wouters et al., 2013) as motivation, whereas we include other motivation theories, such as expectancy-value theory and achievement goal theory. Fifth, the standard (two-level) meta-analytic model used by previous meta-analyses did not consider dependency of effect sizes within studies (e.g., one study reported multiple comparisons, or multiple measurements for the same outcome; Cheung, 2014, 2019; López-López et al., 2018). Instead, we use a three-level meta-analytic model to estimate between- and within-study variance. Sixth, although publication bias (i.e., studies with statistically significant or positive results tend to be published more often than those with not statistically significant or negative results; Rosenthal, 1979) is one of the biggest issues in meta-analyses (Fernández-Castilla et al., 2021), some studies missed many commonly used methods when detecting and correcting publication bias (e.g., Riopel et al., 2020; Sitzmann, 2011), such as the Egger test which has been shown a better correction for publication bias than other methods (Stanley & Doucouliagos, 2014; for detailed comparison of methods, see Fernández-Castilla et al., 2021; Kromrey & Rendina-Gobioff, 2006). Progress in statistical analysis techniques enables us to use more recent and reliable meta-analysis methods, such as meta-regression to detect publication bias and investigate continuous moderators such as sample size.

This meta-analysis systematically synthesizes all experimental studies that applied GBL in K-16 chemistry education by addressing the following questions:

RQ1. Is the effect of GBL in chemistry education on cognition (including retention), motivation, and emotion larger than for non-GBL (media comparison)?

For cognition, GBL changes academic knowledge, which is often measured by immediate and/or delayed tests (Mayer, 2020). As learning and instruction seek to promote the storage of learned knowledge in long-term memory that can be retrieved when needed (Bennett & Rebello, 2012; Paas & Sweller, 2014), delayed tests are advocated to determine the long-lasting impact of GBL (i.e., long-term retention) instead of fleeting knowledge improvement due to arousal (Mayer, 2014b). Furthermore, retention is a key learning outcome in chemistry education (NRC, 2012a, 2014, 2015). For motivation, that GBL are motivating is the most frequent appeal of GBL (Malone, 1981; Plass et al., 2015; Wouters et al., 2013). For emotion, decreasing boredom and increasing enjoyment is another appeal of GBL (Loderer et al., 2020). Previous meta-analyses generally found small to large effect sizes for the different cognition (Karakoç et al., 2020; Lamb et al., 2018; Riopel et al., 2020; Setiawan & Phillipson, 2019; Sitzmann, 2011; Tsai & Tsai, 2020; Wouters et al., 2013) and retention (Riopel et al., 2020; Sitzmann, 2011; Wouters et al., 2013) in favor of GBL relative to non-GBL but disagree about the effects on motivation (Sitzmann, 2011; Vogel et al., 2006; Wouters et al., 2013), and little is known about emotion (see Table 2.1). Fortunately, a meta-analysis on emotions in technology-based learning concludes enjoyment and curiosity are positively related to achievement in GBL (Loderer et al., 2018). Based on this evidence, we hypothesize the following:

Hypothesis 1. GBL in chemistry education yields higher cognition than non-GBL.

Hypothesis 2. GBL in chemistry education yields higher retention than non-GBL.

Hypothesis 3. GBL in chemistry education yields higher motivation than non-GBL.

Hypothesis 4. GBL in chemistry education induces more positive emotions and less negative emotions than non-GBL.

RQ2. Do instruction characteristics (activity level of control group, additional instruction, user grouping, and number of game sessions) and methodology characteristics (randomization, sample size, publication source, and assessment type) moderate the effect?

We identified the following instruction characteristics as moderators based on the aforementioned cognitive foundations of GBL and inconclusive results from previous meta-analyses in GBL (see Table 2.2). First, activity level of control group. According to CTML (Mayer, 2014a, 2014b, 2020) and CTL (Sweller et al., 2019), GBL fosters generative processing in which learners actively engage in selecting, organizing, and integrating new information, but this is also true for non-GBL. Active processing is key to learning; the deeper the processing of information, the more that will be retained and encoded into memory (Craik & Lockhart, 1972); thus, the difference between GBL and non-GBL may decline when non-GBL uses active instead of passive instruction. Second, additional instructions (i.e., instructions that are used together with GBL rather than non-GBL, such as pretraining before GBL). Organizing and integrating information is critical for learning, but these do not occur automatically (Mayer, 2014a, 2014b). Integrating GBL with non-game instructions (e.g., pretraining; Clark et al., 2016; Wouters et al., 2013) may facilitate articulating and integrating new knowledge with prior knowledge, leading to higher recall, transfer, and retention than standalone GBL (Merrill, 2012; Wouters et al., 2008; Wouters et al., 2013; Young et al., 2012). Third, user grouping. Playing games in groups and explaining things to each other may also facilitate knowledge articulation and, thus, the organization and integration of new information. This point is supported by the collaboration principle in multimedia learning, also known as the collective working memory effect (Kirschner et al., 2009, 2011), which states it is better to assign complex learning tasks in groups (van Merriënboer & Kester, 2014). Fourth, number of game sessions. GBL can be complex for novices. When they start playing a game, they must learn technological knowledge—game information (extraneous processing because it does not contribute to learning) and content knowledge (essential and generative processing); thus, they may easily become overwhelmed. Multiple game sessions allow the players to get familiar with the game.

Theoretically, the efficacy of GBL relative to non-GBL may improve in specific learning arrangements—such as additional instruction, group gameplay, multiple sessions, or passive instruction in non-GBL—that facilitate information processing. Empirically, previous meta-analyses only agree on the role of number of game sessions; that is, compared with non-GBL, learners benefited more when GBL involved multiple sessions (Clark et al., 2016; Wouters et al., 2013). However, the meta-studies are unsure regarding additional instruction (Clark et al., 2016; Sitzmann, 2011; Wouters et al., 2013), user grouping (Tsai & Tsai, 2020; Vogel et al., 2006; Wouters et al., 2013), and activity level of control group (Riopel et al., 2020; Sitzmann, 2011; Wouters et al., 2013). Therefore, we only formulated a hypothesis regarding the number of game sessions. For other variables, we investigated whether and to what extent they moderate the overall effect.

Hypothesis 5. Relative to non-GBL, GBL in chemistry education with multiple game sessions yield higher learning than those with one session.

To check study quality, we included the following methodology characteristics as moderators: randomization, sample size, publication source, and assessment type (see Table 2.2). Ideally, large sample sizes and randomized controlled trials (RCTs) are recommended by review organizations, such as What Works Clearinghouse (WWC, 2019). Practically, effect sizes were found to be systematically higher in quasi-experiments designs (QEDs) than in RCTs, in smaller than larger studies, and in published studies than gray literature due to methodological weaknesses and small-study effects (Cheung & Slavin, 2016; Slavin, 2008; Slavin & Smith, 2009). Although closed assessment (e.g., multiple-choice questions) is easier to implement than non-closed assessment (e.g., open-ended questions), effect sizes seem larger when using non-closed assessment (Tsai & Tsai, 2018). Again, previous GBL meta-analyses provide inconclusive results on the moderating effects of these methodology characteristics (Karakoç et al., 2020; Riopel et al., 2020; Sitzmann, 2011; Tsai & Tsai, 2018; Wouters et al., 2013). Therefore, it is valuable to check publication bias. We further evaluate the extent of bias, estimate unbiased effects, and suggest improvements for future research; simply excluding unpublished studies would ignore publication bias and overestimate the overall effects.

RQ3.Which game design or instructional design features improve GBL in chemistry education (value-added comparison)?

As displayed in Table 2.3, value-added research pinpoints design features that promote GBL by reducing extraneous processing (e.g., redundancy), managing essential processing (e.g., modality), and/or fostering generative processing (e.g., personalization; Mayer, 2020). Previous meta-analyses have suggested that some instructional design features such as modality, personalization, feedback (Tsai & Tsai, 2020; Wouters & van Oostendorp, 2013), competition (Chen et al., 2020), and/or enhanced scaffolding (e.g., personalized scaffolding based on individual learner needs; Clark et al., 2016) enhance GBL. However, other features remain unsettled, including game design features such as narrative (integrate a storyline; Wouters & van Oostendorp, 2017) or immersion (use VR), and instructional design features such as collaboration (play in groups; Clark et al., 2016), learner control (allow learners to choose game levels), or segmenting (break the materials into parts; Mayer, 2020). Given the limited research evidence, we do not formulate a hypothesis but explore the efficacy of these features in GBL in chemistry education.

Table 2.2

Coding for basic information, learning outcomes and moderator variables

Variables	Categories
Basic information	Author, publication year, grade, country, comparison, game name, chemistry topic, and assessment method
Game genre	1 = puzzle game with logic thinking, pattern recognition, objects matching, or questions answering (e.g., trivia)
	2 = action game with combat and physical challenges such as shooting
	3 = adventure game with exploring, gathering items, and solving puzzles driven by story
	4 = strategy game with system thinking and via a decision tree
	5 = role-playing game when players assume the roles of characters
	6 = simulation game which simulates reality
Learning outcomes	1 = cognition incl. factual, conceptual, procedural, and/or strategic knowledge
	2 = retention when cognition was measured in a delayed test
	3 = motivation in chemistry incl. learning attitude, interest, intrinsic motivation, self-determination, achievement goal, task value, flow, presence, and/or self-efficacy
	4 = emotion incl. enjoyment, pride, hope, anxiety, anger, shame, boredom, or hopelessness
Activity level of control group	1 = active incl. doing experiments, computer-based tutorials, assignments, or exercises
	2 = passive incl. reading textbooks, listening to lectures, or watching videos
Additional instruction	1 = the game with additional instructions [a] (e.g., pretraining before GBL, debriefing after GBL)
	2 = the game without additional instructions (i.e., GBL is standalone)
User grouping	1 = one: play the game individually
	2 = multiple: play the game in groups
No. of game sessions	1 = one: play the game once
	2 = multiple: play the game repeatedly
Sample size	The actual number of participants
Publication source	1 = gray literature incl. theses or conference proceedings
	2 = published incl. books or peer-reviewed journals
Randomization	1 = random controlled trial: randomly assign the participants (not the class) to groups
	2 = quasi-experiment design: no random assignment
Assessment type	1 = closed incl. only multiple-choice questions
	2 = non-closed incl. short-answer questions or open-ended questions with or without multiple-choice questions
	3 =mix incl. closed questions together with non-closed questions

Note. na = unknown was coded when relevant information is missing.

[a] For example, if a lecture is given before GBL, it counts as additional instruction (e.g., pretraining); if it is given without GBL, it counts as non-GBL.

Table 2.3

Features that promote GBL

Features	Descriptions	Cognitive processing
Pretraining	Provide trainings on key concepts and characteristics before gameplay	Essential processing
Modality	Present words in spoken rather than written forms	
Personalization	Use conversational rather than formal styles	Generative processing
Feedback	Add explanations and advice to corrective feedback	
Self-explanation	Provide prompts to self-explain the performance during gameplay	
Competition	Play the game against virtual components, time, or other players	
Redundancy	Eliminate redundant information, such as written text from spoken texts or pictures	Extraneous processing

Method

Literature Search

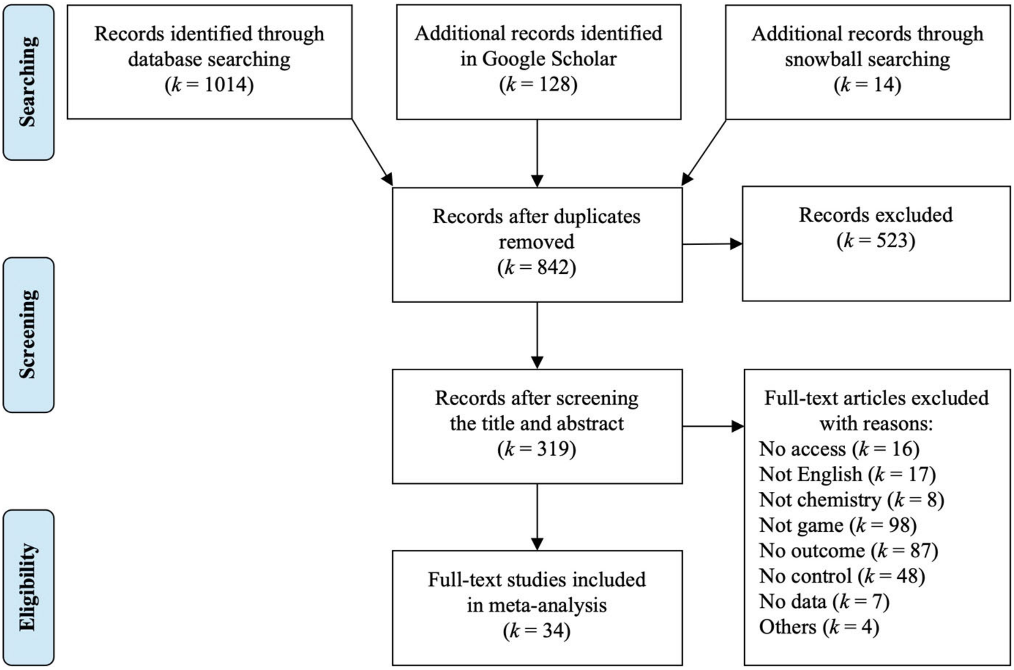
The PRISMA flow diagram (Moher et al., 2009) in Figure 2.2 summarizes the process of the literature search and selection. The search comprised three parts. First, searching databases: Web of Science, Scopus, Eric, and PsycINFO. The search terms were combined in English: "(chemistry or chemical) AND (game or immersive learning or virtual reality or virtual environment or augmented reality or augmenting reality or mixed reality) AND (learning or cognit* or achievement or interest or attitude or engage* or motivation* or involvement or enjoy* or emotion* or affective)", which should be in the title, abstract, or the list of keywords. In case the term "game" was not in the title, abstract, or keywords, we included immersive learning technologies commonly used in games as the search terms, such as VR. The period searched was from 2000, when GBL studies changed dramatically after that due to technological development (Parker et al., 2008) and GBL started becoming popular in chemistry, to January 2020. Second, a Google Scholar search for gray literature (Haddaway et al., 2015). Finally, a snowball search in the reference lists and citations of the aforementioned meta-analyses and systematic reviews. Overall, 1156 records came out, and 842 remained after removing duplicates.

Inclusion and Exclusion Criteria

Studies were evaluated based on (a) language, (b) subject, (c) participants, (d) accessibility, (e) comparison, (f) independent variable, (g) dependent variable, (h) data, and (i) others (Table 2.4). Study inclusion followed two stages (Figure 2.2). First, in the screening stage, all the titles and abstracts were screened based on a, b, f, and h, which led to 319 records being included and 523 excluded. If the study did not state whether it fulfilled these four criteria, the researchers included it and made a further decision based on its full text in the next stage. To assess the inter-rater reliability (IRR), 91 records (>10% of 842) were randomly chosen and screened independently by the first two authors, with Cohen's $k = .83$. Any disagreements were resolved by discussing and consulting with the third author. Second, the full texts of 319 records were retrieved and evaluated using all nine criteria. The first two authors randomly selected 34 full texts (>10% of 319) and assessed them independently. Despite two disagreements regarding the reason for exclusion, a perfect IRR was reached (Cohen's $k = 1$). In total, we included 34 articles.

Figure 2.2

PRISMA flow diagram of literature search and records selection process



Coding

To conduct a quantitative analysis and provide a qualitative description of all the included studies, we collected the following data: basic information, game genre, learning outcomes, and moderator variables (Table 2.2). The coding process was performed in a standardized way. First, a trial coding with five studies was run to evaluate whether all possible situations for a moderator variable were covered by a category. Then, a random sample of four studies (>10% of 34) was coded independently by the first two authors. A satisfactory IRR with Cohen's k = 1 was reached for all variables, except for number of game sessions (Cohen's k = .5). Any disagreements were discussed until agreement was reached. For this variable, another four articles were randomly chosen and coded by the first two authors with IRR of Cohen's k = 1. Finally, the first author coded the remaining studies. A detailed coding for moderators is available in Table S2.1 in the supplementary materials.

Table 2.4

Inclusion and exclusion criteria

Criteria	Inclusion	Exclusion
a. Language	English	Language other than English
b. Subject	Chemistry	Integrated science
c. Participants	Nondisabled K-16 students	Preservice teachers, student teachers, or employees
d. Accessibility	The full text of the study is accessible	No access via on internet or contacting authors
e. Comparison	Game vs. nongame or game with vs. without specific features	No control group, control group without learning the same subject matter
f. Independent variable	Game-based learning or serious games with the term "game" appearing in the article	Games for entertainment, or educational applications, technologies, or tools (e.g., VR, AR, MR) without using the term "game"
g. Dependent variable	Cognition, motivation, or emotion	Introduction, assessment, motivation, or perception of the game
h. Data	Sufficient data to calculate effect size	Case study, no empirical data, or enough data
i. Others		The study published in conferences or theses was updated and later published in journals

Calculating Effect Sizes

Standardized mean difference (mean difference between experiment groups and control groups divided by the pooled standard deviation) was adopted as effect size (Borenstein et al., 2009), and *g* (Hedges, 1981) was calculated using Comprehensive Meta-Analysis (CMA v3; Borenstein et al., 2013). The effect size for each study was computed in a hierarchical order: the first is raw data (mean and standard deviation) and the second is data from inferential statistics (e.g., *t* or *F* value; Wouters et al., 2013).

The following complex cases were evaluated cautiously. First, when studies used a pretest–posttest control group design ($k = 23$), the preexisting difference between the experimental and control groups should be considered, and thus, the effect size was estimated on both pretest and posttest data (Morris, 2008). In practice, effect sizes were computed by post-pre mean difference of experimental group minus post-pre mean difference of control group divided by the pooled standard deviation of post-scores. In addition, when different sample sizes in the pretest and posttest were reported, the smaller sample size was adopted (da Silva Júnior et al., 2018).

Second, when studies used multiple experimental or control groups ($k = 5$), the recommended solution was to form multiple pairwise comparisons and calculate multiple effect sizes (e.g., GBL vs. concept mapping and GBL vs. conventional lecture; Okonkwo, 2012).

Third, when studies recorded multiple measurements of the same outcome ($k = 8$), multiple effect sizes were calculated, as suggested by Cheung (2014, 2019) and López-López et al. (2018)). For instance, knowledge comprehension and knowledge application that were assessed separately were calculated separately as the indicator for cognition (e.g., Chen et al., 2014; Chen & Liao, 2015). Intrinsic motivation, self-determination, self-efficacy, grade motivation, and career motivation were calculated separately as the indicator for motivation in chemistry (e.g., Meesuk & Srisawasdi, 2014; Srisawasdi & Panjaburee, 2019).

Another special case was Johnson-Glenberg et al. (2014) using the AB-BA design: Two groups were given GBL intervention (A) and regular instruction (B) in different sequences. Group 1 received pretest, SMALLab GBL intervention, mid-test, regular instruction, and posttest, while group 2 received regular instruction first after pretest and SMALLab GBL intervention after mid-test. When the mid-test was conducted, groups 1 and 2 had only received GBL intervention (SMALLab) and regular instruction, respectively. Therefore, the treatment before the mid-test was considered as a single-pair comparison (group 1 as GBL group and group 2 as control group) and the mid-test instead of the "posttest" at the end of the study was taken as the posttest.

Data Analysis

All statistical analyses were run via the "metafor" package (version 3.1.8; Viechtbauer, 2010) in R (version 4.1.0), except the distribution of variance across levels that was run via the "dmetar" package (Harrer et al., 2019). Because some studies reported multiple measures of the same construct (e.g., cognition was measured by knowledge comprehension and knowledge application) or multiple comparisons (e.g., two GBL groups vs. one non-GBL group), multiple effect sizes could arise per study and these effect sizes are dependent within studies. Separate meta-analyses were performed for cognition, retention, and motivation using the random-effects three-level meta-analytic model (Cheung, 2014, 2019; López-López et al., 2018). The three-level meta-analytic model includes sampling variance (level 1), within-study variance (level 2), and between-study variance (level 3). Heterogeneity was assessed using Cochran's Q test, and I^2 and τ^2 statistics. We used the Knapp and Hartung adjustment (Knapp & Hartung, 2003) to control the Type I error rate (Viechtbauer et al., 2015) and the restricted maximum likelihood method (López-López et al., 2014).

Following previous meta-analyses, the interpretation of the magnitude of the overall effect size was based on the benchmark identified by Cohen (1988): .2 = small, .5 = medium, and .8 = large, although his standard has some limitations such as not considering methodological features (Cheung & Slavin, 2016; Lipsey et al., 2012). Considering these limitations of Cohen's (1988) criteria, the magnitude of effect size of individual studies was also evaluated based on the benchmarks identified by Cheung and Slavin (2016): .30 = average for studies with small sample size (< 250) and .16 = average for studies with large sample size (≥250).

Sensitivity analysis was conducted by checking whether the study's confidence interval overlaps with that of the pooled effect size and calculating standard deviations ($z \leq -3$ or $z \geq 3.0$ are outliers). Publication bias was visualized by plotting the observed standardized mean differences against their standard errors and tested by funnel plot test (using sample size as a predictor of effect sizes; Macaskill et al., 2001), Begg's rank correlation test (using variance and sample size as a predictor of effect sizes; Begg & Mazumdar, 1994), trim-and-fill method (using L0+ as the number of unavailable effect sizes due to publication bias; Duval & Tweedie, 2000a, 2000b), and an adapted version of Egger's regression test (using sampling variance as a predictor of effect sizes; de Jong et al., 2021; Egger et al., 1997; Fernández-Castilla et al., 2021; Knapp et al., 2017; Sterne et al., 2011; Viechtbauer, 2017). The existence of publication bias is indicated by a statistically significant test and L0+ > 3 (see Fernández-Castilla et al., 2021). The adapted version of Egger's regression test models a quadratic relationship between the standard errors and the standardized mean differences and the intercept of this model is the overall effect size free of publication bias (see Stanley & Doucouliagos, 2014).

Following de Jong et al. (2021) and Knapp et al. (2017), a three-level mixed-effects model was run for moderator analysis. Two groups of moderators were analyzed to decrease the risk of making a Type I error or Type II error caused by testing moderators individually or simultaneously (Jansen et al., 2019; van Alten et al., 2019). Due to missing values, the activity level of control group, additional instruction, and assessment type were omitted from group analysis and analyzed individually.

Results

Descriptive Findings

This meta-analysis included 34 studies published from 2006 to 2020. Their characteristics are in Table 2.5 in the supplementary materials. The sample sizes ranged from 40 to 470. Thirty studies used media comparisons, while only three made value-added comparisons. For learning outcomes, no study reported emotions, nine reported motivation, 33 reported cognition, and six reported both cognition and motivation, but only three reported both an immediate test and a delayed test (retention). For educational level, all studies were implemented in secondary schools (k = 21) and universities (k = 13). Regarding country, one third was conducted in the United States (k = 11) and one in eight in China (k = 4). For chemistry content, the most common topics were nomenclature (k = 8), periodic table (k = 4), and organic chemistry (k = 4). Regarding assessment methods, tests and questionnaires were the most frequent measures for cognition and motivation; only five studies adopted mixed-method research (e.g., tests combined with interviews), among which one retrieved log data. Regarding game genre, the most used genres were puzzle (k = 12), simulation (k = 7), and role-playing games (k = 6). A detailed example of GBL activities for each game genre is available in Table S2.2 in supplementary materials.

Table 2.5

Characteristics of included studies in the meta-analysis (n=34)

Study	*N*	Outcome	Comparison	Grade	Country	Game name	Game genre	Topic	Assessment method
Akkuzu & Uyulgan (2016)	62	Cognition	Media	Higher	Turkey	OrCheTaboo	Puzzle	Functional group	Test Interview
Cahyana et al. (2017)	40	Cognition	Media	Secondary	Indonesia	na	na	Reaction rate	Test
Cha et al. (2017)	198	Cognition Motivation	Media	Higher	Malaysia	Brainteaser	Puzzle	Organic chemistry	Test Questionnaire
Chee & Tan (2012)	77	Cognition	Media	Secondary	Singapore	Legends of Alkhimia	Role-playing	Properties of substances	Test
Chen & Liao (2015)	76	Cognition	Dynamic-AR Strategy	Secondary	China	Manufacturing Iron Man	Simulation	Chemical cell	Test
Chen et al. (2014)	105	Cognition Motivation	Worked example	Secondary	China	The Alchemist's Fort	Role-playing Adventure	Chemical reactions	Test Questionnaire
Chimeno et al. (2006)	40	Cognition	Media	Higher	US	The Rainbow Wheel The Rainbow Matrix	Puzzle	Nomenclature	Test
da Silva Júnior et al. (2018)	246	Cognition	Media	Secondary	Brazil	Say My Name	Puzzle	Organic nomenclature	Test
Daubenfeld & Zenker (2015)	46	Cognition	Media	Higher	Germany	na	Adventure	Equilibria	Test
Fatokun et al. (2016)	96	Cognition Retention	Media	Secondary	Nigeria	Element Card I Atomic Radius Card Ionization Card Group Fixing SPD – Game Sorting – Out Transition Element Card Throw and Answer	Puzzle	Periodicity (Periodic table)	Test
Gupta (2019)	67	Cognition Retention	Media	Higher	US	Molebots	Action	Nomenclature	Test Questionnaire
Halpern et al. (2012)	136	Cognition	Media	Higher	US	Operation ARA	Role-playing	na	Test

Study	*N*	Outcome	Comparison	Grade	Country	Game name	Game genre	Topic	Assessment method
Hodges et al. (2018)	351	Cognition	Media	Secondary	US	Blended reality environment	Simulation Role-playing	Redox reaction	Test Interview Data logs
Jagodziński & Wolski (2015)	200	Cognition Retention	Media	Secondary	Poland	Virtual chemical laboratory	Simulation	Inorganic acids	Test
Joag (2014)	104	Cognition	Media	Secondary	India	na	Puzzle	Periodic table	Test
Johnson-Glenberg et al. (2014)	51	Cognition	Media	Secondary	US	SMALLab	Simulation	Titration	Test
Kavak (2012)	49	Cognition	Media	Secondary	Turkey	ChemOkey	Puzzle	Nomenclature	Test
Lay & Osman (2018)	138	Cognition Motivation	Media	Secondary	Malaysia	MyKimDG	na	Precipitation reaction	Test Questionnaire
le Maire et al. (2018)	210	Cognition	Media	Higher	Belgium	Clash of Chemists	na	Stoichiometry	Test
Low (2010)	75	Cognition	Media	Secondary	Singapore	SynTactic©	Strategy	Organic synthesis	Test Questionnaire Interview Observation
Martin & Shen (2014)	70 61 68 69	Cognition	Media Aesthetic Choice Competition	Higher	US	Element Solitaire	Puzzle	Periodic table	Test
Martinez-Hernandez (2010)	40	Cognition	Media	Higher	US	Electrolysis room Ammonia synthesis The hidden key Marble blocks Light sensor challenge Processing room	Adventure	State of matter Stoichiometry Chemical equilibrium Neutralization Redox reaction	Test Questionnaire Interview
Meesuk & Srisawasdi (2014)	87	Motivation	Media	Secondary	Thailand	SAGOI The IE war	Action	Ionization	Questionnaire
Merchant et al. (2013)	382	Cognition	Media	Higher	US	Second Life® The Molecule Game Chemist as an Artist The Tower of VSEPR Theory	Simulation	VSEPR theory	Test
Okonkwo (2012) 1 2	 234 233	Cognition Motivation	Media	Secondary	Nigeria	Simulation-game	Role-playing	Pollution and waste management	Test Questionnaire

Study	N	Outcome	Comparison	Grade	Country	Game name	Game genre	Topic	Assessment method
Rastegarpour & Marashi (2012)	105	Cognition	Media	Secondary	Iran	na	na	Nomenclature	Test
Renner (2014)	78	Cognition	Media	Secondary	US	na	Simulation	Alpha, Beta and Gamma radiation	Test
Sousa Lima et al. (2019)	144	Cognition	Media	Secondary	Brazil	Chemical Nomenclature	Puzzle	Organic nomenclature	Test
Srisawasdi & Panjaburee (2019)	62	Cognition Motivation	Media	Secondary	Thailand	Factory Game	Role-playing	Properties of substances	Test Questionnaire
Stringfield & Kramer (2014)	120	Cognition Motivation	Media	Higher	US	Who Wants an A in General Chem	Puzzle	General organic biochemistry	Test Questionnaire
Su & Cheng (2019)	72	Cognition	Media	Secondary	China	Virtual chemical laboratory	Simulation	CO_2 gas collection	Test
Sugiyarto et al. (2018)	64	Cognition	Media	Secondary	Indonesia	Chemondro	Puzzle	Nomenclature	Test
Weng et al. (2015)	135	Cognition Motivation	Media	Secondary	China	na	Action	Periodic table	Test Questionnaire
Wood & Donnelly-Hermosillo (2019)	470	Cognition	Media	Higher	US	Topinomica	Puzzle	Nomenclature	Test Questionnaire Observation

Note. N = total sample size; g = effect size; na = not available; 1 = game vs. concept mapping; 2 = game vs. conventional lecture.

Research Question 1: Media Comparison

Sensitivity Analysis

One comparison by Okonkwo (2012)—GBL vs. conventional lecture—is an outlier based on its extremely large effect size (g = 5.34 and g = 3.13 for cognition and motivation) and 3 standard deviations larger than the mean (z = 5.5 for cognition and z = 3.9 for motivation). Furthermore, its 95% CI does not overlap with that of the summary effect. Thus, the study was excepted for further analysis.

Distribution of Effect Sizes

Effect sizes of cognition and motivation for the individual studies and their distribution are presented in forest plots (Figures 2.3 and 2.4). No results were found for emotion. As displayed in Table 2.6, regarding cognition, the effect sizes (k = 30, *#ES* = 57) vary substantially across the studies, from −.62 to 1.84. Among the 53 positive outcomes, 40 are statistically significant. Among the five large-scale studies (sample size ≥250), three reported an equal or above average effect size (≥.16) and among the 52 small-scale studies (sample size <250), 46 reported above average effect size (≥.3), according to Cheung and Slavin (2016). The mean effect size is statistically significant and medium (g = .70; 95% CI [.51; .89]), according to Cohen (1988). Regarding retention, a similar effect (g = .59; 95% CI [.35; .83]; k = 20, *#ES* = 31) is found. Regarding motivation, the effect sizes (k = 7, *#ES* = 21) vary from −.09 to 1.18. Among the 19 positive outcomes, eight are statistically significant. Only two reported a negative but not statistically significant effect; the mean effect is statistically significant but small (g = .35; 95% CI [.19; .50]). The model fit of this three-level model was statistically significantly better than the two-level model than the two-level model that does not consider within-study variance (cognition: χ^2 = 17.20, p < .001; retention: χ^2 = 5.88, p = .01; and motivation: χ^2 = 6.88, p = .009).

Table 2.6

Results of random-effects meta-analysis in media comparisons

Variable	k (N)	*#ES*	g	*SE*	95% CI	Q	τ^2_{level2}	τ^2_{level3}	I^2_{level2}	I^2_{level3}
Cognition	30 (4155)	57	.70	.10	[.51; .89]	415.2*	.10	.16	33%	53%
Retention	20 (2860)	31	.59	.12	[.35; .83]	240.4*	.06	.22	18%	70%
Motivation	7 (974)	21	.35	.08	[.19; .50]	49.14*	.05	.01	49%	12%

Note. k = number of studies; N = total sample size; *#ES* = number of effect sizes; g = mean effect size; *SE* = standard error; CI = confidence interval; Q = heterogeneity value; τ^2_{level2} = within-study variance; τ^2_{level3} = between-study variance; I^2_{level2} = within-study heterogeneity index (%); I^2_{level3} = between-study heterogeneity index (%); $^*p$ < .05.

Figure 2.3

Forest plot for cognition

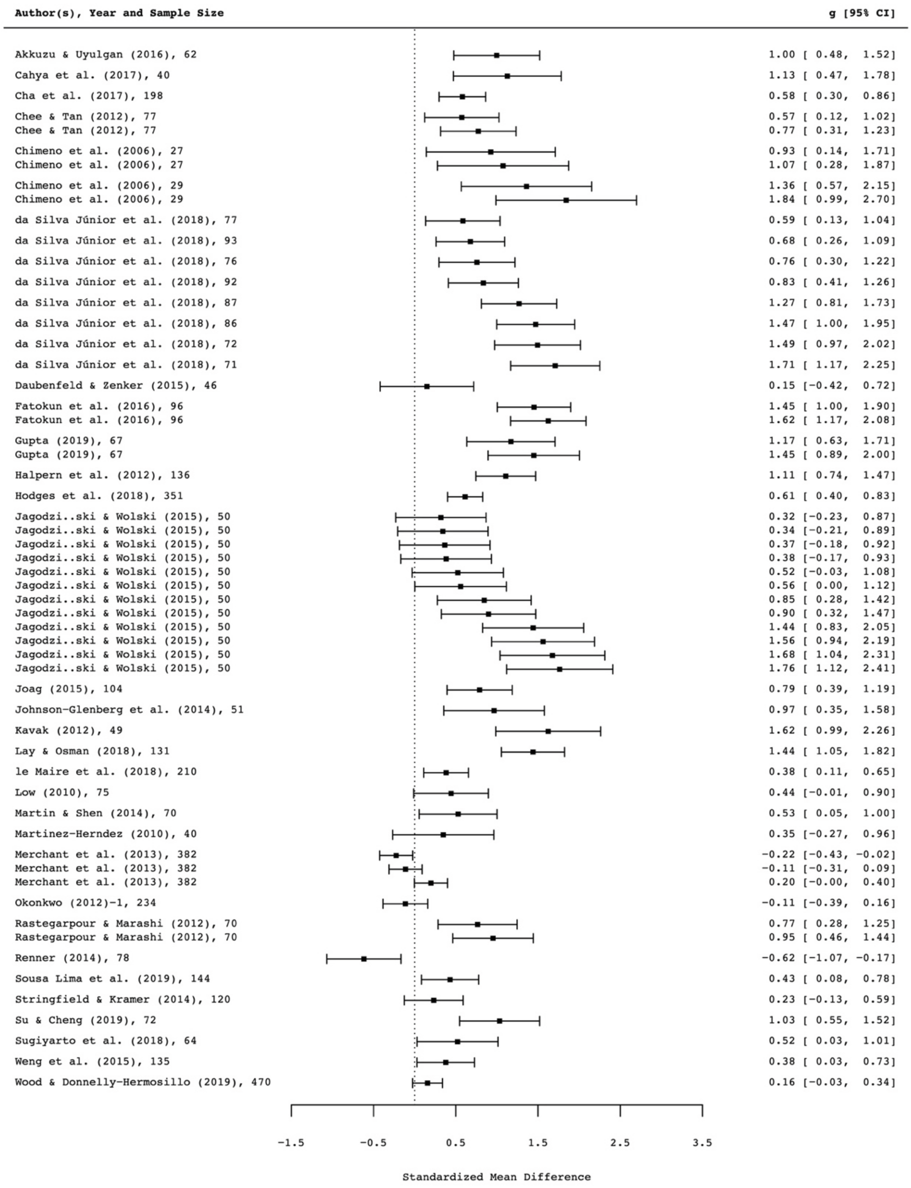
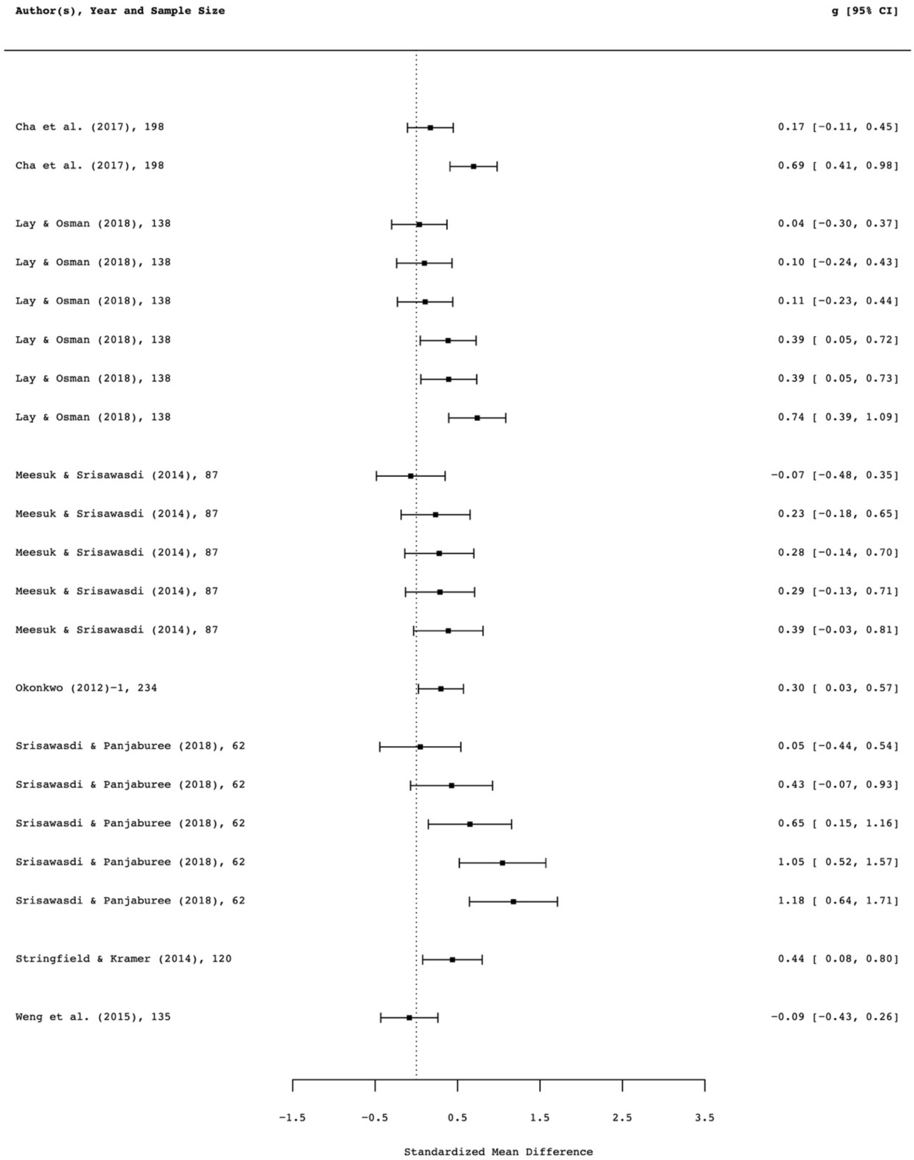


Figure 2.4

Forest plot for motivation



Heterogeneity

As displayed in Table 2.6, the statistically significant Q values reveal that effect sizes vary. For cognition, The I^2 reflects that sampling variance, within-study variance, and between-study variance can explain 14%, 33%, and 53% of the observed variance, respectively (Borenstein, 2019). The similar results are found for retention. For motivation, I^2 indicates that sampling variance, within-study variance, and between-study variance could explain 39%, 49%, and 12% of the observed variance, respectively. Thus, moderator analysis is required to inspect sources for heterogeneity for cognition but not for motivation since it is not informative to analyze only seven studies.

Publication Bias

Similarly, the risk of publication bias was assessed for cognition (Banks et al., 2012; Sterne et al., 2011). In the funnel plot of Figure S2.1 in supplementary materials, an obvious asymmetry was found: The outlying studies were distributed in the middle and upper parts, but studies in the bottom of the left side were missing (Borenstein et al., 2009). This was supported by funnel plot test ($p = .0017$), Begg's rank correlation test (using variance, $p = .0009$; using sample size, $p = .02$), trim-and-fill method (L0+ = 6), and the adapted Egger's regression test ($p < .0001$, see Figure S2.2 in supplementary materials; Fernández-Castilla et al., 2021). The effect free of publication bias can be estimated by the intercept of this model (Knapp et al., 2017; Stanley & Doucouliagos, 2014), that is, .09 (95% CI [−.22; .40]), which is not statistically significant and substantially smaller than the original estimate ($g = .70$; 95% CI [.51; .89]). Furthermore, moderator analyses on publication source and sample size indicate a statistically significant larger effect in published studies than in gray literature and in small-scale studies than in large-scale studies (small-study effects; Banks et al., 2012; Borenstein et al., 2009; Sterne et al., 2011; Table 2.7). Overall, we conclude that small-study effects were likely to happen and probably caused by publication bias (Borenstein, 2019; Borenstein et al., 2009).

Research Question 2: Moderator Analysis

As displayed in Table 2.7, at least one moderator in methodology characteristics exhibits a statistically significant relationship with the effect size ($p < .0001$). Specifically, sample size is negatively related to effect size (estimated coefficient = −.003, $p < .0001$), and publication source is positively related to effect size (estimated coefficient = .62, $p = .001$), while accounting for all the other variables in the model, that is, small-scale studies or published studies are associated with larger effects. Aside from these two variables, no other moderator exhibits a statistically significant relationship with the effect size.

Research Question 3: Value-added Comparison

As displayed in Table 2.8, only three studies reported six effect sizes for cognition, ranging from −0.11 to .46, but none reached statistical significance. The only statistically significant but very small effect is in the study that compares the effect of worked examples on motivation. A meta-analysis was impossible as the comparisons differed strongly, ranging from manipulations of aesthetic, choice, competition, worked example, guiding strategy, and type of AR.

Table 2.7

Results of moderator analysis for cognition

Moderator	*k*	*#ES*	Estimate	*SE*	p^a	95% CI	p^b
Instruction characteristics							.76
Intercept	30	57	.73	.24	.01	[.09; 1.02]	
User grouping							
Multiple (reference)	8	17					
One	22	40	.04	.20	.86	[–.36; .44]	
No. of game sessions							
Multiple (reference)	8	18					
One	22	39	.17	.22	.45	[–.27; .60]	
Activity level of control group							.55
Intercept	24	49	.60	.17	.0003	[.27; .92]	
Active (reference)	10	16					
Passive	14	33	.13	.22	.55	[–.30; .57]	
Additional instruction							.38
Intercept	26	51	.87	.27	.001	[.33; 1.41]	
No (reference)	4	5					
Yes	22	46	–.26	.39	.38	[–.84; 0.32]	
Methodology characteristics							<.0001
Intercept	30	57	.49	.19	.01	[.11; 0.87]	
Sample size	30	57	–.003	0	<.0001	[–.004; –.001]	
Publication source							
Gray literature (reference)	6	7					
Published	24	50	.62	.19	.001	[.25; 1.00]	
Randomization							
Quasi-experiment design (reference)	24	49					
Random controlled trial	6	8	.14	.18	.43	[–.22; .50]	
Assessment type							.63
Intercept	22	48	.59	.14	<.0001	[.31; .87]	
Closed (reference)	14	37					
Mix	3	3	.25	.37	.51	[–.48; .98]	
Non-closed	5	8	.24	.30	.42	[–.34; .82]	

Note. *k* = number of studies; *#ES* = number of effect sizes; *SE* = standard error; CI = confidence interval; p^a tests 1) the null hypothesis that the mean effect size in the reference group is zero (for the intercept row) and 2) the null hypothesis that the difference between that subgroup's mean effect and the reference group's mean effect is zero (for the individual subgroup row); p^b tests the null hypothesis that none of the moderators is related to effect size.

Table 2.8

Results of studies with vs. without specific features in value-added comparisons

Study name	Type of Feature	Comparison	Outcome	*N*	*g*	*SE*
Martin & Shen (2014)	Game feature	Aesthetic vs. no aesthetic	Cognition	61	.457	.256
	Instructional feature	Choice vs. no choice	Cognition	68	.143	.240
	Instructional feature	Competition vs. no competition	Cognition	69	–.112	.239
Chen & Liao (2015)	Instructional feature	Procedure-guided vs. question-guided strategy	Cognition	76	.437	.230
	Game feature	Static-AR vs. dynamic-AR	Cognition	76	.397	.230
Chen et al. (2014)	Instructional feature	Worked example vs. no worked example	Cognition	105	.257	.195
			Motivation	105	.144*	.195

Note. N = sample size; g = effect size; SE = standard error; $*p < .05$

Discussion

This meta-analysis indicates that GBL may address the unique characteristics of chemistry. Essential game design features such as interactivity, challenges, play, and feedback may address the challenges in chemistry education such as low performance, low level of motivation, and the occurrence of negative emotion. Our three-level random-effects model showed that overall, the effect of GBL in chemistry education on cognitive and motivation is larger than for non-GBL. Among instruction characteristics (activity level of control group, additional instruction, user grouping, and number of game sessions) and methodology characteristics (randomization, sample size, publication source, and assessment type), publication source and sample size moderate the effect. Evidence for emotion and game design features and instructional design features that improve GBL in chemistry education is insufficient.

Research Question 1: Media Comparison

Our first goal was to examine whether GBL in chemistry education has a larger effect on cognition, motivation, and emotion than non-GBL (RQ1). Most studies focused on cognition, providing promising evidence that GBL enhances chemistry learning, but did not include motivation, providing moderate evidence that GBL motivates interest in chemistry. No evidence is available on whether GBL increases positive emotions or decreases negative emotions as no study reported emotion. Compared with previous GBL meta-analyses, this study is the first meta-analysis that uses a three-level random-effects model to consider the dependency of effect sizes within studies and the first that emphasizes emotion in GBL.

First, this study confirms GBL in chemistry education is more effective for cognition (Hypothesis 1) and retention (Hypothesis 2) than non-GBL. The mean effects for cognition (g = .70) and retention (g = .59) reveal a statistically significant medium (g > .5; Cohen, 1988). In other words, the score of the average person in GBL in chemistry education would be .6 SD above non-GBL, exceeding 73% of students in non-GBL (Coe, 2002). The effect for cognition is larger than most previous GBL meta-analyses across all subjects in general and math and science in particular, but equal or smaller than those in English (Table 2.1). This effect is also comparable with previous meta-analyses particular to chemistry with other educational interventions (cooperative learning: g = .59, Apugliese & Lewis, 2017; g = .68, Warfa, 2016; cooperative learning, collaborative learning, problem-based learning, process oriented guided inquiry learning, peer-lead team learning, and flipped instruction: d = .62, Rahman & Lewis, 2020). Most importantly, the effect for retention is larger than all previous meta-analyses. This implies that .59 could be a benchmark for a meaningful effect in chemistry education.

Three reasons could explain the differences in the magnitude of the overall effects between current and prior meta-analyses. One reason is that chemistry has a special relationship to GBL: GBL better align with the key characteristics of chemistry education (see Introduction) than other subjects. If that is the case, policymakers and practitioners should implement GBL in chemistry education. Second, technology development: more sophisticated technologies improve learning. Studies included in this meta-analysis, published from 2006 to 2020, are more recent than those from previous meta-analyses, ranging from 1990 to 2012 (Clark et al., 2016; Wouters et al., 2013). During the past decade, new technologies have emerged (Chen, Wang, et al., 2018). Our included studies applied many sophisticated technologies. For instance, based on voice recognition, eye movement, and brain wave analysis, Natural User Interface is used in gaming consoles (Jagodziński & Wolski, 2015); real-time data capture system is used in blended reality environment (Hodges et al., 2018); different types of automated tutoring based on student performances are combined with interactive dialogs with avatars (Halpern et al., 2012); VR simulates experiential learning (Su & Cheng, 2019); and MR benefits embodied learning (Johnson-Glenberg et al., 2014). These sophisticated technologies may better support GBL in chemistry education. Furthermore, students now have better access to technologies, leading to less difficulty playing chemistry games. Third, with the development of instructional design, current GBL in chemistry education may be better embedded in learning theories than older ones. More attention is paid on integrating game design and instructional design when designing GBL in chemistry education (e.g., Mayer, 2014b; NRC, 2011a; Plass et al., 2015) as most studies are from a later period. Nevertheless, GBL in chemistry education can enhance cognition, and the effect lasts over time.

This study also suggests GBL in chemistry education is more motivating than non-GBL (Hypothesis 3). Different from Wouters et al. (2013; d = .26, p > .05), a small but statistically significant effect (g = .35; g > .2; Cohen, 1988) for motivation was found. In other words, the motivation score of the average person in GBL in chemistry education would be .4 SD above non-GBL, exceeding 62% of students in non-GBL (Coe, 2002).

This finding seems to refute the critique that GBL may attract students, but higher motivation does not necessarily mean higher learning. Even though students report liking or having interest in the medium (the game), they tend to perceive that it provides an easier path to learning and invest less mental effort and time (Salomon, 1984), resulting in less learning compared with learning without the medium (Clark & Feldon, 2014). In our case, two included studies confirm this critique: students prefer GBL to study guides (Wood & Donnelly-Hermosillo, 2019) or traditional lectures (Stringfield & Kramer, 2014), but no difference in achievement was found. However, two other studies support our finding that GBL promotes both achievement and motivation to learn chemistry (Cha et al., 2017; Srisawasdi & Panjaburee, 2019). Nevertheless, given that only seven studies reported motivation, this result should be interpreted with caution.

The cognitive and motivational benefits of GBL in chemistry education cannot prove a causal relationship between cognition and motivation. In the studies, only five reported both outcomes and their research methods, one-time pre-posttest design or posttest-only design, may not provide the required evidence. Instead, the cross-lagged panel model aims to detect causal or reciprocal relationships between variables, analyzing longitudinal data collected by testing or recording subjects at multiple points over time (Hamaker et al., 2015; Mulder & Hamaker, 2021; Selig & Little, 2012). In our included studies, no such method was used. Thus, whether cognition causes higher motivation in GBL in chemistry education and whether motivation causes higher cognition remains open questions.

Research Question 2: Moderator Analysis

Our second goal was to examine the possible moderating effects of instruction and methodology characteristics, that is, the conditions under which GBL is more effective relative to non-GBL (RQ2). We found some evidence that methodology characteristics moderate the effects, particularly sample size and publication source. Compared with previous GBL meta-analyses, this study uses more advanced methods to detect and correct publication bias and includes a continuous moderator (i.e., sample size).

Larger effects may be associated more with published studies than gray literature and with smaller studies than larger ones. The small-study effects, particularly publication bias, tent to exist. Researchers in chemistry education and GBL should attend to this issue, given that similar findings were also reported by previous meta-analyses particular to chemistry with other educational interventions (e.g., Rahman & Lewis, 2020; Warfa, 2016) and by meta-analyses in GBL (e.g., Lamb et al., 2018; Riopel et al., 2020; Sitzmann, 2011). However, more standardized methods with high statistical power are needed to assess and control how they impact main effects (e.g., the trim-and-fill method imputes adjusted effect size) and other aspects in multilevel meta-analyses (P. Cuijpers, personal communication, April 20, 2020). For instance, should we add sample size or publication source as covariate of the main effect? How and to what extent do small-study effects influence the moderator analysis?

Other moderators did not reveal statistically significant effects. Effect sizes of cognition were equal between non-GBL with active vs. passive instructions, GBL with vs. without additional instructions, GBL with one vs. multiple sessions (Hypothesis 5), GBL individually vs. in groups, RCTs versus QEDs, or with closed question vs. non-closed questions. Given the small number of studies under moderator categories, these results should never be interpreted as evidence that the effects are the same across subgroups or that there is no relation between the effects and included moderators (Borenstein et al., 2009). Instead, further studies are needed for more reliable evidence. Take randomization, for example, it is premature to conclude that larger effects are associated with RCTs than QEDs based on six RCTs versus 24 QEDs. Moreover, it is difficult to explain why we did not find statistically significant results for those moderators due to the limitations of all meta-analyses.

Other variables may help explain the potential sources of between-study variance. Unfortunately, the number of studies in total or under each subgroup was too low to conduct a moderator analysis. Instead, we performed an explanatory analysis based on findings from specific studies. One potential moderator is game genre. A specific game genre may suit specific chemistry content (Wouters et al., 2013). For instance, puzzle games may help build factual knowledge (e.g., nomenclature, $g = 1.8$; Chimeno et al., 2006) through strengthening and weakening associations (reinforcement theory; Skinner, 1938); simulation games may help build conceptual knowledge (e.g., redox reaction, $g = .61$; Hodges et al., 2018) through constructing a schema of the cause-and-effect system (schema theory; Paas & Sweller, 2012); simulation games with MR or VR may help build procedural knowledge (e.g., titration, g = .97; Johnson-Glenberg et al., 2014) through deliberate practice with feedback (automaticity theory; Fitts & Posner, 1967; Mayer, 2014b). However, which genres suit which types of chemistry knowledge for which types of learners and under which contexts remains to be explored.

Another potential moderator is individual difference, such as gender (e.g., Steegh et al., 2021), prior knowledge (e.g., Lou & Jaeggi, 2019), and prior game experience. Among our included studies, compared with non-GBL, (1) girls outperformed boys but were not more motivated in GBL in chemistry education (Okonkwo, 2012), whereas others found no gender difference (Hodges et al., 2018; Merchant et al., 2013; Weng et al., 2015); (2) students with lower prior knowledge experienced greater learning gains from GBL than those with higher prior knowledge (Merchant et al., 2013; Wood & Donnelly-Hermosillo, 2019), but others found no difference (Sousa Lima et al., 2019); and (3) students with game experience achieved slightly higher learning gains than those without game experience (Merchant et al., 2013).

Research Question 3: Value-added Comparison

Our third goal was to identify the more effective game design and instructional design features for GBL in chemistry education (RQ3). However, studies that used value-added comparisons of GBL with or without specific features (k = 3) are too few to perform a meta-analysis. This lack of studies confirms that the study of effective design features of GBL (value-added research) is often underestimated compared with media comparison research (Boyle et al., 2016; Clark et al., 2016; Young et al., 2012). In line with previous meta-analyses, more evidence from value-added research is required for researchers and practitioners.

First, value-added research may provide design guidelines for GBL in chemistry education, especially for practitioners such as game developers who create games for learning and teachers who implement GBL (Mayer, 2014b). GBL can be complex and require well-designed guidelines (e.g., Eastwood & Sadler, 2013). There are little evidence-informed guidelines for developers to integrate instructional design with game design features. Most game developers are familiar with game design but not instructional design. However, most teachers can only change the GBL environment by instructional design, not the game environment per se. Second, game researchers must first conduct value-added research to refine GBL environments before comparing GBL with non-GBL. Without optimizing GBL through value-added comparisons, it is unpromising to compare learning with poorly designed games versus other media (Plass et al., 2020).

According to one side of the Clark-Kozma “media-effects” debate, media comparison studies come with two challenges. Conceptually, research may confound media (games) with methods; it is not the medium but the method that causes learning (Clark, 1983, 1991, 1994a, 1994b, 2007; Clark et al., 2008; Kirschner & Hendrick, 2020; Mayer, 2014b). Methodologically, GBL and non-GBL groups may differ in dimensions (e.g., instructions, learning materials) other than the game, making it unclear what makes a difference in learning (Clark, 2007; Clark et al., 2008; Kirschner & Hendrick, 2020; Mayer, 2014b; NRC, 2011a). Therefore, it is difficult to attribute learning effects to games, instructional methods, or other factors (e.g., Daubenfeld & Zenker, 2015).

One solution is to focus on value-added research within one game. The other side of the Clark-Kozma “media-effects” debate argues that it is unnecessary to separate instructional methods from games, as together they cause learning (Kozma, 1991, 1994a, 1994b; Parker et al., 2008). Instead of separating them, a good GBL design integrates instructional design with game design. Cognitive benefits are not the sole potential of GBL in chemistry education as games complement learning experiences with other aforementioned unique potentials (see Introduction). To employ these potentials, the focus should be less on media comparisons regarding learning effects, and more on improving GBL via value-added comparisons.

However, this does not mean media comparisons are meaningless and should be abandoned completely; they are still valuable, especially when testing GBL superiority claims (GBL is more effective than learning other media; Mayer, 2014b), justifying the reward, the effort, and cost of developing games for learning, and verifying whether certain instruction methods work specifically for GBL but not for non-GBL. Furthermore, with media comparison research, another solution is to equate GBL and non-GBL in all variables except for the game (Mayer, 2014b). Before that, however, we need high-quality GBL. Again, value-added research comes into play. Since value-added research and media comparison research serve different functions, researchers must first conduct value-added comparisons to create a well-designed GBL and then, if necessary, conduct media comparisons using a rigorous experimental design.

Limitations

The following concerns may affect the study findings. First, some studies fail to report background information. For example, because six studies reported limited or no information about additional instructions and activity levels of control groups (e.g., Fatokun et al., 2016; Rastegarpour & Marashi, 2012), their moderating effect is unknown. Although the considered moderators captured part of high heterogeneity, there is clearly unexplained variance. Missing information prevents us from including other potential moderators, such as game experience, educational research experience, or duration of intervention (Wouters et al., 2013; Wouters & van Oostendorp, 2013). Missing information affects the selection, coding, and/or analysis of moderators. Furthermore, missing information might affect our assessment regarding study quality. For instance, GBL adopts debriefing, whereas this information is missing in non-GBL, or two groups may use different ways to present the learning content. Thus, research may be contaminated (Kirschner & Hendrick, 2020) as there are more differences between the comparison groups other than just the game (Clark, 1983). Again, we cannot include or control this influence because of the missing information.

Unfortunately, we had to exclude many studies because essential information was missing. Out of 842 screened articles, only 34 met our criteria, indicating that many GBL in chemistry education have been developed but are not well-reported and/or well-tested. The increasing popularity of games stimulated a flood of publication (Hwang & Wu, 2012; Tobias et al., 2011), but most of the excluded studies only describe the game without testing its effectiveness on learning outcomes (Tobias & Fletcher, 2012); report students' subjective opinions, satisfaction, or conceptions of the game (i.e., the usability test) without an objective assessment; or measure learning outcomes without a control group. Although post hoc power analysis is not recommended, it indicates we have sufficient power for cognition, retention, and motivation benefits of GBL in chemistry education.

Moreover, our broad definition of cognition and motivation may bias the main effects. Studies vary regarding outcome measures (Cooper, 2015), and the limited number of studies made it impossible to categorize them further into different constructs (e.g., interest, self-efficacy). As all studies are different, the focus is less on sameness and more on difference: what makes effect sizes varied (Hattie, 2013). In this sense, for motivation, the meta-analysis indicates the effects of the interventions on motivation did not vary across type of motivation (Lazowski & Hulleman, 2016). For cognition, previous meta-analyses on GBL also imply the effects did not vary across type of cognition: knowledge vs. skills (Wouters et al., 2013), declarative vs. procedural knowledge (Riopel et al., 2020; Sitzmann, 2011), or intrapersonal vs. cognitive learning outcomes (Clark et al., 2016; see Table 2.1).

Implications

We make the following recommendations regarding the practices and theories of GBL in chemistry education. For practitioners, the positive effect of GBL in chemistry education on cognition, retention, and motivation suggests implementing GBL in chemistry education. The overall effect size provides the benchmark of GBL in chemistry education interventions for further research, and the distribution of effect sizes may help researchers anticipate the effects before their study.

For researchers, we agree with previous meta-analyses that it is time to move beyond "whether or not GBL in chemistry education works" (media comparison) to "what works for GBL in chemistry education and why it enhances learning and others do not" (value-added comparison; Chen et al., 2020; Young et al., 2012) and conduct more research to provide design guidelines for implementing GBL in chemistry education. This meta-analysis suggests that GBL may address the unique characteristics of chemistry. To confirm this, more GBL meta-analyses on subjects other than chemistry are needed.

Regarding learning outcomes, more considerations are needed: (1) for cognition, more delayed tests measuring retention (Mayer, 2014b; Wood & Donnelly-Hermosillo, 2019) to avoid the novelty effect (Clark, 1983); (2) more research in motivation to learn chemistry, which could be a common but questionable appeal of GBL in chemistry education (Clark & Feldon, 2014); (3) more research into emotions (e.g., Raker et al., 2019) since the most desirable instruction is that learners learn most from what they enjoy most (Clark, 1982); (4) further research regarding which game genre is best for which type of learning outcome; and (5) more research on the relationships between cognition, motivation, and emotion (e.g., Gibbons & Raker, 2019) in GBL.

Regarding methodology, more high-quality intervention research in GBL in chemistry education is required to identify what works, for whom, and under which conditions. Considering the small-study effects, we suggest researchers to use power analysis (e.g., G*Power; Faul et al., 2007) to estimate the minimum number of participants needed (Ellis, 2010) when planning primary studies as statistical power and effect size depend on sample size (Simpson, 2017). Considering the contextual factors, mixed methods are promising to evaluate the effectiveness of GBL in chemistry education; tests or questionnaires should be combined with observations, interviews, and/or log data (e.g., Hodges et al., 2018; Wood & Donnelly-Hermosillo, 2019). Regarding assessments, all assessments of the included studies were taken postintervention using separate tests, such as self-reports on motivation after gameplay when motivation might decrease (Wouters et al., 2013). We advocate more embedded tests (e.g., stealth assessment for adaptivity; Shute et al., 2017) or real-time overt assessments (e.g., eye tracking, physiological monitoring, heart rate, blood pressure; Mayer, 2020; Wouters et al., 2013) focusing on learning processes.

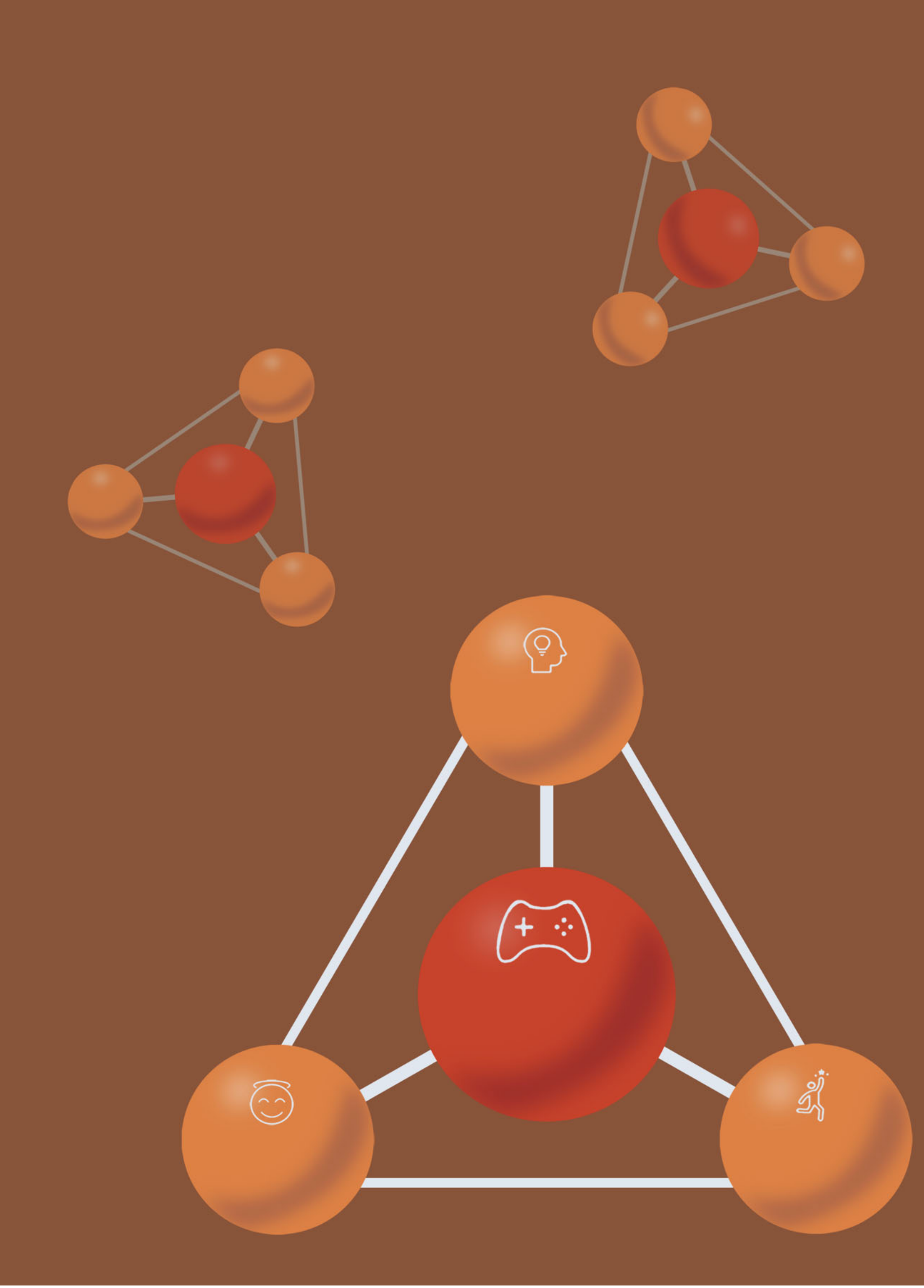
On the theoretical aspect, more evidence is needed regarding how people learn chemistry with games (learning mechanics) and how to support GBL in chemistry education (instructional support). First, we lack studies on how chemistry games affect learning processes and outcomes;

how they affect motivation or emotions; what the roles of motivation and emotions are; and how motivation, cognition, and emotions interact with each other. Second, we lack studies on how people learn better with chemistry games. As GBL is part of multimedia learning, game researchers can refer to multimedia principles from CTML (Mayer, 2014a). In certain subjects, some principles enhance GBL (e.g., modality), whereas some do not (e.g., redundancy; Mayer, 2020). Future studies should explore the learning effects of these multimedia principles in GBL in chemistry education.

Conclusion

This meta-analysis suggests that GBL may address the unique characteristics of a single subject. For example, essential game design features such as interactivity, challenges, play, and feedback may address the challenges in chemistry education such as low performance, motivation, and emotion. More GBL meta-analyses on subjects other than chemistry are needed. We systematically reviewed 34 studies on the cognitive, motivational, and emotional effects of GBL in chemistry. Compared with previous GBL meta-analyses, this study is the first meta-analysis that uses a three-level random-effects model to consider the dependency of effect sizes within studies. Generally, we found GBL in chemistry education is more effective not only for cognition and retention but also motivation than non-GBL. Publication source and sample size possibly moderate this effect. The substantial heterogeneity between studies underscores how GBL in chemistry education is implemented, particularly sample size and publication source. This study used more advanced methods to detect and correct publication bias and is the first GBL meta-analysis that includes sample size as a continuous moderator. We found that there may be the small-study effects, particularly publication bias. Furthermore, this study is also the first meta-analysis that emphasizes emotions in GBL. Unfortunately, studies assessing learner's emotions in GBL in chemistry education are absent. More robust research is required to provide a clear understanding of their true effects. Similarly, more value-added research is needed to identify more effective game design features and instructional design features and provide design guidelines for GBL in chemistry education. We advocate conducting well-developed value-added research to optimize GBL before comparing it with non-GBL.

In closing, GBL has good chemistry with chemistry education in media comparison research—GBL in chemistry education holds the right formula for improved learning and motivation; they need more value-added research before getting married; and design is the key in this relationship.



Chapter 3 Effects of timing of information presentation on students' mental effort, performance, achievement goals, and achievement emotions in game-based learning

This chapter is based on:

Hu, Y., Wouters, P., van der Schaaf, M., & Kester, L. (2022). Timing of information presentation matters: Higher motivation, more enjoyment, but less learning in game-based learning. *Manuscript submitted for publication.* Department of Education, Utrecht University

Acknowledgement of author contributions

All authors designed the study. Yuanyuan Hu recruited participants, collected, and analyzed the data, and drafted the manuscript. All authors contributed to critical revision of the manuscript. Pieter Wouters, Marieke van der Schaaf, and Liesbeth Kester supervised the study.

Acknowledgment

The authors would like to thank Rebecca Kuiper and Kyle Lang (from Utrecht University) for sharing their statistical expertise.

Abstract

Learning with games requires two types of information, namely domain-specific information, and game-specific information. Presenting these two types of information together with gameplay may pose a heavy demand on cognitive resources. This study investigates 1) how timing of information presentation affects cognition (i.e., mental effort, performance), motivation (i.e., achievement goals), and emotion (i.e., achievement emotions) and whether mental effort, perceived competence, perceived control, and perceived value mediate the effects. Secondary school students (N = 145) participated in a 2 × 2 factorial experiment with two factors Timing of domain-specific information presentation and Timing of game-specific information presentation, either before or during gameplay. Multiple regression and robust regression revealed that presenting domain-specific information before gameplay promoted higher approach goals, higher avoidance goals, more positive achievement emotions, but lower test performance than presenting it during gameplay. There was no difference between presenting game-specific information before gameplay and during gameplay except for on performance-avoidance goals. Results support productive failure but do not support the just-in-time principle and question the pre-training principle. Structural equation modeling revealed no mediating effects. We conclude that timing of information presentation affects cognitive, motivational, and emotional processes and outcomes differently, and that students feel more motivated and enjoy the learning environment from which they learn less. Educators may change the timing of information presentation based on the learning goal, such as promoting motivation, emotion, or cognition.

Keywords: Game-based learning; Timing of information presentation; Performance; Achievement goals; Achievement emotions

Introduction

Game-based learning (GBL) is a type of learning environment with gameplay, accompanied by learning goals, learning outcomes, game goals, and game outcomes, in which a game is the medium for learning (Plass, Mayer et al., 2020). In practice, stakeholders, in particular educators, expect that GBL is motivating, enjoyable, and effective. Previous meta-analyses agreed that GBL promotes cognitive processes and outcomes (e.g., test or game performance), but GBL needs effective design features (see Author, 2022 for an overview of all meta-analyses). One of the design features is scaffolding, which is defined as support on performing learning tasks that learners otherwise cannot perform (van de Pol et al., 2010; Wood et al., 1976). Previous meta-analyses have shown that in general providing scaffolding is effective in GBL, but not all types of scaffolding do so equally (Cai et al., 2022; Author, 2013). Thus, more attention should be paid on the design of scaffolding, such as the timing of scaffolding (e.g., Rahimi et al., 2022).

In complex learning environments, such as GBL, scaffolding can involve multiple dimensions. Take the scaffolding of information presentation for example. Gameplay in GBL needs *domain-specific information* - information about the domain, such as chemistry, and *game-specific information* - information about the game, which may relate to cognitive scaffolds (on knowledge content concepts, such as chemistry) and procedural scaffolds (on how to use tools or resources, such as in a game) of multidimensional scaffolding (Hou, 2022), respectively. According to cognitive load theory (Sweller et al., 2019), a challenge is that studying both types of information and playing the game together may cause cognitive overload, which could hamper learning. Furthermore, from a learning perspective, because learners must deal with the new learning environment and acquiring new domain knowledge, this kind of cognitive overload is likely to occur. This implies that the timing of presenting the domain-specific and the game-specific information matters. When timing of information presentation is not right, learners cannot optimally learn from GBL. Learners may even feel frustrated and become demotivated, which could further hamper learners' motivation, emotion, and learning.

In our definition multimedia is always part of computers games because often the verbal (text, spoken words) and visual (animations, pictures, and video) modality are combined. In complex learning environments, such as GBL, this can detract from the learning effect. The application of multimedia principles, such as the pre-training principle and just-in-time principle, may solve this instructional design and learning challenge (see also: Mayer, 2021, 2014). The *pre-training principle* from the cognitive theory of multimedia learning (i.e., presenting the names and characteristics of components before learning helps learning; Mayer & Fiorella, 2021) suggests that domain-specific information should be presented before gameplay to manage cognitive load, whereas the *just-in-time principle* of the four-component instructional design model (i.e., different types of information can best be presented precisely when learners need it; van Merriënboer & Kirschner, 2018) suggests game-specific information should be presented precisely when needed during gameplay to reduce load.

However, we found inconclusive evidence regarding these two well-known principles. The just-in-time principle was investigated by three non-GBL studies but with mixed results: some supported information presentation during learning (Author, 2004a; Noroozi et al., 2012), some supported information presentation before learning (Author, 2004b), and others supported information presentation piece-by-piece instead of simultaneously (Author, 2006a, 2006b). Similarly, the pre-training principle was investigated by eight GBL studies with inconclusive results: Some supported that pre-training helps learning (Barzilai & Blau, 2014; Fiorella & Mayer, 2012; Leutner, 1993; Mayer et al., 2002; Rahimi et al., 2022), but others not (Pilegard & Mayer, 2016, 2018; Tsai et al., 2022). Furthermore, the control groups of some studies did not receive the same information that was presented in the pre-training group (e.g., Fiorella & Mayer, 2012), which might have caused a bias (Kirschner & Hendrick, 2020). Thus, it is unclear whether these two types of information should be presented before or during learning.

In addition, although motivation and emotion are essential in learning (Author, 2022), evidence on how instructional design principles affect them is sparse but critical to draw a full picture of complex learning (Schrader et al., 2021). For example, some studies even found pre-training seems to detriment motivation to learn (Charsky & Ressler, 2011). Particularly in GBL, most studies only focused on cognition and motivation (Mayer, 2020), while there is some evidence that the use of educational technology, such as GBL, can affect cognition, motivation, and emotions, and have an impact on a person's cognitive, emotional, physical, and social well-being (Melo et al., 2020).

Taken together, this study investigates how timing of information presentation affects cognitive, motivational, and emotional processes and outcomes in GBL. The major empirical contribution of this study is to explore the optimal timing of information presentation, particularly its effects on cognition, motivation, and emotion. The major theoretical contribution of this study is to explore potential instructional design principles that benefit cognition, motivation, and emotion, particularly reconsidering the pre-training principle and the just-in-time principle. The major practical contribution of this study is to convince educators that with this design guideline, learners will feel motivated and enjoyed and learn effectively in GBL.

Effect of Timing of Information Presentation on Learning

Instructional design features, such as timing of domain-specific and game-specific information presentation, may influence GBL in three ways: by affecting cognitive processes, by affecting motivation, and/or by affecting emotion (see Author, 2022 for a detailed discussion). Among the numerous relevant theories, this study is based on performance or competence-related theories, that is, cognitive load theory (CLT), achievement goal theory (AGT), and the control-value theory (CVT) of achievement emotions. These theories were chosen because of their critical role when assessing cognition, motivation, and emotion in complex learning, such as GBL (Loderer et al., 2020; Mayer, 2020; Plass, Mayer et al., 2020). Figure 3.1 shows our theoretical propositions.

Cognition and Timing of Information Presentation

From a cognitive perspective, instructional design features, such as timing of information presentation, affect cognitive load and consequently, test and game performance (Figure 3.1 line c_{c1} or c_{c2}). One way to estimate cognitive load is the amount of mental effort that learners exert in a task (Paas, 1992). We will focus on two types of cognitive load here, considering the critics of germane load, such as germane load, may not be distinguished from intrinsic load and extraneous load (de Jong, 2010). According to CLT, the overall cognitive load varies, depends on *element interactivity* (i.e., the number of elements that must be proceeded simultaneously), and consists of *intrinsic load* - the load caused by cognitive processes or activities that are relevant for learning and performing the task and *extraneous load* - the load caused by cognitive processes or activities that are unnecessary for learning and performing the task (Sweller et al., 2019). In GBL, accordingly, domain-specific information imposes intrinsic load, whereas game-specific information is irrelevant to learning goals and thus imposes extraneous load (Kalyuga & Plass, 2009). Regarding the timing, presenting domain-specific and game-specific information at the wrong time may impose too high intrinsic load and extraneous load, respectively.

Timing of information presentation may optimize cognitive load in two ways (Figure 3.1 line a, Table 1). First, when domain-specific information is presented during gameplay or together with game-specific information, it can be difficult to simultaneously process this new information and play the game or study game-specific information. In contrast, as suggested by the pre-training principle (Mayer & Fiorella, 2021), when domain-specific information is presented before gameplay and separately from game-specific information, all working memory resources are available to process this complex information at that time and construct relevant schemata (i.e., knowledge structures of integrated elements, such as here a separator schema which includes a variety of separators, and the properties of materials determining which separator is used) in long-term memory (Sweller et al., 1998). In turn, schemata can be retrieved during gameplay as a single entity to solve problems. Retrieving information from long-term memory is usually less cognitively demanding than processing new information in working memory (Sweller et al., 1998). In this way, element interactivity may be reduced and thus, intrinsic load may be managed during gameplay (Mayer & Fiorella, 2021).

Second, game-specific information is context-bound, that is, specific game information is required in specific game situations (e.g., higher game levels may require more sophisticated game actions, or some information depends on a choice made by the player). Information could be irrelevant for learners who receive these before the game starts (e.g., a learner may not reach the higher levels). The cognitive capacity used to study this information is wasted and thus can be qualified as extraneous load. In contrast, as suggested by the just-in-time principle (Author, 2021), when game-specific information is presented during gameplay, at a time when it is needed, extraneous load may be reduced. If game-specific information has low element interactivity, processing new game-specific information together with playing the game would not cause overload. In this way, extraneous load will be reduced.

In short, timing of information presentation may affect cognitive load in GBL: optimal timing of information presentation may manage intrinsic load, reduce extraneous load of complex learning tasks (lower overall cognitive load, lower mental effort), and thus, yield higher test and game performance (Figure 3.1 line a – b_{c1} or b_{c2}). After all, too high cognitive load (i.e., cognitive overload) can hamper learning (Schnotz & Kürschner, 2007).

Table 3.1

The effects of timing of information presentation on cognitive load

Domain-specific info	Game-specific info	Do learners study domain-specific info and game-specific info separately?	Do learners study game-specific info during gameplay?	Intrinsic load	Extraneous load
Before	Before	No	No	Not managed	Not reduced
Before	During	Yes	Yes	Managed	Reduced
During	Before	Yes	No	Managed	Not reduced
During	During	No	Yes	Not managed	Reduced

Motivation and Timing of Information Presentation

Cognitive processes have a reciprocal relationship with motivation in a task (Paas et al., 2005; Schnotz et al., 2009). As such, instructional design features in GBL, such as timing of information presentation, that manipulate cognitive load may also change motivation, such as achievement goals, as displayed in Figure 3.1 line c_{m1} and c_{m2}. According to AGT, *achievement goal* is the purpose of competence-related behavior in achievement settings, in which *achievement or competence* indicates whether one is doing poorly or well (Elliot & Hulleman, 2017). The 2 × 2 achievement goal model (Elliot & Hulleman, 2017) distinguishes four achievement goals: *mastery-approach goa*ls - striving for task- or self-based competence, such as learning as much as possible, *performance-approach goals* - striving for other-based competence, such as performing better than others, *mastery-avoidance goals* - striving to avoid task-based or self-based incompetence, such as avoiding learning less than one possibly could, and *performance-avoidance goals* - striving to avoid other-based incompetence, such as avoiding performing worse than others.

As suggested by AGT, instructional design in GBL aims to promote higher approach goals and lower avoidance goals (Elliot & Hulleman, 2017). Approach goals focus on success, promote task engagement, and consequently link to positive learning outcomes, whereas avoidance goals focus on failure, undermine task engagement, and consequently link to negative learning outcomes (Elliot, 1999). Previous meta-analyses have confirmed that approach goals generally link to more positive emotions, higher cognitive, motivational, and behavioral outcomes than avoidance goals (Baranik et al., 2010; Carpenter, 2007; Celler et al., 2011; Huang, 2011; Hulleman et al., 2010; Payne et al., 2007; van Yperen et al., 2015; Wirthwein et al., 2013).

Considering that limited research connects timing of information presentation with achievement goals directly, we propose cognitive load and perceived competence may jointly mediate the effects of timing of information presentation on achievement goals (Figure 3.1 line $a - d_m - b_{m1}$ or b_{m2}). AGT posits that competence-related variables, such as perceived competence, are the major antecedents of achievement goals and factors that affect these antecedents also affect achievement goal adoption (Elliot & Hulleman, 2017). *Perceived competence* is the self-confidence to be able to accomplish the task at hand (Law et al., 2012) or the need for challenge and feeling of mastery (Ryan & Deci, 2017). High (or low) perceived competence orients learners to the possibility of success (or failure), commits to approaching success (or avoiding failure), and thus, promotes more approach (or avoidance) goals adoption, particularly mastery-approach goals (Senko et al., 2011). Previous meta-analyses confirmed that perceived competence (or self-efficacy) positively links to approach goals and negatively links to avoidance goals (Baranik et al., 2010; Cellar et al., 2011; Huang, 2016). Moreover, cognitive load may affect perceived competence (Figure 3.1 line d_m): Lower cognitive load in complex learning may predict higher perceived competence (Feldon et al., 2018).

Timing of information presentation may affect cognitive load, shape learners' perceived competence and trigger the adoption of achievement goals. Compared with learners who experience cognitive overload, those who experience less cognitive load in complex tasks may perceive higher competence and thus, adopt higher approach goals and lower avoidance goals (Senko & Hulleman, 2013). Based on these assumptions, timing of information presentation may indirectly affect achievement goals.

Emotion and Timing of Information Presentation

From an emotional perspective, instructional design features in GBL, such as timing of information presentation, affect achievement emotions (Loderer et al., 2020), as displayed in Figure 3.1 line c_{e1} and c_{e2}. According to the CVT, *achievement emotions* are emotions relating to achievement activities (e.g., studying) or outcomes (i.e., success or failure) in achievement settings (Pekrun, 2006). In this study, we focus on two positive achievement emotions (i.e., enjoyment and pride) and four negative achievement emotions (i.e., anger, anxiety, boredom, and shame) because these six achievement emotions are most frequently reported in educational settings (Pekrun et al., 2009).

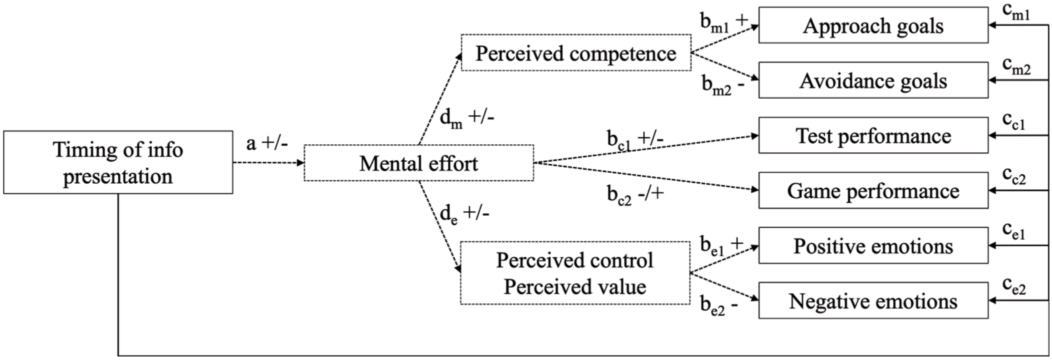
As suggested by CVT, instructional design in GBL aims to induce more positive emotions and less negative emotions (Linnenbrink-Garcia et al. 2016; Loderer et al., 2020). Achievement emotions in general reciprocally affect cognitive and motivational processes, which in turn, affect performance over time (Pekrun & Perry, 2014). Positive (negative) achievement emotions generally increase (decrease) task attention, preserve (consume) working memory resources, facilitate flexible (rigid) information processing, promote (undermine) motivation (e.g., intrinsic motivation) and consequently, increase (decrease) performance. Negative achievement emotions sometimes increase extrinsic motivation (e.g., learners who experience test anxiety may invest more effort to avoid failure) and increase performance. Previous meta-analyses have confirmed that in general, performance and motivation positively link to enjoyment but negatively link to anger and boredom (Camacho-Morles et al., 2021; Loderer et al., 2018; Tze et al., 2016).

Considering that limited research connects timing of information presentation with achievement emotions directly, we propose cognitive load together with perceived control and perceived value may mediate the effects of timing of information presentation on achievement emotions (Figure 3.1 line $a - d_e - b_{e1}$ or b_{e2}). CVT posits that perceived control and perceived value are the major antecedents of achievement emotions and factors that affect these antecedents also affect achievement emotions (Pekrun & Perry, 2014). *Perceived control* involves feeling in control of achievement activities and outcomes and includes action-control expectancies that one can perform an action (self-efficacy; Bandura, 1977) and action-outcome expectancies that the action will lead to desired outcomes (academic control; Perry et al., 2001). *Perceived value* involves perceived importance of achievement activities and outcomes, such as valuing chemistry because it is interesting (intrinsic value), useful (extrinsic or utility value), or important to get good marks (achievement or attainment value; Putwain et al., 2021). Learners experience positive achievement emotions when both perceived control and perceived value are high and negative achievement emotions when perceived control and/or perceived value is low (Loderer et al., 2020).

Timing of information presentation may affect cognitive load, shape learners' perceived control and perceived value, and thus induce correspondent achievement emotions. Compared with learners who experience cognitive overload, those who experience less cognitive load in complex learning may perceive higher control and value, and thus experience more positive emotions (e.g., enjoyment; Camacho-Morles et al., 2021) and less negative emotions. Otherwise, if cognitive load is too high or too low, perceived value can be low and negative emotions, such as boredom, are likely to be triggered (Pekrun, Hall, et al., 2014). Based on these assumptions, timing of information presentation may indirectly affect achievement emotions.

Figure 3.1

Summary of theoretical propositions for timing of information, cognitive load (mental effort), perceived competence, perceived control, perceived value, performance, achievement goals, and achievement emotions



Note. Arrows and c represent total effect; Dotted arrows represent direct effect; Rectangles represent independent and dependent variables; + represents positive relations; - represents negative relations; +/- represents there is an optimal cognitive load: when below it, the higher value the better, when above it, the higher value the worse; a, d, and b in a serial mediation represent the direct effect of independent variable – first mediator, first mediator – second mediator, and second mediator – dependent variable, respectively; For simplicity, direct effects of first mediator – dependent variable are not displayed; $_{c}$, $_{m}$, and $_{e}$ represent cognition, motivation, and emotion, respectively; The numbers are further explained in the text.

Present Study

This study investigates **RQ1**) how timing of (domain-specific and game-specific) information presentation affects 1) cognition (i.e., mental effort, test performance, and game performance), 2) motivation (i.e., achievement goals), and 3) emotion (i.e., achievement emotions); and **RQ2**) whether mental effort, perceived competence, perceived control, and perceived value mediate the effects.

RQ1 (Main effects): Presenting domain-specific information before gameplay and game-specific information during gameplay yields **H1a)** lower mental effort, **H1c1**) higher test performance, **H1c2**) higher game performance (i.e., lower time-on-task), **H1m1**) higher approach goals (i.e., mastery-approach and performance-approach goals) and **H1m2**) lower avoidance goals (i.e., mastery-avoidance and performance-avoidance goals), and **H1e1**) more positive achievement emotions (i.e., enjoyment and pride) and **H1e2**) less negative achievement emotions (i.e., anger, anxiety, shame, and boredom) than other conditions (Figure 3.1 line a, c_{c1}, c_{c2}, c_{m1}, c_{m2}, c_{e1}, c_{e2}).

RQ2 (Mediation): H2c) Mental effort mediates the effects of timing of information presentation on test performance and game performance (Figure 3.1 line a – b_{c1} or b_{c2}); **H2m)** Mental effort and perceived competence jointly mediate the effects of timing of information presentation on achievement goals (Figure 3.1 line a – d_m – b_{m1} or b_{m2}); **H2e)** Mental effort, perceived control, and/or perceived value jointly mediate the effects of timing of information presentation on achievement emotions (Figure 3.1 line a – d_e – b_{e1} or b_{e2}).

Methods

Participants

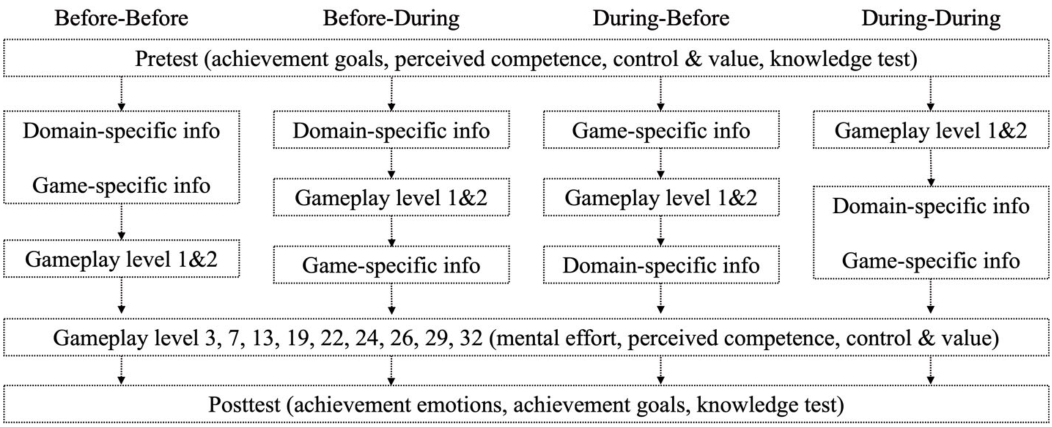
Secondary school students (n = 292) from six schools in the Netherlands participated. Every student received a pack of snacks, and every teacher received a 25 euros voucher. We excluded participants who played the wrong game levels and included 145 participants (77 male, 60 female, 8 neutral, age range = 14-18, M = 16.3 years, SD = 1.10). Dutch was used for all instructions and materials.

Design

A 2 × 2 factorial design with two factors was used: *Timing of domain-specific information presentation* (before vs. during gameplay) and *Timing of game-specific information presentation* (before vs. during gameplay). All participants were randomly assigned to one of the four groups: 1) presenting domain-specific and game-specific information before gameplay (n = 36); 2) presenting domain-specific information before gameplay and game-specific information during gameplay (n = 43); 3) presenting domain-specific information during gameplay and game-specific information before gameplay (n = 34); and 4) presenting domain-specific and game-specific information during gameplay (n = 32; Figure 3.2).

Figure 3.2

The procedure and measures



Note. Level 1 and 2 were preinstalled with receptors; Level 3 is the first complete task that learners need to build a whole recycling chain including a conveyor, a separator, and receptors.

Materials and Measures

The Game — CosmiClean

CosmiClean (https://recyclegame.eu/) was designed by LuGus Studios (https://www.lugus-studios.be/) to teach 14-23 year-old students the principles for separation processes of recycling materials (see Table S3.1 in supplementary materials for its learning objectives, game levels, characters, and scenes; see Figure S3.1 in supplementary materials for the scene of level 16). The chemistry learning content includes the functions of the nine separators (including a sieve, a magnet, a melter, a shredder, a stream separator, a non-ferrous separator, a boiler, a dissolver, and a centrifuge) and the eight properties (including size, phase, melting point, boiling point, magnetic metal, non-ferrous metal, solubility, and density) of 12 materials (including iron, plastics, glass, concrete, water, wood, sand, copper, salt, gold, solvent, and fuel). Players play a series of game levels with a certain mixture of materials in a spaceship cargo. The goal is to make a recycling chain, including a conveyor (for transporting the materials), one or more separators (for separating material based on different properties), and receptors (for collecting the recycled materials). Participants completed the 11 game levels that introduce nine separators individually and once.

The Manual

All conditions received two online manuals that contained two types of information (see Figure S3.2 in supplementary materials) either immediately after the pretest for before groups or after level 2 for during groups (see Figure 3.2). Game-specific information includes key game concepts (i.e., container), the general procedure of how to play the game (e.g., "Step 1 Let's open a Container!"), and an example (i.e., separate plastics, concrete and iron from plastic basket and concrete with iron fillings in level 16). Domain-specific information includes how the processors work (e.g., The sieve separates materials by size), the general procedure of recycling in real life (e.g., "Step 1 Check the materials you get!"), and an example (i.e., separate plastics, concrete and iron from plastics and concrete with iron fillings in real life). To distinguish both types of information, domain-specific information excluded game-specific information, and vice versa.

In practice, we use common and plain language in the game-specific information and use jargon and domain-specific terms in the domain-specific information. For example, in Figure S3.2, the concept "chunk" (i.e., 'each composed of one material') in game-specific information will be "pure substance" in domain-specific information and the concept "a particle" (i.e., 'single "chunk" composed of various materials') in game-specific information will be "a mixture" in domain-specific information. When reading this game-specific info, students will not read any information about the properties of materials. They may imagine the way materials may be integrated together, but this imagination may have no effect on learning from the intervention. This imagination may be also different for the two types of information. The schemata of "a particle" in game-specific information would be like the picture in Figure S3.2 and the schemata "a mixture" in domain-specific information would be like the picture of 'concretes with iron fillings' in real life, which is given as an example of "a mixture".

To connect both types of information, the same example was used. To prevent the spatial split-attention effect (Schroeder & Cenkci, 2018), game screenshots were used to illustrate game-specific information and this information was also placed close to the relevant parts of the screenshots. The manuals were written in active language and small information units as suggested by Author (2021). The manuals were developed by the first author who was a high school chemistry teacher and checked by the participating chemistry teachers.

Demographic Information

Demographic information comprised age, sex, and game experience (e.g., "*Do you have experiences with playing computer games?*" *"If yes, how many hours do you play on a school day?*"; "*If yes, how many hours did you play on a weekend day?*").

Knowledge Test

The knowledge test was developed by the authors and evaluated by three participating chemistry teachers. It assesses Remember (5 multiple-choice questions), Apply (5 multiple-choice questions), and Evaluate (3 open-ended questions) based on the Bloom taxonomy (Anderson & Krathwohl, 2001). For example, a Remember question is: "*Which property is used by stream separator?*". An Apply question is: "*Gold is not dissolvable. Which processor can be used to separate copper and gold from their mixture?*". An Evaluate question is: "*To separate water and plastics, your teacher will select between steam separator and dissolver. Explain which one is more proper.*". The prior and post knowledge test include the same test items but in a different order. Participants got tested with mostly domain-specific information and very few game-specific information (e.g., the word 'processors'). After we removed the items that had low or high item difficulty (proportion correct) or low item discrimination based on the pilot study, all multiple-choice questions had acceptable item difficulty ranging from .2 - .8 and item discrimination greater than .2 (Cohen et al., 2018). The knowledge tests were reliable (pretest: greatest lower bound = .60; posttest: greatest lower bound = .69). Prior knowledge tests do not usually measure the same underlying construct (nine separators instead of one), so a reliability value lower than .70 is normal (Taber, 2018).

Mental Effort, Perceived Competence, Perceived Control, and Perceived Value

As recommended by Sweller (2010), we used Paas' (1992) scale to measure how much mental effort was invested in a task as an indicator of overall cognitive load that learners experienced in every game level (1 = very, very low mental effort, 9 = very, very high mental effort). Perceived competence was operationalized as self-efficacy, which was measured by "*How confident were you that you had mastered the chemistry skills like recycling being taught in the game*?" (1 = very, very unconfident, 9 = very, very confident; adapted from van Harsel et al. (2020)). Following Bieg et al. (2013) and Goetz et al. (2010), perceived control and perceived value were measured by "*I have the impression that playing the game is under my control*" and "*Playing the game is important for me*", respectively (1 = strongly disagree, 5 = strongly agree).

Game Performance

Log data was collected to evaluate game performance, that is, time-on-task (start time of the game - end time of the game - the time for the manual).

Achievement Goals Questionnaire

The achievement goals questionnaire (AGQ; adopted from Bipp and Van Da (2014)) was based on the Achievement Goals Questionnaire-Revised (Elliot & Murayama, 2008; 1 = strongly disagree, 5 = strongly agree) and measures mastery-approach goals, performance-approach goals, mastery-avoidance goals, and performance-avoidance goals. The keywords 'class' and 'course' were changed to 'game'. An example of an item was "My goal was to learn as much as possible in the game". All subscales were reliable (greatest lower bound ranging from .79 - .94).

Achievement Emotions Questionnaire

The achievement emotions questionnaire (AEQ; adapted from Donker et al. (2021)) was based on the Achievement Emotions Questionnaire (Pekrun et al., 2011; 1 = strongly disagree, 5 = strongly agree) and measures enjoyment (4 items), pride (3 items), anger (4 items), anxiety (3 items), boredom (3 items), and shame (3 items). The instruction of AEQ asked participants to describe how they felt during the game. All subscales were reliable (greatest lower bound ranging from .77 - .94) except for pride (greatest lower bound ranging = .67) and anxiety (greatest lower bound ranging = .66) and thus these two subscales were removed from further analysis.

Procedure

Pilot Study

We run a pilot study with six participants to explore whether they could distinguish the game-specific information and domain-specific information in the manuals, whether students understood all the materials, or whether they had technical problems with the game and the online experiment in Qualtrics.

Main Study

The study was done in two hours in one session. After giving informed consent (parental consent is needed if participants are younger than 16 years old), participants were randomly assigned (see Figure 3.2). Then, they received an introduction about 1) the study, the number of sections, and the duration and 2) the rules, such as work individually, and completed AGQ, knowledge test, perceived competence, perceived control, and perceived value (pretest) in 15 mins. After that, all groups read the manual(s) but at different times and finished the game in one hour, during which they rated mental effort, perceived competence, perceived control, and perceived value for each game level. Then they completed AEQ, AGQ, perceived competence, perceived control, perceived value, and knowledge test (posttest) in 20 mins.

Scoring, Data Preparation, and Data Analysis

We treated data from Likert scales with five or more categories as continuous data instead of ordinal data and used means instead of medians (Sullivan & Artino, 2013). For the AEQ and AGQ, we first calculated the scale means of each item. For the knowledge test, we calculated a sum score of 10 multiple-choice questions (1 point per question) and 3 open-ended questions (3 points per question, partial credit is allowed), resulting in a maximum score of 19 points. For the three open-ended questions, we developed a coding schema. Two raters scored 10% of the pretest and posttest for each question independently (inter-rater reliability: Cohen's $k = 1$), and then they scored the remainder ($k = .95$). For the mental effort, perceived competence, perceived control, and perceived value, we calculated the means of all game levels. For game experience, we summarized the hours they played per weekday and weekends and calculated how much time they spent on game in hours.

For RQ1, data were analyzed by multiple linear regression in R studio (R Studio Team, 2021). Prior knowledge, game experience, prior mastery-approach goals, prior performance-approach goals, prior mastery-avoidance goals, and prior performance-avoidance goals correlated with relevant dependent variables and thus were added as covariates. We checked missing data (less than 5%), outliers (| standardized residuals | > 3), and assumptions of normality of residuals

(Shapiro-Wilk test), homogeneity of variances (Levene's test), independence of residuals, linearity, and no multicollinearity (Field et al., 2012). As some variables had non-normally distributed residuals, we ran robust methods using lmrob function (Field & Wilcox, 2017) from the robustbase package (Maechler et al., 2021).

For RQ2, data were analyzed by structural equation modeling (SEM) in *Mplus 8.6* (Muthén & Muthén, 2017). To reduce complexity, separate analyses were run for cognition, motivation, and emotion. Manifest variables (indicators) were included in the model. The estimator is maximum likelihood estimation with robust standard errors (MLR). The indirect effect of mediation was accessed by standardized path coefficients. In evaluating model fit we focused primarily on the comparative fit index (CFI: ≥ .90 = acceptable; ≥ .95 = excellent), as it is less sensitive to the model and data characteristics than other fit indexes, such as chi-square (Asparouhov & Muthén, 2018; Kenny et al., 2015; Marsh et al., 1988). All our models had acceptable fit, ranging from .91 to .97.

Results

Table 3.2 shows the means and standard deviations for the dependent variables and covariates. Table 3.3 shows multiple linear regression and robust regression results for main effects and interaction effects. Figure 3.3 shows path analysis for direct effects of mediation.

RQ1: The Effect of Timing of Information Presentation on Learning

There was no statistically significant difference between groups on pretest performance, prior achievement goals, perceived competence, perceived control, and perceived value. There was no statistically significant interaction effect on all learning processes and outcomes (see Table 3.3). The main effects were explained below.

Mental Effort and Performance

Robust regression revealed no statistically significant main effect of timing of information presentation on mental effort after controlling for prior knowledge and on time-on-task after controlling for game experience.

Multiple linear regression revealed a statistically significant main effect for timing of domain-specific information presentation on test performance with a medium effect size ($t = 2.14$, $p \leq .05$, $f^2 = .2$) after controlling for prior knowledge. Presenting domain-specific information during gameplay yield higher test performance than presenting it before gameplay. There was no statistically significant main effect of timing of game-specific information presentation on test performance.

A paired t-test showed that there is a statistically significant difference between pretest performance and posttest performance ($t = 6.28$; $df = 144$; $p < .001$) with a medium effect size (Cohen’s $d = .58$), so, learning happened.

Achievement Goals

Multiple linear regression and robust regression revealed a statistically significant main effect for timing of domain-specific information presentation on approach goals (mastery-approach goals: $t = 2.27$, $p < .05$, $f^2 = 1.3$; performance-approach goals: $t = 3.24$, $p < .05$, $f^2 = 1.5$) and avoidance goals (mastery-avoidance goals: $t = 2.47$, $p < .05$, $f^2 = 1.0$; performance-avoidance goals: $t = 2.44$,

$p < .05$, $f^2 = 1.4$) with very large effect sizes, and a statistically significant main effect of timing for game-specific information presentation on performance-avoidance goals ($t = 1.98$, $p < .05$, $f^2 = 1.4$) with a very large effect size after controlling for prior achievement goals. Presenting domain-specific information before gameplay promoted higher approach goals and avoidance goals than presenting it during gameplay. Presenting game-specific information during gameplay promoted higher performance-avoidance goals than presenting it before gameplay.

Achievement Emotions

Robust regression revealed a statistically significant main effect for timing of domain-specific information presentation on positive emotions with a small effect size (enjoyment: $t = 2.26$, $p \leq .05$, $f^2 = .1$). Presenting domain-specific information before gameplay reported more positive emotions than presenting it during gameplay. There was neither a statistically significant main effect of timing of domain-specific information presentation on negative emotions, nor a main effect of timing of game-specific information presentation on achievement emotions.

Table 3.2

Mean and standard deviation for dependent variables and covariates

	Timing of domain-specific info						Timing of game-specific info					
	Before			During			Before			During		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Pretest performance (0-19)	79	5.10	2.13	66	5.44	1.95	70	5.43	2.15	75	5.09	1.96
Test performance (0-19)	79	6.14	2.46	66	7.08	2.52	70	6.46	2.79	75	6.67	2.28
Mental effort (1-9)	78	3.58	1.59	66	3.79	1.52	69	3.85	1.66	75	3.52	1.45
Time-on-task (s)	79	2933	723	66	2934	820	70	2939	830	75	2928	707
Prior mastery-approach goals (1-5)	79	3.85	.65	66	3.76	.70	70	3.76	.67	75	3.85	.68
Prior performance-approach goals (1-5)	79	3.52	.85	66	3.41	.85	70	3.43	.79	75	3.51	.91
Prior mastery-avoidance goals (1-5)	79	3.62	.73	66	3.66	.64	70	3.55	.71	75	3.72	.65
Prior performance-avoidance goals (1-5)	79	3.46	.89	66	3.52	.82	70	3.47	.74	75	3.50	.95
Mastery-approach goals (1-5)	79	3.76	.80	66	3.46	.85	70	3.53	.82	75	3.71	.84
Performance-approach goals (1-5)	79	3.46	.97	66	3.00	1.08	70	3.15	1.02	75	3.34	1.06
Mastery-avoidance goals (1-5)	79	3.62	.92	66	3.36	.81	70	3.47	.90	75	3.54	.87
Performance-avoidance goals (1-5)	79	3.34	.98	66	3.08	.94	70	3.12	.91	75	3.32	1.01
Enjoy (1-5)	79	3.58	.81	66	3.33	.81	70	3.43	.74	75	3.49	.88
Anger (1-5)	79	2.45	1.10	66	2.57	1.05	70	2.57	1.13	75	2.44	1.02
Boredom (1-5)	79	2.78	.94	66	2.94	.86	70	2.88	.89	75	2.83	.93
Shame (1-5)	79	1.73	.81	66	1.76	.72	70	1.83	.81	75	1.67	.72

Note. n = sample size per condition per outcome; *M* = mean; *SD* = standard deviation.

Table 3.3

Multiple linear regression and robust regression results for main effects

Variables	β	SE	t	R^2	Adjusted R^2
Outcome: Posttest performance [a]				.18	.16
(Intercept)	3.58	.59	6.09***		
Pretest performance	.46	.09	4.90***		
Timing of domain-specific info	.81	.38	2.14*		
Timing of game-specific info	.41	.38	1.09		
Outcome: Mental effort [b]				.07	.05
(Intercept)	4.65	.49	9.48***		
Pretest performance	-.19	.06	-2.81**		
Timing of domain-specific info	.32	.26	1.25		
Timing of game-specific info	-.35	.26	-1.35		
Outcome: Time-on-task [b]				.06	.04
(Intercept)	2427.8	258.3	9.40		
Prior game experience	358.9	160.4	2.24*		
Timing of domain-specific info	-106.3	130.1	-.82		
Timing of game-specific info	26.9	141.1	.19		
Outcome: Mastery-approach goals [a]				.56	.55
(Intercept)	.24	.28	.86		
Prior mastery-approach goals	.90	.07	12.91***		
Timing of domain-specific info	-.21	.09	-2.27*		
Timing of game-specific info	.09	.09	.98		
Outcome: Performance-approach goals [a]				.60	.59
(Intercept)	.20	.25	.80		
Prior performance-approach goals	.91	.07	13.99***		
Timing of domain-specific info	-.36	.11	-3.24**		
Timing of game-specific info	.10	.11	.93		
Outcome: Mastery-avoidance goals [a]				.50	.48
(Intercept)	.47	.29	1.56***		
Prior mastery-avoidance goals	.89	.08	8.47***		
Timing of domain-specific info	-.30	.11	-2.83*		
Timing of game-specific info	-.09	.11	-.90		
Outcome: Performance-avoidance goals [b]				.59	.58
(Intercept)	.25	.27	.91*		
Prior performance-avoidance goals	.86	.08	11.22***		
Timing of domain-specific info	-.27	.11	-2.44*		
Timing of game-specific info	.22	.11	1.98*		
Outcome: Enjoyment [b]				.06	.04
(Intercept)	3.69	.09	42.22***		
Timing of domain-specific info	-.26	.12	-2.17*		
Timing of game-specific info	.13	.11	1.23		
Outcome: Anger [b]				.007	-.007
(Intercept)	2.48	.17	14.07***		
Timing of domain-specific info	.12	.19	.66		
Timing of game-specific info	-.14	.19	-.73		
Outcome: Boredom [b]				.007	-.004
(Intercept)	2.80	.13	20.86***		
Timing of domain-specific info	.16	.16	1.04		
Timing of game-specific info	-.04	.16	-.45		

Variables	β	SE	t	R^2	$Adjusted\ R^2$
Outcome: Shame [b]				.01	-.002
(Intercept)	1.64	.13	12.82***		
Timing of domain-specific info	.09	.12	.75		
Timing of game-specific info	-.11	.12	-.94		

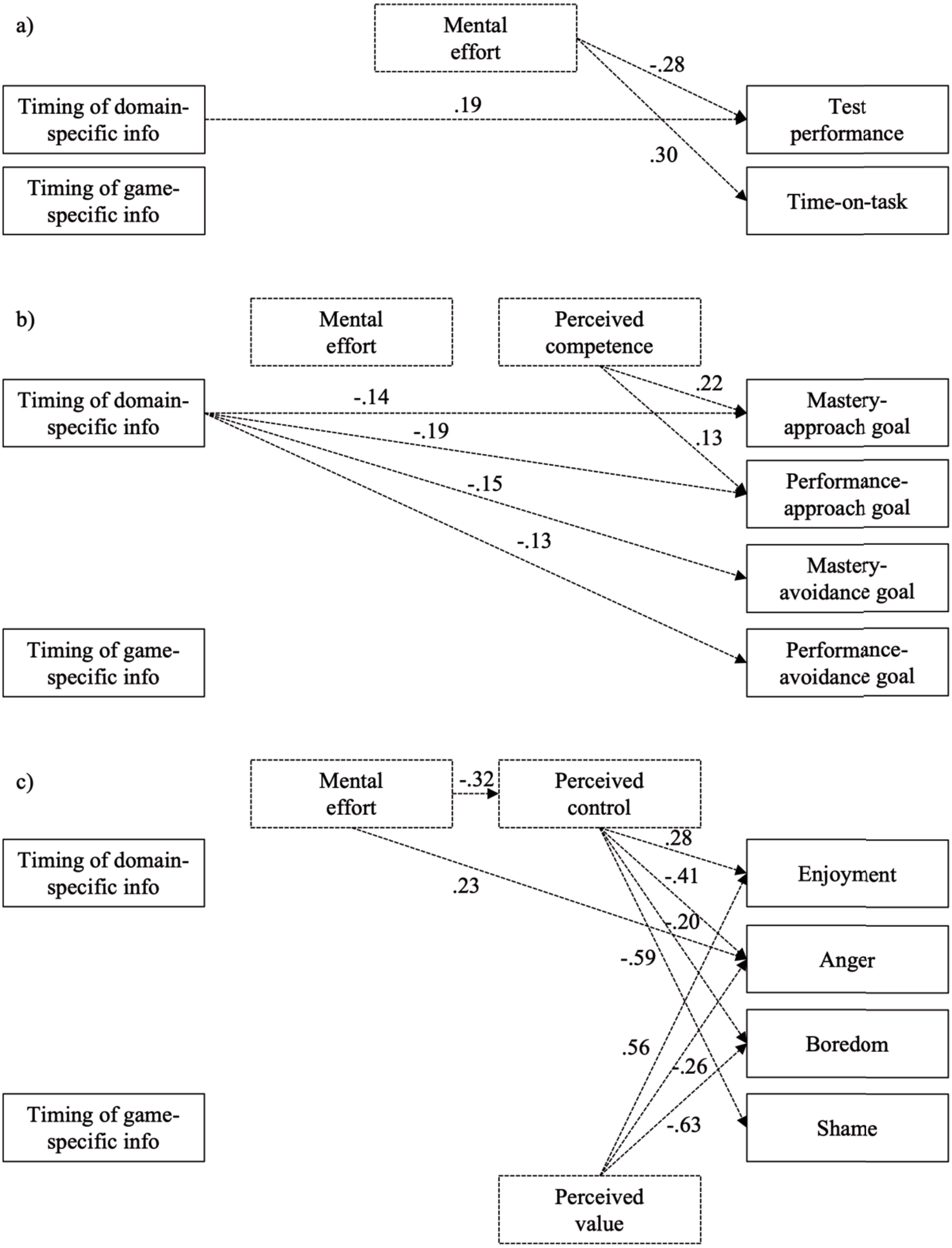
Note. N = 145; β = regression coefficient; SE = standard error; t = test whether β is significantly differs from zero; R^2 = the proportion of the variance of the outcome that is explained by the model; Because the interaction effect between timing of domain-specific information presentation and timing of game-specific information presentation is not statistically significant, all models do not include the interaction; a the model is run by multiple regression; b the model is run by robust methods; Cohen's $f^2 = R^2/(1-R^2)$, ≥.02, ≥.15, and ≥.35 means small, medium, and large effect size; Timing of info: Before = 0, During = 1; $^*p < .05$; $^{**}p < .01$; $^{***}p < .001$.

RQ2: Mediated Effect of Timing of Information Presentation on Learning

SEM revealed that indirect effects of timing of information presentation on performance, achievement goals, and achievement emotions were not statistically significant (Table S3.2 in supplementary materials). The path analysis (Figure 3.3) revealed that timing of domain-specific information presentation had a statistically significant direct effect on test performance (β = .19, p < .05), mastery-approach goals (β = -.14, p < .05), performance-approach goals (β = -.19, p < .01), mastery-avoidance goals (β = -.15, p < .05), and performance-avoidance goals (β = -.13, p < .05). There were no statistically significant direct effects of timing of domain-specific information presentation on game performance and achievement emotions or of timing of game-specific information presentation.

Figure 3.3

Path analysis for a) performance, b) achievement goals, and c) achievement emotions



Note. Only statistically significant direct effects are displayed. Timing of before and during were coded 1 and 2, respectively.

Discussion

Timing of domain-specific and game-specific information presentation matters in GBL. Studying both types of information and playing the game simultaneously may cause cognitive overload, which may result in frustration and demotivation. This study tested how timing of information presentation affects performance, achievement goals, and achievement emotions and whether mental effort, perceived competence, perceived control, and/or perceived value mediate the effects. The pre-training principle and just-in-time principle seems to conflict with each other in terms of their effects on learning.

RQ1: Effect of Timing of Information Presentation on Learning

We found that timing of domain-specific information presentation and timing of game-specific information presentation did not interact, timing of domain-specific information presentation affected test performance, approach goals, avoidance goals, and positive achievement emotions, and timing of game-specific information presentation affected performance-avoidance goals.

Main Effects of Timing of Domain-specific Information Presentation

Mental Effort and Performance

Hypothesis 1a is not supported: Why doesn't timing of information presentation affect mental effort? This may be due to a floor effect: students reported on average rather low mental effort (M = 3.68, SD = 1.56). This indicates that GBL in this study may not have been complex enough to cause cognitive overload during gameplay.

In contrast to our hypothesis 1c1 and Author (2004b) but in line with Author (2004a) and Noroozi et al. (2012), presenting domain-specific information during gameplay promotes higher test performance than presenting it before gameplay. Receiving domain-specific information is the instruction and the gameplay before presenting domain-specific information is the problem-solving. This initial gameplay may create a necessary challenge, encourage learners to try new things (Plass, Mayer et al., 2020), and could provide a meaningful context (Author, 2004a) that allow learners to better understand later domain-specific information. Thus, this result may support *productive failure* (i.e., the process of problem-solving first without additional instruction may frequently result in initial failure but can be a productive exercise for later learning; Kapur, 2008): *problem-solving followed by instruction* leads to higher learning than *instruction followed by problem-solving* (Sinha & Kapur, 2021). However, caution is needed when comparing the results of this study with those of previous studies: Supportive information (i.e., domain models, cognitive strategies; e.g., information about statistical testing in general and circumstances under which a Chi-square test is called for) and procedural information (i.e., rules and procedures, prerequisite knowledge; e.g., info about the exact form of the Chi-square test formula and definitions of the elements in the formula) in previous studies on the just-in-time principles (e.g., Author, 2004a) is different from domain-specific information (i.e., information about the domain; which includes mainly supportive information and slightly procedural information) and game-specific information (i.e., information about the game; which includes procedural information) in this study.

This result may question the pre-training principle. One possible explanation is that the effect of pre-training might be dependent on *the type of information* – the elements (e.g., components of a wet-cell battery), the interactions between these elements (e.g., how a wet-cell battery works), or both. For example, Pilegard and Mayer (2016) found that a pre-training group of experiment 2 receiving a worksheet that asked for the elements and their interactions (i.e., label the diagram of a wet-cell battery and explain how a wet-cell battery works) did not outperform the no-training group. In our study, domain-specific information includes both the elements and their interactions, which may yield a different effect than only including the elements or their relationships. Thus, the pre-training principle may only work when presenting the elements before learning and presenting their interactions during learning.

Another possible explanation is that it might depend on *the timing of training*. For example, Pilegard and Mayer (2016) found that the group who received the same information during learning (during-training) outperformed the pre-training group and those without pre-training (no-training). In our study, students who received domain-specific information during gameplay (during-training) performed better than those who received domain-specific information before gameplay (pre-training). Thus, a pre-training group performs better than a no-training group in general, but compared to a during-training group, the benefit of pre-training may disappear. Further research on the type of training (the elements versus their interactions versus the elements and their interactions) and the timing of training (pre-training versus during-training versus no-training) are needed to confirm these explanations.

Hypothesis 1c2 is not supported: Why doesn't timing of domain-specific information presentation affect game performance? One possible explanation is that the result of mental effort indicates that GBL may not have been complex enough in this study, so students presented with domain-specific information during gameplay did not become cognitively overloaded and not have required different time-on-task. This result does not support the just-in-time principle.

Achievement Goals

In line with our hypothesis 1m1 but in contrast to our hypothesis 1m2, presenting domain-specific information before gameplay promotes higher approach goals and higher avoidance goals than presenting it during gameplay. This implies that presenting domain-specific information before gameplay promotes higher (approach and avoidance) motivation. One possibility is that students receiving domain-specific information before gameplay get an idea of what to expect and therefore may feel less frustrated, more certain, and thus more motivated than those receiving it during gameplay. Among previous studies, Charsky and Ressler (2011) found that pre-training decreased motivation, Fiorella and Mayer (2012) found that pre-training increased satisfaction, and Barzilai and Blau (2014) found that pre-training had no effects on flow. As mentioned before, however, these studies compared mostly pre-training with no-training and thus their results might be not comparable to our results.

A follow-up question is: Why does presenting domain-specific information before gameplay promote higher motivation but results in less test performance than when presenting it during gameplay? On one hand, this result may support that higher motivation does not guarantee higher learning (Schrader et al., 2021). In our study, some students suggested that they prefer to receive domain-specific and/or game-specific information before gameplay because it makes it easier for them to play. Thus, one possible explanation is that even though students are more motivated in their preferred type of learning, they tend to perceive that their preferred type of learning provides an easier learning path and may invest less effort and thus learn less (e.g., TV; Salomon, 1984; Clark et al., 2006).

On the other hand, this result may question that higher motivation in general leads to higher performance (Feldon et al., 2021). Although a meta-analysis suggests positive relations between motivation and performance (Lazowski & Hulleman, 2016), most included studies are observational (non-experimental). Specifically, most meta-analyses that report positive approach goals-performance relations focused on observational studies (e.g., Celler et al., 2011; Huang, 2012; Hulleman et al., 2010; van Yperen et al., 2014; Wirthwein et al., 2013) and only three on experimental studies (i.e., Noordzij et al., 2021; Utman, 1997; van Yperen et al., 2015). It is unclear whether the positive motivation-performance relations hold for cross-sectional experimental studies. Observational and experimental studies differ in aspects, such as the duration, the way of measuring learning outcomes, or the strictness of the setting. Take the duration for example, CLT (Sweller et al., 2019) suggests that task performance is determined by working memory constraints only when time is limited (e.g., in the cross-sectional experimental studies, participants often perform tasks within a fixed time). For the same task, the longer time it takes, the less cognitive load it may impose. Thus, one possibility for the contradictory positions is the study design: cross-sectional experimental studies with negative motivation-performance relations versus longitudinal observational studies with positive motivation-performance relations.

Achievement Emotions

In line with our hypothesis 1e1 and in contrast to hypothesis 1e2, presenting domain-specific information before gameplay promotes more positive achievement emotions but not less negative achievement emotions than presenting it during gameplay. Only two previous studies on the pre-training or just-in-time principle focused on emotions: Barzilai and Blau (2014) found that pre-training had no effects on enjoyment, but Rahimi et al. (2022) found that a pre-training group reported more enjoyment and lower frustration than a no-training group. Again, these studies compared pre-training with no-training and thus their results might be not comparable to our results. A similar study by Muis et al. (2015) reports similar findings: Compared with a no feedback condition, the feedback condition reported less enjoyment but equal boredom in study 1 (groups differ in only positive emotions) and reported more boredom but equal enjoyment in study 2 (groups differ in only negative emotions). Thus, one possible explanation is that instructional design features may only affect either positive emotions or negative emotions but not both.

But why does presenting domain-specific information before gameplay promote more positive achievement emotions but less test performance than presenting it during gameplay? On the one hand, this may support that students often enjoy the learning environment from which they learn less (Clark, 1982), the so-called *suppression hypothesis* (Knörzer et al., 2016; Schrader et al., 2021). There are two possible explanations. First, emotion may act as a source of extraneous load (Plass & Kalyuga, 2019) as positive and negative emotions decrease task-related processing resources, compared with neutral emotions (Meinhardt & Pekrun, 2003). This mechanism may play a role in our study, but future research that measures extraneous load is needed to confirm this explanation. Second, emotions arise regarding how learners are doing well or poorly: Positive emotion may signify things are going better than needed and learners may reduce effort and attend to something else, while negative emotion may signify things are going worse than needed and learners may increase effort and work harder to catch up (Carver, 2003). In our study, students with more positive emotions may perceive they are doing better than needed, reduce effort, and thus perform worse than those with less positive emotions.

On the other hand, this result may question that positive emotions positively relate to cognitive processes and outcomes, the so-called *facilitation hypothesis* (Knörzer et al., 2016, Schrader et al., 2021). Although a meta-analysis on emotions and performance suggests positive relations between enjoyment and performance (Camacho-Morles et al., 2021), most included studies are observational. It is unclear whether the positive enjoyment-performance relations hold for cross-sectional experimental studies. Among the four studies that reported negative enjoyment-performance relations, three are cross-sectional experimental studies (i.e., Behrens et al., 2019; Chevrier et al., 2019; Muis et al., 2015, study 1). Like motivation, thus, one possibility for the contradictory positions is the study design: Cross-sectional experimental studies with negative positive emotions-performance relation versus longitudinal observational studies with positive emotions-performance relations.

Main Effects of Timing of Game-specific Information Presentation

In contrast to our hypotheses, timing of game-specific information presentation did not affect test performance, game performance, approach goals, or achievement emotions except for performance-avoidance goals. This implies that it does not matter when game-specific information is presented. Even when domain-specific and game-specific information are presented together, this may not hamper learning. Further research with high cognitive load information could better address this.

RQ2: Mediated Effect of Timing of Information Presentation on Learning

In contrast to our second hypothesis, mental effort, perceived competence, perceived control, and/or perceived value did not mediate the effects of timing of information presentation. There are two possible explanations. First, instructional design features, such as timing of information presentation, may have only direct effects on learning outcomes in GBL, independent of any mediational processes. If this is the case, it supports that context factors could directly affect achievement goals (Elliot, 1999) or achievement emotions (Pekrun & Perry, 2014). Second, untested mediational processes of the effects of instructional design features on learning in GBL may play a role, as suggested by the decision tree of mediation (Zhao et al., 2010). More research on untested mediators is needed to uncover the processes through which instructional design features affect learning.

Implications

Practically, this study provides rationale that practitioners (e.g., educators and game designers) should attend to the timing of domain-specific information presentation in GBL. Considering that most practitioners pursue learning environments with higher motivation, enjoyment, and learning, a question they may ask is: What is learning goal in terms of cognition, motivation, and emotion? Our findings suggest that practitioners may change the timing of information presentation based on the learning goal. If the goal is to promote cognition, educators should present domain-specific information during gameplay, whereas if the goal is to promote motivation and emotion, practitioners should present domain-specific information before gameplay. However, it may be difficult for educators to change the elements of the game in the implementation phase of GBL in the class, such as removing or adding domain-specific information in the game. We suggest that game designers afford users to adapt certain features in the game, such as creating personal versions of the game. Future research should assess whether similar effects of domain-specific information presentation hold in learning environments other than GBL or instructional design features other than timing of information presentation. Furthermore, our results showed that students perceived higher motivation and more enjoyment in the GBL environments from which they learned less. Future research should assess whether it is possible to design GBL environments that promote motivation, positive emotions, and learning.

Theoretically, this study adds to the research base showing that instructional design features, such as timing of information presentation, may affect performance, achievement goals, and achievement emotions differently. However, it seems that the existing theories, such as CLT, AGT, and CVT may not fully capture the complexity of cognition-motivation-emotion relations. CLT may not sufficiently cover how learners process the learning materials (Gerjets & Scheiter, 2003), such as the interaction between non-cognitive variables (e.g., motivation or emotion) and cognitive load (Feldon et al., 2019). AGT may not sufficiently cover what the learning materials look like. CVT may not sufficiently cover how instructional design features directly affect emotions or how discrete emotions affect cognitive processes and outcomes. This study is one of few studies that focused on the effects of instructional design features on cognition, motivation, and emotion, their mediational processes, and cognition-motivation-emotion relations, more similar research (e.g., Muis et al., 2015) could provide more evidence and move these theories forward.

Methodologically, this short-term experimental study successfully manipulated cognition and further motivation and emotion. Considering the substantial differences between state and trait achievement goals and achievement emotions, the mixed results on cognition-motivation-emotion relations may be due to the study design (observational vs. experimental) or study duration (cross-sectional vs. longitudinal). To confirm this, we call for more short-term and long-term experimental studies that directly manipulate motivation (e.g., Pekrun, Cusack et al., 2014) and emotion (e.g., Plass, Homer et al., 2020) and more meta-analyses (e.g., Wong & Adesope, 2021).

Limitations

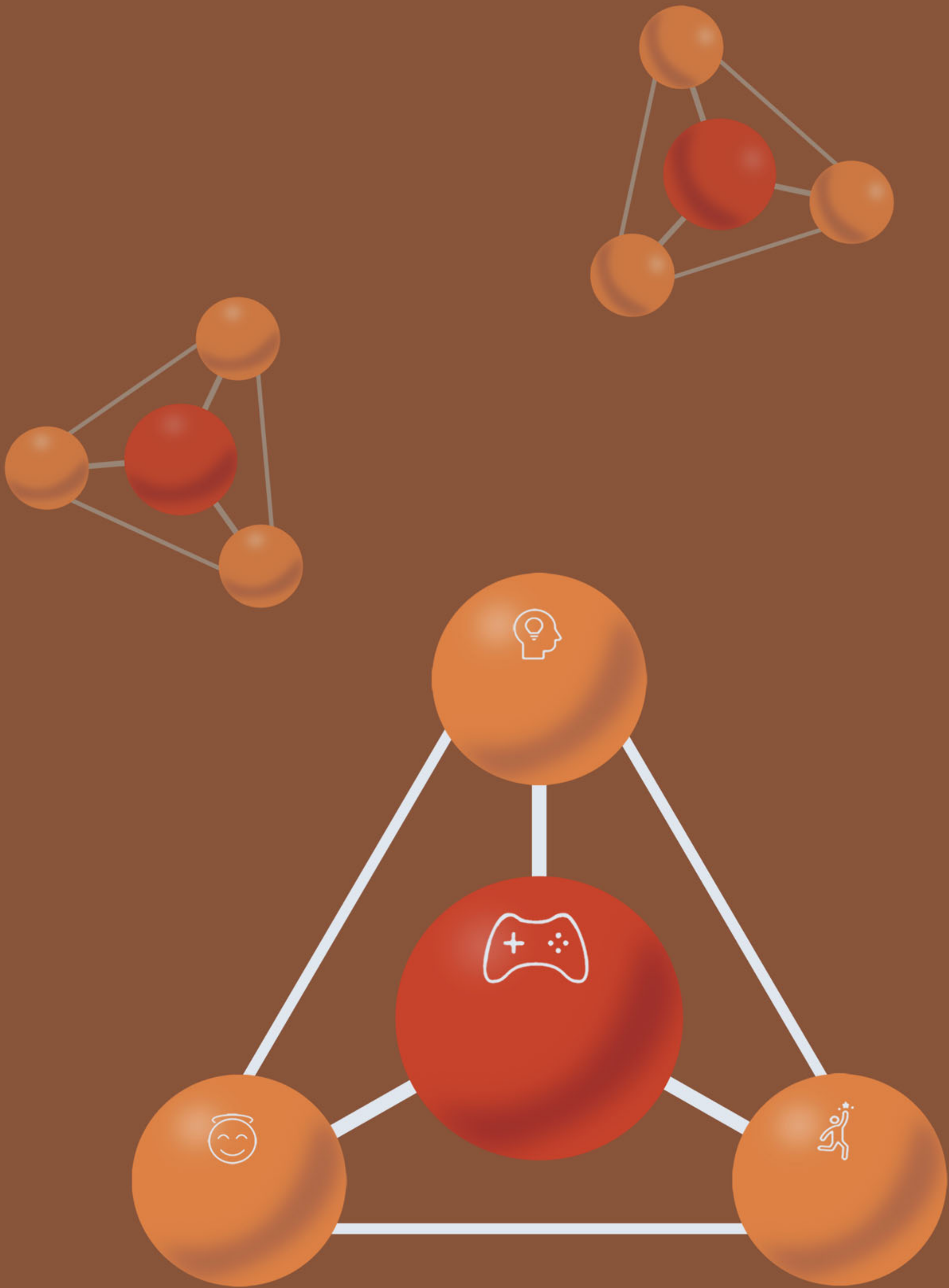
Some limitations need attention when generalizing our results. First, we used a specific GBL environment with a specific game genre (i.e., puzzle game, strategy game) and chemistry content (i.e., separation processes), which makes it a complex learning environment. We may get different results if testing the same content in a virtual reality game or testing the same game genre in chemistry lab skills or in a language course. We call for similar research on other game genres, other chemistry content, or other subjects to confirm our findings.

Second, we only included participants who finished all the game levels and the posttest. We could not distinguish those who did not finish due to their ability from those who did not finish due to other reasons, such as other appointments. The included participants reported on average rather low mental effort, which indicates that the game may have been not too complex for them. For those who did not finish due to their ability, the game may have been too complex and thus they may have reported higher mental effort and perform worse. Therefore, timing of information presentation may affect mental effort and time-on-task. Future research needs to distinguish these two types of participants when collecting data.

Third, motivation and emotion were measured only in a self-report measure at the end of the study. Therefore, it is possible that because there was more time between domain-specific information learning in the "before" group, the report of more positive emotions could contribute to conflation of playing the game than of the overall learning experience. That is, the students were reporting more on feelings because they have been playing the game for a longer amount of time than students who were given training materials in the middle of gameplay, thus breaking up the "fun". Further research should measure the dynamics of motivation and emotion and their change throughout the entire experiment. For example, using the think-aloud protocol to understand behavioral patterns of players' specific actions and more in-depth interviews to understand players' detailed feelings during gameplay (e.g., emote-aloud; Graesser et al., 2014).

Conclusion

We conclude that timing of information presentation may affect learners' cognitive, motivational, and emotional processes and outcomes in GBL differently and that students feel more motivated and enjoy the learning environment from which they learn less. Consequently, educators may change the timing of information presentation based on the learning goal, such as promoting cognition, motivation, and/or emotion. This study is one of the few that attends to information presentation, which shows that timing of information presentation matters. This study is one of the first to focus on cognitive, motivational, and emotional processes and outcomes, which shows that research on instructional design features should attend all three outcomes instead of just one or two. This study is also one of the first to integrate the pre-training principle and the just-in-time principle, which shows that more research is needed before we implement these two principles. In this way, the results of this study on the specific conditions (regarding the learning outcomes and methodological dimensions) under which both principles can be effective may advance the cognitive theory of multimedia learning and the four-component instructional design model.



Chapter 4 Effects of achievement goal instructions on students’ achievement goals, mental effort, performance, and achievement emotions in game-based learning

This chapter is based on:

Hu, Y., Wouters, P., van der Schaaf, M., & Kester, L. (2023). The effects of achievement goal instructions in game-based learning on students’ achievement goals, performance, and achievement emotions. *Manuscript submitted for publication*. Department of Education, Utrecht University

Acknowledgement of author contributions

All authors designed the study. Yuanyuan Hu recruited participants, collected, and analyzed the data, and drafted the manuscript. All authors contributed to critical revision of the manuscript. Pieter Wouters, Marieke van der Schaaf, and Liesbeth Kester supervised the study.

Acknowledgments

We would like to thank Andrew Elliot (from University of Rochester) for comments on the study and Rebecca Kuiper, Mirjam Moerbeek, and Huub van den Bergh (from Utrecht University) for sharing their statistical expertise.

Abstract

Achievement goal instructions can induce specific goals in learning, but it is unclear which achievement goal instruction is best for motivation, cognition, and emotion. in game-based learning. The purpose of this paper is to investigate 1) how achievement goal instructions affect motivation (i.e., achievement goals), cognition (i.e., mental effort and performance), and emotion (i.e., achievement emotions) in chemistry game-based learning and 2) whether prior achievement goals moderate the effects of achievement goal instructions. Achievement goals instructions are instructions that assign the learners' achievement goals beforehand, such as mastery-approach goal instructions that emphasize "to learn as much as possible" and performance-approach goal instructions that emphasize "to be the best player". Secondary school students (N = 450) were randomly assigned to one of the four groups, namely, mastery-approach goal instructions, performance-approach goal instructions, combined mastery-approach and performance-approach goal instructions, and no goal instructions. Robust regression analysis revealed that mastery-approach goal instructions promoted higher mental effort but had no effects on induced mastery-approach goals and posttest performance. Performance-approach goal instructions promoted higher induced performance-approach goals and higher mental effort but lower posttest performance. Prior mastery-approach goals moderated the effects of achievement goal instructions on mental effort. We conclude that achievement goal instructions in game-based learning affect cognitive and motivational outcomes differently. Educators would do well to consider achievement goals instructions and learners' prior mastery-approach goals. The paper may advance the field of achievement goal theory, instructional design, and game-based learning.

Keywords: Goal instructions; Achievement goals; Achievement emotions; Performance; Game-based learning

Introduction

Stakeholders, in particular educators, expect that game-based learning (GBL) is effective (i.e., facilitating cognitive processes and outcomes), motivating (i.e., facilitating motivational processes and outcomes), and enjoyable (i.e., facilitating emotional processes and outcomes; see Author, 2022a for an overview). GBL needs effective instructional design features (Plass et al., 2020) to achieve this. In this study, we focus on the instructional design feature: *goal instructions* - instructions that assign the learners' specific goals beforehand (Erhel & Jamet, 2016). Practically, compared with other instructional design features, goal instructions can be easily implemented in GBL. Theoretically, goal instructions may affect at least the cognitive aspects, such as cognitive load, in learning (Hawlitschek & Joeckel, 2017). Empirically, previous studies have shown that goal instructions can induce specific goals. However, the results of goal instructions on cognitive processes and outcomes are inconsistent, and research on the effects on motivation or emotion is sparse or even lacking (Erhel & Jamet, 2013, 2016, 2019; Hawlitschek & Joeckel, 2017; Miller et al., 1999; Nebel et al., 2017; Vandercruysse et al., 2015).

This study investigates how achievement goal instructions affect motivational, cognitive, and emotional processes and outcomes in chemistry GBL and whether prior achievement goals moderate the effects of achievement goal instructions. *Achievement goals* are goals that relates to competence-relevant behavior inside or outside of achievement settings (Elliot & Hulleman, 2017). They can be categorized into two types: *Mastery-approach goals* that emphasize mastery and *performance-approach goals* that emphasize outperformance. It is known that achievement goals influence GBL (Plass et al., 2020) and researchers have suggested that *multiple goals* - a combination of mastery-approach goals and performance-approach goals - have a stronger positive influence on learning than either mastery-approach goals or performance-approach goals (Barron & Harackiewicz, 2001). However, research on *achievement goal instructions* is lacking (Elliot & Hulleman, 2017). In addition, the effects of achievement goal instructions may depend on *prior achievement goals* - achievement goals that are set prior to learning (Niemivirta, 2002). When students enter learning settings with different prior achievement goals, prior achievement goals may interact with achievement goal instructions (Murayama & Elliot, 2009). However, it is yet unclear whether prior achievement goals moderate the effects of achievement goal instructions (*moderation effect)*.

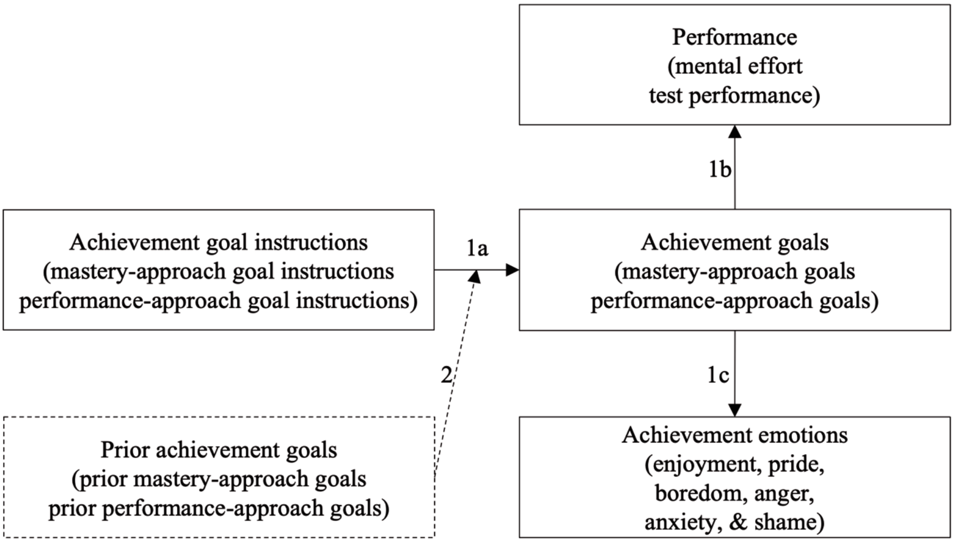
So, the major theoretical contribution of this study will be insight in the effect of achievement goal instructions on not only cognition but also motivation and emotion, and the role therein of prior achievement goals. The major practical contribution is to guide the practitioners, such as educators and game designers, to induce achievement goals that promote motivating, effective, and enjoyable learning experiences in educational settings, such as GBL, and consider learners' individual characteristics, such as prior achievement goals.

Theoretical Background and Empirical Research

We assume that achievement goal instructions influence GBL in three ways: by affecting motivational processes, by affecting cognitive processes, and/or by affecting emotional processes (see Author, 2022 for a detailed theoretical framework). This study focuses on achievement-relevant theories, namely, achievement goal theory (AGT; Elliot & Hulleman, 2017), cognitive load theory (CLT; Sweller et al., 2019), and control-value theory of achievement emotions (CVT; Pekrun & Linnenbrink-Garcia, 2014) to substantiate this assumption. These theories and their constructs are chosen because of their critical role when assessing motivation, cognition, and emotion in complex learning, such as GBL (Loderer et al., 2020; Plass et al., 2020). Figure 4.1 illustrates the theoretical propositions.

Figure 4.1

Overview of the relations between achievement goal instructions, achievement goals, performance, achievement emotions, and prior achievement goals



Note. Rectangles represent independent and dependent variables; Arrows represent causal relations; Dotted rectangles represent moderators; Dotted arrows represent moderation. To reduce the complexity, the direct effects of covariates including prior achievement goals and pretest performance are not shown but are tested.

Achievement Goal Instructions and Achievement Goals

From a motivational perspective, achievement goal instructions may promote *achievement goals* (i.e., achievement goals that are induced by achievement goal instructions, Noordzij et al., 2021). Achievement goals instructions are examples of *achievement goals interventions* – interventions that aim to manipulate achievement goals. According to AGT (Elliot & Hulleman, 2017), *achievement goals* indicate whether one is doing poorly or well, relative to the demands of a task (task-based), past performance or future potential (self-based or intrapersonal), and/or others (other-based or normative). The 2 × 2 achievement goal model (Elliot & Murayama, 2008) introduces four categories of achievement goals: 1) *mastery-approach goals:* striving for task-based or self-based competence, 2) *mastery-avoidance goals:* avoiding task-based or self-based incompetence, 3) *performance-approach goals:* striving for other-based competence, and 4) *performance-avoidance goals:* avoiding other-based incompetence. This study focuses only on approach goals, considering that approach goals are more beneficial for learning than avoidance goals (see 1.1.2 and 1.1.3 for more elaboration) and avoidance goals, especially mastery-avoidance goals, may not be as prevalent among secondary school students as theory expected (see e.g., Bong, 2009; Strunk et al., 2020).

According to a meta-analysis on achievement goal interventions (Bardach et al., 2020), mastery-approach goal instructions that emphasize mastery, understanding, and improvement may mostly induce mastery-approach goals, performance-approach goal instructions that emphasize competition may mostly induce performance-approach goals, and multiple goal instructions may mostly induce multiple goals (Figure 4.1 line 1a). Based on the results of four similar studies, it is unclear whether achievement goal interventions work: Three studies reported that achievement goal interventions worked for both mastery-approach goals and performance-approach goals (Pahljina-Reinić & Kolić-Vehovec, 2017) or for either mastery-approach goals (Linnenbrink, 2005) or performance-approach goals (Muis et al., 2013), and one did not seem to check this (Barron & Harackiewicz, 2001). Based on this disquisition, we expect that achievement goal instructions promote corresponding achievement goals (Figure 4.1 line 1a). If this is the case, educators can use achievement goal instructions to induce achievement goals in GBL.

Achievement Goals, Mental Effort, and Performance

From a cognitive perspective, achievement goal instructions may affect cognitive load and performance via achievement goals (Figure 4.1 coupling of lines 1a and 1b; Murayama & Elliot, 2009). According to the updated CLT (Sweller et al., 2019), three types of cognitive load can be defined: 1) *intrinsic load* - the load caused by processing task-relevant information; the more complex this information the higher the intrinsic cognitive load and the more knowledgeable the learner the lower the intrinsic cognitive load; 2) *extraneous load* - the load caused by cognitive processes or activities that are unnecessary for learning and performing the task; for example, searching for task-relevant information because of ill-structured text or ill-designed diagrams; and 3) germane load – the load caused by cognitive processes or activities that are relevant for learning and performing the task; for example, distinguishing main and side points in task-relevant information. The overall cognitive load - intrinsic load plus extraneous load - varies and can be estimated by the amount of mental effort that learners exert in a task (Sweller et al., 2019). Germane load arises when mental effort is invested in learning from task-relevant information on top of merely processing it (intrinsic load) instead of investing effort in processes irrelevant for learning (extraneous load). So, germane load does not add to the overall cognitive load but redistributes the mental effort investment from extraneous to intrinsic task aspects. Given the methodological weakness of measuring intrinsic load and extraneous load separately (Krieglstein et al., 2022), this study focuses on mental effort as the indicator of overall cognitive load.

Although research that investigates the relation between achievement goal instruction and mental effort is sparse, we propose that mastery-approach goals and performance-approach goals may increase mental effort, relative to no goals. Especially, mastery-approach goals that focus on learning and improving, are likely to stimulate learners to invest more mental effort in the learning process (i.e., intrinsic and germane load). In addition, performance-approach goals focus on outperforming, and learners are likely to invest more mental effort on, for instance, winning (i.e., extraneous load). Thus, mastery-approach goals and performance-approach goals have a similar effect on invested mental effort albeit through different mechanisms. However, investing more mental effort does not necessarily mean higher performance due to the different sources of cognitive load: Relative to no goals, higher mental effort may come from intrinsic and germane load for mastery-approach goals and from extraneous load for performance-approach goals.

Regarding the relation between achievement goals and performance, AGT proposes that mastery-approach goals and performance-approach goals focus on success and promote task engagement and performance, relative to no goals (Elliot & Hulleman, 2017). Previous meta-analyses on achievement goals found that mastery-approach goals generally promote higher performance than no goals, whereas performance-approach goals do not promote higher performance than no goals (Noordzij et al., 2021; Utman, 1997; van Yperen et al., 2015). An example of the achievement goal instructions of the studies included in the meta-analysis is “the purpose of this session was to teach you a new way of doing math, to adopt a learning goal as you go through the session, and to focus on how well the new techniques can help you develop and improve you math skills” (mastery-approach goal instructions) and “the purpose of the session is to evaluate how well you could perform match problem using a new way of doing match, to adopt a performance goal as you go through the session, and to focus on how well the techniques can help you perform and solve more math problems than other students” (performance-approach goal instructions; Barron & Harackiewicz, 2001). Based on these findings, we expect that mastery-approach goals are associated with higher cognitive outcomes, such as mental effort and performance, than no goals, while performance-approach goals are associated with higher mental effort and higher or equivalent performance than no goals (Figure 4.1 line 1b).

Moreover, regarding multiple goal instructions, the *multiple goals perspective* (Barron & Harackiewicz, 2001) suggests that multiple goals are more beneficial for learning than either mastery-approach goals or performance-approach goals. This can be explained by four mechanisms: 1) mastery-approach goals and performance-approach goals affect different outcomes (specialized effects); 2) the most relevant goal in a given context produces an effect (selective effects); 3) each goal is beneficial for one outcome, independent of other goals (additive effects), or 4) mastery-approach goals and performance-approach goals interact (interactive effects; Muis et al., 2013). In addition, the effects of performance-approach goals depend on many factors, such as the age of the learners, the type of the performance-approach goals measure, and/or the definition of performance-approach goals (Hulleman et al., 2010; Senko & Dawson, 201). Based on this disquisition, we expect that multiple goals are associated with the highest mental effort and performance, relative to mastery-approach goals, performance goals, or no goals.

Achievement Goals and Achievement Emotions

From an emotional perspective, achievement goal instructions may affect achievement emotions via achievement goals (e.g., Pekrun et al., 2006, 2009; Pekrun et al., 2014; Figure 4.1 coupling of line 1a and 1c). According to CVT (Pekrun & Linnenbrink-Garcia, 2014), *achievement emotions* are emotions that relate to competence-relevant activities (e.g., attending class) and/or outcomes (i.e., success or failure) in achievement settings. Depending on the object of emotions, achievement emotions can be distinguished as *activity emotions* associated with achievement-relevant activities or tasks, such as enjoyment, boredom, and anger, and *outcome emotions* associated with the outcomes of those activities, such as pride, hope, anxiety, shame, and anger. Combined with the valence of emotions (positive or negative), achievement emotions can be categorized into four categories: 1) *positive activity emotions,* such as enjoyment; 2) *negative activity emotions,* such as boredom and anger; 3) *positive outcome emotions* related to success, such as hope and pride; and 4) *negative outcome emotions* related to failure, such as anxiety and shame. Achievement emotions can help or harm learning (Pekrun & Linnenbrink-Garcia, 2014), so the goal of instructional design in GBL is to induce achievement emotions that help learning. This study focuses on learning-related emotions, including enjoyment, pride, anger, anxiety, boredom, and shame, as these emotions are typical in GBL (Loderer et al., 2020).

The relations between achievement goals and achievement emotions may depend on the type of achievement goals (mastery goals or performance goals) and the type of achievement emotions (activity emotions or outcome emotions). Theoretically, mastery-approach goals may focus students' attention on the activity itself and thereby influence activity emotions, such as enjoyment or boredom, whereas performance-approach goals may focus students' attention on the outcomes and thereby influence outcome emotions, such as pride or shame (Pekrun et al., 2014). Empirically, these links between mastery goals and activity emotions and between performance goals and outcome emotions have been supported by studies on achievement goal interventions (e.g., mastery-approach goals feedback or performance-approach goals feedback; Pekrun et al., 2014). In addition, a meta-analysis on personal achievement goals (Huang, 2011) suggested that mastery-approach goals and performance-approach goals are mostly positively associated with positive emotions and negatively associated with negative emotions.

Based on this disquisition, we expect that mastery-approach goals are more associated with positive activity emotions and less with negative activity emotions than no goals, and performance-approach goals are more associated with positive outcome emotions and less with negative outcome emotions than no goal instructions (Figure 4.1 line 1c). Based on multiple goal perspective mentioned earlier, we expect that multiple goals are associated with the highest positive emotions and lowest negative emotions, relative to mastery-approach goals, performance-approach goals, or no goals.

Moderation Effect of Prior Achievement Goals on the Effect of Achievement Goal Instructions

Prior achievement goals may function as moderators if they either strengthen or weaken the effects of achievement goal instructions on achievement goals (Figure 4.1 lines 2). According to research on person-environment interactions, learning processes and outcomes are expected to be optimal when the characteristics of the person are congruent with those of the social environment (Eccles et al., 1993). Specifically, we expect that achievement goal instructions may interact with the characteristic of the person, such as prior achievement goals (Murayama & Elliot, 2009).

According to the *match hypothesis* (Murayama & Elliot (2009), achievement goal instructions may have optimal effects on learning processes and outcomes when they match prior achievement goals. For example, mastery-approach goal instructions (or performance-approach goal instructions) may have most positive influences on achievement goals, mental effort, performance, and achievement emotions when learners hold higher prior mastery-approach goals (or prior performance-approach goals). According to the *mismatch hypothesis* (Murayama & Elliot (2009), if achievement goal instructions mismatch prior achievement goals, the effects of achievement goal instructions may be *vibrated,* that is, a beneficial effect is weakened, *migrated,* that is, a detrimental effect is weakened, or *exacerbated,* that is, a detrimental effect is strengthened. For example, mastery-approach goal instructions may have weaker positive effects when learners hold higher prior performance-approach goals (a vibration effect); mastery-avoidance goal instructions may have weaker negative effects when learners hold higher prior performance-approach goals (a migration effect); or performance-avoidance goal instructions may have stronger negative effects when learners hold higher prior performance-approach goals (an exacerbation effect).

Empirically, evidence from seven previous studies is inconclusive. Five studies supported the match hypothesis, but only for some learning processes and outcomes (e.g., effort) and not for others (e.g., course grade) and only one out of four studies supported the mismatch hypothesis: A positive correlation between performance-approach goals and intrinsic motivation is vibrated by mastery goal structures (Hofverberg & Winberg, 2020; Lau & Nie, 2008; Linnenbrink, 2005; Muis et al., 2013; Murayama & Elliot, 2009; Wolters et al., 2004). In addition, five studies were non-experimental and only two manipulated achievement goals (Linnenbrink, 2005; Muis et al., 2013). Based on this disquisition, we expect that prior achievement goals moderate the effects of achievement goal instructions. If this is the case, educators would do well to consider students' prior achievement goals.

Present Study

This study investigates how mastery-approach goal instructions and performance-approach goal instructions affect motivation (i.e., achievement goals), cognition (i.e., mental effort and performance), and emotion (i.e., achievement emotions) in chemistry GBL (**RQ1**) and whether prior achievement goals moderate the effects (**RQ2**).

RQ1 (Main effects and interaction effects): Mastery-approach goal instructions promote higher mastery-approach goals (**H1mot1**), higher mental effort and posttest performance (**H1cog1**), and higher positive (enjoyment) and lower negative activity emotions (boredom, anger) (**H1emo1**) than no mastery-approach goal instructions.

Performance-approach goal instructions promote higher performance-approach goals (**H1mot2**), higher mental effort and higher or equivalent posttest performance (**H1cog2**), and higher positive (pride) and lower negative outcome emotions (anxiety, shame) (**H1emo2**) than no performance-approach goal instructions.

Multiple goal instructions promote highest approach goals (**H1mot3**), highest mental effort and posttest performance (**H1cog3**), and highest positive emotions (enjoyment and pride) and lowest negative emotions (boredom, anger, anxiety, and shame) (**H1emo3**) than other groups.

RQ2 (Moderation): The higher the prior mastery-approach goals or the lower the prior performance-approach goals, the higher the effects of mastery-approach goal instructions on mastery-approach goals (**H2mot1**), mental effort, performance (**H2cog1**), and achievement emotions (**H2emo1**).

The higher the prior performance-approach goals or the lower the prior mastery-approach goals, the higher the effects of performance-approach goal instructions on performance-approach goals (**H2mot2**), mental effort, performance (**H2cog2**), and achievement emotions (**H2emo2**).

Method

Participants

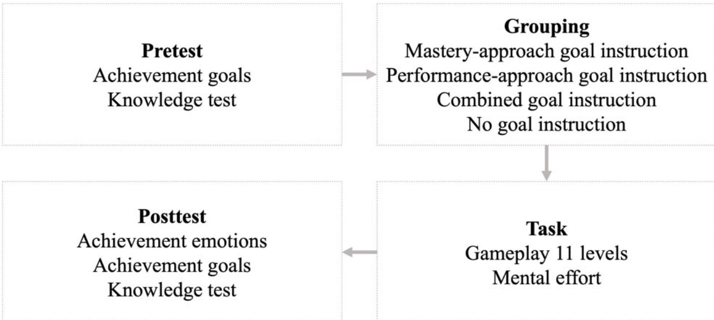
Based on a prior power analysis in G*power 3.1 (Faul et al., 2007),our minimum sample size is 128 (effect size f = .25, alpha = .05, power = .80) We approached chemistry teachers from 113 secondary schools in the Netherlands. Seven chemistry teachers from seven schools were interested in participating with at least one class. A total of 583 students from nine schools participated. At the end of the experiment, each student received a pack of snacks, and each teacher received a 25 euros voucher. We excluded 133 participants who played the wrong game levels and/or who did not finish the game due to other obligations, such as other appointments. Thus, the missing data were unrelated to any of the study variables. In total, 450 participants were included (243 male, 201 female, 6 neutral, age range = 13-18, M = 15.1 years, SD = 1.1). Post-hoc sensitivity analysis indicated that with this sample size we can detect a small effect size of .19 (alpha = .05; power = .80).

Design

The experiment used a 2 (mastery-approach goal instructions: yes or no) × 2 (performance-approach goal instructions: yes or no) design. We randomly assigned all participants to one of the four groups: 1) mastery-approach goal instructions (n = 110); 2) performance-approach goal instructions (n = 111); 3) multiple goal instructions (n = 112); and 4) no goal instructions (n = 117; Figure 4.2).

Figure 4.2

The procedure and measures



Materials and Measures

The materials and measures, except achievement goal instructions, were based on a previous study (Hu et al., 2022a). All the materials and measures were in Dutch.

The materials and measures, except for the achievement goal instructions, were based on a previous study (Author, 2022b). All the materials and measures were in Dutch.

Achievement Goal Instructions

Three achievement goal instructions were created. The *mastery-approach goal instruction* emphasized mastery and learning: "*When you play the game, try to master the chemistry content. For example, when you are in the game level of the sieve, try to learn what is a sieve, when and how to use a sieve. The goal is to learn as much as possible.*". The *performance-approach goal instruction* emphasized outperformance and winning: "*When you play the game, try to be the first to finish the game. For example, when you are in the game level of the sieve, try to finish the level as fast as possible. The goal is to be the best player.*". The *multiple-goal instruction* emphasized mastery, learning, outperformance, and winning: "*When you play the game, try to master the chemistry content and to be the first to finish the game. For example, when you are in the game level of the sieve, try to learn what is a sieve, when and how to use a sieve, and to finish the level as fast as possible. The goal is to learn as much as possible and to be the best player.*". The *no goal instructions* received no goal instructions. All groups received the same general introduction: "*In the next hour, you will play a chemistry game and explore separation techniques to recycle* *materials. In the game, you will complete 11 game levels (1, 2, 3, 7, 13, 19, 22, 24, 26, 29, and 32) that introduce you to 9 processors, such as a sieve, a melter, a magnet, a shredder, a non-ferrous separator, a stream separator, a boiler, a dissolver, and a centrifuge.*".

The mastery-approach goal instructions focused on task-based goals instead of self-based goals as it was difficult to ask participants to adopt self-based goals in our GBL environment, and thus task-based goals were more relevant than self-based goals. These instructions were written as texts on Gorilla (https://app.gorilla.sc). Participants received these instructions after the pretest but before the gameplay. Participants who finished the study were required to leave and other participants could see who finished and who did not.

Immediately after the achievement goal instructions, participants were asked to choose "Yes" or "No" to one of the following questions: "*Based on the previous instructions, should you play the game to learn as much as possible? (Mastery-approach goal instructions)*"; "*Based on the previous instructions, should you play the game to be the best player? (Performance-approach goal instructions)*"*; or* "*Based on the previous instructions, should you play the game to learn as much as possible and be the best player?"(Multiple goal instructions)*". The no goal instructions group received no questions. This question was to ensure that participants read and understood the achievement goal instructions. After this question but before gameplay, we again showed participants the correct answer to reconfirm the manipulated achievement goals.

The game – CosmiClean

CosmiClean (recyclegame.eu) was designed by LuGus Studios (www.lugus-studios.be) to teach secondary school and university students the principles for separation processes of recycling materials. The chemistry learning content includes the functions of the nine separators (including the sieve, the melter, the magnet, the shredder, the non-ferrous separator, the stream separator, the boiler, the dissolver, and the centrifuge) and the eight properties (including size, density, phase, melting point, boiling point, solubility, magnetic metal, and non-ferrous metal) of 12 materials (including iron, plastics, concrete, wood, glass, sand, copper, water, salt, fuel, gold, and solvent). Players complete a series of game levels with different mixtures of materials in a spaceship cargo. The goal is to make a recycling chain, including a conveyor (for transporting the materials), one or more separators (for separating material based on different properties), and receptors (for collecting the recycled materials).

Participants completed the 11 game levels individually and once. Levels 1 and 2 are preinstalled with receptors. Level 3 is the first complete task in which players are required to build a complete recycling chain. From level 3 onwards, each level introduces a new separator, and a higher level does not necessarily have to be any more challenging than lower levels. Participants were asked to play from low to high levels and could not skip levels.

Knowledge Test

The pre- and post-knowledge test included the same test items, but in a different order. The knowledge test assesses Remember (5 multiple-choice questions), Apply (7 multiple-choice questions), and Evaluate (3 open-ended questions) levels based on the Bloom taxonomy (Anderson & Krathwohl, 2001). Each multiple-choice question had three alternatives. For example, a Remember question is: "*Which property is used by stream separator?"*. An Apply question is: "*Gold is not dissolvable. Which processor can be used to separate copper and gold from their* *mixture?*”. An Evaluate question is: “*To separate water and plastics, your teacher will select between a steam separator and a dissolver. Explain which is more appropriate*.”. The knowledge tests were reliable (pretest: greatest lower bound = .55; posttest: greatest lower bound = .70). Prior knowledge tests do not usually measure the same underlying construct (nine separators instead of one), and a reliability value lower than .70 is normal (Taber, 2018). All multiple-choice questions had acceptable item discrimination (greater than .2) and item difficulty (ranging from .2 - .8; Cohen et al., 2018).

Demographic Information

Demographic information measured age and sex.

Mental Effort

The original Paas’ (1992) scale was used to measure how much mental effort was invested in each game level (1 = very, very low mental effort, 9 = very, very high mental effort; see Sweller et al., 2019).

Achievement Goals Questionnaire

The achievement goals questionnaire (AGQ) was adopted from Bipp and Van Dam (2014), based on the Achievement Goals Questionnaire-Revised (Elliot & Murayama, 2008; 1 = strongly disagree, 5 = strongly agree) and measures mastery-approach goals and performance-approach goals. Because motivational constructs, such as achievement goals, vary across contexts (cf. e.g., Bong, 2001), the items were framed as goals for the game. An example of an item is “*My goal is to learn as much as possible in the game*”. The items of the mastery-approach goals focused on task-based goals instead of self-based goals. The AGQ was administered twice, once before the intervention started and once after it, with the same items but in a different order. The first AGQ was used to determine prior achievement goals and the second one to measure achievement goals after the achievement goal instructions. All subscales were reliable (greatest lower bound ranging from .77 - .92).

Achievement Emotions Questionnaire

The achievement emotions questionnaire (AEQ) was adapted from Donker et al. (2021), based on the Achievement Emotions Questionnaire (Pekrun et al., 2011). The 20 items assess enjoyment (4 items), pride (3 items), anger (4 items), anxiety (3 items), boredom (3 items), and shame (3 items; 1 = strongly disagree, 5 = strongly agree). The instruction of the questionnaire asked participants to describe how they felt during the game. All subscales were reliable (greatest lower bound ranging from .79 to .94), except for anxiety (greatest lower bound = .57) which was removed from further analysis.

Procedure

Pilot Study

We ran a pilot study with 57 participants to check whether they understood all the materials or whether they experienced technical problems with the online experiment in Gorilla (app.gorilla.sc) and Qualtrics (www.qualtrics.com). The unclear materials were changed. Participant indicated that they understood the achievement goal instructions.

Main Study

After giving informed consent (note that parental consent was needed and obtained if participants are younger than 16 years old), participants received a written introduction about the number of sections, the duration, and the rules (e.g., answer carefully) of the experiment and completed the AGQ and prior knowledge test (pretest) in 15 mins (see Figure 4.2). Afterwards, they were randomly assigned to one of the four groups, read the achievement goal instructions, and finished gameplay within 1 hour, during which they rated mental effort for each game level. Then they completed the demographic information, AEQ, AGQ, and post knowledge test (posttest) in 20 mins.

Scoring, Data Preparation, and Data Analysis

For the knowledge test, we calculated a sum score of 12 multiple-choice questions (1 point per correct answer per question) and three open-ended questions (3 points per correct answer per question; partial credit is allowed), with a maximum score of 21 points. For the three open-ended questions, we developed a coding schema. Two raters first scored independently 10% of the pretest and posttest answers to each question to resolve disagreements (inter-rater reliability Cohen's k = .90) and then scored the remainder (k = .89). For AEQ and AGQ, we first calculated the scale means of all items. For the mental effort, we calculated the means of all game levels.

For RQ1 and RQ2, data were analyzed by robust regression using the lmrob function from the robustbase package (Maechler et al., 2021) in R studio (R Studio Team, 2022). As suggested by Field & Wilcox (2017), robust regression was used because all variables violated the assumptions of normality of residuals and multivariate normality (Shapiro-Wilk test) and there were outliers (| standardized residuals | > 3). The assumptions of homogeneity of variances (Levene's test), homogeneity of covariance matrices (Box's test), independence of residuals, linearity, and no multicollinearity (Field et al., 2012) were all met. We checked for missing data (less than 5%). We will run a robust multivariate regression with mastery-approach goal instructions (yes or no) and performance-approach goal instructions (yes or no) as the factors, mastery-approach goals, performance-approach goals, mental effort, posttest performance, enjoyment, pride, anger, boredom, and shame as the dependent variables, prior mastery-approach goals and performance-approach goals as moderators, and prior achievement goals and pretest performance as the covariates (see Figure S4.1 in supplementary materials for Pearson correlation). However, since robust multivariate regression is currently unavailable in robustbase (M. Maechler, personal communication, January 20, 2023), we used robust regression for each dependent variable.

Participants (level 1) were nested within teachers (level 2). Our Intraclass Correlation Coefficients were low (< .08). However, the sample size at level 2 (n = 7) was too small for multilevel regression. As suggested by McNeish and Stapleton (2016), we dummy coded teachers and included them as predictors.

Results

Table 4.1 presents the means and standard deviations for the dependent variables and covariates.

RQ1: Effect of Achievement Goal Instructions on Learning

There were no statistically significant interaction effects on all learning processes and outcomes. The main effects are explained below (see Table 4.2).

Achievement Goals

Robust regression revealed a statistically significant main effect of the performance-approach goal instructions on performance-approach goals (t = 1.97, p = .049, f^2 = .59) with a large effect size after controlling for teacher, pretest performance, and prior achievement goals. Participants receiving the performance-approach goal instructions reported higher performance-approach goals than those who did not receive the performance-approach goal instructions. There were no statistically significant main effects on mastery-approach goals after controlling for teacher, pretest performance, and prior achievement goals. As a result, mastery-approach goal instructions were not included in the further analysis.

Mental Effort and Performance

After controlling for teacher, pretest performance, and prior achievement goals, robust regression revealed a statistically significant main effect for performance-approach goal instructions on mental effort with a very small effect size (t = 2.05, p = .041, f^2 = .002) and posttest performance with a small effect size (t = -2.04, p = .041, f^2 = .06). Participants receiving the performance-approach goal instructions reported higher mental effort but achieved lower posttest performance than those who did not receive the performance-approach goal instructions.

A paired t-test shows that there is a statistically significant difference between pretest performance and posttest performance (t = 8.19; df = 449; p < .001) with a small effect size (Cohen's d = .42), so, learning happened.

Achievement Emotions

Robust regression revealed no statistically significant main effects on achievement emotions.

RQ2: Moderation Effect of Prior Achievement Goals

Robust regression revealed that there was a statistically significant interaction effect between prior mastery-approach goals and the performance-approach goal instructions (t = -2.00, p = .046) on mental effort (see Table 4.2 and Figure S4.2 in supplementary materials). With higher prior mastery-approach goals, participants who received the performance-approach goal instructions reported lower mental effort than those who did not receive the performance-approach goal instructions. There were no statistically significant moderating effects on achievement goals, posttest performance, and achievement emotions.

Table 4.1

Mean and standard deviation for dependent variables and covariates

	Mastery-approach goal instructions (n = 110)		Performance-approach goal instructions (n = 111)		Multiple goal instructions (n = 112)		No goal instructions (n = 117)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Prior mastery-approach goals (1-5)	3.56	.65	3.72	.57	3.65	.65	3.63	.65
Prior performance-approach goals (1-5)	3.22	.85	3.32	.87	3.33	.85	3.34	.81
Mastery-approach goals (1-5)	3.53	.75	3.60	.68	3.55	.77	3.46	.80
Performance-approach goals (1-5)	2.98	.96	3.23	.97	3.18	.95	3.14	.98
Pretest (0-21)	5.58	2.46	5.68	2.06	5.54	1.92	5.64	2.12
Posttest (0-21)	6.81	2.81	6.39	2.90	6.44	2.63	6.92	2.82
Mental effort (1-9)	3.66	1.32	3.49	1.29	3.92	1.56	3.56	1.65
Enjoyment (1-5)	3.20	.94	3.17	1.00	3.09	.96	3.16	.99
Pride (1-5)	3.66	.65	3.60	.80	3.56	.83	3.52	.76
Anger (1-5)	2.40	.92	2.48	1.06	2.48	1.02	2.56	1.18
Anxiety (1-5)	1.99	.56	1.96	.64	1.96	.58	1.92	.64
Boredom (1-5)	2.90	.96	2.89	.98	2.95	.95	2.93	1.04
Shame (1-5)	1.71	.55	1.76	.65	1.76	.71	1.95	.72

Note. n = sample size per condition per outcome; M = mean; SD = standard deviation.

Table 4.2

Robust regression results for main effects and moderation

Variables	β	*SE*	*t*	*p*	R^2	*Adjusted* R^2
Outcome: Mastery-approach goals					.44	.43
(Intercept)	.86	.24	3.54	<.001		
Teacher	-.01	.01	-.93	.353		
Pretest performance	.03	.01	1.99	.047		
Prior mastery-approach goals	.69	.06	11.82	<.001		
Prior performance-approach goals	.03	.04	.72	.474		
Mastery-approach goal instructions	.06	.05	1.25	.210		
Performance-approach goal instructions	.02	.05	.32	.747		
Outcome: Performance-approach goals					.62	.61
(Intercept)	-.17	.22	-.76	.449		
Teacher	.00	.02	.00	.997		
Pretest performance	.00	.02	-.16	.870		
Prior mastery-approach goals	.19	.06	3.25	.001		
Prior performance-approach goals	.80	.04	18.47	<.001		
Mastery-approach goal instructions	-.06	.06	-1.04	.300		
Performance-approach goal instructions	.12	.06	1.97	.049		
Outcome: Posttest performance					.23	.22
(Intercept)	3.18	.92	3.44	<.001		
Teacher	-.17	.07	-2.39	.017		
Pretest performance	.51	.06	8.09	<.001		
Prior mastery-approach goals	.32	.24	1.35	.178		
Prior performance-approach goals	.12	.17	.70	.487		
Performance-approach goal instructions	-.49	.24	-2.04	.041		
Outcome: Mental effort					.05	.04
(Intercept)	2.88	.71	4.04	<.001		
Teacher	.00	.04	.01	.991		
Pretest performance	-.05	.04	-1.47	.143		
Prior performance-approach goals	-.32	.09	-3.46	<.001		
Prior mastery-approach goals	.55	.18	3.09	.002		
Performance-approach goal instructions	2.08	1.01	2.05	.041		
Prior mastery-approach goals*Performance-approach goal instructions	-.54	.27	-2.00	.046		
Outcome: Enjoyment					.22	.21
(Intercept)	.60	.31	1.90	.058		
Teacher	-.01	.02	-.33	.743		
Pretest performance	.05	.02	2.36	.019		
Prior mastery-approach goals	.55	.07	7.89	<.001		
Prior performance-approach goals	.13	.05	2.72	.007		
Performance-approach goal instructions	-.11	.08	-1.28	.202		
Outcome: Pride					.11	.10
(Intercept)	2.30	.26	8.80	<.001		
Teacher	.01	.01	.65	.518		
Pretest performance	-.01	.01	-.71	.480		
Prior mastery-approach goals	.30	.07	4.09	<.001		
Prior performance-approach goals	.09	.05	1.83	.068		
Performance-approach goal instructions	.00	.06	-.02	.981		

Variables	β	*SE*	*t*	*p*	R^2	*Adjusted* R^2
Outcome: Anger					.05	.04
(Intercept)	3.77	.38	9.81	<.001		
Teacher	.01	.02	.45	.657		
Pretest performance	-.02	.02	-.93	.353		
Prior mastery-approach goals	-.37	.10	-3.81	<.001		
Prior performance-approach goals	.01	.07	.21	.837		
Performance-approach goal instructions	.02	.10	.22	.830		
Outcome: Boredom					.11	.11
(Intercept)	4.99	.35	14.27	<.001		
Teacher	.00	.02	-.18	.854		
Pretest performance	-.04	.02	-1.55	.123		
Prior mastery-approach goals	-.52	.08	-6.59	<.001		
Prior performance-approach goals	.01	.06	.19	.846		
Performance-approach goal instructions	.04	.09	.49	.627		
Outcome: Shame					.02	.01
(Intercept)	2.08	.26	8.09	<.001		
Teacher	-.04	.02	-2.67	.008		
Pretest performance	-.01	.02	-.89	.376		
Prior mastery-approach goals	-.06	.06	-.89	.375		
Prior performance-approach goals	.02	.04	.56	.578		
Performance-approach goal instructions	.02	.06	.30	.768		

Note. $N = 450$; β = regression coefficient; *SE* = standard error; t = test whether β is significantly differ from zero; R^2 = the proportion of the variance of the outcome that is explained by the model; Only statistically significant independent variables (mastery-approach goal instruction and performance-approach goal instruction), covariates (pretest performance, prior mastery-approach goals, or prior performance-approach goals) and moderators (prior mastery-approach goals or prior performance-approach goals) are kept in the models; Because the interaction effects between mastery-approach goal instruction and performance-approach goal instruction is not statistically significant, all models do not include the interaction effects; Cohen's $f^2 = R^2/(1-R^2)$, ≥.02, ≥.15, and ≥.35 means small, medium, and large effect size; Dummy coding: yes = 1, no = 0.

Discussion

Research on achievement goal instructions and their effects on secondary school students' achievement goals, mental effort, performance, and achievement emotions is sparse. To the best of our knowledge, this study is one of the first to present an integrated view on achievement goals, performance, and achievement emotions, and it appears to be the first to test how achievement goal instructions affect these three learning outcomes moderated by prior achievement goals.

RQ1: The Effect of Achievement Goal Instructions on Learning

We found that mastery-approach goal instructions and performance-approach goal instructions did not interact, mastery-approach goal instructions did not induce mastery-approach goals, while performance-approach goal instructions induced performance-approach goals, increased mental effort, and decreased posttest performance.

Achievement Goals

Consistent with our hypothesis, performance-approach goal instructions promoted higher performance-approach goals than no performance-approach goal instructions. Contrary to our hypotheses and Linnenbrink (2005) but consistent with Muis et al. (2013), mastery-approach goal instructions did not promote higher mastery-approach goals than no mastery-approach goal instructions. These results seem to indicate that performance-approach goal instructions work for secondary school students (i.e., young adolescents) in chemistry GBL, but mastery-approach goal instructions not. These results are consistent with previous studies on achievement goal interventions: Mastery goal interventions appear to be effective for primary school students (i.e., children) but not for undergraduates (i.e., late adolescents), whereas performance goal interventions appear to be effective for undergraduates but not for primary school students (e.g., Linnenbrink, 2005; Muis et al., 2013).

One explanation for these results is that there are age-related differences in the adoption of achievement goals. For example, children at primary school are more likely to adopt mastery-approach goals, whereas adolescents are more likely to adopt performance-approach goals (Bong, 2009). When evaluating competence, children at primary school are more likely to concentrate on the tasks and improving skills, whereas adolescents and young adults are more likely to concentrate on the performance of others (Dweck & Leggett, 1988). This age-related difference may be a result of the developmental stage of students (Bong, 2009) or of learning environments that emphasize social comparison in adolescents and young adults (Urdan & Midgley, 2003).

Mental Effort and Performance

Consistent with our hypothesis, performance-approach goal instructions promoted higher mental effort than no performance-approach goal instructions. Contrary to our hypothesis but consistent with previous studies (e.g., Erhel & Jamet, 2016; Muis et al., 2013), performance-approach goal instructions promoted lower posttest performance than no goal instructions. These results imply that performance-approach goal instructions may hamper learning.

These results can be interpreted by considering the source of mental effort. Participants in the performance-approach goal instructions group were required to finish as fast as possible, so, their attention may have been diverted from the game activities that are directly related to learning to interfering thoughts (Sarason et al., 1986), such as winning, or to coping with the imposed time pressure. These are both processes irrelevant for learning. According to CLT (Sweller et al., 2019), they might have experienced higher extraneous load, that is, they exerted higher mental effort, but got lower posttest performance.

Achievement Emotions

Contrary to our hypothesis, performance-approach goal instructions did not promote higher positive outcome emotions and lower negative outcome emotions than no performance-approach goal instructions. One possibility is that the achievement goal instructions may indeed affect specific achievement emotions, but the measurement method we used may not be sensitive enough to detect this effect. Research on the temporal dynamics of emotional experiences during complex learning suggested that based on duration, emotions can be classified as persistent states that last for a while (e.g., boredom, engagement/flow, and confusion), transitory states that dissipate almost immediately (e.g., delight and surprise), and intermediate states (e.g., frustration; D'Mello & Graesser, 2011). If this categorization also applies to achievement emotions, then the timing of measuring emotions may need to change accordingly. For example, transitory achievement emotions should be measured in real time rather than delayed (e.g., after gameplay in this study). Future studies could use real-time measures (Pekrun & Linnenbrink-Garcia, 2014), such as experience sampling method, to capture emotional trajectories during tasks.

RQ2: Moderation Effect of Prior Achievement Goals

Contrary to our hypothesis, prior performance-approach goals did not moderate the effects of performance-approach goal instructions. Consistent with our hypothesis, prior mastery-approach goals moderated the effects of performance-approach goal instructions on mental effort. For the performance-approach goal instructions group, the higher the prior mastery-approach goals, the lower the mental effort. This result implies that performance-approach goal instructions have weaker positive influences on learning when learners hold higher prior mastery-approach goals than lower prior mastery-approach goals. This is also in line with the vibration effect from the mismatch hypothesis (i.e., a beneficial effect is weakened, if achievement goal instructions mismatch prior achievement goals; Murayama & Elliot, 2009). Overall, our results do not support the match hypothesis and partially support the mismatch hypothesis (vibration effect).

Combining with previous research, the evidence for the match and mismatch hypotheses seems inconclusive. Similar to our results, most of the supportive evidence is limited to certain learning processes and outcomes, such as effort (Wolters, 2004), metacognitive self-regulation (Muis et al., 2013), or intrinsic motivation (Murayama & Elliot, 2009) but not others, such as test result (Linnenbrink, 2005), course grade (Muis et al., 2013; Wolters, 2004), or test anxiety (Wolters, 2004). This implies that prior achievement goals may moderate the effects of achievement goal interventions on some but not all learning processes and outcomes, which needs more research in the future.

Limitations

Some limitations need attention to evaluate our results. First, we focused only on achievement goals. Considering that learners enter a learning environment with many goals, such as getting a high grade in chemistry, future research should assess whether goal instructions have a similar effect on goals other than achievement goals.

Second, we used a specific GBL environment with a specific game genre, that is, CosmiClean is a combination of a computer game, puzzle game, and strategy game. The game has no in-game scores or badges. The indicator of whether participants do better than others (performance-approach goals) is whether they finish faster than others at the end of gameplay. We might get different findings if performance-approach goal instructions were investigated through other indicators in the game. For example, if there was a leaderboard after each level, then students might know more frequently if they were doing better than others, and thus the effects of performance-approach goal instructions may be even stronger than we found in present study. We call for similar research on other games with other performance related indicators to confirm our findings.

Third, this study may be influenced by *pretest effects*, that is, a pretest might cause learning before the learning materials are presented (i.e., a testing effect; Mayer, 2021) and may direct participants' attention to the relevant topics that researchers are interested in and alert them to the posttest and interventions (i.e., a priming effect; Kim & Willson, 2010). According to the meta-analysis on pretest effects in experimental studies, the effect size of a pretest on a posttest is small if there is less one day between pretest and posttest (Willson & Putnam, 1982). To minimize the impact of a potential pretest effect, we added pretest performance as a covariate. To prevent pretest effects, future research may use Solomon four-group design: a) two groups (control 1 and treatment 1) receive pretest and posttest and another two groups (control 2 and treatment 2) receive only posttest and b) the effects of the pretest and intervention can be evaluated by comparing control groups and comparing treatment groups, respectively (Navarro & Siegel, 2018).

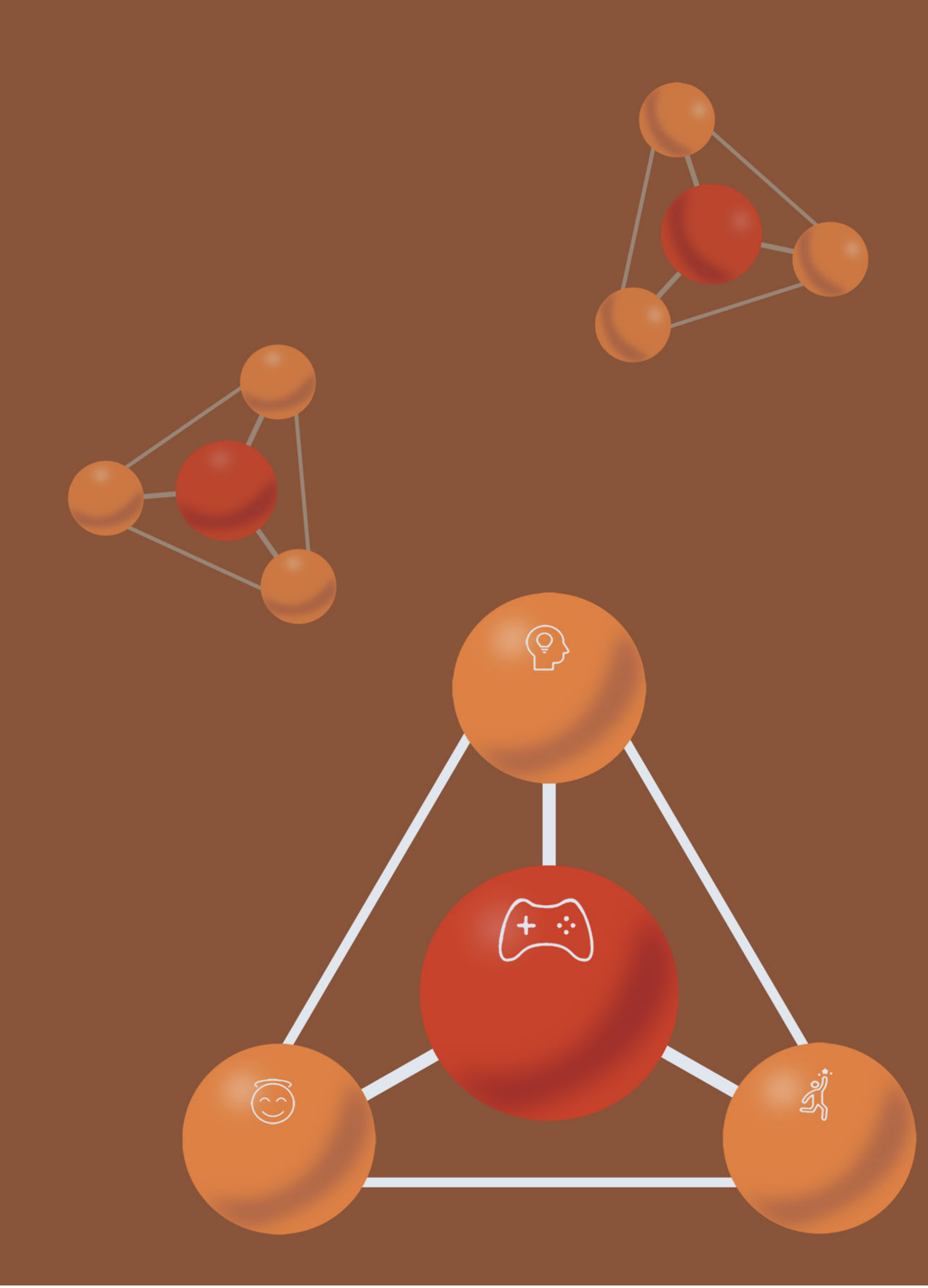
Implications

Practically, educators aim to implement instructions that foster learners' cognitive, motivational, and positive emotional processes and outcomes. First, considering that achievement goals are frequently used in learning environments, we suggest that educators can use achievement goal instructions to induce some achievement goals (e.g., performance-approach goals) but not others (e.g., mastery-approach goals) in chemistry GBL. Second, our findings suggest that educators may change achievement goal instructions for different learning goals. If the learning goal is to promote posttest performance in chemistry GBL, we suggest that educators should not use performance-approach goal instructions. Third, when using achievement goal instructions, educators would do well to take students' prior achievement goals into account.

Theoretically, instructional design features that manipulate motivation, such as performance-approach goal instructions, can affect not only motivation, such as performance-approach goals, but also cognition, such as mental effort and posttest performance. Currently, we do not fully understand how motivation, cognition, and emotion interact in learning environments (Dietrich et al., 2022; Mayer, 2021; Pekrun & Linnenbrink-Garcia, 2014; Sweller et al., 2019). Future research on the interactions between motivation, emotion and cognition will help us understand and change learning processes and outcomes.

Conclusion

We conclude that performance-approach goal instructions affect cognitive and motivational processes and outcomes differently in GBL. Consequently, educators may design the achievement goal instructions based on learner's prior achievement goals and the learning goals, such as promoting posttest performance. This study is one of the few to attend to achievement goal instructions in GBL. In addition, this study is one of the first to manipulate multiple goals and to focus on motivational, cognitive, and emotional processes and outcomes. This study is the first experimental study that supports the vibration effect of the mismatch hypotheses. The results of this study may advance the field of achievement goal theory and instructional design by our findings that achievement goal instructions partially affect learners' achievement goals, mental effort, and performance, and that prior mastery-approach goals moderate the effects of performance-approach goal instructions on learning processes and outcomes (mental effort).



Chapter 5 Effects of peers' achievement emotions on students' achievement emotions, achievement goals, mental effort, and performance in game-based learning

This chapter is based on:

Hu, Y., Wouters, P., van der Schaaf, M., Elliot, A. J., Pekrun, R., & Kester, L. (2023). Effects on peers' achievement emotions on students' achievement emotions, achievement goals, and performance in game-based learning. *Manuscript submitted for publication*.

Acknowledgement of author contributions

All authors designed the study. Yuanyuan Hu recruited participants, collected, and analyzed the data, and drafted the manuscript. All authors contributed to critical revision of the manuscript. Pieter Wouters, Marieke van der Schaaf, and Liesbeth Kester supervised the study.

Acknowledgement

We would like to thank Vincent Hoogerheide (from Utrecht University), Jeremy Jamieson (from University of Rochester), and Alyssa P. Lawson (from University of California, Santa Barbara) for sharing their expertise on emotions and Rebecca Kuiper (from Utrecht University) for sharing her expertise on statistical analysis.

Abstract

Emotional contagion often occurs, but limited research has been done on it in education. Given the frequent interactions among teachers and students, not only teachers' emotions but also peers' emotions may influence students' learning. This preregistered study investigated how peers' achievement emotions (operationalized as either enjoyment, neutral state, or frustration) affect students' emotion, motivation, and cognition in game-based learning. University students (N = 210) played a game that included watching a video in which a peer model played the game and displayed either enjoyment, neutral state, or frustration. We included 131 participants who correctly identified the peers' emotions being displayed. The data were analyzed by random intercept cross-lagged panel models (RI-CLPMs) with Bayesian estimation and generalized order-restricted information criterion approximation (GORICA). Students exposed to peers' enjoyment reported higher enjoyment, relaxation, mastery-based goals, and game performance and lower frustration, anger, boredom, mental effort, and posttest performance than those exposed to peers' frustration. We conclude that peers' achievement emotions affect students' achievement emotions, mastery-based goals, mental effort, and performance differently. Educators and researchers should consider emotional contagion among students, the role of social contagion in education, and the interactions between emotion, motivation, and cognition.

Keywords: Emotional contagion; Achievement emotions; Achievement goals; Performance; Game-based learning

Introduction

Emotions can play a key role in game-based learning (GBL; Plass et al., 2020). For example, GBL often involves collaboration in which group players express different emotions. Emotions can either help or hurt learning (Pekrun, 2006). For example, enjoyment increases learning, whereas boredom decreases learning. So, it is critical to design learning environments in which emotions help learning.

According to social contagion theory, emotions can spread among people (i.e., *emotional contagion*; Hatfield et al., 2014). Imagine that a class is discussing a game used to help learning when one student says, '*I really enjoy the game. Can I play more*?'. Another student says, '*I am so frustrated. Can I stop playing?*'. A third student says, '*The game is so-so, nothing in particular.*'. How would other students react? Understanding emotional contagion may help us to answer this question and to regulate emotions in GBL, such as promoting (avoiding) the contagion of positive (negative) emotions. However, there is limited research on emotional contagion in the field of education (Burgess et al. 2018). For example, research has shown that teachers can transmit enjoyment to their students (Frenzel et al., 2018), but to date, we are unaware of any experimental studies that have tested emotional contagion from student to student.

In research on emotional design principles in multimedia learning (e.g., GBL), most studies have focused on emotion design features of learning materials that can carry emotions (e.g., game characters) or emotion induction before or during learning (e.g., listening to happy or sad music) and how they affect students' emotions and/or learning (Loderer et al., 2020; Plass & Hovey, 2021). The emotions of teachers and students and their effects on students' learning are also an important area of research. For example, teachers' happiness, content, boredom, or frustration can affect students' emotion and motivation (Lawson & Mayer, 2022). Thus far, experimental studies on how teachers' emotions affect students' learning are limited, and studies testing the impact of peers' emotions on students' learning are lacking. In the present study, we test how peers' achievement emotions affect students' emotional, motivational, and cognitive processes and outcomes in science GBL. In the classroom, peers are the students' classmates. Since this study is the first of this kind, we use a virtual peer that the participants do not know as a starting point.

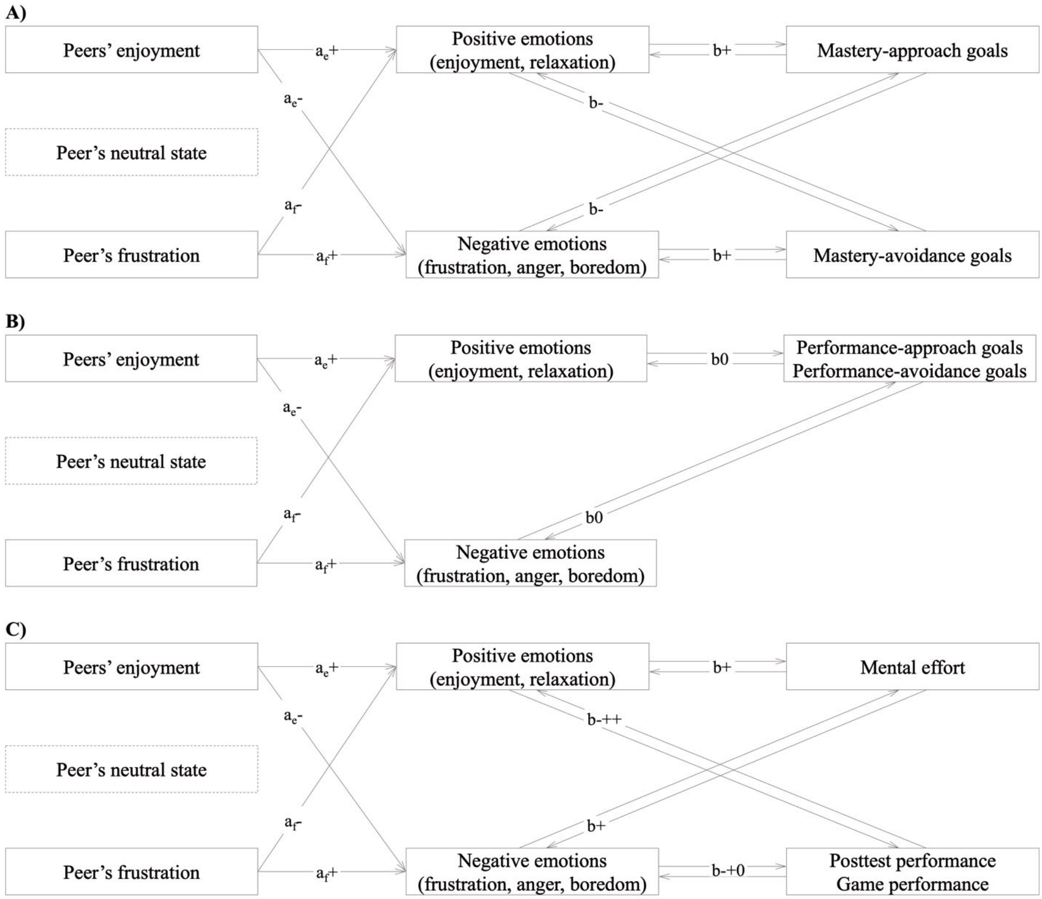
Furthermore, in most previous studies using intensive longitudinal data (i.e., two or more variables are repeatedly measured at multiple time points) to investigate effects of instructional design features, between-person relations were typically of interest (e.g., Hu et al., 2022a, 2022b). Researchers often used the classic cross-lagged panel model (CLPM) to estimate reciprocal effects over time. However, the CLPM has been criticized because it does not separate between- and within-person relations, and the results from between- and within-person relations can diverge substantially (Hamaker et al., 2015). In the present analysis, we use the random-intercept cross-lagged panel model (RI-CLPM; Mulder & Hamaker, 2021) to de-confound between- and within-person relations.

The Effect of Peers' Achievement Emotions and Students' Learning

This study is based on achievement-relevant theories, namely, the control-value theory of achievement emotions (CVT; Pekrun & Perry, 2014), achievement goal theory (AGT; Elliot & Hulleman, 2017), and cognitive load theory (CLT, Sweller et al., 2019). Figure 5.1 illustrates the relations among the focal achievement-relevant variables.

Figure 5.1

Overview of the predicted relations between peers' achievement emotions and students' achievement emotions, achievement goals, mental effort, and performance



Note. The control group (peers' neutral state) was the reference group; e and f represents peers' enjoyment group and frustration group, relative to the control group; a_e (a_f) represents the mean difference in a variable between the enjoyment (frustration) group and the control group; b represents the effect of a mediator on an outcome variable; Arrows represent direct effect; + represents positive relations; - represents negative relations; 0 represents null relations.

Peers' Achievement Emotions and Students' Emotions

According to CVT, *achievement emotions* are emotions related to competence-relevant activities (e.g., attending class) and/or outcomes (i.e., success or failure) in achievement settings (Pekrun, 2006). Depending on the object of emotions, achievement emotions can be distinguished as *activity emotions* related to achievement-relevant activities or tasks (e.g., enjoyment, relaxation, frustration, boredom, and anger) and *outcome emotions* related to the outcomes of these activities (e.g., pride, hope, relief, anxiety, shame, anger, and hopelessness). Furthermore, depending on the valence (positive/negative or pleasant/unpleasant) and activation (physiologically activating/deactivating) of emotions, achievement emotions can be distinguished as *positive activating emotions* (e.g., enjoyment, pride), *positive deactivating emotions* (e.g., relaxation, relief), *negative activating emotions* (e.g., frustration, anxiety), and *negative deactivating emotions* (e.g., boredom, hopelessness; Pekrun & Perry, 2014).

Because this is the first study of this kind and GBL often involves achievement-relevant activities or tasks, we focus on only activity emotions. Specifically, peers' enjoyment and frustration are selected as the representatives of peers' positive and negative emotions because they are typical emotions in GBL (Loderer et al., 2020) and differ only in the valence of the emotions. We use *neutral state* ("feeling nothing in particular and no preference of one over the other") as a reference point for positive and negative emotions (Gasper et al., 2019).

The phenomenon of people catching the emotions of others is often called emotional contagion (Hatfield et al., 1994), emotional transmission (Frenzel et al., 2018), emotional transfer (Parkinson, 2011), emotional crossover (Westman et al., 2013), or emotion diffusion (Peters & Kashima, 2015). We use the most used term, emotional contagion, in this study. Three possible mechanisms have been proposed to explain emotional contagion. One mechanism is primitive emotional contagion (Hatfield et al., 2014): The observer may mimic the (nonverbal) emotional (facial, vocal, and/or postural) expressions of the expresser (i.e., emotional mimicry) and this mimicry may trigger the same emotional states in the observer (i.e., afferent feedback). The second mechanism is emotion categorization (Peters & Kashima, 2015): The observer categorizes the (verbal and nonverbal) emotional expressions of the expresser as an emotional state (i.e., categorization) and this categorization activates the same emotional state in the observer (i.e., activation). The third mechanism is social appraisal (Bruder et al., 2014): The observer may interpret the situation based on information inferred from the (verbal and nonverbal) emotional expressions of the expresser (i.e., appraisal) and this appraisal may trigger the same emotional state in the observer.

For example, when someone else shows a smiley face, one may mimic the smiley face in return, may categorize the smiley face as a feeling of enjoyment, and/or may evaluate the smiley face as indicating that someone else being happy with the situation, all of which may trigger his/her own enjoyment. This implies that these three mechanisms – primitive emotional contagion, emotion categorization, and social appraisal – may not be mutually exclusive and will similarly affect the outcomes, their agreement on the existence of emotional contagion supports our expectation that peers' achievement emotions affect students' achievement emotions, that is, emotion contagion occurs from peers to students (Figure 5.1 arrows a).

Students' Achievement Emotions and Students' Motivation: Reciprocal Effects

CVT (Pekrun, 2006) proposes that achievement emotions affect motivational processes and performance outcomes. Motivational processes and performance, in turn, impact achievement emotions as they influence the appraisals of control and value related to achievement. Thus, emotions, motivation, and performance are expected to be linked by reciprocal effects over time (Linnenbrink & Pintrich, 2002).

According to AGT, the 2 × 2 achievement goal model (Elliot & Hulleman, 2017) distinguishes between four achievement goals: *mastery-approach goals* (striving for task- or self-based competence, such as learning as much as possible), *mastery-avoidance goals* (striving to avoid task-based or self-based incompetence, such as avoiding learning less than one possibly could), *performance-approach goals* (striving for other-based competence, such as performing better than others), and *performance-avoidance goals* (striving to avoid other-based incompetence, such as avoiding performing worse than others). This study focuses on all four of these achievement goals as the motivational processes and outcomes.

Regarding the relationship between students' achievement emotions and achievement goals, mastery-based goals may focus students' attention on the activity itself, thereby influencing activity emotions and performance-based goals may focus students' attention on outcomes, thereby influencing outcome emotions (Pekrun et al., 2009). Thus, performance-based goals may be more associated with outcome emotions than activity emotions. Since this study focused only on activity emotions, we expect that peers' achievement emotions affect mastery-based goals but not performance-based goals (Figure 5.1A arrows a).

Regarding the relations between achievement emotions and mastery-based goals, mastery-approach goals may focus attention on the positive value of the activity and help feeling in control, thereby positively influencing positive activity emotions and negatively influencing negative activity emotions (Pekrun et al., 2009). Although Pekrun et al. (2009) did not establish a link between mastery-avoidance goals and achievement emotions, we propose that, similar to performance-avoidance goals, the avoidance component makes mastery-avoidance goals focus attention on the negative value of the activity and promote feeling a lack of control and therefore mastery-avoidance goals negatively influence positive activity emotions and positively influence negative activity emotions. Although there are no meta-analyses on the relationship between induced achievement emotions and achievement goals, a meta-analysis on personal achievement goals has suggested that mastery-approach goals are more strongly associated with positive emotions than negative emotions, whereas mastery-avoidance goals are more strongly associated with negative emotions than positive emotions (Huang, 2011). Therefore, we expect positive relations between students' mastery-approach goals and their positive achievement emotions and positive relations between students' mastery-avoidance goals and their negative achievement emotions (Figure 5.1B arrows b).

Students' Achievement Emotions and Students' Cognition: Reciprocal Effects

CVT (Pekrun, 2006) proposes that achievement emotions affect cognitive processes and performance outcomes. Cognitive processes and performance, in turn, impact achievement emotions as they influence the appraisals of control and value related to achievement. Thus, emotions, cognitive processes, and performance are expected to be linked by reciprocal effects over time (Pekrun et al., 2023).

According to CLT, overall cognitive load involves two types of cognitive load: *intrinsic load* (load caused by cognitive processes or activities related to learning and performing the task) and *extraneous load* (load caused by cognitive processes or activities that are unnecessary for learning and performing the task; Sweller et al., 2019). Overall cognitive load can be estimated by the *mental effort* that students exert on a task (Paas, 1992). This study focuses on mental effort and performance as the cognitive processes and outcomes.

Regarding the relationships between students' achievement emotions, mental effort, and performance, our expectations were guided by three possible assumptions. The first assumption is related to extraneous load: Compared with neutral state, *the emotions-as-suppressor-of-learning* *hypothesis* suggests that, based on CLT, positive and negative emotions may impose extraneous load (e.g., via task-irrelevant thinking such as thinking about the consequences of failure; Meinhardt & Pekrun, 2003), thereby decreasing performance. The second assumption is based on motivation: *The emotions-as-facilitator-of-learning hypothesis* suggests that positive emotions may increase intrinsic motivation and negative emotions may increase effort in learning to improve their emotions and extrinsic motivation, both of which increase performance, compared with neutral state (Plass & Kalyuga, 2019).

The third assumption combines attention and motivation: Based on CVT, positive and negative emotions may have different effects on performance (Pekrun & Linnenbrink-Garcia, 2012), as shown in Table 5.1. As noted above, the present study focuses on activity emotions. Positive activating emotions (e.g., enjoyment of learning) increase both task attention and motivation to invest effort, thereby facilitating performance. Negative deactivating emotions (e.g., boredom) decrease both task attention and motivation to invest effort, thereby impairing performance. Furthermore, positive deactivating emotions and negative activating emotions have variable effects on performance. Positive deactivating emotions (e.g., relaxation) broaden the focus - decrease attention on details and increase attention on the broad picture - but also decrease motivation to invest effort. Negative activating emotions (e.g., frustration, anger) decrease task attention and intrinsic motivation but also increase extrinsic motivation to invest effort to avoid failure. A previous meta-analysis on activity emotions has suggested a positive relation between performance and enjoyment, a negative relation for anger and boredom, and a near null relation for frustration (Camacho-Morles et al., 2021). Therefore, we expect, based on extraneous load assumption, that students' achievement emotions positively relate to students' mental effort, and based on all three of the assumptions, that students' achievement emotions relate to performance differently (Figure 5.1B arrows b).

Table 5.1

Effects of activity emotions in short-term tasks (Pekrun & Linnenbrink-Garcia, 2012)

	Example	Task attention and cognitive resources	Motivation to invest effort	Performance
Positive activating	Enjoyment	Increase	Increase	Increase
Positive deactivating	Relaxation	Variable	Decrease Variable	Variable
Negative activating	Frustration Anger	Decrease	Variable	Variable
Negative deactivating	Boredom	Decrease	Decrease	Decrease

Joint Influence of Peers' Achievement Emotions and Students' Emotion, Motivation, and Cognition: Reciprocal Effects

Given that emotions and achievement goals reciprocally influence each other (e.g., Linnenbrink & Pintrich, 2002; see Figure 5.1B arrows b), instructional design features, such as peers' achievement emotions, may affect students' achievement goals (Figure 5.1A arrows a) via students' achievement emotions (indirect effects; Figure 5.1B coupling of arrows a – b). Specifically, we anticipate that peers' enjoyment positively influences students' positive emotions and negatively influences students' negative emotions. In contrast, we anticipate that peers' frustration negatively influences students' positive emotions and positively influences students' negative emotions. For both peers' enjoyment and peers' frustration, we anticipate that students' achievement emotions and students' mastery-approach goals or mastery-avoidance goals are linked by reciprocal effects over time, compared with peers' neutral state.

Similarly, given that emotion and cognition reciprocally influence each other over time (Pekrun et al., 2023; see Figure 5.1B double-headed arrows b), instructional design features, such as peers' achievement emotions, may affect students' mental effort and performance (Figure 1A arrows a) via students' achievement emotions (indirect effects; Figure 5.1B coupling of arrows a – b). Specifically, we anticipate that peers' enjoyment positively influences students' positive emotions and negatively influences students' negative emotions, whereas peers' frustration negatively influences students' positive emotions and positively influences students' negative emotions. In both cases, we anticipate that students' achievement emotions and students' mental effort or performance are linked by reciprocal effects over time, compared with peers' neutral state.

Present Study

This study investigates how peers' achievement emotions (enjoyment/frustration/neutral state) affect students' achievement emotions (i.e., positive emotions and negative emotions), achievement goals (i.e., mastery-based and performance-based goals), and cognition (i.e., mental effort and performance). The hypotheses below are stated broadly in terms of positive/negative emotions during GBL (without considering the different types of emotions within these categories). We formulate three different hypotheses for performance based on three different assumptions.

RQ (Main effects): Students exposed to peers' enjoyment report higher positive achievement emotions (**H1emo1**; Figure 5.1A: $a_e > 0 > a_f$) and lower negative achievement emotions (**H1emo2**; Figure 5.1A: $a_e < 0 < a_f$) than those exposed to peers' neutral state, followed by those exposed to peers' frustration.

Students exposed to peers' enjoyment report higher mastery-approach goals (**H1mot1**; $a_e > 0 > a_f$) and lower mastery-avoidance goals (**H1mot2**; $a_e < 0 < a_f$) than those exposed to peers' neutral state, followed by those exposed to peers' frustration, and students from these three groups report equal performance-approach goals and performance-avoidance goals (**H1mot3**; $a_e = 0 = a_f$).

Students exposed to peers' enjoyment and those exposed to peers' frustration report higher mental effort than those exposed to peers' neutral state (**H1cog1**; $a_e = a_f > 0$).

Students exposed to peers' enjoyment and those exposed to peers' frustration report lower game and posttest performance than those exposed to peers' neutral state (**H1cog2**; $a_e = a_f < 0$); or students exposed to peers' enjoyment and those exposed to peers' frustration report higher game and posttest performance than those exposed to peers' neutral state ($a_e = a_f > 0$); or students exposed to peers' enjoyment report higher game and posttest performance than those exposed to peers' frustration and those exposed to peers' neutral state ($a_e > 0 = a_f$).

Method

Participants

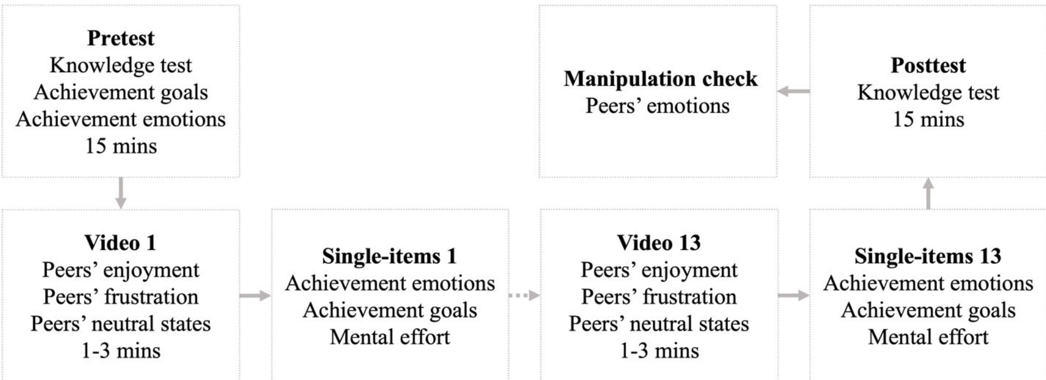
Participants were recruited from Prolific (https://www.prolific.co/) and received 10 euros as compensation. The inclusion criteria were English as first language, being an undergraduate, and majoring in something other than chemical engineering/chemistry (because we used a game about this subject). As suggested by Kline (2015), structural equation modelling needs at least 10 participants per indicator. We had 12 indicators, so our minimum sample size target was 120. Based on our manipulation check, we excluded 79 participants who did not correctly identify the peers' emotions being displayed in the videos. In total, 131 participants were included (68 male, 63 female, $M = 20.5$ years old, $SD = 1.3$). This study's design and data analysis were preregistered (https://osf.io/52dr3), where materials and analysis code for this study are available.

Design

We randomly assigned all participants to one of three groups: 1) peers' enjoyment ($n = 62$), 2) peers' frustration ($n = 33$), and 3) peers' neutral state ($n = 36$). This unequal sample distribution was due to the higher number of excluded participants in the peers' frustration and neutral state groups (see Figure 5.2).

Figure 5.2

The procedure and measures



Materials and Measures

All the materials and measures were in English and presented in Qualtrics (www.qualtrics.com).

The Game – CosmiClean

LuGus Studios (https://www.lugus-studios.be/) designed CosmiClean (https://recyclegame.eu/) to teach secondary school and university students the principles of separation processes for recycling materials with nine separators (e.g., a melter). The aim of the game is to make a sequence of separators and collect recycled materials.

Videos

The expressers' emotions can be displayed as dynamic expressions (e.g., films) or static expressions (e.g., texts; Herrando & Constantinides, 2021). Dynamic expressions may be more contagious than static expressions (Sato et al., 2008). Meta-analyses also confirm that film/video clips are an effective method to induce emotions (Fernández-Aguilar et al., 2019; Lench et al., 2011; Joseph et al., 2020) and thus are suggested as standardized stimuli for facilitating emotional contagion (Hatfield et al., 2014).

The intervention materials consisted of three recorded videos with 13 game levels. In the corresponding level of each video, a peer model played the game with the same script and different target emotions, that is, enjoyment (link: here), frustration (link: here), or neutral state (link: here). First, the peer model showed the materials to be recycled and the properties of the materials. Then the multiple-answer questions asked participants to choose one, two, or three separators to recycle the materials. These questions were to mimic the real gameplay scenarios: the player needs to select certain separators to succeed in a game level. After they answered, they got corrective feedback (i.e., whether their answer was correct or not). After they received the feedback, the peer model explained the correct separators and ran the recycling. We uploaded the videos in Edpuzzle (www.edpuzzle.com). Participants could pause and rewatch but not skip the videos. Each video lasted approximately 23 mins. The peer model was a male actor in training. He varied his facial, postural, and vocal expressions according to the target emotion. We filmed the videos via first-person perspective (Fiorella et al., 2017). The instructions for watching the videos were: '*You will play the game by watching a video, in which another student will play the game together with you. Please watch the video carefully'*.

Achievement Emotions Questionnaire

Because our experiment lasted about 1 hour and achievement emotions and achievement goals might decay over time, we used an experience sampling method (e.g., Goetz et al., 2016) with 13 timepoints per person rather than a one-time measure method in order to capture their dynamics. Due to time constrains, we used single-item measures, which have been shown to be as reliable as multiple-item measures (Gogol et al., 2014). The achievement emotions questionnaire was based on Goetz et al. (2016): '*At this moment I am experiencing enjoyment/relaxation/frustration/anger/boredom*' (1 = strongly disagree; 5 = strongly agree). The instructions asked participants to describe their emotions and goals at that moment.

Achievement Goals Questionnaire

The achievement goals questionnaire was adopted from the Achievement Goals Questionnaire-Revised (Elliot & Murayama, 2008; 1 = strongly disagree, 5 = strongly agree). The items were framed as goals for the game: "*At this moment my goal is to learn as much as possible in the game*" (mastery-approach), "*At this moment my goal is to avoid learning less than I possibly could in the game*" (mastery-avoidance); "*At this moment my goal is to perform better than the other participants*" (performance-approach), and "*At this moment my goal is to avoid performing poorly compared to other participants*" (performance-avoidance).

Mental Effort

Paas's (1992) scale was used to measure mental effort: "*How much mental effort did this game level require from you*" (1 = very, very low mental effort; 9 = very, very high mental effort).

Game Performance

Game performance was measured by the total number of correct separators that were chosen in the multiple-answer questions in the videos (one question per video).

Knowledge Test

The prior and post knowledge test assessed the same chemistry content but with different items. The knowledge tests assessed Remember (5 multiple-choice questions), Apply (5 multiple-choice questions), and Evaluate (3 open-ended questions) based on the Bloom taxonomy (Anderson & Krathwohl, 2001). For example, a Remember question was "*Which processor can separate glass and iron?";* an Apply question was "*Glass is not metal. Iron, copper, and gold are metal. Which materials can be separated by a non-ferrous separator?*"; and an Evaluate question was "*To separate fuel and copper, your teacher will select either a non-ferrous separator or a dissolver. Explain which one is more proper.*". The knowledge tests were developed and validated in previous studies (Hu et al., 2022a, 2022b), and were reliable (i.e., prior knowledge test: greatest lower bound = .66; post knowledge test: greatest lower bound = .78). Knowledge tests often do not measure the same underlying concept (here, nine separators instead of one) and a reliability value lower than .70 is normal (Taber, 2018).

Manipulation Check

We included a manipulation check on whether participants correctly identified the peers' emotions being displayed. To avoid interrupting or disturbing the automatic process of emotional contagion, after the posttest (rather than after watching the videos) participants chose one of three alternatives to indicate the emotion being displayed: "*The student in the video: a) enjoyed the game, b) was frustrated by the game, c) neither enjoyed nor was frustrated by the game*". We included only the participants who correctly identified the emotions being displayed.

Procedure

Pilot Study

We ran a small pilot test with 11 participants to check on the fluidity and effectiveness of the procedure. This test (descriptively) revealed that participants could correctly identify the peers' emotions, that there were learning gains from the prior to the postgame knowledge test, that the game was neither too complex nor too easy, and that there were no comprehension or technical problems.

Main Study

After providing informed consent, participants received instruction about the number of sections and their duration. Then they completed the prior knowledge test and reported their achievement goals and achievement emotions (pretest; see Figure 5.2). They were then randomly assigned to one of the three experimental conditions, and they watched the corresponding videos for 30 minutes. After each game level, they reported their achievement emotions, achievement goals, and mental effort. Immediately after watching the complete set of videos, they completed the postgame knowledge test (posttest), and the manipulation check.

Scoring, Data Preparation, and Data Analysis

For the knowledge test, we calculated a sum of correct scores of 10 multiple-choice questions (1 point per question) and three open-ended questions (2 or 4 points per question), resulting in a maximum score of 20 points. For the three open-ended questions, we developed a coding schema. Two raters scored 10% of the answers for each question independently with extremely high inter-rater reliability (Cohen's $k = .836$; disagreements resolved through discussion) and then the first author scored the remainder. For game performance, we calculated a sum score of the questions (1, 2 or 3 points per question, maximum: 20 points).

Although our sample size met the minimal requirement, it was relatively modest and led us to adjust our preregistered analysis plan, accordingly. Specifically, we utilized a Bayesian and generalized order-restricted information criterion approximation (GORICA) rather than Null Hypothesis Significance Testing (NHST) approach (see below for further details). The data were analyzed by RI-CLPMs (Mulder & Hamaker, 2021) with the package blavaan (Merkle & Rosseel, 2018) in R studio (R Studio Team, 2022) using Bayesian estimation with two Markov chains, and the number of iterations set at 2000. We used RI-CLPMs to accommodate the dynamics of achievement emotions, achievement goals, and mental effort. In RI-CLPMs, the data were decomposed into a grand mean (i.e., the means over individuals and over time), between-person components (i.e., trait-like, stable deviations from the grand means) and within-person components (i.e., state-like, temporal deviations from the individual mean; Figure 5.3). At the between level, we specified random intercepts, which indicate the stable differences between individuals. We added peers' achievement emotions (i.e., enjoyment, frustration, or neutral state) as the grouping variable; enjoyment, relaxation, frustration, anger, or boredom as mediators; mastery-approach goals, performance-approach goals, mastery-avoidance goals, performance-avoidance goals, mental effort, posttest performance, and game performance as outcome variables; and prior enjoyment, prior relaxation, prior frustration, prior anger, prior boredom, prior mastery-approach goals, prior performance-approach goals, prior mastery-avoidance goals, prior performance-avoidance goals, and pretest performance as covariates.

At the within level, we specified the autoregression effects and cross-lagged effects. The autoregression effects indicate how the deviations from an individual's mean on one variable at one timepoint predict the deviations from the individual's mean on the same variable at the next timepoint. For example, a positive autoregressive effect of enjoyment implies that an individual who experiences deviation from their mean on enjoyment at the current game level is likely to experience deviation from their mean on enjoyment at the next game level. The cross-lagged effects indicate how the deviations from an individual's mean on one variable at one timepoint predict the deviations from the individual's mean on another variable at the next timepoint. For example, a positive cross-lagged effect from enjoyment to mental effort implies that an individual who experiences a deviation from their mean on enjoyment at the current game level is likely to experience a deviation from their mean on mental effort at the next game level.

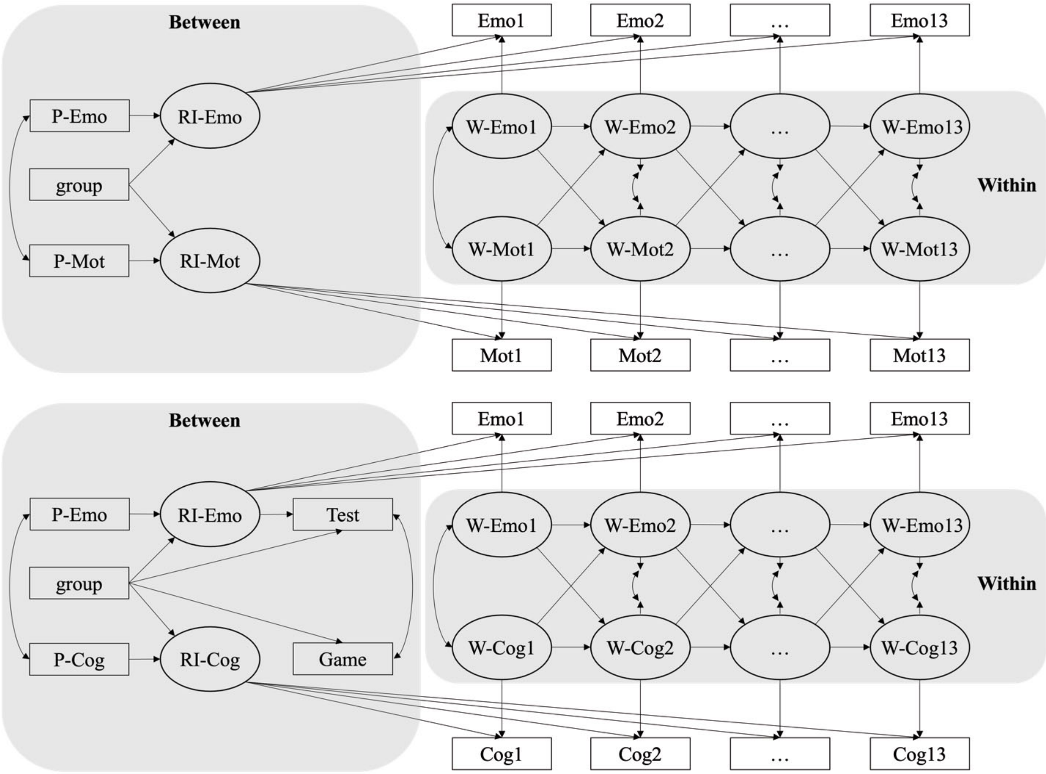
We used Bayesian estimation to accommodate small samples (McNeish, 2016). Because we had no prior knowledge about our parameter estimates from previous research, we used default noninformative priors so that priors had little influence on the analysis and parameter estimates were determined solely by the data (Gelman et al., 2014). To make sure that our complex models converged, we used bivariate models with one variable from five achievement emotions and one variable from four achievement goals and mental effort, all of which were measured 13 times during gameplay. In this way, each model included one variable from achievement emotions. This resulted in five RI-CLPM models for each variable and 25 RI-CLPM models in total. The trace plots and Rhats values (< 1.01) indicated that our models had good convergence. All the points on the When-to-Worry-and-How-to-Avoid-the-Misuse-of-Bayesian-Statistics checklist (the WAMBS checklist; Depaoli & van de Schoot, 2017) were addressed. Missing data were managed by full information maximum likelihood (FIML). The time intervals between consecutive measurement points did not vary. In evaluating model fit we focused primarily on the comparative fit index (CFI: ≥ .900 = acceptable; ≥ .950 = excellent), as it is less sensitive to the model and data characteristics than other fit indexes, such as chi-square (Asparouhov & Muthén, 2018; Kenny et al., 2015; Marsh et al., 1988). All our models had acceptable fit, ranging from .900 to .951. We also conducted sensitivity analysis to test the robustness of the models. Specifically, we compared RI-CLPMs with traditional CLPMs (i.e., random-intercepts and covariance between random-intercepts of RI-CLPMs were constrained to zero) and RI-CLPMs with constrained lagged-effects over time (i.e., lagged-effect were constrained as time-invariant). Model comparison was based on the Chi-bar-square ($\bar{X}2$) difference test. All our models had a better fit than traditional CLPMs and RI-CLPMs with constrained lagged-effects over time (see Table S5.1 in supplementary materials).

For our research question on main effects, we were interested in directly evaluating hypotheses containing inequality constraints (e.g., $a_e > 0 > a_f$), also called informative hypotheses (Hoijtink, 2011). However, the traditional NHST is not appropriate because of its limitations, such as a *p*-value cannot quantify the evidence in favor of one hypothesis, NHST cannot test multiple hypotheses simultaneously, and NHST cannot evaluate the hypotheses containing equality (e.g., $a_e = a_f$) and/or inequality constraints (Altinisik et al., 2021; Wasserstein et al., 2019). A newly developed alternative approach that meets our needs is GORICA (Kuiper, 2022; Kuiper et al., 2011). As an extension of Akaike-type information criterion (AIC; Akaike, 1974), GORICA can select the best hypothesis from a set of hypotheses (see Figure S5.1 in supplementary materials).

We used GORICA to evaluate whether our hypothesis of interest has more support than its complement (i.e., all possible hypotheses except the hypothesis of interest). The log likelihood indicates the compatibility of the hypothesis with the data. The values of GORICA and the log likelihood themselves are not interpretable but only comparable and thus GORICA weights and log likelihood weights are computed. We used the ratio of the GORICA weights and the ratio of log likelihood weights of two competing hypotheses to find which hypothesis is best supported. For the hypothesis of interest without equality constraints, if the ratio of GORICA weights is larger than 1 and the ratio of log likelihood weights is larger than 1.5, the hypothesis of interest is more supported than its complement. For the hypothesis of interest with equality constraints, there is no need to check the ratio of log likelihood weights. If our hypothesis of interest was not supported, the best alternative hypothesis was explored. Because each variable was estimated in five models, if different models supported different hypotheses for the same variable, then the overlapping part of these hypotheses was more supported.

Figure 5.3

Random-intercept cross-lagged panel models for achievement emotions and achievement goals (top) and achievement emotions and mental effort (bottom)



Note. Emo = achievement emotions (enjoyment, frustration, boredom, anger, relaxation); Mot = motivation (mastery-approach goals, performance-approach goals, mastery-avoidance goals, performance-avoidance goals); Cog = cognition (mental effort); Test = test performance; Game = game performance; RI-Emo, RI-Mot, RI-Cog = random intercepts for emotion, motivation, and cognition, respectively; P-Emo = prior achievement emotions; P-Mot = prior achievement goals; P-Cog = prior test performance; prior test performance is the covariate of test performance and game performance, which are not showed due to complexity of the models; W-Emo, W-Mot, W-Cog = within-person emotion, motivation, and cognition, respectively; Numbers indicate number of timepoints; Residual covariances are not showed.

Specifically, when comparing the hypothesis of interest with its complement, we first checked the ratio of GORICA weights. If the ratio of GORICA weights was not larger than 1, this indicated that the hypothesis of interest was not supported, and we explored all possible hypotheses. If the ratio of GORICA weights was larger than 1, then we checked the ratio of log likelihood weights. If the ratio of log likelihood weights was larger than 1.5, the hypothesis of interest was supported. If the ratio of log likelihood weights was not larger than 1.5, the hypothesis of interest was not supported, and we explored all possible hypotheses. If one of these possible hypotheses had higher GORICA weights than others, this indicated that there was support for this hypothesis and we further compared this hypothesis with its complement to confirm the support. If two or more hypotheses had higher GORICA weights than others, this indicated that there was support for the overlapping part of these hypotheses and we further compared this overlapping part with its complement to confirm the support.

Results

RQ1: Effects of Peers' Achievement Emotions on Students' Learning

Achievement Emotions

As shown in Table 5.2, for enjoyment and frustration, all models showed that students exposed to peers' enjoyment reported higher enjoyment and lower frustration than those exposed to peers' neutral state, followed by those exposed to peers' frustration (enjoyment: $a_e > 0 > a_f$; frustration: $a_e < 0 < a_f$). For relaxation and anger, all models showed that students exposed to peers' enjoyment and those exposed to peers' neutral state reported higher relaxation and lower anger than those exposed to peers' frustration (relaxation: $a_e = 0 > a_f$; anger: $a_e = 0 < a_f$). For boredom, all models showed that students exposed to peers' enjoyment reported lower boredom than those exposed to peers' neutral state and those exposed to peers' frustration ($a_e < 0 = a_f$).

Achievement Goals

For mastery-approach goals and mastery-avoidance goals, all models showed that students exposed to peers' enjoyment reported higher mastery-approach goals and higher mastery-avoidance goals than those exposed to peers' neutral state and those exposed to peers' frustration ($a_e > 0 = a_f$). For performance-approach goals and performance-avoidance goals, all models showed that there were no differences among groups ($a_e = 0 = a_f$).

Mental Effort and Performance

For mental effort, all models showed that students exposed to peers' enjoyment and those exposed to peers' neutral state reported lower mental effort than those exposed to peers' frustration ($a_e = 0 < a_f$). For game performance and posttest performance, all models showed that students exposed to peers' enjoyment reported higher game performance and lower posttest performance than those exposed to peers' neutral state and those exposed to peers' frustration (game performance: $a_e > 0 = a_f$; posttest performance: $a_e < 0 = a_f$).

Within-Person Correlations, Between-Person Correlations, Autoregressive Effects, and Cross-Lagged Effects

As shown in Table 5.3, the within-person correlations between positive emotions and achievement goals were mostly positive, whereas the within-person correlations between negative emotions and achievement goals were mostly negative. The within-person correlations between positive emotions and mental effort were mostly negative, whereas the within-person correlations between negative emotions and mental effort were mostly positive.

The between-person correlations between positive emotions and mastery-approach goals, mastery-avoidance goals, or performance-approach goals were positive, whereas the between-person correlations between negative emotions and mastery-approach goals, mastery-avoidance goals, or performance-approach goals were negative. The between-person correlations between enjoyment, relaxation, anger, or boredom and performance-avoidance goals were positive, whereas the between-person correlations between frustration and performance-avoidance goals were negative. The between-person correlations between enjoyment, anger, or boredom and mental effort were negative, whereas the between-person correlations between relaxation or frustration and mental effort were positive.

As shown in Table 5.4, most of achievement emotions, achievement goals, and mental effort have substantial within-person autoregressive effects over time (enjoyment: ß = .142 to .722; relaxation: ß = .051 to .542; frustration: ß = -.011 to 1.028; anger: ß = .074 to .835; boredom: ß = .211 to .739; mastery-approach goals: ß = -.101 to .613; mastery-avoidance goals: ß = -.185 to 1.329; performance-approach goals: ß = -.079 to .613; performance-avoidance goals: ß = .093 to .838; mental effort: ß = .066 to .697).

For enjoyment, there were mostly positive cross-lagged effects between enjoyment and achievement goals or mental effort. For relaxation, there were mostly positive cross-lagged effects between relaxation and mastery-approach goals, performance-approach goals, performance-avoidance goals, or mental effort, whereas there were positive and negative cross-lagged effects between relaxation and mastery-avoidance goals or mental effort. For frustration, there were mostly positive cross-lagged effects between frustration and mental effort, whereas there were mostly negative cross-lagged effects between frustration and performance-approach goals or performance-avoidance goals. Frustration has mostly negative cross-lagged effects on mastery-approach goals and mastery-avoidance goals, whereas mastery-approach goals have positive and negative cross-lagged effects on frustration, and mastery-avoidance goals have positive cross-lagged effects on frustration. For anger, there were mostly positive cross-lagged effects between anger and mastery-approach goals or performance-approach goals, whereas there were mostly negative cross-lagged effects between anger and performance-avoidance goals. Anger has mostly negative cross-lagged effects on mastery-avoidance goals or mental effort, whereas mastery-avoidance goals or mental effort has positive cross-lagged effects on frustration. For boredom, there were mostly negative cross-lagged effects between boredom and achievement goals or mental effort except that boredom has mostly negative cross-lagged effects on mental effort.

Table 5.2

Main effects of peers' achievement emotions on students' learning after controlling for the covariates

	Loglik weights H1	Loglik weights Hc	Loglik ratio H1/Hc	GORICA weights H1	GORICA weights Hc	GORICA ratio H1/Hc
Enjoyment and mental effort model						
Enjoyment H: $a_e > 0 > a_f$	.602	.398	1.513	.784	.216	3.630
Mental effort H: $a_e = a_f > 0$				.390	.610	.639
Mental effort Ha: $a_e = 0 < a_f$				.708	.292	2.425
Posttest performance H: $a_e = a_f < 0$				.670	.330	2.030
Posttest performance H: $a_e = a_f > 0$				.500	.500	1.000
Posttest performance H: $a_e > 0 = a_f$				.500	.500	1.000
Posttest performance Ha: $a_e = 0 < a_f$				.680	.320	2.125
Game performance H: $a_e = a_f < 0$				.179	.821	.218
Game performance H: $a_e = a_f > 0$				.259	.741	.350
Game performance H: $a_e > 0 = a_f$				.730	.270	2.704
Enjoyment and mastery-approach goals model						
Enjoyment H: $a_e > 0 > a_f$	.604	.396	1.525	.789	.211	3.739
Mastery-approach goals H: $a_e > 0 > a_f$	.504	.496	1.016	.708	.292	2.425
Mastery-approach goals Ha: $a_e > 0 = a_f$				.728	.272	2.676
Enjoyment and mastery-avoidance goals model						
Enjoyment H: $a_e > 0 > a_f$	.619	.381	1.625	.795	.205	3.878
Mastery-avoidance goals H: $a_e < 0 < a_f$	.128	.872	.147	.258	.742	.348
Mastery-avoidance goals Ha: $a_e > 0 = a_f$				.727	.273	2.663
Enjoyment and performance-approach goals model						
Enjoyment H: $a_e > 0 > a_f$	.601	.399	1.506	.787	.213	3.695
Performance-approach goals Ha: $a_e = 0 = a_f$				.806	.174	4.632
Enjoyment and performance-avoidance goals model						
Enjoyment H: $a_e > 0 > a_f$	.669	.331	2.021	.831	.169	4.917
Performance-avoidance goals H: $a_e = 0 = a_f$				.831	.169	4.917
Relaxation and mental effort model						
Relaxation H: $a_e > 0 > a_f$	.529	.471	1.123	.732	.268	2.731
Relaxation Ha: $a_e = 0 > a_f$				.708	.292	2.427
Mental effort H: $a_e = a_f > 0$				.295	.705	.418
Mental effort Ha: $a_e = 0 < a_f$				.715	.285	2.509

	Loglik weights H1	Loglik weights Hc	Loglik ratio H1/Hc	GORICA weights H1	GORICA weights Hc	GORICA ratio H1/Hc
Posttest performance H: $a_e = a_f < 0$				.668	.332	2.012
Posttest performance H: $a_e = a_f > 0$				.495	.505	.980
Posttest performance H: $a_e > 0 = a_f$				.495	.505	.980
Posttest performance Ha: $a_e < 0 = a_f$				.684	.316	2.165
Game performance H: $a_e = a_f < 0$				.146	.854	.171
Game performance H: $a_e = a_f > 0$				.232	.678	.342
Game performance H: $a_e > 0 = a_f$				.730	.270	2.704
Relaxation and mastery-approach goals model						
Relaxation H: $a_e > 0 > a_f$	.502	.498	1.008	.708	.294	2.408
Relaxation Ha: $a_e = 0 > a_f$				.730	.270	2.704
Mastery-approach goals H: $a_e > 0 > a_f$	.508	.492	1.033	.712	.288	2.472
Mastery-approach goals Ha: $a_e > 0 = a_f$				.725	.275	2.636
Relaxation and mastery-avoidance goals model						
Relaxation H: $a_e > 0 > a_f$	.500	.500	1.000	.707	.293	2.413
Relaxation Ha: $a_e = 0 > a_f$				.731	.269	2.717
Mastery-avoidance goals H: $a_e < 0 < a_f$	.130	.870	.149	.264	.736	.359
Mastery-avoidance goals Ha: $a_e > 0 = a_f$				.730	.270	2.704
Relaxation and performance-approach goals model						
Relaxation H: $a_e > 0 > a_f$	.503	.497	1.012	.715	.285	2.509
Relaxation Ha: $a_e = 0 > a_f$				.729	.271	2.690
Performance-approach goals Ha: $a_e = 0 = a_f$				.825	.175	4.714
Relaxation and performance-avoidance goals model						
Relaxation H: $a_e > 0 > a_f$	.500	.500	1.000	.705	.295	2.390
Relaxation Ha: $a_e = 0 > a_f$				.731	.269	2.717
Performance-avoidance goals Ha: $a_e = 0 = a_f$				.834	.166	5.024
Frustration and mental effort model						
Frustration H: $a_e < 0 < a_f$	.735	.265	2.774	.869	.131	6.634
Mental effort H: $a_e = a_f > 0$				.281	.719	.391
Mental effort Ha: $a_e = 0 < a_f$				.693	.307	2.257
Posttest performance H: $a_e = a_f < 0$				.670	.330	2.030
Posttest performance H: $a_e = a_f > 0$				.502	.498	1.008
Posttest performance H: $a_e > 0 = a_f$				.502	.498	1.008
Posttest performance Ha: $a_e < 0 = a_f$				.686	.314	2.185
Game performance H: $a_e = a_f < 0$				.155	.845	.183

	Loglik weights H1	Loglik weights Hc	Loglik ratio H1/Hc	GORICA weights H1	GORICA weights Hc	GORICA ratio H1/Hc
Game performance H: $a_e = a_f > 0$				.251	.749	.335
Game performance H: $a_e > 0 = a_f$				.730	.270	2.704
Frustration and mastery-approach goals model						
Frustration H: $a_e < 0 < a_f$	.699	.301	2.322	.851	.149	5.711
Mastery-approach goals H: $a_e > 0 > a_f$	.502	.498	1.008	.710	.290	2.448
Mastery-approach goals Ha: $a_e > 0 = a_f$				.730	.270	2.704
Frustration and mastery-avoidance goals model						
Frustration H: $a_e < 0 < a_f$	.635	.365	1.740	.806	.194	4.155
Mastery-avoidance goals H: $a_e < 0 < a_f$	.143	.856	.167	.289	.711	.406
Mastery-avoidance goals Ha: $a_e > 0 = a_f$				.731	.269	2.717
Frustration and performance-approach goals model						
Frustration H: $a_e < 0 < a_f$	.728	.272	2.676	.869	.131	6.634
Performance-approach goals Ha: $a_e = 0 = a_f$				.845	.155	5.452
Frustration and performance-avoidance goals model						
Frustration H: $a_e < 0 < a_f$	.785	.215	3.651	.900	.100	9.000
Performance-avoidance goals Ha: $a_e = 0 = a_f$				.835	.165	5.061
Anger and mental effort model						
Anger H: $a_e < 0 < a_f$	.516	.484	1.066	.710	.290	2.448
Anger Ha: $a_e = 0 < a_f$				.718	.282	2.546
Mental effort H: $a_e = a_f > 0$				.290	.710	.408
Mental effort Ha: $a_e = 0 < a_f$				.710	.290	2.448
Posttest performance H: $a_e = a_f < 0$				.673	.327	2.058
Posttest performance H: $a_e = a_f > 0$				.488	.512	.953
Posttest performance H: $a_e > 0 = a_f$				.488	.512	.953
Posttest performance Ha: $a_e < 0 = a_f$				.682	.318	2.145
Game performance H: $a_e = a_f < 0$				.313	.687	.456
Game performance H: $a_e = a_f > 0$				.520	.480	1.083
Game performance H: $a_e > 0 = a_f$				.713	.287	2.484
Anger and mastery-approach goals model						
Anger H: $a_e < 0 < a_f$	.531	.469	1.132	.729	.271	2.690
Anger Ha: $a_e = 0 < a_f$				.699	.303	2.307
Mastery-approach goals H: $a_e > 0 > a_f$	.505	.495	1.020	.708	.292	2.425
Mastery-approach goals Ha: $a_e > 0 = a_f$				.727	.273	2.663

	Loglik weights H1	Loglik weights Hc	Loglik ratio H1/Hc	GORICA weights H1	GORICA weights Hc	GORICA ratio H1/Hc
Anger and mastery-avoidance goals model						
Anger H: $a_e < 0 < a_f$	.530	.470	1.128	.728	.272	2.676
Anger Ha: $a_e = 0 < a_f$				.707	.293	2.413
Mastery-avoidance goals H: $a_e < 0 < a_f$	.157	.843	.186	.310	.690	.449
Mastery-avoidance goals Ha: $a_e > 0 = a_f$				.728	.272	2.676
Anger and performance-approach goals model						
Anger H: $a_e < 0 < a_f$	.520	.480	1.083	.718	.282	2.546
Anger Ha: $a_e = 0 < a_f$				.717	.283	2.534
Performance-approach goals Ha: $a_e = 0 = a_f$				.872	.128	6.813
Anger and performance-avoidance goals model						
Anger H: $a_e < 0 < a_f$	.524	.476	1.101	.722	.278	2.597
Anger Ha: $a_e = 0 < a_f$				.714	.286	2.497
Performance-avoidance goals Ha: $a_e = 0 = a_f$				.840	.160	5.250
Boredom and mental effort model						
Boredom H: $a_e < 0 < a_f$	.493	.507	.972	.701	.299	2.344
Boredom Ha: $a_e < 0 = a_f$				.725	.275	2.636
Mental effort H: $a_e = a_f > 0$				.349	.651	.536
Mental effort Ha: $a_e = 0 < a_f$				.710	.290	2.448
Posttest performance H: $a_e = a_f < 0$				.668	.332	2.012
Posttest performance H: $a_e = a_f > 0$				.473	.527	.896
Posttest performance H: $a_e > 0 = a_f$				.473	.527	.896
Posttest performance Ha: $a_e < 0 = a_f$				.685	.315	2.175
Game performance H: $a_e = a_f < 0$				.159	.841	.189
Game performance H: $a_e = a_f > 0$				.237	.763	.311
Game performance H: $a_e > 0 = a_f$				.730	.270	2.704
Boredom and mastery-approach goals model						
Boredom H: $a_e < 0 < a_f$	.495	.505	.980	.701	.299	2.344
Boredom Ha: $a_e < 0 = a_f$				.727	.273	2.663
Mastery-approach goals H: $a_e > 0 > a_f$	.500	.500	1.000	.706	.294	2.401
Mastery-approach goals Ha: $a_e > 0 = a_f$				.731	.269	2.717
Boredom and mastery-avoidance goals model						
Boredom H: $a_e < 0 < a_f$	.500	.500	1.000	.710	.290	2.448
Boredom Ha: $a_e < 0 = a_f$				.731	.269	2.717
Mastery-avoidance goals H: $a_e < 0 < a_f$	.130	.870	.149	.262	.738	.355

	Loglik weights H1	Loglik weights Hc	Loglik ratio H1/Hc	GORICA weights H1	GORICA weights Hc	GORICA ratio H1/Hc
Mastery-avoidance goals Ha: $a_e > 0 = a_f$				.728	.272	2.676
Boredom and performance-approach goals model						
Boredom H: $a_e < 0 < a_f$	.493	.507	.972	.701	.299	2.344
Boredom Ha: $a_e < 0 = a_f$				.726	.274	2.650
Performance-approach goals Ha: $a_e = 0 = a_f$				.821	.179	4.587
Boredom and performance-avoidance goals model						
Boredom H: $a_e < 0 < a_f$	.515	.485	1.062	.718	.282	2.546
Boredom Ha: $a_e < 0 = a_f$				.719	.281	2.559
Performance-avoidance goals Ha: $a_e = 0 = a_f$				.871	.129	6.752

Note. The control group (peers' neutral state) is the reference group; a_e (a_f) represents the mean difference in a variable between the enjoyment (frustration) group and the control group; H = hypothesis of interest; Hc = the complement hypotheses of H or not H; Ha = the best alternative hypothesis when H is not more supported than Hc; loglik = log likelihood; GORICA = generalized order-restricted information criterion approximation; weights = the relative likelihood of a hypothesis given the data and the set of hypotheses; GORICA H/Hc = GORICA weights of H/ GORICA weights of Hc; For example, H of GORICA weights = .718 and Hc of GORICA weights = .282 means that H has (.718/.282 = 2.542 > 1 times) more support than Hc. For the hypothesis of interest with equality restriction, there is no need to check loglik.weights, so they are not shown.

Table 5.3

Within-person correlations (correlations between within-person centered variables) and between-person correlations (correlations of random intercepts)

	Enjoyment	Relaxation	Frustration	Anger	Boredom
Between person correlations					
Mental effort	-.023 (.110)	.100 (.097)	.099 (.104)	-.121 (.106)	-.189 (.128)
Mastery-approach goals	.308 (.073)	.238 (.054)	-.256 (.065)	-.278 (.069)	-.405 (.079)
Mastery-avoidance goals	.199 (.072)	.138 (.055)	-.027 (.059)	-.061 (.056)	-.206 (.079)
Performance-approach goals	.209 (.067)	.193 (.052)	-.120 (.058)	-.174 (.062)	-.210 (.072)
Performance-avoidance goals	.049 (.052)	.157 (.045)	-.022 (.048)	.021 (.044)	.006 (.054)
Within-person correlations					
Mental effort					
T1	.002 (.107)	-.131 (.108)	.256 (.133)	.038 (.113)	-.064 (.124)
T2	-.086 (.058)	-.096 (.045)	.296 (.075)	.103 (.043)	-.051 (.056)
T3	-.045 (.063)	.010 (.047)	.221 (.085)	.073 (.071)	.015 (.067)
T4	-.062 (.055)	.017 (.048)	.099 (.065)	.030 (.050)	.073 (.054)
T5	.011 (.061)	-.012 (.052)	.123 (.081)	.220 (.068)	-.133 (.057)
T6	-.094 (.056)	-.012 (.044)	.159 (.073)	.114 (.054)	.003 (.056)
T7	-.217 (.073)	-.014 (.050)	.204 (.077)	.074 (.051)	-.032 (.054)
T8	-.194 (.066)	-.225 (.060)	.384 (.098)	.232 (.064)	.151 (.065)
T9	-.063 (.048)	-.010 (.040)	.012 (.058)	.065 (.040)	.021 (.047)
T10	-.128 (.055)	-.103 (.043)	.164 (.074)	.088 (.064)	.069 (.058)
T11	-.061 (.052)	-.221 (.068)	.149 (.074)	.160 (.061)	.082 (.059)
T12	.006 (.065)	-.137 (.071)	.165 (.097)	.111 (.072)	.111 (.073)
T13	-.007 (.062)	-.144 (.078)	.132 (.099)	.123 (.058)	-.046 (.058)
Mastery-approach goals					
T1	.284 (.069)	.134 (.056)	-.267 (.078)	-.187 (.062)	-.357 (.073)
T2	.051 (.035)	.015 (.025)	-.098 (.042)	-.089 (.025)	-.075 (.036)
T3	.126 (.036)	.029 (.024)	-.061 (.043)	-.056 (.035)	-.080 (.033)
T4	.071 (.032)	.060 (.028)	-.065 (.038)	-.057 (.029)	-.078 (.031)
T5	.044 (.028)	.004 (.023)	-.030 (.035)	-.061 (.030)	-.018 (.025)
T6	.010 (.029)	.022 (.023)	-.032 (.033)	.033 (.026)	-.032 (.029)
T7	.124 (.038)	.017 (.026)	-.143 (.040)	-.051 (.026)	.008 (.028)
T8	.029 (.028)	.034 (.024)	-.083 (.035)	-.023 (.025)	.028 (.025)
T9	.020 (.027)	.012 (.021)	-.062 (.032)	-.001 (.022)	-.039 (.025)
T10	.056 (.022)	-.009 (.018)	.006 (.030)	.031 (.026)	.015 (.024)

	Enjoyment	Relaxation	Frustration	Anger	Boredom
T11	.082 (.029)	.044 (.037)	-.022 (.046)	-.090 (.034)	-.060 (.033)
T12	.056 (.025)	.050 (.028)	-.082 (.035)	-.051 (.026)	-.108 (.028)
T13	.016 (.028)	.016 (.037)	-.032 (.045)	-.043 (.028)	.008 (.026)
Mastery-avoidance goals					
T1	.201 (.078)	.078 (.064)	-.069 (.076)	-.073 (.064)	-.235 (.081)
T2	.050 (.036)	-.033 (.026)	.040 (.042)	.005 (.026)	-.027 (.037)
T3	.023 (.034)	.047 (.023)	-.027 (.044)	-.053 (.036)	-.020 (.034)
T4	.052 (.034)	-.018 (.030)	-.008 (.041)	.020 (.032)	-.053 (.033)
T5	.065 (.034)	.003 (.029)	-.046 (.040)	.001 (.035)	.036 (.032)
T6	.011 (.026)	-.005 (.021)	.001 (.032)	.004 (.026)	-.040 (.027)
T7	.052 (.031)	.037 (.022)	-.108 (.034)	-.044 (.024)	-.023 (.024)
T8	-.043 (.026)	-.017 (.024)	.046 (.029)	.043 (.026)	.056 (.025)
T9	.093 (.031)	-.003 (.022)	-.062 (.032)	.024 (.020)	-.057 (.028)
T10	.025 (.026)	.011 (.021)	-.022 (.032)	.028 (.025)	.046 (.026)
T11	.065 (.028)	.045 (.034)	-.021 (.046)	-.032 (.036)	-.070 (.033)
T12	.047 (.020)	-.002 (.022)	-.034 (.026)	-.018 (.022)	-.027 (.020)
T13	.011 (.033)	.056 (.042)	.031 (.047)	.001 (.028)	-.028 (.030)
Performance-approach goals					
T1	.227 (.078)	.079 (.067)	-.006 (.091)	-.112 (.070)	-.216 (.081)
T2	.102 (.040)	.066 (.029)	-.062 (.047)	-.042 (.031)	-.006 (.040)
T3	.095 (.038)	.049 (.028)	-.092 (.047)	-.049 (.039)	-.013 (.037)
T4	.043 (.031)	.070 (.029)	-.020 (.037)	.011 (.030)	-.051 (.030)
T5	.067 (.031)	-.025 (.024)	-.045 (.038)	-.065 (.031)	-.014 (.027)
T6	.073 (.035)	.016 (.027)	.001 (.039)	.058 (.033)	-.037 (.033)
T7	.131 (.041)	.042 (.030)	-.150 (.045)	-.078 (.031)	-.018 (.032)
T8	.056 (.026)	.038 (.023)	-.030 (.034)	-.010 (.024)	.031 (.024)
T9	.036 (.027)	-.026 (.022)	-.042 (.030)	.022 (.022)	-.062 (.025)
T10	.058 (.025)	.015 (.020)	.031 (.033)	-.021 (.027)	-.047 (.026)
T11	.061 (.026)	.020 (.032)	.022 (.043)	-.074 (.032)	-.006 (.030)
T12	.065 (.030)	.071 (.032)	-.151 (.043)	-.093 (.033)	-.055 (.032)
T13	.054 (.026)	.038 (.037)	-.096 (.045)	-.052 (.026)	-.018 (.026)
Performance-avoidance goals					
T1	-.026 (.057)	.063 (.051)	.068 (.066)	.051 (.044)	.082 (.055)
T2	.018 (.034)	-.017 (.025)	.002 (.039)	.006 (.025)	.086 (.036)
T3	-.009 (.033)	-.017 (.023)	-.012 (.042)	-.021 (.034)	.022 (.037)
T4	-.008 (.035)	-.005 (.031)	.017 (.040)	-.046 (.030)	-.071 (.036)
T5	.040 (.033)	.020 (.027)	-.072 (.042)	-.098 (.035)	.004 (.030)

	Enjoyment	Relaxation	Frustration	Anger	Boredom
T6	.060 (.028)	.040 (.023)	-.037 (.033)	-.090 (.028)	-.096 (.029)
T7	.126 (.036)	.053 (.026)	-.163 (.040)	-.065 (.025)	-.083 (.029)
T8	.115 (.034)	.101 (.030)	-.117 (.047)	-.111 (.034)	-.013 (.033)
T9	.026 (.030)	.014 (.025)	-.032 (.035)	-.025 (.026)	-.061 (.030)
T10	.009 (.028)	.003 (.024)	.023 (.042)	.006 (.035)	-.025 (.030)
T11	.128 (.032)	.085 (.039)	-.108 (.052)	-.115 (.037)	-.033 (.037)
T12	.100 (.030)	.054 (.031)	-.110 (.042)	-.061 (.033)	-.060 (.033)
T13	.018 (.024)	.041 (.033)	-.091 (.040)	-.036 (.024)	.001 (.022)

Table 5.4

Mean difference between conditions, effects of covariates on random intercepts, autoregressive effects, and cross-lagged effects.

	Enjoyment and mental effort model			
Mean differences	Enjoyment	Mental effort	Posttest Perf.	Game Perf.
Enjoyment vs. Neutral state (a_e)	.173 (.157)	-.158 (.332)	-.851 (.587)	1.108 (.621)
Frustration vs. Neutral state (a_f)	-.157 (.178)	.447 (.388)	-.483 (.688)	-.099 (.711)
Effects of covariates on random intercepts	.524 (.076)	-.211 (.056)	.678 (.100)	.427 (.104)
Autoregressive effects	Enjoyment	Mental effort		
T1 -> T2	.327 (.097)	.667 (.062)		
T2 -> T3	.288 (.094)	.581 (.082)		
T3 -> T4	.145 (.106)	.460 (.072)		
T4 -> T5	.376 (.111)	.376 (.104)		
T5 -> T6	.434 (.098)	.253 (.101)		
T6 -> T7	.322 (.120)	.301 (.108)		
T7-> T8	.449 (.090)	.268 (.128)		
T8 -> T9	.604 (.109)	.475 (.084)		
T9 -> T10	.532 (.082)	.361 (.123)		
T10 -> T11	.521 (.082)	.235 (.112)		
T11 -> T12	.722 (.085)	.180 (.151)		
T12 -> T13	.669 (.081)	.225 (.102)		
Cross-lagged effects	Enjoyment -> mental effort	Mental effort -> enjoyment		
T1 -> T2	.035 (.112)	-.039 (.053)		
T2 -> T3	.276 (.138)	.008 (.054)		
T3 -> T4	.294 (.135)	.015 (.055)		
T4 -> T5	.484 (.158)	.014 (.072)		
T5 -> T6	.056 (.143)	.008 (.063)		
T6 -> T7	.121 (.140)	.082 (.085)		
T7-> T8	-.479 (.143)	.046 (.077)		
T8 -> T9	-.283 (.125)	-.086 (.071)		
T9 -> T10	-.137 (.131)	-.183 (.070)		
T10 -> T11	-.129 (.146)	-.033 (.057)		
T11 -> T12	-.288 (.189)	-.001 (.062)		
T12 -> T13	-.026 (.142)	-.114 (.051)		

	Enjoyment and Map goals model		Enjoyment and Mav goals model	
Mean differences	Enjoyment	Map goals	Enjoyment	Mav goals
Enjoyment vs. Neutral state (a_e)	.183 (.156)	.275 (.184)	.214 (.160)	.283 (.186)
Frustration vs. Neutral state (a_f)	-.165 (.180)	-.036 (.200)	-.176 (.179)	-.041 (.212)
Effects of covariates on random intercepts	.403 (.069)	.459 (.082)	.402 (.078)	.394 (.073)
Autoregressive effects	Enjoyment	Map goals	Enjoyment	Mav goals
T1 -> T2	.274 (.112)	.513 (.071)	.321 (.101)	.610 (.054)
T2 -> T3	.291 (.100)	.327 (.083)	.293 (.095)	.534 (.070)
T3 -> T4	.063 (.119)	.518 (.094)	.152 (.111)	.487 (.077)
T4 -> T5	.392 (.126)	.580 (.083)	.431 (.119)	.381 (.084)
T5 -> T6	.461 (.093)	.200 (.092)	.419 (.093)	.181 (.074)
T6 -> T7	.315 (.110)	-.090 (.113)	.324 (.113)	.395 (.101)
T7-> T8	.434 (.097)	.436 (.134)	.405 (.086)	.032 (.100)
T8 -> T9	.617 (.094)	.243 (.105)	.618 (.109)	.217 (.164)
T9 -> T10	.546 (.069)	.224 (.117)	.547 (.077)	-.185 (.128)
T10 -> T11	.568 (.084)	.468 (.165)	.525 (.081)	.123 (.140)
T11 -> T12	.716 (.089)	.440 (.089)	.715 (.092)	.226 (.076)
T12 -> T13	.664 (.081)	.525 (.104)	.685 (.084)	.343 (.180)
Cross-lagged effects	Enjoyment -> Map goals	Map goals -> enjoyment	Enjoyment -> Mav goals	Mav goals -> enjoyment
T1 -> T2	.001 (.076)	.077 (.108)	-.028 (.071)	.104 (.081)
T2 -> T3	.055 (.076)	-.022 (.108)	.111 (.071)	.200 (.089)
T3 -> T4	.004 (.081)	.221 (.132)	.064 (.082)	.152 (.104)
T4 -> T5	.011 (.076)	.165 (.128)	.255 (.094)	.086 (.108)
T5 -> T6	.028 (.069)	.095 (.116)	.084 (.064)	.252 (.101)
T6 -> T7	.116 (.068)	.080 (.176)	.086 (.060)	.018 (.178)
T7-> T8	.056 (.063)	-.085 (.199)	.034 (.058)	-.134 (.145)
T8 -> T9	.103 (.063)	.299 (.156)	-.022 (.078)	-.018 (.212)
T9 -> T10	.120 (.052)	.470 (.151)	.057 (.064)	.183 (.147)
T10 -> T11	-.119 (.078)	-.111 (.167)	-.006 (.076)	-.170 (.146)
T11 -> T12	.046 (.071)	.018 (.109)	.048 (.060)	-.030 (.118)
T12 -> T13	.184 (.064)	.224 (.129)	.151 (.076)	-.176 (.198)

	Enjoyment and Pap goals model		Enjoyment and Pav goals model	
Mean differences	Enjoyment	Pap goals	Enjoyment	Pav goals
Enjoyment vs. Neutral state (a_e)	.196 (.158)	.069 (.177)	.179 (.148)	.088 (.160)
Frustration vs. Neutral state (a_f)	-.162 (.180)	.188 (.203)	-.201 (.170)	.163 (.182)
Effects of covariates on random intercepts	.422 (.071)	.462 (.060)	.458 (.070)	.570 (.061)
Autoregressive effects	Enjoyment	Pap goals	Enjoyment	Pav goals
T1 -> T2	.299 (.105)	.482 (.061)	.338 (.099)	.476 (.082)
T2 -> T3	.263 (.103)	.434 (.084)	.339 (.095)	.293 (.107)
T3 -> T4	.142 (.112)	.558 (.074)	.179 (.117)	.294 (.141)
T4 -> T5	.377 (.126)	.448 (.080)	.498 (.117)	.428 (.115)
T5 -> T6	.437 (.095)	.010 (.120)	.533 (.083)	.398 (.084)
T6 -> T7	.299 (.115)	-.021 (.122)	.371 (.108)	.439 (.105)
T7-> T8	.395 (.094)	.391 (.102)	.373 (.091)	.568 (.126)
T8 -> T9	.615 (.100)	.391 (.117)	.524 (.111)	.298 (.096)
T9 -> T10	.542 (.071)	.274 (.109)	.486 (.069)	.691 (.122)
T10 -> T11	.511 (.079)	.498 (.132)	.504 (.085)	.507 (.097)
T11 -> T12	.684 (.093)	.237 (.122)	.631 (.105)	.481 (.099)
T12 -> T13	.629 (.083)	.326 (.095)	.556 (.102)	.660 (.079)
Cross-lagged effects	Enjoyment -> Pap goals	Pap goals -> enjoyment	Enjoyment -> Pav goals	Pav goals -> enjoyment
T1 -> T2	.032 (.078)	.016 (.082)	-.066 (.063)	-.047 (.107)
T2 -> T3	.073 (.082)	.114 (.096)	-.153 (.066)	-.313 (.125)
T3 -> T4	.072 (.077)	-.080 (.103)	-.117 (.085)	-.354 (.175)
T4 -> T5	-.020 (.084)	.055 (.111)	-.010 (.084)	-.209 (.156)
T5 -> T6	-.011 (.083)	.121 (.120)	.078 (.062)	.093 (.120)
T6 -> T7	.116 (.080)	.040 (.160)	.117 (.065)	-.026 (.169)
T7-> T8	.117 (.058)	.138 (.160)	.250 (.074)	.277 (.143)
T8 -> T9	.131 (.066)	.192 (.172)	.155 (.081)	.315 (.117)
T9 -> T10	.030 (.055)	.358 (.132)	.179 (.066)	.572 (.118)
T10 -> T11	.032 (.066)	.137 (.143)	.245 (.083)	.155 (.091)
T11 -> T12	.171 (.087)	.119 (.118)	.282 (.098)	.171 (.099)
T12 -> T13	.262 (.065)	.312 (.119)	.256 (.069)	.298 (.112)

	Relaxation and mental effort model			
Mean differences	Relaxation	Mental effort	Posttest Perf.	Game Perf.
Enjoyment vs. Neutral state (a_e)	.066 (.137)	-.137 (.342)	-.835 (.592)	1.131 (.610)
Frustration vs. Neutral state (a_f)	-.140 (.153)	.474 (.389)	-.465 (.691)	-.066 (.706)
Effects of covariates on random intercepts	.488 (.062)	-.213 (.059)	.677 (.101)	.428 (.104)
Autoregressive effects	Relaxation	Mental effort		
T1 -> T2	.542 (.071)	.677 (.059)		
T2 -> T3	.363 (.068)	.588 (.085)		
T3 -> T4	.325 (.123)	.460 (.072)		
T4 -> T5	.303 (.096)	.390 (.101)		
T5 -> T6	.267 (.096)	.276 (.096)		
T6 -> T7	.422 (.115)	.337 (.110)		
T7-> T8	.338 (.105)	.392 (.114)		
T8 -> T9	.509 (.109)	.529 (.077)		
T9 -> T10	.467 (.073)	.306 (.105)		
T10 -> T11	.051 (.153)	.139 (.122)		
T11 -> T12	.290 (.100)	.066 (.189)		
T12 -> T13	.332 (.122)	.151 (.106)		
Cross-lagged effects	Relaxation -> mental effort	Mental effort -> relaxation		
T1 -> T2	.015 (.102)	.013 (.038)		
T2 -> T3	.186 (.144)	.046 (.037)		
T3 -> T4	.474 (.174)	-.032 (.047)		
T4 -> T5	.174 (.167)	.092 (.055)		
T5 -> T6	.004 (.174)	.030 (.047)		
T6 -> T7	.100 (.198)	.053 (.058)		
T7-> T8	-.248 (.189)	-.079 (.058)		
T8 -> T9	-.153 (.145)	-.009 (.057)		
T9 -> T10	-.398 (.153)	-.085 (.048)		
T10 -> T11	-.241 (.200)	-.192 (.081)		
T11 -> T12	-.287 (.199)	-.113 (.082)		
T12 -> T13	-.213 (.177)	-.135 (.070)		

	Relaxation and Map goals model		Relaxation and Mav goals model	
Mean differences	Relaxation	Map goals	Relaxation	Mav goals
Enjoyment vs. Neutral state (a_e)	.016 (.138)	.262 (.182)	-.002 (.134)	.291 (.188)
Frustration vs. Neutral state (a_f)	-.176 (.154)	-.043 (.210)	-.191 (.152)	-.019 (.215)
Effects of covariates on random intercepts	.460 (.054)	.479 (.086)	.427 (.059)	.405 (.074)
Autoregressive effects	Relaxation	Map goals	Relaxation	Mav goals
T1 -> T2	.489 (.077)	.432 (.067)	.499 (.072)	.595 (.053)
T2 -> T3	.315 (.079)	.325 (.092)	.317 (.072)	.536 (.069)
T3 -> T4	.239 (.135)	.505 (.095)	.224 (.133)	.501 (.076)
T4 -> T5	.331 (.101)	.572 (.082)	.274 (.105)	.426 (.082)
T5 -> T6	.266 (.092)	.199 (.097)	.288 (.090)	.182 (.074)
T6 -> T7	.448 (.116)	-.050 (.120)	.422 (.115)	.380 (.104)
T7-> T8	.360 (.104)	.497 (.124)	.350 (.103)	.032 (.106)
T8 -> T9	.487 (.093)	.295 (.111)	.482 (.093)	.228 (.151)
T9 -> T10	.496 (.075)	.280 (.119)	.495 (.076)	-.102 (.128)
T10 -> T11	.234 (.131)	.431 (.148)	.218 (.138)	.194 (.131)
T11 -> T12	.402 (.088)	.471 (.085)	.403 (.087)	.237 (.073)
T12 -> T13	.408 (.113)	.573 (.100)	.448 (.115)	.425 (.176)
Cross-lagged effects	Relaxation -> Map goals	Map goals -> relaxation	Relaxation -> Mav goals	Mav goals -> relaxation
T1 -> T2	.222 (.064)	.026 (.074)	.005 (.065)	.084 (.058)
T2 -> T3	-.052 (.084)	-.045 (.081)	.177 (.081)	-.043 (.063)
T3 -> T4	-.186 (.107)	-.037 (.115)	-.077 (.117)	-.013 (.087)
T4 -> T5	-.108 (.080)	-.209 (.106)	.075 (.101)	-.128 (.087)
T5 -> T6	-.021 (.087)	-.126 (.101)	-.029 (.085)	.134 (.082)
T6 -> T7	.035 (.101)	-.103 (.123)	-.022 (.088)	-.120 (.126)
T7-> T8	.083 (.075)	.078 (.158)	-.019 (.079)	-.064 (.135)
T8 -> T9	.099 (.074)	.194 (.122)	-.051 (.077)	.040 (.159)
T9 -> T10	.127 (.068)	.106 (.114)	-.017 (.085)	.005 (.109)
T10 -> T11	.046 (.095)	.065 (.187)	-.059 (.096)	-.138 (.171)
T11 -> T12	.084 (.061)	-.031 (.114)	.077 (.050)	-.159 (.124)
T12 -> T13	.080 (.071)	.291 (.158)	.057 (.082)	.149 (.237)

	Relaxation and Pap goals model		Relaxation and Pav goals model	
Mean differences	Relaxation	Pap goals	Relaxation	Pav goals
Enjoyment vs. Neutral state (a_e)	.021 (.140)	.082 (.174)	.007 (.128)	.120 (.157)
Frustration vs. Neutral state (a_f)	-.180 (.158)	.190 (.200)	-.187 (.145)	.143 (.185)
Effects of covariates on random intercepts	.436 (.058)	.479 (.064)	.405 (.056)	.560 (.059)
Autoregressive effects	Relaxation	Pap goals	Relaxation	Pav goals
T1 -> T2	.513 (.075)	.481 (.054)	.462 (.071)	.473 (.085)
T2 -> T3	.322 (.077)	.460 (.082)	.296 (.078)	.259 (.120)
T3 -> T4	.263 (.127)	.600 (.075)	.157 (.134)	.208 (.187)
T4 -> T5	.313 (.103)	.442 (.079)	.272 (.113)	.376 (.129)
T5 -> T6	.257 (.097)	-.033 (.137)	.328 (.097)	.390 (.094)
T6 -> T7	.440 (.119)	-.017 (.133)	.533 (.110)	.458 (.109)
T7-> T8	.355 (.104)	.436 (.099)	.415 (.100)	.672 (.125)
T8 -> T9	.452 (.093)	.397 (.111)	.480 (.098)	.364 (.095)
T9 -> T10	.490 (.074)	.260 (.117)	.484 (.071)	.778 (.122)
T10 -> T11	.163 (.132)	.507 (.132)	.252 (.135)	.565 (.088)
T11 -> T12	.360 (.088)	.280 (.118)	.396 (.090)	.574 (.076)
T12 -> T13	.361 (.121)	.432 (.106)	.353 (.114)	.791 (.066)
Cross-lagged effects	Relaxation -> Pap goals	Pap goals -> relaxation	Relaxation -> Pav goals	Pav goals -> relaxation
T1 -> T2	.178 (.072)	.032 (.058)	.006 (.064)	.015 (.084)
T2 -> T3	.008 (.090)	.040 (.069)	-.062 (.081)	-.206 (.094)
T3 -> T4	-.059 (.100)	.031 (.092)	-.186 (.125)	-.265 (.159)
T4 -> T5	-.053 (.081)	-.044 (.091)	.079 (.097)	-.384 (.140)
T5 -> T6	-.032 (.105)	-.009 (.110)	.106 (.084)	-.140 (.102)
T6 -> T7	-.031 (.119)	-.065 (.118)	.090 (.100)	-.070 (.125)
T7-> T8	.155 (.073)	.084 (.138)	.260 (.090)	.138 (.122)
T8 -> T9	.137 (.075)	.273 (.135)	.083 (.085)	.172 (.091)
T9 -> T10	.029 (.071)	.010 (.103)	.138 (.084)	.219 (.090)
T10 -> T11	.034 (.090)	.177 (.179)	.338 (.104)	.131 (.104)
T11 -> T12	.199 (.077)	.070 (.125)	.263 (.070)	.140 (.091)
T12 -> T13	.028 (.078)	.321 (.151)	.123 (.065)	.378 (.115)

	Frustration and mental effort model			
Mean differences	Frustration	Mental effort	Posttest Perf.	Game Perf.
Enjoyment vs. Neutral state (a_e)	-.201 (.142)	-.203 (.333)	-.838 (.602)	1.115 (.603)
Frustration vs. Neutral state (a_f)	.249 (.167)	.480 (.385)	-.461 (.701)	-.080 (.718)
Effects of covariates on random intercepts	.226 (.053)	-.188 (.057)	.679 (.100)	.428 (.106)
Autoregressive effects	Frustration	Mental effort		
T1 -> T2	.270 (.093)	.658 (.063)		
T2 -> T3	-.011 (.103)	.572 (.091)		
T3 -> T4	.149 (.113)	.412 (.083)		
T4 -> T5	.217 (.107)	.380 (.122)		
T5 -> T6	.527 (.091)	.250 (.103)		
T6 -> T7	.420 (.102)	.317 (.106)		
T7-> T8	.432 (.100)	.292 (.130)		
T8 -> T9	.326 (.099)	.527 (.084)		
T9 -> T10	.370 (.093)	.324 (.116)		
T10 -> T11	.358 (.108)	.272 (.088)		
T11 -> T12	.444 (.090)	.123 (.143)		
T12 -> T13	.344 (.103)	.223 (.109)		
Cross-lagged effects	Frustration -> mental effort	Mental effort -> frustration		
T1 -> T2	-.016 (.090)	.059 (.066)		
T2 -> T3	-.184 (.117)	.185 (.081)		
T3 -> T4	.057 (.118)	.163 (.078)		
T4 -> T5	-.098 (.133)	.048 (.096)		
T5 -> T6	.272 (.112)	-.061 (.077)		
T6 -> T7	-.091 (.115)	-.128 (.094)		
T7-> T8	.278 (.120)	-.132 (.102)		
T8 -> T9	.025 (.094)	.232 (.086)		
T9 -> T10	.282 (.115)	.222 (.091)		
T10 -> T11	.031 (.092)	-.100 (.100)		
T11 -> T12	.002 (.134)	-.040 (.094)		
T12 -> T13	-.012 (.143)	.046 (.077)		

	Frustration and Map goals model		Frustration and Mav goals model	
Mean differences	Frustration	Map goals	Frustration	Mav goals
Enjoyment vs. Neutral state (a_e)	-.192 (.150)	.285 (.179)	-.149 (.143)	.283 (.186)
Frustration vs. Neutral state (a_f)	.230 (.174)	-.023 (.212)	.270 (.161)	-.006 (.211)
Effects of covariates on random intercepts	.231 (.051)	.535 (.085)	.208 (.054)	.426 (.076)
Autoregressive effects	Frustration	Map goals	Frustration	Mav goals
T1 -> T2	.377 (.095)	.492 (.068)	.263 (.092)	.602 (.056)
T2 -> T3	.099 (.091)	.345 (.082)	.072 (.099)	.544 (.068)
T3 -> T4	.204 (.105)	.521 (.086)	.251 (.098)	.505 (.077)
T4 -> T5	.223 (.107)	.580 (.078)	.248 (.105)	.964 (.183)
T5 -> T6	.487 (.089)	.192 (.089)	.549 (.087)	.165 (.074)
T6 -> T7	.330 (.101)	-.053 (.116)	.368 (.092)	.318 (.122)
T7-> T8	.329 (.104)	.445 (.120)	.391 (.101)	.311 (.096)
T8 -> T9	.323 (.089)	.274 (.108)	.398 (.087)	.432 (.089)
T9 -> T10	.381 (.104)	.215 (.113)	.407 (.100)	-.109 (.089)
T10 -> T11	.336 (.108)	.406 (.146)	.347 (.107)	.212 (.117)
T11 -> T12	.432 (.082)	.471 (.083)	.427 (.075)	.231 (.079)
T12 -> T13	.391 (.106)	.572 (.100)	1.028 (.242)	.176 (.185)
Cross-lagged effects	Frustration -> Map goals	Map goals -> frustration	Frustration -> Mav goals	Mav goals -> frustration
T1 -> T2	-.035 (.052)	.216 (.121)	.059 (.053)	.000 (.095)
T2 -> T3	.038 (.056)	.159 (.133)	-.141 (.055)	.138 (.119)
T3 -> T4	.060 (.062)	-.058 (.138)	.060 (.056)	.010 (.118)
T4 -> T5	.069 (.056)	.094 (.149)	-.057 (.079)	.215 (.131)
T5 -> T6	-.048 (.056)	.199 (.137)	-.017 (.049)	.007 (.118)
T6 -> T7	.108 (.057)	-.061 (.188)	-.004 (.048)	-.146 (.209)
T7-> T8	-.053 (.053)	-.057 (.234)	.006 (.043)	.330 (.220)
T8 -> T9	-.029 (.051)	-.496 (.182)	-.005 (.048)	-.176 (.154)
T9 -> T10	-.100 (.056)	-.115 (.207)	.017 (.051)	-.126 (.174)
T10 -> T11	-.003 (.058)	.246 (.259)	.002 (.054)	.172 (.209)
T11 -> T12	-.007 (.046)	.030 (.146)	-.025 (.034)	.080 (.143)
T12 -> T13	.001 (.054)	-.105 (.185)	-.106 (.060)	.135 (.348)

	Frustration and Pap goals model		Frustration and Pav goals model	
Mean differences	Frustration	Pap goals	Frustration	Pav goals
Enjoyment vs. Neutral state (a_e)	-.223 (.144)	.063 (.181)	-.211 (.133)	.120 (.163)
Frustration vs. Neutral state (a_f)	.231 (.166)	.159 (.206)	.253 (.153)	.146 (.183)
Effects of covariates on random intercepts	.224 (.051)	.506 (.066)	.223 (.051)	.595 (.064)
Autoregressive effects	Frustration	Pap goals	Frustration	Pav goals
T1 -> T2	.285 (.096)	.509 (.057)	.276 (.092)	.411 (.085)
T2 -> T3	.095 (.095)	.467 (.078)	.086 (.098)	.177 (.125)
T3 -> T4	.237 (.107)	.598 (.073)	.285 (.103)	.138 (.154)
T4 -> T5	.234 (.110)	.465 (.078)	.267 (.108)	.347 (.131)
T5 -> T6	.528 (.089)	.013 (.114)	.561 (.091)	.358 (.099)
T6 -> T7	.364 (.096)	.024 (.133)	.389 (.093)	.517 (.109)
T7-> T8	.414 (.107)	.468 (.095)	.405 (.107)	.725 (.130)
T8 -> T9	.415 (.085)	.416 (.106)	.382 (.085)	.379 (.075)
T9 -> T10	.401 (.097)	.181 (.114)	.422 (.096)	.814 (.119)
T10 -> T11	.338 (.105)	.459 (.144)	.355 (.103)	.647 (.080)
T11 -> T12	.435 (.080)	.240 (.124)	.434 (.085)	.646 (.073)
T12 -> T13	.401 (.116)	.381 (.115)	.317 (.109)	.826 (.062)
Cross-lagged effects	Frustration -> Pap goals	Pap goals -> frustration	Frustration -> Pav goals	Pav goals -> frustration
T1 -> T2	.005 (.058)	-.027 (.093)	.044 (.050)	-.115 (.137)
T2 -> T3	-.038 (.063)	.235 (.115)	-.026 (.052)	.414 (.185)
T3 -> T4	.111 (.060)	.087 (.115)	.089 (.068)	.046 (.214)
T4 -> T5	.047 (.057)	.067 (.133)	-.137 (.063)	.199 (.225)
T5 -> T6	-.068 (.062)	.239 (.139)	-.138 (.055)	.003 (.169)
T6 -> T7	-.006 (.063)	-.013 (.169)	-.058 (.056)	-.130 (.182)
T7-> T8	-.005 (.049)	.157 (.202)	-.069 (.070)	-.030 (.190)
T8 -> T9	-.072 (.046)	-.352 (.179)	-.129 (.051)	-.382 (.114)
T9 -> T10	-.100 (.055)	-.226 (.191)	-.096 (.066)	-.209 (.156)
T10 -> T11	-.096 (.052)	-.170 (.258)	-.175 (.061)	-.132 (.124)
T11 -> T12	-.085 (.054)	-.254 (.163)	-.079 (.055)	-.115 (.105)
T12 -> T13	-.026 (.061)	-.036 (.208)	-.077 (.050)	-.443 (.128)

	Anger and mental effort model			
Mean differences	Anger	Mental effort	Posttest Perf.	Game Perf.
Enjoyment vs. Neutral state (a_e)	-.049 (.136)	-.209 (.343)	-.827 (.575)	1.053 (.592)
Frustration vs. Neutral state (a_f)	.187 (.160)	.468 (.397)	-.468 (.679)	.297 (.694)
Effects of covariates on random intercepts	.366 (.072)	-.201 (.057)	.679 (.098)	.406 (.103)
Autoregressive effects	Anger	Mental effort		
T1 -> T2	.230 (.123)	.697 (.060)		
T2 -> T3	.223 (.173)	.621 (.083)		
T3 -> T4	.522 (.095)	.496 (.071)		
T4 -> T5	.465 (.103)	.412 (.100)		
T5 -> T6	.540 (.102)	.234 (.100)		
T6 -> T7	.566 (.108)	.326 (.112)		
T7-> T8	.537 (.113)	.319 (.114)		
T8 -> T9	.573 (.114)	.528 (.085)		
T9 -> T10	.567 (.117)	.259 (.118)		
T10 -> T11	.562 (.099)	.150 (.113)		
T11 -> T12	.582 (.096)	.112 (.170)		
T12 -> T13	.763 (.077)	.154 (.113)		
Cross-lagged effects	Anger -> mental effort	Mental effort -> anger		
T1 -> T2	-.277 (.144)	.045 (.043)		
T2 -> T3	-.247 (.240)	.014 (.062)		
T3 -> T4	-.060 (.135)	.089 (.046)		
T4 -> T5	.062 (.137)	.063 (.070)		
T5 -> T6	.395 (.123)	.003 (.059)		
T6 -> T7	.113 (.151)	-.058 (.061)		
T7-> T8	.320 (.161)	.084 (.063)		
T8 -> T9	.034 (.127)	.135 (.061)		
T9 -> T10	.257 (.150)	.058 (.081)		
T10 -> T11	.261 (.138)	.069 (.071)		
T11 -> T12	.010 (.170)	.029 (.080)		
T12 -> T13	.114 (.143)	.123 (.052)		

	Anger and Map goals model		Anger and Mav goals model	
Mean differences	Anger	Map goals	Anger	Mav goals
Enjoyment vs. Neutral state (a_e)	-.078 (.140)	.252 (.183)	-.065 (.134)	.254 (.182)
Frustration vs. Neutral state (a_f)	.146 (.164)	-.033 (.213)	.173 (.154)	-.037 (.208)
Effects of covariates on random intercepts	.402 (.066)	.568 (.085)	.389 (.072)	.428 (.075)
Autoregressive effects	Anger	Map goals	Anger	Mav goals
T1 -> T2	.294 (.108)	.500 (.069)	.163 (.109)	.607 (.058)
T2 -> T3	.186 (.173)	.356 (.090)	.195 (.184)	.532 (.076)
T3 -> T4	.542 (.102)	.513 (.086)	.559 (.104)	.516 (.077)
T4 -> T5	.540 (.103)	.613 (.077)	.516 (.101)	.998 (.170)
T5 -> T6	.556 (.075)	.239 (.091)	.599 (.074)	.199 (.079)
T6 -> T7	.539 (.084)	-.055 (.106)	.589 (.079)	.383 (.136)
T7-> T8	.552 (.095)	.419 (.128)	.617 (.085)	.313 (.127)
T8 -> T9	.649 (.082)	.234 (.114)	.660 (.084)	.461 (.092)
T9 -> T10	.587 (.100)	.212 (.125)	.625 (.091)	1.329 (.402)
T10 -> T11	.572 (.091)	.339 (.162)	.477 (.108)	.257 (.112)
T11 -> T12	.612 (.080)	.455 (.082)	.605 (.076)	.207 (.099)
T12 -> T13	.799 (.073)	.563 (.103)	.777 (.074)	.387 (.190)
Cross-lagged effects	Anger -> Map goals	Map goals -> anger	Anger -> Mav goals	Mav goals -> anger
T1 -> T2	-.048 (.091)	.059 (.076)	.015 (.096)	-.108 (.064)
T2 -> T3	-.036 (.128)	-.139 (.120)	-.296 (.131)	-.008 (.100)
T3 -> T4	-.068 (.077)	.003 (.116)	-.056 (.074)	.018 (.092)
T4 -> T5	.097 (.059)	.149 (.131)	-.085 (.088)	.100 (.105)
T5 -> T6	-.082 (.063)	.189 (.104)	-.017 (.053)	.120 (.093)
T6 -> T7	.071 (.073)	.045 (.125)	-.002 (.064)	-.264 (.144)
T7-> T8	.021 (.067)	.205 (.171)	.217 (.068)	.243 (.145)
T8 -> T9	.092 (.069)	-.061 (.136)	.027 (.048)	.144 (.128)
T9 -> T10	.050 (.068)	.085 (.183)	-.003 (.114)	.022 (.149)
T10 -> T11	.155 (.074)	.143 (.202)	.032 (.064)	1.161 (.287)
T11 -> T12	.080 (.060)	.289 (.111)	.090 (.057)	.250 (.136)
T12 -> T13	.025(.064)	.031 (.118)	-.088 (.069)	.334 (.217)

	Anger and Pap goals model		Anger and Pav goals model	
Mean differences	Anger	Pap goals	Anger	Pav goals
Enjoyment vs. Neutral state (a_e)	-.050 (.132)	.008 (.181)	-.052 (.129)	.105 (.153
Frustration vs. Neutral state (a_f)	.174 (.157)	.133 (.207)	.208 (.149)	.119 (.179)
Effects of covariates on random intercepts	.402 (.071)	.510 (.063)	.379 (.071)	.595 (.065)
Autoregressive effects	Anger	Pap goals	Anger	Pav goals
T1 -> T2	.074 (.136)	.510 (.061)	.206 (.115)	.363 (.088)
T2 -> T3	.275 (.197)	.472 (.087)	.243 (.188)	.119 (.124)
T3 -> T4	.585 (.098)	.613 (.070)	.541 (.101)	.114 (.158)
T4 -> T5	.532 (.096)	.489 (.075)	.502 (.106)	.278 (.135)
T5 -> T6	.610 (.076)	.107 (.130)	.608 (.089)	.417 (.109)
T6 -> T7	.607 (.083)	.087 (.129)	.584 (.092)	.473 (.109)
T7-> T8	.639 (.088)	.512 (.096)	.580 (.099)	.840 (.113)
T8 -> T9	.716 (.076)	.431 (.108)	.727 (.081)	.410 (.075)
T9 -> T10	.677 (.090)	.200 (.110)	.631 (.088)	.884 (.100)
T10 -> T11	.665 (.074)	.366 (.164)	.601 (.077)	.693 (.078)
T11 -> T12	.924 (.105)	.185 (.149)	.660 (.082)	.658 (.071)
T12 -> T13	.835 (.065)	.331 (.109)	.782 (.071)	.838 (.056)
Cross-lagged effects	Anger -> Pap goals	Pap goals -> anger	Anger -> Pav goals	Pav goals -> anger
T1 -> T2	-.020 (.128)	-.148 (.069)	.123 (.095)	-.099 (.089)
T2 -> T3	-.120 (.149)	.032 (.108)	-.081 (.112)	.001 (.163)
T3 -> T4	-.035 (.075)	-.003 (.093)	-.135 (.081)	-.101 (.174)
T4 -> T5	.029 (.061)	.059 (.104)	-.230 (.075)	.034 (.185)
T5 -> T6	-.070 (.069)	.196 (.107)	-.140 (.070)	.030 (.131)
T6 -> T7	.029 (.078)	.017 (.113)	-.208 (.075)	-.031 (.121)
T7-> T8	.093 (.058)	.241 (.135)	.079 (.085)	-.103 (.119)
T8 -> T9	.043 (.062)	.036 (.130)	-.155 (.070)	.024 (.080)
T9 -> T10	.012 (.067)	-.039 (.161)	-.062 (.072)	-.136 (.119)
T10 -> T11	.059 (.057)	.437 (.193)	-.085 (.068)	-.184 (.081)
T11 -> T12	.056 (.071)	.183 (.145)	-.085 (.067)	.078 (.081)
T12 -> T13	.035 (.061)	.288 (.112)	-.081 (.054)	-.118 (.072)

	Boredom and mental effort model			
Mean differences	Boredom	Mental effort	Posttest Perf.	Game Perf.
Enjoyment vs. Neutral state (a_e)	-.412 (.174)	-.155 (.337)	-.858 (.585)	1.124 (.624)
Frustration vs. Neutral state (a_f)	-.049 (.205)	.466 (.386)	-.469 (.704)	-.073 (.717)
Effects of covariates on random intercepts	.472 (.084)	-.205 (.057)	.676 (.099)	.429 (.105)
Autoregressive effects	Boredom	Mental effort		
T1 -> T2	.482 (.092)	.661 (.059)		
T2 -> T3	.508 (.083)	.543 (.086)		
T3 -> T4	.389 (.089)	.439 (.079)		
T4 -> T5	.502 (.091)	.387 (.112)		
T5 -> T6	.548 (.117)	.302 (.104)		
T6 -> T7	.322 (.100)	.278 (.109)		
T7-> T8	.518 (.116)	.362 (.125)		
T8 -> T9	.642 (.109)	.471 (.081)		
T9 -> T10	.514 (.088)	.333 (.118)		
T10 -> T11	.567 (.101)	.224 (.115)		
T11 -> T12	.655 (.085)	.195 (.157)		
T12 -> T13	.649 (.075)	.223 (.104)		
Cross-lagged effects	Boredom -> mental effort	Mental effort -> boredom		
T1 -> T2	-.138 (.099)	-.023 (.051)		
T2 -> T3	-.068 (.123)	-.081 (.055)		
T3 -> T4	-.082 (.114)	-.007 (.058)		
T4 -> T5	-.247 (.149)	-.054 (.061)		
T5 -> T6	.152 (.163)	-.050 (.068)		
T6 -> T7	.015 (.140)	-.009 (.068)		
T7-> T8	.415 (.196)	.039 (.070)		
T8 -> T9	.340 (.131)	-.021 (.064)		
T9 -> T10	.186 (.141)	.139 (.069)		
T10 -> T11	.054 (.151)	.028 (.069)		
T11 -> T12	.037 (.172)	-.114 (.070)		
T12 -> T13	-.093 (.146)	.017 (.049)		

	Boredom and Map goals model		Boredom and Mav goals model	
Mean differences	Boredom	Map goals	Boredom	Mav goals
Enjoyment vs. Neutral state (a_e)	-.408 (.174)	.295 (.183)	-.418 (.179)	.281 (.185)
Frustration vs. Neutral state (a_f)	-.040 (.199)	-.003 (.212)	.010 (.211)	-.037 (.212)
Effects of covariates on random intercepts	.418 (.072)	.545 (.079)	.430 (.087)	.383 (.076)
Autoregressive effects	Boredom	Map goals	Boredom	Mav goals
T1 -> T2	.211 (.144)	.483 (.081)	.417 (.104)	.593 (.055)
T2 -> T3	.523 (.097)	.252 (.092)	.500 (.091)	.527 (.071)
T3 -> T4	.402 (.097)	.429 (.091)	.373 (.089)	.462 (.076)
T4 -> T5	.526 (.105)	.601 (.082)	.465 (.100)	.409 (.088)
T5 -> T6	.658 (.096)	.249 (.091)	.597 (.104)	.194 (.075)
T6 -> T7	.479 (.089)	-.101 (.104)	.369 (.096)	.397 (.100)
T7-> T8	.601 (.095)	.387 (.133)	.527 (.112)	-.034 (.106)
T8 -> T9	.739 (.091)	.195 (.115)	.663 (.111)	.110 (.185)
T9 -> T10	.656 (.076)	.200 (.121)	.584 (.079)	-.154 (.128)
T10 -> T11	.657 (.084)	.475 (.150)	.596 (.093)	.229 (.130)
T11 -> T12	.685 (.079)	.494 (.080)	.645 (.084)	.284 (.069)
T12 -> T13	.655 (.069)	.551 (.098)	.651 (.075)	.513 (.167)
Cross-lagged effects	Boredom -> Map goals	Map goals -> boredom	Boredom -> Mav goals	Mav goals -> boredom
T1 -> T2	-.049 (.086)	-.350 (.129)	-.039 (.068)	-.170 (.083)
T2 -> T3	-.181 (.073)	.071 (.118)	-.075 (.068)	-.054 (.091)
T3 -> T4	-.169 (.065)	-.024 (.123)	-.145 (.072)	-.081 (.099)
T4 -> T5	.047 (.068)	-.203 (.113)	-.046 (.090)	-.177 (.100)
T5 -> T6	.047 (.070)	-.186 (.117)	-.058 (.072)	-.105 (.102)
T6 -> T7	-.039 (.060)	.078 (.140)	.083 (.058)	.137 (.140)
T7-> T8	.025 (.063)	.234 (.191)	.052 (.074)	.180 (.148)
T8 -> T9	.012 (.066)	-.337 (.151)	.079 (.084)	-.037 (.224)
T9 -> T10	-.115 (.052)	-.004 (.165)	-.067 (.067)	.262 (.139)
T10 -> T11	-.019 (.067)	.111 (.177)	.006 (.078)	-.005 (.156)
T11 -> T12	.059 (.055)	-.080 (.110)	.058 (.047)	.042 (.120)
T12 -> T13	-.139 (.059)	-.008 (.112)	-.081 (.071)	-.201 (.157)

	Boredom and Pap goals model		Boredom and Pav goals model	
Mean differences	Boredom	Pap goals	Boredom	Pav goals
Enjoyment vs. Neutral state (a_e)	-.426 (.167)	.077 (.179)	-.344 (.164)	.056 (.150)
Frustration vs. Neutral state (a_f)	-.045 (.196)	.201 (.208)	.064 (.185)	.068 (.176)
Effects of covariates on random intercepts	.453 (.075)	.523 (.064)	.486 (.074)	.578 (.061)
Autoregressive effects	Boredom	Pap goals	Boredom	Pav goals
T1 -> T2	.416 (.114)	.511 (.061)	.422 (.105)	.381 (.093)
T2 -> T3	.497 (.088)	.451 (.079)	.437 (.107)	.093 (.160)
T3 -> T4	.345 (.093)	.573 (.068)	.481 (.096)	.204 (.174)
T4 -> T5	.524 (.101)	.439 (.078)	.576 (.080)	.454 (.118)
T5 -> T6	.644 (.099)	-.004 (.117)	.706 (.088)	.474 (.085)
T6 -> T7	.429 (.091)	-.079 (.116)	.437 (.091)	.496 (.110)
T7-> T8	.598 (.100)	.409 (.110)	.506 (.107)	.701 (.130)
T8 -> T9	.698 (.091)	.406 (.117)	.604 (.095)	.415 (.084)
T9 -> T10	.659 (.083)	.198 (.118)	.548 (.086)	.761 (.121)
T10 -> T11	.645 (.092)	.539 (.133)	.610 (.099)	.616 (.092)
T11 -> T12	.635 (.077)	.358 (.108)	.647 (.085)	.673 (.076)
T12 -> T13	.662 (.071)	.429 (.091)	.600 (.069)	.802 (.061)
Cross-lagged effects	Boredom -> Pap goals	Pap goals -> boredom	Boredom -> Pav goals	Pav goals -> boredom
T1 -> T2	.023 (.081)	-.029 (.084)	.105 (.064)	.151 (.127)
T2 -> T3	-.122 (.073)	.185 (.091)	.130 (.080)	.458 (.175)
T3 -> T4	-.111 (.065)	.139 (.095)	.060 (.076)	-.089 (.199)
T4 -> T5	.014 (.074)	-.094 (.098)	-.027 (.074)	-.121 (.135)
T5 -> T6	.078 (.086)	-.200 (.121)	-.140 (.065)	-.100 (.116)
T6 -> T7	-.008 (.070)	.095 (.133)	-.091 (.065)	-.131 (.134)
T7-> T8	.046 (.070)	.101 (.147)	-.173 (.094)	-.262 (.133)
T8 -> T9	-.049 (.064)	-.284 (.157)	-.106 (.076)	-.286 (.097)
T9 -> T10	-.094 (.059)	.209 (.151)	-.201 (.077)	-.124 (.126)
T10 -> T11	-.062 (.066)	-.113 (.167)	-.141 (.088)	-.160 (.095)
T11 -> T12	.066 (.066)	-.340 (.113)	-.016 (.071)	-.143 (.084)
T12 -> T13	-.175 (.060)	-.116 (.103)	-.153 (.053)	-.231 (.079)

Note. posterior standard deviation (SD) in parentheses; Map = mastery-approach; Pap = performance-approach; Mav = mastery-avoidance; Pav = performance-avoidance; Perf. = performance.

Discussion

Emotional contagion has primarily been studied in social psychology, rather than educational psychology. Given the frequent interactions between teachers and students and between students and peers in learning environments, not only teachers' emotions but also peers' emotions may influence students' learning. This research aims to study learning environments that may enhance learning by focusing on emotional contagion among students. To our knowledge, this investigation is the first experimental study to test how peers' achievement emotions affect students' learning and to analyze the mediating effects of students' achievement emotions in GBL. Furthermore, this study is one of the first to decompose within- and between-person relations using an experimental design with intensive longitudinal data (Hamaker et al., 2021).

RQ: Effects of Peers' Achievement Emotions on Students' Learning

We found that peers' achievement emotions affected students' achievement emotions, mastery-based goals, mental effort, and performance but not performance-based goals.

Achievement Emotions

Overall, students exposed to peers' enjoyment report higher positive achievement emotions and lower negative achievement emotions than those exposed to peers' frustration. This is consistent with our hypothesis and the sparse research on emotional contagion from teachers to students, which found that students exposed to happy and content instructors reported higher positive emotions (i.e., happy and content) and lower negative emotions (i.e., boredom and frustration) than those exposed to bored and frustrated instructors (e.g., Horovitz & Mayer, 2021; Lawson & Mayer, 2022). These results imply that achievement emotions are contagious from peers to students, but that the strength and degree of influence depend on the type of emotions.

Achievement Goals

Overall, peers' achievement emotions affect mastery-approach and mastery-avoidance goals but, as we expected, not performance-based goals. First, also consistent with our hypotheses, students exposed to peers' enjoyment reported higher mastery-approach goals than those exposed to peers' frustration, and students from all three groups reported similar performance-approach goals and performance-avoidance goals. Second, contrary to our hypotheses, students exposed to peers' enjoyment reported higher mastery-avoidance goals than those exposed to peers' neutral state and those exposed to peers' frustration.

These results may be interpreted by considering the nature of mastery-based goals. Mastery-based goals are a combination of two components – how competence is defined (definition component) and how competence is valenced (valence component; Elliot & Hulleman, 2017). Regarding the definition component, mastery-based goals are defined in terms of the task itself (task-based) or one's own personal trajectory (self-based). Regarding the valence component (approach/avoidance or positive/negative distinction), mastery-approach goals have a positive valence and mastery-avoidance goals have a negative valence. From our results, the definition component of mastery-based goals appears to be responsive to viewing the emotions of others, while the valence component is not. This may be because the definition component of mastery-based goals is so strongly linked to peers' achievement emotions that it overwhelms the valence component (approach/avoidance distinction).

Mental Effort and Performance

Overall, students exposed to peers' enjoyment showed higher game performance but lower mental effort and posttest performance than those exposed to peers' frustration. First, contrary to our hypothesis, students exposed to peers' enjoyment reported lower mental effort than those exposed to peers' frustration. This differs from research on emotion induction, which found that students with induced positive emotions reported equivalent mental effort to those with induced negative emotions (Knörzer et al., 2016). Second, contrary to our hypothesis, students exposed to peers' enjoyment showed lower posttest performance than those exposed to peers' frustration. This result differs from research on emotional contagious from teachers to students, which found that students exposed to happy and content instructors showed equivalent immediate and delayed posttest performance than those exposed to bored and frustrated instructors (e.g., Horovitz & Mayer, 2021; Lawson & Mayer, 2022), and research on emotional design of learning materials, which found that positive emotional designs and negative emotional designs increased posttest performance relative to neutral designs (e.g., Kumar, 2019; Stark et al., 2018). However, this result is consistent with research on emotion induction, which found that students with induced positive emotions showed lower posttest performance than those with induced negative emotions (Knörzer et al., 2016). Third, consistent with our hypothesis and the attention and motivation assumption, students exposed to peers' enjoyment showed higher game performance than those exposed to peers' frustration.

These results seem to indicate that compared with the frustration group, the enjoyment group was less triggered to actively process the information presented in the videos, particularly integrating this information into the long-term memory. This had no consequences for game performance because this information was still available during gameplay (making it unnecessary to process and integrate this information into long-term memory) and was sufficient to answer the questions of game performance. However, it had consequences for posttest performance because this information was absent in the posttest, and learners had to rely on long-term memory to answer the questions in the posttest. The less active processing may also explain the lower mental effort of the enjoyment group. However, this raises the question: Why were there less active processing in the enjoyment group than in the frustration group?

A possible explanation is that emotions influence the scope of attention: Compared with negative emotions, positive emotions can broaden the scope of attention, thereby causing a reduction of attention to the original (learning) task and a decrease in the perceived need to actively process new information (Fredrickson, 2013). Another possible explanation is that emotions can influence the interpretation of a situation, such as the performance of a task: Compared with negative emotions, positive emotions can lead to the assumption that the task is not difficult, which reduces the perceived need to actively process new information (Bruder et al., 2014; Raaijmakers et al., 2017; Salomon, 1984). Finally, emotions signify whether task performance is going well or poorly: Compared with negative emotions, positive emotions are more likely to indicate that task performance is going well (Carver, 2003), suggesting that attention can be shifted to things other than the task, and that active processing is not necessary.

As a whole, students exposed to peers' enjoyment showed higher positive emotions, game performance, and achievement motivation and lower negative emotions, mental effort, and posttest performance than those exposed to peers' frustration. As in previous research (e.g., Hu et al., 2022a), this may imply that learners often feel enjoyed and motivated by the learning environment from which they learn least (Clark, 1982; Feldon et al., 2019; Graesser, 2017; Schrader et al., 2021; Salomon, 1984).

Null Results

Contrary to all our hypotheses on the peers' neutral state group, students exposed to peers' neutral state reported no differences in relaxation, anger, and mental effort than those exposed to peers' enjoyment and no differences in frustration, mastery-based goals, and performance than those exposed to peers' frustration. Previous research has shown that the evidence regarding the effects of neutral state group is inconsistent. A meta-analysis found that compared with neutral designs, positive emotional designs increased mental effort (Wong & Adesope, 2020), but another meta-analysis did not (Brom et al., 2018). Additionally, both meta-analyses found that, compared with neutral designs, positive emotional designs increased positive emotions, intrinsic motivation, and posttest performance in multimedia learning, but effect sizes vary across studies. These inconsistent results suggest that the effect of a neutral design on learning requires further research.

Within-Person Correlations, Between-Person Correlations, Autoregressive Effects, and Cross-Lagged Effects

First, overall, the within-person correlations and between-person correlations between emotions and mastery-approach goals, mastery-avoidance goals, performance-approach goals, performance-avoidance goals, or mental effort are consistent: Positive for positive emotions and negative for negative emotions. However, there are two exceptions for the between-person correlations: Anger and boredom positively correlate to performance-avoidance goals; and mental effort positively correlates to frustration and negatively correlates to enjoyment. This result support our usage of RI-CLPMs to decompose within- and between-person relations.

Second, most of students' achievement emotions, achievement goals, and mental effort have substantial positive within-person autoregressive effects over time. The positive effects suggest that there are positive carry-over effects (i.e., inertia) from game level to game level, indicating that most of the deviations in students' achievement emotions, achievement goals, and mental effort from the individual person average were positively related to deviations in the same variables at the next game level. This result implies that students' achievement emotions, achievement goals, and mental effort tend to be stable over time before returning to the person average.

Third, reciprocal effects can be estimated at the within level and the between level. Reciprocal effects at the between level are indicated by the random intercepts. Reciprocal effects at the within level are indicated by the cross-lagged effects. Overall, there were mostly positive cross-lagged effects between students' positive emotions and achievement goals, whereas there were mostly negative cross-lagged effects between students' negative emotions and achievement goals. There were mostly positive cross-lagged effects between most of students' emotions and mental effort. This implies that students' emotions and achievement goals or mental effort are linked by reciprocal effects over time at the within level.

Limitations

We acknowledge several limitations of the present research. First, many of the participants from the peers' frustration and neutral state groups were excluded due to failing the manipulation check. Even though the male peer model was carefully and extensively trained, a noteworthy number of participants were not able to distinguish between his frustration and a neutral state, relative to his enjoyment. This is consistent with research on emotional recognition: Recognizing positive emotions appears to be easier than recognizing neutral states and negative emotions. For example, people tend to perceive neutral faces as negative rather than positive (Rollins et al., 2021), and it is difficult to distinguish between expressionless faces that display neutral emotions and frowning faces that display negative emotions (Park et al., 2015). To distinguish more clearly between frustration and neutral state, future research could, for example, include verbal cues, such as sentences about emotions (e.g., "I am so frustrated").

Second, because we had to exclude many participants, we ended up with a small sample that necessitated us to use bivariate models when analyzing the data. We propose to replicate the study with a larger sample size. With a larger sample size, we would be able to model more variables in one model (e.g., emotions, achievement goals, mental effort, and performance) to accommodate the interactions between them.

Third, individual differences (e.g., in the tendency to "catch" others' emotions) and social factors (e.g., the relationship between the expresser and the observer including perceived likability, desire for affiliation, superior and inferior power differences, ingroup versus outgroup membership, and competitive versus cooperative relations; Fischer & Hess, 2017) may play a role in emotional contagion in the classroom. For example, if observers have a high tendency to be influenced by others' emotions, to like the expresser, or to perceive the expresser as an ingroup member or as cooperative, emotional contagion may be more likely to happen. In the present study, the peer was similar in age and education level to the participants, and they encountered the game together. Thus, students may have felt similar power to the peer and may have perceived him as an ingroup member and as cooperative. Future research would do well to vary the match between the model and the perceivers and, more generally, to take individual difference and social factors into account.

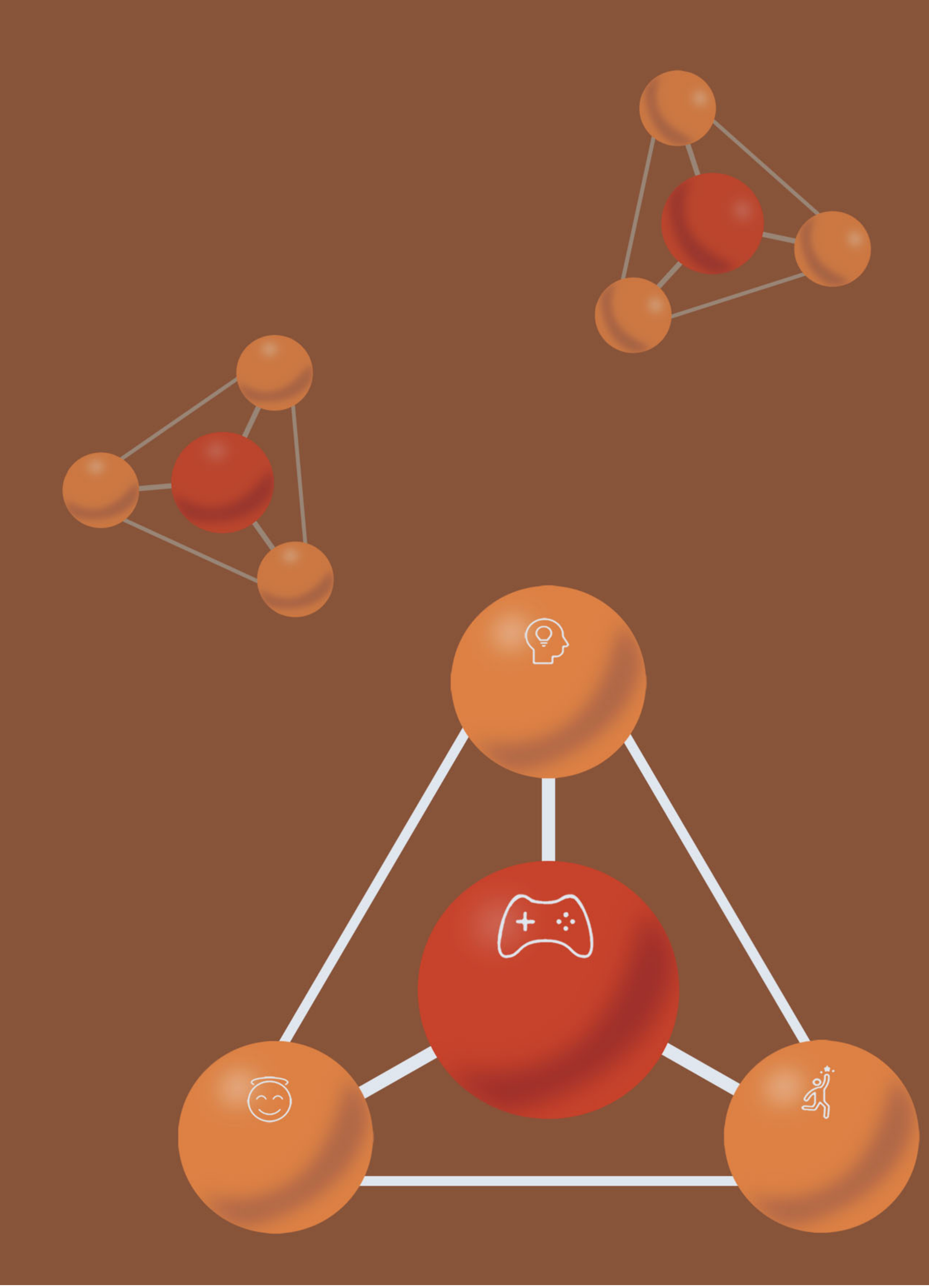
Implications

The present study has several implications for practice and theory. First, our findings suggest that educators would do well to attend to the emotions of their students, considering that collaboration and communication become more prevalent in education (e.g., as some of the interpersonal 21st century competencies; NRC, 2012) and thus the emotions involved in these processes. Second, our findings suggest that instructional design features that primarily target emotions can also affect motivation and cognition. Considering that researchers often focus on cognitive, motivational, or emotional theories (Hu et al., 2021), researchers may consider focusing on their interactions and revealing the underlying mechanisms, such as how emotions can best be used to indirectly stimulate motivation/cognition or, conversely, how to recognize emotions that can thwart motivation/cognition and mitigate their influence. Third, our findings suggest that higher motivation and positive emotions may not always associate with higher cognition, although educators and researchers often aim to design learning environments that are engaging, enjoyable, and effective.

Fourth, our findings extend recent work on emotional design in multimedia learning, from emotional design of learning materials, emotion induction, and emotional contagion from teachers to students to emotional contagion among students. Fifth, our findings also extend work on social contagion in education, from contagion of intrinsic and extrinsic motivation (Friedman et al., 2010) or goals (e.g., Aarts et al., 2004) to emotions. Social contagion theory proposes that most psychological states (e.g., emotions, motivation, behaviors, values, norms) can be spread among others (Levy & Nail, 1993). In our view, it remains open whether contagion of other psychological states also occurs in education (e.g., effortful or disruptive behavior; intrinsic or extrinsic values).

Conclusion

The present study is one of the first to manipulate students' achievement emotions with a focus on emotional, motivational, and cognitive processes and outcomes. We conclude that peers' achievement emotions may differentially affect students' emotional, motivational, and cognitive processes and outcomes in GBL. In general, the results of this study advance the field of emotional contagion, achievement emotions, achievement goals, instructional design, and, particularly, their interconnection.



Chapter 6 General discussion

Game-based learning (GBL) uses a game as the medium for learning. Well-designed GBL should and can promote cognitive processes, motivation to learn, and positive emotions, all of which contribute to learning. However, from a societal perspective, it is still unclear whether teachers and students should use GBL, and which design features improve GBL. From a scientific perspective, evidence for the effects of GBL in comparison to non-GBL was inconsistent and research on instructional design features that improve GBL were limited. GBL has the potential to address the challenges in a single subject, such as chemistry. For example, the CHARMING project takes on this challenge by developing games for learning chemistry and chemical engineering. Most GBL research focused mainly on cognition, less on motivation, and rarely on emotion, let alone interconnections between them. Therefore, the focus of this thesis was:

1. Is the effect of GBL in chemistry education on cognition, motivation, and emotion (i.e., cognitive, motivational, and emotional processes and outcomes) larger than for non-GBL? (Media comparison research).
2. Which design features improve the effectiveness of GBL and how? (Value-added research).

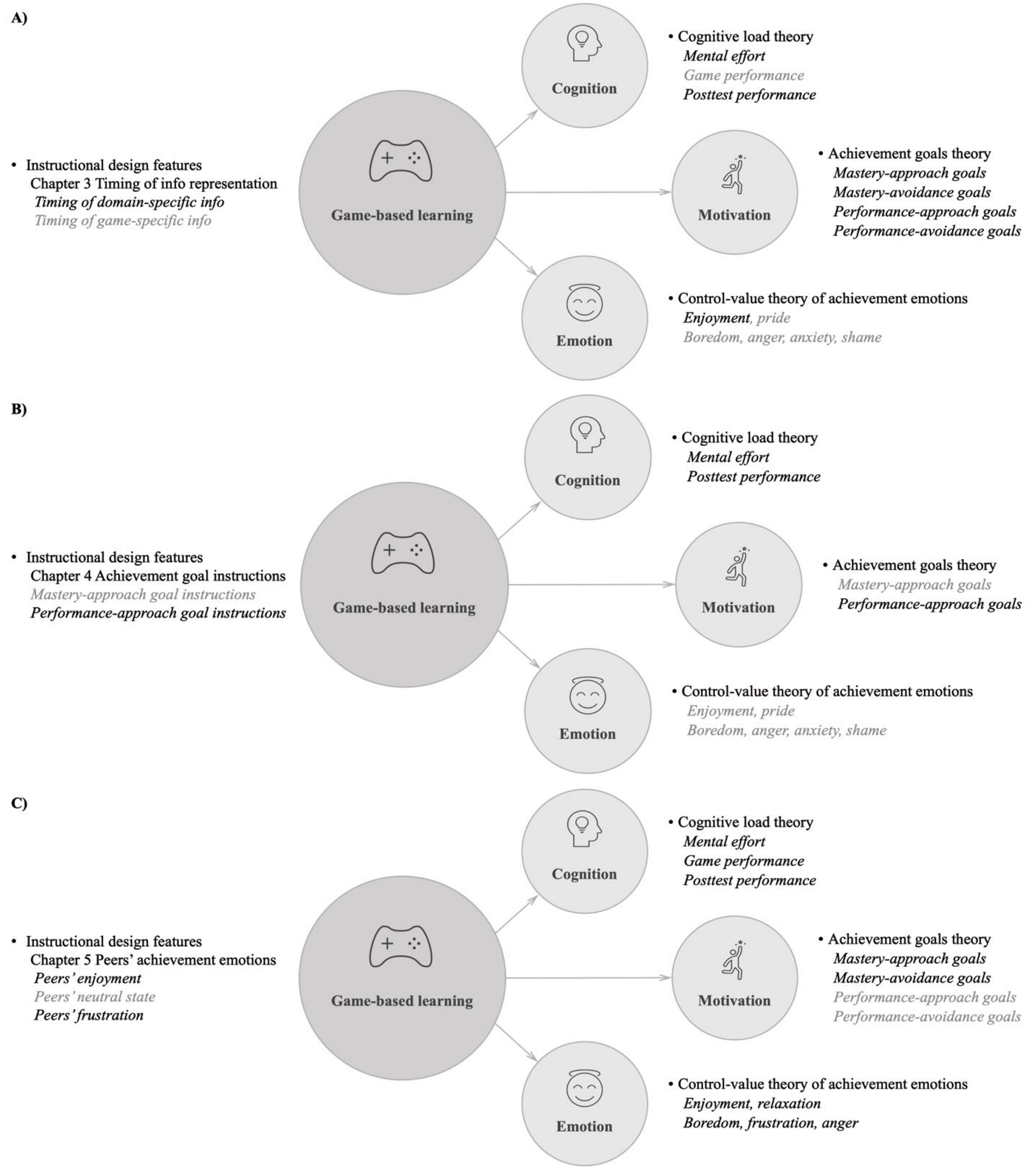
This thesis systematically reviewed empirical studies on the effects of GBL in comparison to non-GBL (Chapter 2) and investigated not only the effects of three instructional design features that manipulated students' cognition (i.e., timing of information presentation; Chapter 3), motivation (i.e., achievement goals; Chapter 4), and emotion (i.e., peers' achievement emotions; Chapter 5) on three learning processes and outcomes, namely, cognition (i.e., mental effort and performance), motivation (i.e., achievement goals), and emotion (i.e., achievement emotions), but also their interconnection in GBL in chemistry education, as displayed in Figure 6.1. Moreover, this thesis suggested how cognition, motivation and emotion could be optimized in GBL.

Review of the Results

The research questions in Chapter 2 were: Is the effect of GBL in chemistry education on cognition (including retention), motivation, and emotion larger than for non-GBL (media comparison)? Do instruction characteristics (activity level of control group, additional instruction, user grouping, and number of game sessions) and methodology characteristics (randomization, sample size, publication source, and assessment type) moderate the effect? And which game design or instructional design features improve GBL in chemistry education (value-added research)? We systematically reviewed 34 studies on GBL in chemistry education from 2006-2022. We found: 1) GBL in chemistry education was more effective not only for cognition and retention but also for motivation than non-GBL, but there was significant heterogeneity in effect sizes between studies and studies on emotions were absent; 2) sample size and publication sources moderated the effects; and 3) studies on value-added research were too few to conduct a proper meta-analysis.

Figure 6.1

Summary of the findings on main effects of instructional design features in Chapter 3 (A), 4 (B), and 5 (C)



Note. Single-headed arrows represent causal relations; Words in grey means no main effects.

Chapter 3 focused on an instructional design feature that aims to optimize cognitive load: timing of (domain-specific and game-specific) information presentation. The research questions were: How does timing of (domain-specific and game-specific) information presentation affect cognition (i.e., mental effort, posttest performance, and game performance), motivation (i.e., achievement goals), and emotion (i.e., achievement emotions) in GBL in chemistry education? and Do mental effort, perceived competence, perceived control, and perceived value mediate the effects? We found that presenting domain-specific information before gameplay promoted lower posttest performance and higher mastery-/performance-approach goals, mastery-/performance-avoidance goals, and positive achievement emotions than presenting it during gameplay. There was no difference on mental effort. There was no difference between presenting game-specific information before gameplay and during gameplay except for performance-avoidance goals. Mental effort, perceived competence, perceived control, and/or perceived value did not mediate the effects of timing of information presentation on performance, achievement goals, and achievement emotions.

Chapter 4 focused on an instructional design feature that aims to increase approach goals: achievement goal instructions (mastery-approach goals and performance-approach goals). The research questions were: How do achievement (mastery-approach and performance-approach) goal instructions affect motivation (i.e., achievement goals), cognition (i.e., mental effort and posttest performance), and emotion (i.e., achievement emotions) in GBL in chemistry education? Do prior achievement goals moderate the effects? And do achievement goals mediate the effects? We found that mastery-approach goal instructions promoted higher mental effort but that it had no effects on mastery-approach goals and posttest performance. Performance-approach goal instructions promoted higher performance-approach goals and mental effort but lower posttest performance. Only prior mastery-approach goals moderated the effects of achievement goal instructions on mental effort. There were no mediating effects of achievement goals on mental effort, performance, and achievement emotions.

Chapter 5 focused on an instructional design feature that aims to induce positive achievement emotions: emotional contagion of peers' achievement emotions (enjoyment, frustration, and neutral state). The research questions were: How do peers' achievement emotions (enjoyment/frustration/neutral state) affect students' emotion (i.e., achievement emotions), motivation (i.e., achievement goals), and cognition (i.e., mental effort, posttest performance, and game performance) in GBL in chemistry education? And do students' achievement emotions mediate the effects? We found that students exposed to peers' enjoyment reported higher enjoyment, relaxation, mastery-approach goals, mastery-avoidance goals, and game performance and lower frustration, anger, boredom, mental effort, and posttest performance than those exposed to peers' frustration. There were no mediating effects of students' achievement emotions except that students' boredom mediated the effects of peers' achievement emotions on mastery-approach goals and mastery-avoidance goals.

Theoretical Implications

First, the findings of this thesis have implications for theories of learning and instruction design. We found that instructional design features that manipulated cognition, motivation, or emotions could all affect cognition, motivation, and emotion in GBL in secondary and higher chemistry education. The studies of Chapter 3, 4, and 5 were one of the first that attend to cognition (i.e., mental effort and performance), motivation (i.e., achievement goals), and emotion (i.e., achievement emotions) simultaneously. An instructional design feature that manipulated cognition affected not only posttest performance but also achievement goals and achievement emotions (Chapter 3). An instructional design feature that manipulated achievement goals affected not only achievement goals but also mental effort and posttest performance (Chapter 4). An instructional design feature that manipulated achievement emotions affected not only achievement emotions but also mental effort, posttest performance, and game performance (Chapter 5). The results suggest that cognition, motivation, and emotion may interact with each other. The results of this thesis emphasize the need to integrate cognition, motivation, and emotion in theories of learning and instructional design.

For example, the findings of this thesis have implications for multimedia learning. Most GBL is a type of multimedia learning because it involves words and pictures (Mayer, 2014). The dominant theory of multimedia learning is the cognitive theory of multimedia learning (CTML, Mayer, 2021). There are some attempts to integrate motivation and emotion in CTML, such as the integrated model of multimedia learning and motivation (IMMLM, Astleitner & Wiesner, 2004), the cognitive-affective theory of learning with media (CATLM; Moreno & Mayer, 2007), the integrated cognitive affective model of learning with multimedia (ICALM; Plass & Kaplan, 2015), and the cognitive-affective-social theory on (digital) learning environments (CASTLE; Schneider et al., 2022). However, most multimedia principles in these extended multimedia theories are based on cognitive foundations and the role of motivation and emotion in multimedia learning is still lacking a solid theoretical and empirical foundation (c.f. Schrader et al., 2021). This thesis used cognitive load theory, achievement goal theory, and control-value theory of achievement emotions as an example of theoretical and empirical foundations and aimed to connect cognitive load, achievement goals, and achievement emotions. The results contribute to motivational and emotional design principles of multimedia learning.

Second, the findings of this thesis have implications for the relationship between cognition, motivation, and emotion. We found that high approach and avoidance goals and more positive achievement emotions were associated with low posttest performance. The studies in Chapter 3, 4, and 5 indicated that instructional design features affected cognition, motivation, and emotion differently. Presenting domain-specific information before gameplay promoted higher approach and avoidance goals, and positive achievement emotions, but lower posttest performance than presenting it during gameplay (Chapter 3). A performance-approach goal instruction promoted higher performance-approach goals but lower posttest performance (Chapter 4). Students exposed to peers' enjoyment reported higher positive emotions, mastery-based goals, and lower negative emotions but lower posttest performance than those exposed to peers' frustration (Chapter 5). This result may contradict the common desire to create effective, engaging, and enjoyable learning environments.

For example, the results contradict the argument that higher motivation and/or positive emotions lead to more learning, which has been challenged forty years ago (Clark, 1982) and more recently (Feldon et al., 2021), but has not been researched systematically yet. This contradiction may be because of the study design: High motivation and more positive emotions may be associated with high performance in non-experimental studies and long-term learning settings (e.g., over one semester), but with low performance in experimental studies and short-term learning settings (e.g., over one hour; Hu et al., 2022a). In most meta-analyses reporting positive relations between motivation and performance or between positive emotions and performance (e.g., Camacho-Morles et al., 2021; Lazowski & Hulleman, 2016), the included studies are mostly non-experimental and few experimental studies. Non-experimental and experimental studies differ in some aspects, such as duration or the settings. In terms of duration, for example, cognitive load theory (Sweller et al., 2019) suggests that performance is determined by working memory constraints only when time is limited. For the same task, with the same high motivation and positive emotions, the shorter the duration, the higher the cognitive load, and the lower the performance. Conversely, in the long run, high motivation and positive emotions may stimulate learners to invest more time and effort in learning and thus improve performance. If this is the case, then the interaction between cognition, motivation, and emotions needs to be systematically studied in both experimental and non-experimental studies.

Practical Implications

First for educators and learners, the finding that GBL in chemistry education is more effective for cognition and retention but also motivates more than non-GBL suggests that educators can implement GBL in chemistry education. The finding that instructional design features affect cognition, motivation, and emotion in GBL differently suggests educators and learners to use GBL differently depending on the learning goals: Should performance, motivation and/or emotion be fostered? Specifically, timing of information presentation may affect learners' cognition, motivation, and emotion in GBL differently (Chapter 3). The finding that learners who received domain-specific information during gameplay reported higher posttest performance than those who received domain-specific information before gameplay suggests that if the learning goal is to increase performance, domain-specific information should be presented during gameplay. Mastery-approach goal instructions and performance-approach goal instructions may affect cognition and motivation in GBL differently (Chapter 4). The finding that learners who received mastery-approach goal instructions reported higher mental effort but equivalent posttest performance than those who received no mastery-approach goal instructions and learners who received performance-approach goal instructions reported higher mental effort but lower posttest performance than those who received no performance-approach goal instructions suggests that if the learning goal is to increase performance, educators should not use performance-approach goal instructions and learners should not adopt performance-approach goals. Peers' achievement emotions may affect learners' emotion, motivation, and cognition in GBL differently (Chapter 5). The finding that learners exposed to peers' frustration reported higher mental effort and posttest performance than those exposed to peers' enjoyment suggests that if the learning goal is to increase performance, educators should allow and monitor (because it is unclear how the peers' frustration affects peers' performance) peers' frustration and learners should pay attention to peers' frustration.

Second for researchers, the finding of small-study effects (especially publication bias) suggests that researchers would do well to conduct sample size planning before data collection (e.g., what is the minimum number of participants required). The overall effect size of the meta-analysis provides researchers a benchmark of the interventions on GBL in chemistry education, which help them form new hypotheses. For example, we found the overall effect size for cognition is medium (Cohen's $d = .6$) so the researchers can expect at least a medium effect size of comparing GBL in chemistry education vs. non-GBL in the future.

Third, for studies in Chapter 3, 4, and 5, the experimental interventions could not be implemented in the game itself but had to take place outside the game. For example, when testing the instructional design feature in the study of Chapter 3, we had to put domain-specific information on a separate website outside of the game. This shows the relevance of adaptability of games in terms of being able to make simple adjustments by its users. Otherwise, introducing educational technology, such as games, can leave teachers feel like that they have no say about the learning processes of their students if they cannot change anything in the applications.

Limitations and Future Research

This thesis acknowledges several limitations and offers accompanying suggestions for future research. First, the studies in Chapter 3, 4, and 5 used mental effort as the indicator of overall cognitive load. For this indicator, in relation to performance, we found that timing of information presentation affected posttest performance but has no effect on mental effort. Next, we showed that performance-approach goal instructions promoted higher mental effort but lower posttest performance than no performance-approach goal instructions. Further we found that students exposed to peers' enjoyment reported higher game performance and lower mental effort and posttest performance than those exposed to peers' frustration. Taken together, the results on the relationship between mental effort and performance appear to be inconsistent. These results may contradict to our expectation that higher mental effort is associated with higher performance. This may be mainly because the source of mental effort is unclear. Under some circumstances, higher mental effort may come from higher extraneous load, thereby decreasing performance, whereas under other circumstances, higher mental effort may come from higher intrinsic load, thereby increasing performance. Future research could confirm these explanations by mapping two types of cognitive load with the instructional design features and measuring two types of cognitive load separately (e.g., via the cognitive load questionnaire; Krieglstein et al., 2023). For example, mastery-approach goal instructions may be more associated with intrinsic load and performance-approach goal instructions with extraneous load.

Second, the studies in Chapter 4 and 5 found that the instructional design features affected either mastery-based goals or performance-based goals but not both. Specifically, achievement goal instructions affect performance-approach goals but not mastery-approach goals (Chapter 4). Peers' achievement emotions affect mastery-based goals but not performance-based goals (Chapter 5). It is possible that mastery-approach goals are less relevant than performance-approach goals in our age group (e.g., secondary school students), that performance-based goals are less relevant than mastery-based goals in our study setting (e.g., activity emotions), or that the game used in our studies lacks game features (e.g., leaderboards) that could stimulate players to pursue performance-based goals. Future research on achievement goal interventions could replicate the studies using different type of achievement emotions (e.g., outcome emotions), age groups (e.g., primary school students), and games (e.g., a game with leaderboards).

Third, the studies in Chapter 3 and 5 found that instructional design features had similar effects on approach goals and avoidance goals. Specifically, compared with presenting domain-specific information during learning, presenting domain-specific information before learning yielded higher not only mastery-approach goals and performance-approach goals but also higher mastery-avoidance goals and performance-avoidance goals (Chapter 3). Compared with peers' frustration, peers' enjoyment promoted higher students' mastery-approach goals and mastery-avoidance goals (Chapter 5). These results contrast with the theoretical hypotheses that instructional design features should increase approach goals and decrease avoidance goals. These results may be interpreted by the nature of achievement goals. Achievement goals are a combination of two components – how competence is defined and how competence is valenced (Elliot & Hulleman, 2017). Regarding the definition component, mastery-based goals are defined in terms of the task itself (task-based) or one's own personal trajectory (self-based) and performance-based goals are defined in terms of others' performance (other-based). Regarding the valence component (approach/avoidance or positive/negative distinction), mastery-approach goals and performance-approach goals have a positive valence, and mastery-avoidance goals and performance-avoidance goals have a negative valence. From our results, it seems that the definition component of achievement goals is responsive to our instructional design features, while the valence component is not. This may be because that the definition component of achievement goals is so strongly linked to our instructional design features that it overwhelms the approach/avoidance distinction (the valence component). Future research is needed to confirm whether other instructional design features have the similar effects on approach and avoidance goals.

Fourth, the studies in Chapter 3 and 4 found the effects of instructional design features on some achievement emotions (enjoyment) but not others (pride, anger, anxiety, boredom, and shame). It is possible that emotions change over time during the course of a learning activity and these studies measured achievement emotions once only at the end of learning. Therefore, the study in Chapter 5 measured emotions multiple times during learning to capture the dynamics of emotions and did find the effects on all interested achievement emotions (enjoyment, relaxation, frustration, anger, and boredom). However, measurements at multiple time points during learning via self-reports may interrupt the flow of learning, distract learners' attention, and/or induce negative emotion, which may affect the accuracy of their answers (Pekrun, 2020). Future research could consider the dynamics of emotions, such as via dynamic structural equation modeling (DSEM) to analyze intensive longitudinal data (Hamaker et al., 2021), and minimize the impact of the measurement, such as via automatic facial coding to automatically code emotions based on emotional facial expressions (Höfling et al., 2020).

Fifth, compared with other types of multimedia learning, the key advantage of GBL is that it can integrate cognition, motivation, and emotion in one learning system. However, this also bring extra challenges for GBL. One example is the assessment challenge: Participants in our studies indicated that it was more challenging to solve a problem in the knowledge test with texts than solving the same problem in a game. This implies that tests and quizzes may not be well-suited to GBL because they may be not in line with the way of learning in GBL. One solution could be stealth assessment that integrates assessments into learning environments (e.g., Shute & Rahimi, 2021). With stealth assessment, the assessment is implemented during the course of gameplay itself. Another example is the analysis challenge: Game log data has the potential to reveal players' behavioral patterns and what factors impact learning (Owen & Baker, 2021), but because of practical constraints, log data were not analyzed in our studies. One solution could be educational data mining and learning analytics (Baker et al., 2016) that may support personalized GBL environments. Future research is needed to take on these challenges.

Sixth, the studies in Chapter 3, 4, and 5 used a specific game genre (i.e., CosmiClean is a computer game, puzzle game, and strategy game), specific subject (i.e., chemistry), and specific learning content (i.e., separating materials). Different game genres may offer different GBL experiences. For example, the main interaction between players and games is based on the mouse and keyboard for computer games but on the headsets and controllers for virtual reality games. Different subjects may also offer different GBL experiences. For example, chemistry learning emphasizes the central role of the multilevel thinking (Hu et al., 2021) and language learning emphasizes the central role of vocabulary knowledge (Tsai & Tsai, 2018). Besides, different type of learning content may offer different GBL experiences. For example, chemistry knowledge, such as separating materials, is more about supportive information (i.e., domain model and cognitive strategies) than procedural information (i.e., prior knowledge, rules, and procedures), based on the four-component instructional design model (4C/ID model; van Merriënboer & Kirschner, 2018). In contrast, chemistry skills are more about procedural information than supportive information. Thus, we may get different results if testing the three instructional design features in different game genres (e.g., virtual reality games), language courses (e.g., English as a secondary language), or chemistry lab skills (e.g., soap making). Future research could investigate these assumptions.

General Conclusion

This thesis gives more insight into whether learners learn more from GBL in chemistry education than non-GBL and which instructional design features improve GBL in chemistry education, particularly, how cognition, motivation and emotion can be optimized in GBL. We found that GBL enhances chemistry learning more than non-GBL, that more research is needed on emotions and instructional design features (Chapter 2), that timing of information presentation differentially affects cognition, motivation, and emotion in GBL (Chapter 3), that achievement goal instructions differentially affect motivation and cognition in GBL (Chapter 4), and that peers' achievement emotions differentially affect students' emotion, motivation, and cognition in GBL (Chapter 5). Furthermore, prior mastery-approach goals can moderate the effects of achievement goal instructions on cognition (Chapter 4). Overall, we concluded that GBL enhances chemistry learning more than non-GBL and that instructional design features differentially affect cognition, motivation, and emotion in GBL in chemistry education, that is, learners feel more enjoyed and/or motivated by the learning environment from which they learn less. This thesis is one of the first to focus on cognition, motivation, and emotion and their interconnections. With the insights from this thesis, we hope to move the field of GBL, instructional design, and theories of learning forward to a more integrated approach to research and practice.

Summary

The European Training Network for Chemical Engineering Immersive Learning (https://charming-etn.eu/) or CHARMING project for short, aims to develop immersive learning technologies, such as games, in chemistry and chemical engineering for children in primary education, students in secondary and higher education, and employees in chemical industry. As part of CHARMING project, this thesis focuses on game-based learning (GBL) in secondary and higher chemistry education. This thesis describes four studies on the effects of GBL including different instructional design features on the learning processes and outcomes in secondary and higher chemistry education.

Chapter 1 provides an overarching introduction that contains the theoretical background of the research questions of this thesis. First, we introduce the societal and scientific relevance of this thesis. Societally, education for future employees in the workplace in chemistry and chemical engineering has major challenges and GBL has the potential to provide solutions to them. However scientifically, GBL research has not provided concrete answers to the questions whether students learn more with GBL in chemistry education than non-GBL (i.e., media comparison research) and which instructional design features improve GBL and how (i.e., value-added research).

Second, we give a broad definition of GBL and outline the theoretical background of GBL in terms of how different design features of GBL affect cognition, motivation, and emotion and ultimo learning. Based on an extensive analysis of the effectiveness of GBL (Chapter 2), three experimental studies investigate the effects of instructional design features that aim to manage cognitive load (Chapter 3), promote motivation (Chapter 4), and promote positive emotions (Chapter 5) on three learning outcomes, namely, cognition (i.e., mental effort and performance), motivation (i.e., achievement goals), and emotion (i.e., achievement emotions) in GBL in chemistry education. We also pay attention to the interconnection between cognition, motivation, and emotion. The three experimental studies are grounded in cognitive load theory, achievement goal theory, and control-value theory of achievement emotions as these three theories are highly relevant in GBL.

Chapter 2 examines the extent to which GBL adds value over non-GBL and what design features can improve GBL. The research questions are:

1. Is the effect of GBL in chemistry education on cognition (including retention), motivation, and emotion larger than for non-GBL (media comparison research)?
2. Do instruction characteristics (activity level of control group, additional instruction, user grouping, and number of game sessions) and methodology characteristics (randomization, sample size, publication source, and assessment type) moderate the effect?
3. Which game design or instructional design features improve GBL in chemistry education (value-added research)?

We conducted a meta-analysis on 34 empirical studies. Theoretically, GBL can impact chemistry learning by affecting cognition, motivation, and/or emotion. Thus, we hypothesized:

1. GBL in chemistry education would yield higher cognitive outcomes including retention, motivation, and positive emotions, and lower negative emotions than non-GBL.
2. Relative to non-GBL, GBL in chemistry education with multiple game sessions would yield higher cognitive outcomes than those with one session.

The first hypothesis was partially confirmed, and the second hypothesis was not. A three-level meta-analysis revealed that GBL in chemistry education was more effective in terms of cognitive outcomes including retention, and motivation than non-GBL. In our sample, we found a substantial heterogeneity for cognitive outcomes (i.e., the effect sizes vary across studies), small-study effects (i.e., studies with small sample size are more likely to report larger effect sizes than those with large sample size), and particularly publication bias (i.e., studies with statistically significant results are more likely to be published). However, there were no studies on emotion and too few studies comparing GBL with or without specific design features. We concluded that GBL enhances chemistry learning more than non-GBL, and that more research is needed on emotion and instructional design features.

The studies on instructional design features in Chapter 3, 4, and 5 use the same game, CosmiClean (https://recyclegame.eu/), which was developed by LuGus Studios (www.lugus-studios.be). The game teaches secondary school and university students the principles for separation processes of recycling materials. The chemistry learning content includes the functions of the nine separators (including the sieve, the melter, the magnet, the shredder, the non-ferrous separator, the stream separator, the boiler, the dissolver, and the centrifuge) and the eight properties (including size, density, phase, melting point, boiling point, solubility, magnetic metal, and non-ferrous metal) of 12 materials (including iron, plastics, concrete, wood, glass, sand, copper, water, salt, fuel, gold, and solvent). Players complete a series of game levels with different mixtures of materials in a spaceship cargo. The goal is to make a recycling chain, including a conveyor (for transporting the materials), one or more separators (for separating material based on different properties), and receptors (for collecting the recycled materials).

Chapter 3 studies when specific information is best presented (i.e., before or during game play) so that the cognitive load is optimal, and thus learning is best supported. The research questions are:

1. How does timing of domain-specific and game-specific information presentation affect cognition (i.e., mental effort, posttest performance, and game performance), motivation (i.e., achievement goals), and emotion (i.e., achievement emotions) in GBL in chemistry education?
2. Do mental effort, perceived competence, perceived control, and perceived value mediate the effects?

We compared four conditions: Timing of domain-specific information presentation (before/during gameplay) and timing of game-specific information presentation (before/during gameplay). According to cognitive load theory, timing of information presentation may affect cognitive load in GBL: Optimal timing of information presentation may manage intrinsic load, reduce extraneous load of complex learning tasks, and thus yield higher performance. According to achievement goals theory, timing of information presentation may affect cognitive load, shape learners' perceived competence, and thus affect achievement goals. According to control-value theory of achievement emotions, timing of information presentation may affect cognitive load, shape learners' perceived control and perceived value, and thus induce achievement emotions. Thus, we hypothesized:

1. Presenting domain-specific information before gameplay and game-specific information during gameplay would be optimal, namely, highest performance, mastery-approach goals, performance-approach goals, and positive emotions, and lowest mental effort, mastery-

avoidance goals, performance-avoidance goals, and negative emotions in comparison to other conditions.
2. Mental effort would mediate the effects of timing of information presentation on posttest performance and game performance.
3. Mental effort and perceived competence would jointly mediate the effects of timing of information presentation on achievement goals.
4. Mental effort, perceived control, and/or perceived value would jointly mediate the effects of timing of information presentation on achievement emotions.

The hypotheses were partially confirmed. Multiple regression and robust regression revealed that presenting domain-specific information before gameplay promoted higher mastery-approach goals, performance-approach goals, mastery-avoidance goals, performance-avoidance goals, and positive emotions, but lower posttest performance than presenting it during gameplay. There was no effect on mental effort, which may be due to a floor effect (i.e., students reported on average rather low mental effort). There was no difference between presenting game-specific information before gameplay and during gameplay except for performance-avoidance goals. Structural equation modeling revealed no mediating effects. We concluded that timing of information presentation differentially affects cognition, motivation, and emotion in GBL. Specifically, higher mastery-approach goals, performance-approach goals, mastery-avoidance goals, performance-avoidance goals, and positive emotions are associated with lower performance.

Chapter 4 studies the learning effect of specific achievement goal instructions to a learner in GBL. We investigated the research questions:

1. How do mastery-approach and performance-approach goal instructions affect motivation (i.e., achievement goals), cognition (i.e., mental effort and posttest performance), and emotion (i.e., achievement emotions) in GBL in chemistry education?
2. Do prior achievement goals moderate the effects?

We compared four conditions: Mastery-approach goal instructions (yes/no) and performance-approach goal instructions (yes/no). According to achievement goals theory, mastery-approach goal instructions that emphasize mastery, understanding, and improvement mostly induce mastery-approach goals, performance-approach goal instructions that emphasize competition mostly induce performance-approach goals, and combined mastery-approach goal instructions and performance-approach goal instructions mostly induce mastery-approach goals and performance-approach goals. According to cognitive load theory, mastery-approach goals and performance-approach goals may increase mental effort and performance. According to control-value theory of achievement emotions, mastery-approach goals and performance-approach goals mostly correlate to positive emotions: Mastery-approach goals may focus students' attention on the activity itself, thereby influencing activity emotions and performance-approach goals may focus students' attention on the outcomes, thereby influencing outcome emotions. Thus, we hypothesized:

1. Mastery-approach goal instructions would induce higher mastery-approach goals, mental effort, posttest performance, and game performance, more positive (enjoyment) and less negative activity emotions (boredom, anger) than no mastery-approach goal instructions.
2. Performance-approach goal instructions would induce higher performance-approach goals, mental effort, posttest performance, and game performance, higher positive (pride) and lower negative outcome emotions (anxiety, shame) than no performance-approach goal instructions.

3. Combined mastery-approach goal instructions and performance-approach goal instructions would induce highest approach goals, mental effort, posttest performance, game performance, most positive emotions (enjoyment and pride) and least negative emotions (boredom, anger, anxiety, and shame) in comparison to other conditions.
4. Prior achievement goals would moderate the effects of achievement goal instructions.

The hypotheses were partially confirmed. Robust regression revealed that mastery-approach goal instructions did not induce mastery-approach goals. Performance-approach goal instructions promoted higher performance-approach goals and higher mental effort but lower posttest performance. Prior mastery-approach goals moderated the effects of performance-approach goal instructions on mental effort. We conclude that performance-approach goal instructions differentially affect motivation and cognition in GBL. Specifically, higher performance-approach goals are associated with lower performance.

Chapter 5 focuses on emotions and specifically on whether, and if so, to what extent, the emotions of a fellow student performing a learning activity can influence the student's emotions and thus motivation and learning. We investigated the research question:

> How do peers' achievement emotions (enjoyment/frustration/neutral state) affect students' emotions (i.e., achievement emotions), motivation (i.e., achievement goals), and cognition (i.e., mental effort, posttest performance, and game performance) in GBL in chemistry education?

We compared three conditions: peers' enjoyment, peers' frustration, and peers' neutral state. According to control-value theory of achievement emotions and emotional contagion, peers' achievement emotions may affect students' corresponding achievement emotions. According to achievement goals theory, mastery-approach goals and mastery-avoidance goals may focus students' attention on the activity itself, thereby influencing activity emotions, and performance-approach goals and performance-avoidance goals may focus students' attention on outcomes, thereby influencing outcome emotions. Furthermore, mastery-approach goals may focus attention on the positive value of the activity and help feeling in control, thereby positively influencing positive activity emotions, while mastery-avoidance goals may focus attention on the negative value of the activity and promote feeling a lack of control, thereby positively influencing negative activity emotions. According to cognitive load theory, emotions may impose extraneous load, thereby decreasing performance. However, positive activating emotions (e.g., enjoyment) may increase task attention and motivation to invest effort, thereby increasing performance, while negative activating emotions (e.g., frustration, anger) may decrease task attention but also increase effort to avoid failure, thereby increasing or decreasing performance. Thus, we hypothesized:

1. Students exposed to peers' enjoyment would report higher positive emotions and lower negative emotions, higher mastery-approach goals and lower mastery-avoidance goals than those exposed to peers' neutral state, followed by those exposed to peers' frustration; and students from these three groups report equal performance-approach goals and performance-avoidance goals.
2. Students exposed to peers' enjoyment and those exposed to peers' frustration would report higher mental effort than those exposed to peers' neutral state.
3. Students exposed to
    a. peers' enjoyment or peers' frustration report lower game and posttest performance than those exposed to peers' neutral state;

b. peers' enjoyment or peers' frustration report higher game performance and posttest performance than those exposed to peers' neutral state;
c. peers' enjoyment report higher game performance and posttest performance than those exposed to peers' frustration and those exposed to peers' neutral state.

The hypotheses were partially confirmed. Random intercept cross-lagged panel models with Bayesian estimation revealed that students exposed to peers' enjoyment reported more enjoyment and relaxation, higher mastery-approach goals, mastery-avoidance goals, and game performance, less frustration, anger, boredom, and mental effort, higher posttest performance, equal performance-approach goals and performance-avoidance goals than those exposed to peers' frustration. We conclude that peers' achievement emotions may differentially affect students' emotion, motivation, and cognition in GBL. Specifically, higher mastery-approach goals, mastery-avoidance goals, and positive emotions and lower negative emotions are associated with lower performance.

Chapter 6 provides a general discussion of the findings. First, we provide an overview of the results. Overall, GBL in chemistry education was more effective and motivating than non-GBL (i.e., media comparison research) and instructional design features, such as timing of information presentation, achievement goal instructions, and peers' achievement emotions, differentially affected cognition, motivation, and emotion in GBL in chemistry education (i.e., value-added research). Specifically, high mastery-approach goals, performance-approach goals, mastery-avoidance goals, and/or performance-avoidance goals and more positive achievement emotions were associated with low posttest performance.

Second, we provide the theoretical and practical implications of the results. From a scientific perspective, researchers need to integrate cognition, motivation, and emotion in theories of learning and instructional design. For example, our results contribute to motivational and emotional design principles in multimedia learning. We found that high approach and avoidance goals and more positive achievement emotions were associated with low posttest performance. This result contradicts the common desire to create effective, engaging, and enjoyable learning environments. This result also contradicts the argument that higher motivation and/or positive emotions lead to more learning. From a practical perspective, researchers should do sample size planning to avoid small-study effects. Educators can implement GBL in chemistry education, but GBL should be well-designed. They should pay attention to timing of information presentation, achievement goals instructions, and peers' achievement emotions. Game designers should consider leaving room for users, such as teachers, to adapt the game. Educators and learners should use GBL differently depending on the learning goals: performance, motivation, and/or emotion.

Third, this chapter acknowledges some limitations with accompanying suggestions for future research.

1. Mental effort: We used mental effort as the indicator of overall cognitive load also in relation to performance, but the results on the relationship between mental effort and performance appear to be inconsistent. This may be because the source of mental effort is unclear. Future research could differentiate and measure different types of cognitive load (e.g., intrinsic load and extraneous load).
2. Effects of instructional design features on achievement goals: We found that the instructional design features affected either mastery-based goals or performance-based goals but not both. This may be because that mastery-approach goals are less relevant than

performance-approach goals in our age group (e.g., secondary school students), that performance-based goals are less relevant than mastery-based goals in our study setting (e.g., activity emotions), or that the game used in our studies lacks game features (e.g., leaderboards) that could stimulate players to pursue performance-based goals. Future research could explore this using different type of achievement emotions (e.g., outcome emotions), age groups (e.g., primary school students), or games (e.g., a game with leaderboards).

3. Effects of instructional design features on avoidance goals: We found that instructional design features had similar effects on approach goals and avoidance goals. These results contrast with the theoretical hypotheses that instructional design features should increase approach goals and decrease avoidance goals. This may be because that the definition component of achievement goals (task-based, self-based, or other-based) is so strongly linked to our instructional design features that it overwhelms approach/avoidance distinction. Future research should confirm this.
4. Effects of instructional design features on achievement emotions: We did not find the effects of instructional design features on all achievement emotions. This may be because that emotions change over time during the course of a learning activity, and we measured achievement emotions once only at the end of GBL. Future research could use dynamic structural equation modelling to measure and analyze the dynamics of achievement emotions.
5. Challenges for GBL: One example is the assessment challenge: Our participants indicated that it was more challenging to solve a problem in the knowledge test with texts than solving the same problem in a game. This may be because that tests and quizzes may not be well-suited to GBL. Future research could explore stealth assessment that the assessment is implemented during the course of gameplay itself. Another example is the analysis challenge: Because of practical constraints, we did not analyze log data. Game log data has the potential to reveal players' behavioral patterns and herewith what factors impact learning. Future research could explore educational data mining and learning analytics.
6. Other GBL environments: We used a specific game genre (i.e., CosmiClean is a computer game, puzzle game, and strategy game), specific subject (i.e., chemistry), and specific learning content (i.e., separating materials). Different game genres, subjects, and learning contents may offer different GBL experiences. Future research could explore other GBL environments (e.g., virtual reality games), subject areas (e.g., English as a secondary language), or chemistry lab skills (e.g., soap making).

Finally, this chapter finishes with a conclusion: GBL enhances chemistry learning more than non-GBL, and instructional design features differentially affect cognition, motivation, and emotion in GBL in chemistry education. Specifically, students felt more enjoyed and motivated in the learning environment from which they learned less.

Samenvatting (Summary in Dutch)

Het European Training Network for Chemical Engineering Immersive Learning (https://charming-etn.eu/), of kortweg het CHARMING-project, beoogt technologieën voor immersive learning, zoals games, in scheikunde en chemical engineering te ontwikkelen voor kinderen in het basisonderwijs, studenten in het secundair en hoger onderwijs en werknemers in de chemische industrie. Als onderdeel van het CHARMING-project richt dit proefschrift zich op game-based learning (GBL) in het secundair en hoger scheikundeonderwijs. Dit proefschrift beschrijft vier onderzoeken naar de effecten van GBL met verschillende instructieontwerpkenmerken, op de leerprocessen en leerprestaties van studenten in het secundair en hoger scheikundeonderwijs.

Hoofdstuk 1 geeft een overkoepelende inleiding die de theoretische achtergrond van de onderzoeksvragen in dit proefschrift bevat. Eerst introduceren we de maatschappelijke en wetenschappelijke relevantie van dit proefschrift. Maatschappelijk gezien staat het onderwijs voor toekomstige werknemers in de chemie voor grote uitdagingen. GBL heeft het potentieel om hiervoor oplossingen te bieden. Wetenschappelijk gezien heeft GBL-onderzoek echter nog geen concrete antwoorden gegeven op de vraag of studenten in scheikundeonderwijs meer leren met GBL dan zonder GBL (d.w.z. mediavergelijkingsonderzoek; Engels: media comparison research) en welke instructieontwerpkenmerken van GBL, GBL verbeteren en hoe (d.w.z. toegevoegde waarde onderzoek; Engels: value-added research).

Daarna geven we een brede definitie van GBL, schetsen we de theoretische achtergrond van GBL en leggen we uit hoe verschillende ontwerpkenmerken van GBL cognitie, motivatie en emotie en uiteindelijk leren beïnvloeden. Gebaseerd op een uitgebreide analyse van de effectiviteit van GBL (Hoofdstuk 2), onderzoeken we in drie experimentele studies de effecten van verschillende instructieontwerpkenmerken in GBL in het scheikundeonderwijs gericht op het managen van de cognitieve belasting (Hoofdstuk 3), het bevorderen van motivatie (Hoofdstuk 4) en het bevorderen van positieve emoties (Hoofdstuk 5) op drie leerresultaten, namelijk cognitie (d.w.z. mentale inspanning en leerprestaties), motivatie (d.w.z. prestatiedoelen) en emotie (d.w.z. prestatie-emoties). Ook besteden we aandacht aan de samenhang tussen cognitie, motivatie en emotie. De drie experimentele onderzoeken zijn gebaseerd op de cognitieve belasting theorie (Engels: cognitive load theory), de prestatie-doeltheorie (Engels: achievement goal theory) en de controle-waardetheorie van prestatie-emoties (Engels: control-value theory of achievement emotions), aangezien deze drie theorieën zeer relevant zijn in GBL.

Hoofdstuk 2 onderzoekt in hoeverre een leeromgeving met GBL beter is dan een leeromgeving zonder GBL en welke ontwerpkenmerken GBL kunnen verbeteren. De onderzoeksvragen zijn:

1. Is het effect van een leeromgeving met GBL in het scheikundeonderwijs op cognitie (inclusief retentie), motivatie en emotie groter dan dat van een leeromgeving zonder GBL (mediavergelijkingsonderzoek)?
2. Modereren ontwerpkenmerken (bijv. activiteitenniveau controlegroep, aanvullende instructie, gebruikersgroepering en aantal spelsessies) en methodologische kenmerken (bijv. randomisatie, steekproefomvang, publicatiebron en beoordelingstype) het effect?
3. Welke spelontwerp- of instructieontwerpkenmerken verbeteren GBL in scheikundeonderwijs (toegevoegde waarde onderzoek)?

We voerden een meta-analyse uit op 34 empirische studies. Theoretisch kan GBL het leren van scheikunde beïnvloeden door cognitie, motivatie en/of emotie te beïnvloeden. Dus veronderstelden we:

> Een leeromgeving met GBL in scheikundeonderwijs zou hogere cognitieve resultaten, een hogere motivatie, meer positieve emoties, en minder negatieve emoties moeten opleveren dan een leeromgeving zonder GBL.
>
> In vergelijking met een leeromgeving zonder GBL zou een leeromgeving met GBL in het scheikundeonderwijs met meerdere spelsessies hogere cognitieve resultaten moeten opleveren dan die met één sessie.

De eerste hypothese werd gedeeltelijk bevestigd en de tweede hypothese niet. Een meta-analyse op drie niveaus onthulde dat een leeromgeving met GBL in het scheikundeonderwijs effectiever was in termen van cognitieve resultaten (waaronder retentie) en motivatie dan een leeromgeving zonder GBL. In onze steekproef vonden we een substantiële heterogeniteit voor de cognitieve resultaten (d.w.z. de effectgroottes varieerden tussen studies), disproportionele effecten van kleine studies (d.w.z. studies met een kleine steekproefomvang rapporteren eerder grotere effectgroottes dan die met een grote steekproefomvang), en een publicatiebias (d.w.z. studies met statistisch significante resultaten worden eerder gepubliceerd). Er waren geen studies over emotie en er waren amper studies waarin GBL met of zonder specifieke ontwerpkenmerken werd vergeleken. We concludeerden dat een leeromgeving met GBL het leren van scheikunde meer bevordert dan een leeromgeving zonder GBL, en dat er meer onderzoek nodig is naar instructieontwerpkenmerken en emotie.

De onderzoeken naar instructieontwerpkenmerken van GBL in hoofdstuk 3, 4 en 5 gebruiken hetzelfde spel, CosmiClean (https://recyclegame.eu/), ontwikkeld door LuGus Studios (www.lugus-studios.be). Het spel leert middelbare scholieren en studenten in het hogere onderwijs de principes voor scheidingsprocessen van recyclingmaterialen. De leerstof in dit spel omvat de functies van negen scheiders (d.w.z. de zeef, de smelter, de magneet, de shredder, de non-ferroscheider, de stroomscheider, de koker, de oplosser en de centrifuge) en acht eigenschappen (d.w.z. grootte, dichtheid, fase, smeltpunt, kookpunt, oplosbaarheid, magnetisch metaal, en non-ferrometaal) van 12 materialen (d.w.z. ijzer, kunststoffen, beton, hout, glas, zand, koper, water, zout, brandstof, goud en oplosmiddel). Spelers voltooien een reeks spelniveaus met verschillende mengsels van materialen in een ruimteschiplading. Het doel is om een recyclingketen te maken, inclusief een transportband (voor het transporteren van de materialen), een of meer scheiders (voor het scheiden van materiaal op basis van verschillende eigenschappen), en receptoren (voor het verzamelen van de gerecyclede materialen).

Hoofdstuk 3 onderzoekt wanneer specifieke informatie het beste kan worden gepresenteerd (d.w.z. voor of tijdens het spelen van de game), zodat de cognitieve belasting optimaal is en het leren dus het beste wordt ondersteund. De onderzoeksvragen zijn:

1. Welke invloed heeft de timing van domein-specifieke en spel-specifieke informatie presentatie op cognitie (d.w.z. mentale inspanning, leerprestaties en spelprestaties), motivatie (d.w.z. prestatiedoelen) en emotie (d.w.z. prestatie-emoties) bij GBL in het scheikundeonderwijs?
2. Mediëren mentale inspanning, ervaren competentie, ervaren controle, en ervaren waarde de effecten?

We vergeleken vier condities: timing van domein-specifieke informatie presentatie (vóór/tijdens gameplay) en timing van game-specifieke informatie presentatie (vóór/tijdens gameplay). Volgens de cognitieve belasting theorie kan de timing van informatie presentatie de cognitieve belasting in GBL beïnvloeden: Optimale timing van informatie presentatie kan de intrinsieke belasting managen, de irrelevante belasting van complexe leertaken verminderen, en dus betere prestaties opleveren. Volgens de prestatie-doeltheorie kan de timing van informatie presentatie de cognitieve belasting beïnvloeden, de ervaren competentie van studenten vormen, en dus de prestatiedoelen beïnvloeden. Volgens de controle-waardetheorie van prestatie-emoties kan de timing van informatie presentatie de cognitieve belasting beïnvloeden, de ervaren controle en ervaren waarde van studenten vormen, en zo prestatie-emoties beïnvloeden. Dus veronderstelden we:

1. Het presenteren van domein-specifieke informatie voorafgaand aan het spel en spel-specifieke informatie tijdens het spel zou optimaal zijn, namelijk hoogste prestatie, hoogste leer-streef doelen (Engels: mastery-approach goals), hoogste prestatie-streef doelen (Engels: performance-approach goals), meer positieve emoties en laagste mentale inspanning, laagste leer-vermijd doelen (Engels: mastery-avoidance goals), laagste prestatie-vermijd doelen (Engels: performance-avoidance goals) en minder negatieve emoties in vergelijking met andere condities.
2. Mentale inspanning zou de effecten van timing van informatie presentatie op leerprestaties en spelprestaties mediëren.
3. Mentale inspanning en ervaren competentie zouden samen de effecten van timing van informatie presentatie op prestatiedoelen mediëren.
4. Mentale inspanning, ervaren controle en/of ervaren waarde zouden gezamenlijk de effecten van timing van informatie presentatie op prestatie-emoties mediëren.

De hypothesen werden gedeeltelijk bevestigd. Meervoudige regressie en robuuste regressie lieten zien dat het presenteren van domein-specifieke informatie vóór het spelen van games hogere leer-streef doelen, prestatie-streef doelen, leer-vermijd doelen, prestatie-vermijd doelen en meer positieve emoties opleverde, maar lagere leerprestaties dan presentatie tijdens het spelen. Er was geen effect op de mentale inspanning, wat mogelijk te wijten is aan een vloereffect (d.w.z. studenten rapporteerden gemiddeld een vrij lage mentale inspanning). Er was geen verschil tussen het presenteren van spel-specifieke informatie vóór het spelen van het spel en tijdens het spelen, behalve voor prestatie-vermijd doelen. Structurele vergelijkingsmodellering (Engels: structural equation modeling) bracht geen mediërende effecten aan het licht. We concludeerden dat de timing van informatie presentatie een verschillende invloed heeft op cognitie, motivatie en emotie bij GBL. In het bijzonder worden hogere leer-streef doelen, hogere prestatie-streef doelen, hogere leer-vermijd doelen, prestatie-vermijd doelen en meer positieve emoties geassocieerd met lagere prestaties.

Hoofdstuk 4 onderzoekt het leereffect van specifieke doelinstructies aan de student in GBL. We onderzochten de onderzoeksvragen:

1. Hoe beïnvloeden leer-streef doelinstructies en prestatie-streef doelinstructies motivatie (d.w.z. prestatiedoelen), cognitie (d.w.z. mentale inspanning en leerprestaties) en emotie (d.w.z. prestatie-emoties) bij GBL in het scheikundeonderwijs?
2. Modereren vooraf aanwezige prestatiedoelen de effecten?

We vergeleken vier condities: leer-streef doelinstructie (ja/nee) en prestatie-streef doelinstructie (ja/nee). Volgens de prestatie-doeltheorie leiden leer-streef doelinstructies die de nadruk leggen op beheersing, begrip en verbetering meestal tot leer-streef doelen, prestatie-streef doelinstructies die de nadruk leggen op competitie leiden meestal tot prestatie-streef doelen, en gecombineerde doelinstructies leiden meestal tot leer-streef doelen en prestatie-streef doelen. Volgens de cognitieve belasting theorie kunnen leer-streef doelen en prestatie-streef doelen de mentale inspanning en prestaties verhogen. Volgens de controle-waardetheorie van prestatie-emoties correleren leer-streef doelen en prestatie-streef doelen meestal met positieve emoties: Leer-streef doelen kunnen de aandacht van studenten richten op de activiteit zelf, waardoor activiteit-emoties worden beïnvloed en prestatie-streef doelen kunnen de aandacht van studenten richten op de resultaten, waardoor resultaat-emoties worden beïnvloed. Dus veronderstelden we:

1. Leer-streef doelinstructies zouden leiden tot hogere leer-streef doelen, mentale inspanning, leerprestaties en spelprestaties, meer positieve emoties (plezier) en minder negatieve emoties (verveling, woede) dan instructies zonder leer-streef doel.
2. Prestatie-streef doelinstructies zouden leiden tot hogere prestatie-streef doelen, mentale inspanning, leerprestaties en spelprestaties, meer positieve emoties (trots) en minder negatieve emoties (angst, schaamte) dan instructie zonder prestatie-streef doel.
3. Gecombineerde doelinstructies zouden leiden tot de hoogste streefdoelen, mentale inspanning, leerprestatie, spelprestaties, de meeste positieve emoties (plezier en trots), en de minste negatieve emoties (verveling, woede, angst en schaamte) in vergelijking met andere condities.
4. Vooraf aanwezige prestatiedoelen zouden de effecten van instructies voor prestatiedoelen modereren.

De hypothesen werden gedeeltelijk bevestigd. Robuuste regressie onthulde dat leer-streef doelinstructies niet leidden tot leer-streef doelen. Prestatie-streef doelinstructies bevorderden hogere prestatie-streef doelen en hogere mentale inspanning, maar lagere leerprestaties. Vooraf aanwezige leer-streef doelen modereerden de effecten van prestatie-streef doelinstructies op mentale inspanning. We concluderen dat prestatie-streef doelinstructies een verschillende invloed hebben op motivatie en cognitie bij GBL. In het bijzonder worden hogere prestatie-streef doelen geassocieerd met lagere prestaties.

Hoofdstuk 5 onderzoekt of, en zo ja in welke mate, de emoties van een medestudent die een leeractiviteit uitvoert van invloed zijn op de emoties van de toekijkende student en daarmee op motivatie en leren. We onderzochten de onderzoeksvraag:

Welke invloed hebben de prestatie-emoties van medestudenten (plezier/frustratie/neutrale toestand) op de prestatie-emoties, motivatie (d.w.z. prestatiedoelen) en cognitie (d.w.z. mentale inspanning, leerprestaties en spelprestaties) van studenten in GBL in het scheikundeonderwijs?

We vergeleken drie condities: het plezier van medestudenten, de frustratie van medestudenten en de neutrale toestand van medestudenten. Volgens de controle-waardetheorie van prestatie-emoties en emotionele besmettelijkheid (Engels: emotional contagion) kunnen de prestatie-emoties van medestudenten de overeenkomstige prestatie-emoties van studenten beïnvloeden. Volgens de prestatie-doeltheorie kunnen leer-streef doelen en leer-vermijd doelen de aandacht van studenten richten op de activiteit zelf, waardoor ze emoties over de activiteit beïnvloeden, en prestatie-streef doelen en prestatie-vermijd doelen kunnen de aandacht van studenten richten op leeruitkomsten, waardoor ze emoties over de leeruitkomst beïnvloeden. Bovendien kunnen leer-streef doelen de aandacht vestigen op de positieve waarde van de activiteit en het gevoel van controle ondersteunen, waardoor positieve emoties over de activiteit positief worden beïnvloed, terwijl prestatie-vermijd doelen de aandacht kunnen vestigen op de negatieve waarde van de activiteit en het gevoel van gebrek aan controle kunnen bevorderen, waardoor negatieve emoties over de activiteit positief worden beïnvloed. Volgens de cognitieve belasting theorie kunnen emoties een irrelevante belasting veroorzaken voor leren, waardoor de prestatie afneemt. Positief activerende emoties (bijv. plezier) kunnen echter de aandacht voor de taak en de motivatie om moeite te doen vergroten, waardoor de prestaties toenemen, terwijl negatief activerende emoties (bijv. frustratie, woede) de aandacht voor de taak kunnen verminderen, maar ook de inspanning om falen te voorkomen vergroten, waardoor de prestaties kunnen toenemen of afnemen. Dus veronderstelden we:

1. Studenten die worden blootgesteld aan plezier van medestudenten rapporteren hogere positieve emoties en minder negatieve emoties, hogere leer-streef doelen en lagere leer-vermijd doelen dan degenen die worden blootgesteld aan de neutrale toestand van medestudenten, gevolgd door degenen die worden blootgesteld aan frustraties van medestudenten; en studenten uit deze drie groepen rapporteren gelijke prestatie-streef doelen en prestatie-vermijd doelen.
2. Studenten die worden blootgesteld aan het plezier van medestudenten en degenen die worden blootgesteld aan de frustratie van medestudenten, rapporteren een hogere mentale inspanning dan degenen die worden blootgesteld aan de neutrale toestand van medestudenten.
3. Studenten die worden blootgesteld aan:
    a. het plezier van medestudenten of de frustratie van medestudenten rapporteren lagere spel- en leerprestaties dan degenen die worden blootgesteld aan de neutrale toestand van medestudenten;
    b. het plezier van medestudenten of de frustratie van medestudenten rapporteren hogere spel- en leerprestaties dan degenen die worden blootgesteld aan de neutrale toestand van medestudenten;
    c. het plezier van medestudenten rapporteren hogere spel- en leerprestaties dan degenen die zijn blootgesteld aan de frustratie van medestudenten en degenen die zijn blootgesteld aan de neutrale toestand van medestudenten.

De hypothesen werden gedeeltelijk bevestigd. Random intercept cross-lagged panel-modellen met Bayesiaanse schatting lieten zien dat studenten die werden blootgesteld aan het plezier van medestudenten, meer plezier en ontspanning, hogere leer-streef doelen, leer-vermijd doelen en spelprestaties, minder frustratie, woede, verveling en mentale inspanning en lagere leerprestaties, gelijke prestatie-streef doelen en prestatie-vermijd doelen rapporteerden dan degenen die werden blootgesteld aan de frustratie van medestudenten. We concludeerden dat de prestatie-emoties van medestudenten een verschillende invloed kunnen hebben op de emotie, motivatie en cognitie van studenten in GBL. In het bijzonder worden hogere leer-streef doelen en leer-vermijd doelen, meer positieve emoties en minder negatieve emoties geassocieerd met lagere leerprestaties.

Hoofdstuk 6 reflecteert op de bevindingen. Eerst geven we een overzicht van de resultaten. Over het algemeen was een leeromgeving met GBL in scheikundeonderwijs effectiever en motiverender dan een leeromgeving zonder GBL (d.w.z. mediavergelijkingsonderzoek) en beïnvloedden instructieontwerpkenmerken, zoals timing van informatie presentatie, instructies voor prestatiedoelen en prestatie-emoties van medestudenten, cognitie, motivatie en emotie in GBL in scheikundeonderwijs (d.w.z. onderzoek met toegevoegde waarde) op een verschillende manier. In het bijzonder waren hoge leer-streef doelen, prestatie-streef doelen, leer-vermijd doelen en/of prestatie-vermijd doelen en meer positieve prestatie-emoties geassocieerd met lagere leerprestaties.

Ten tweede bespreken we de theoretische en praktische implicaties van de resultaten. Vanuit een wetenschappelijk perspectief zouden onderzoekers cognitie, motivatie en emotie meer moeten integreren in leertheorieën en instructieontwerp. Onze resultaten dragen bij aan zulke motiverende en emotionele ontwerpprincipes bij multimediaal leren. Verder, ontdekten we dat hoge streef doelen en vermijd doelen en meer positieve prestatie-emoties samenhingen met lage leerprestaties. Dit resultaat is in tegenspraak met de algemene wens om effectieve, boeiende en plezierige leeromgevingen te creëren. Dit resultaat is ook in tegenspraak met het argument dat een hogere motivatie en/of meer positieve emoties tot meer leren leiden.

Vanuit een praktisch perspectief zouden onderzoekers de steekproefomvang goed moeten plannen om kleine-studie-effecten te vermijden. Docenten kunnen GBL implementeren in het scheikundeonderwijs, maar GBL moet goed ontworpen zijn. Ze moeten aandacht besteden aan de timing van de informatie presentatie, instructies voor prestatiedoelen en de prestatie-emoties van medestudenten. Spelontwerpers zouden moeten overwegen om ruimte te laten voor gebruikers, zoals docenten, om het spel aan te passen. Docenten en studenten zouden GBL verschillend moeten gebruiken, afhankelijk van de leerdoelen: prestatie, motivatie en/of emotie.

Ten derde noemen we in dit hoofdstuk enkele beperkingen met bijbehorende suggesties voor toekomstig onderzoek.

1. Mentale inspanning: we gebruikten mentale inspanning als indicator voor algehele cognitieve belasting ook in relatie tot prestatie, maar de resultaten over de relatie tussen mentale inspanning en prestatie lijken inconsistent te zijn. Dit kan zijn omdat de bron van de mentale inspanning onduidelijk is. Toekomstig onderzoek zou verschillende soorten cognitieve belasting (bijv. intrinsieke belasting en irrelevante belasting) kunnen onderscheiden en meten.
2. Effecten van instructieontwerpkenmerken op prestatiedoelen: we ontdekten dat de instructieontwerpkenmerken van invloed waren op op leer doelen of op prestatie doelen, maar niet op beide. Dit kan zijn omdat leer doelen minder relevant zijn dan prestatie doelen bij onze doelgroep (middelbare scholieren), dat prestatie doelen minder relevant zijn dan leer doelen in onze studiesetting (bijvoorbeeld emoties over de activiteit), of dat de game die in onze studies wordt gebruikt spelfuncties mist (bijvoorbeeld scoreborden) die spelers zouden kunnen stimuleren om prestatie doelen na te streven. Toekomstig onderzoek zou dit kunnen onderzoeken door gebruik te maken van verschillende soorten prestatie-emoties (bijv. emoties over het resultaat), leeftijdsgroepen (bijv. basisschoolstudenten) of games (bijv. een spel met scoreborden).
3. Effecten van instructieontwerpkenmerken op vermijdingsdoelen: we ontdekten dat instructieontwerpkenmerken vergelijkbare effecten hadden op streefdoelen en vermijd doelen. Deze resultaten staan in contrast met de theoretische hypothesen dat instructieontwerpkenmerken de streefdoelen zouden moeten verhogen en de vermijd

doelen zouden moeten verminderen. Dit kan zijn omdat prestatiedoelen (taakgebaseerd, op zichzelf gebaseerd of op anderen gebaseerd) op zichzelf zo sterk verbonden zijn met de instructieontwerpkenmerken dat dit het onderscheid tussen streef en vermijd te niet doet. Toekomstig onderzoek moet dit bevestigen.

4. Effecten van instructieontwerpkenmerken op prestatie-emoties: we vonden niet de effecten van instructieontwerpkenmerken op alle prestatie-emoties. Dit kan zijn omdat emoties in de loop van de tijd veranderen tijdens een leeractiviteit, en we hebben prestatie-emoties slechts één keer gemeten aan het einde van GBL. Toekomstig onderzoek zou dynamische structurele vergelijkingsmodellering (Engels: dynamic structural equation modeling) kunnen gebruiken om de dynamiek van prestatie-emoties te meten en te analyseren.
5. Uitdagingen voor GBL: Een voorbeeld is de assessment uitdaging: onze deelnemers gaven aan dat het uitdagender was om een probleem op te lossen in de kennistoets met teksten dan hetzelfde probleem op te lossen in een game. Dit kan zijn omdat tests en quizzen mogelijk niet geschikt zijn voor GBL. Toekomstig onderzoek zou stealth-beoordeling kunnen onderzoeken waarin de beoordeling wordt geïmplementeerd tijdens het spelen van het spel zelf. Een ander voorbeeld is de analyse-uitdaging: vanwege praktische beperkingen hebben we geen loggegevens geanalyseerd. Gameloggegevens hebben het potentieel om de gedragspatronen van spelers te laten zien en daarmee welke factoren van invloed zijn op het leren. Toekomstig onderzoek zou educatieve datamining en learning analytics kunnen onderzoeken.
6. Andere GBL-omgevingen: we gebruikten een specifiek spelgenre (d.w.z. CosmiClean is een computerspel, puzzelspel en strategiespel), specifiek onderwerp (d.w.z. scheikunde) en specifieke leerinhoud (d.w.z. materialen scheiden). Verschillende spelgenres, onderwerpen en leerinhouden kunnen verschillende GBL-ervaringen bieden. Toekomstig onderzoek zou andere GBL-omgevingen (bijv. virtual reality-spellen), vakgebieden (bijv. Engels als secundaire taal) of scheikundelabvaardigheden (bijv. zeep maken) kunnen onderzoeken.

De conclusies van dit hoofdstuk zijn dat een leeromgeving met GBL het leren van scheikunde meer verbetert dan een leeromgeving zonder GBL, en dat instructieontwerpkenmerken een verschillende invloed op cognitie, motivatie en emotie hebben bij GBL in het scheikundeonderwijs. Concreet gezegd hadden studenten meer plezier en waren ze meer gemotiveerd in een leeromgeving waarin ze minder leerden.

References

References marked with an asterisk indicate studies included in the meta-analysis in Chapter 2.

Aarts, H., Gollwitzer, P. M., & Hassin, R. R. (2004). Goal contagion: Perceiving is for pursuing. *Journal of Personality and Social Psychology*, 87, 23–37. https://doi.org/10.1037/0022-3514.87.1.23

Acquah, E. O., & Katz, H. T. (2020). Digital game-based L2 learning outcomes for primary through high- school students: A systematic literature review. *Computers & Education*, 143, 103667. https://doi.org/10. 1016/j.compedu.2019.103667

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, 16(3), 183–198. https://doi.org/10.1016/j.learninstruc.2006.03.001

Ainsworth, S. E. (2014). The multiple representations principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 464–486). Cambridge University Press. https:// doi.org/10.1017/CBO9781139547369.024

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. https://doi.org/10.1922/CDH_2120VandenBroucke08

*Akkuzu, N., & Uyulgan, M. A. (2016). How to improve students' comprehension concerning the major terms of functional groups?—In the experiment of OrCheTaboo game. *International Journal of Higher Education*, 5(2), 196–212. https://doi.org/10.5430/ijhe.v5n2p196

Altinisik, Y., Van Lissa, C. J., Hoijtink, H., Oldehinkel, A. J., & Kuiper, R. M. (2021). Evaluation of inequality constrained hypotheses using a generalization of the AIC. *Psychological Methods*, 26(5), 599–621. https://doi.org/10.1037/met0000406

American Chemical Society [ACS]. (2015a). *ACS guidelines for Bachelor's degree programs*. The American Chemical Society.

American Chemical Society [ACS]. (2015b). *ACS guidelines for chemistry in two-year college programs*. The American Chemical Society.

American Chemical Society [ACS]. (2018). *ACS guidelines and recommendations for teaching middle and high school chemistry*. The American Chemical Society.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84, 261–271. https://doi.org/10.1037/0022-0663.84.3.261

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Apugliese, A., & Lewis, S. E. (2017). Impact of instructional decisions on the effectiveness of cooperative learning in chemistry through meta-analysis. *Chemistry Education Research and Practice*, 18, 271–278. https://doi. org/10.1039/C6RP00195E

Artino, A. R., & Jones, K. D. (2012). Exploring the complex relations between achievement emotions and self- regulated learning behaviors in online learning. *The Internet and Higher Education*, 15(3), 170–175. https:// doi.org/10.1016/j.iheduc.2012.01.006

Asparouhov, T., & Muthén, B. (2018). *SRMR in Mplus*. Retrieved December 2, 2022, from https:www.statmodel.com/download/SRMR2.pdf

Astleitner, H., & Wiesner, C. (2004). An integrated model of multimedia learning and motivation. *Journal of Educational Multimedia and Hypermedia*, 13(1), 3–21.

Baker, S., & Inventado, P. S. (2016). Educational data mining and learning analytics: Potentials and possibilities for online education. In G. Veletsianos (Ed.), *Emergence and Innovation in Digital Learning* (83–98). https:// doi.org/10.15215/aupress/9781771991490.01

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.

Banks, G. C., Kepes, S., & Banks, K. P. (2012). Publication bias: The antagonist of meta-analytic reviews and effective policymaking. *Educational Evaluation and Policy Analysis*, 34(3), 259–277. https://doi.org/10.3102/0162373712446144

Bardach, L., Oczlon, S., Pietschnig, J., & Lüftenegger, M. (2020). Has achievement goal theory been right? A meta-analysis of the relation between goal structures and personal achievement goals. *Journal of Educational Psychology*, 112(6), 1197–1220. https://doi.org/10.1037/edu0000419

Barron, K. E., & Harackiewicz, J. M. (2001). Achievement goals and optimal motivation: Testing multiple goal models. *Journal of Personality and Social Psychology, 80,* 706–722. https://doi.org/10.1037/0022-3514.80.5.706

Baranik, L. E., Stanley, L. J., Bynum, B. H., & Lance, C. E. (2010). Examining the construct validity of mastery-avoidance achievement goals: A meta-analysis. *Human Performance*, 23, 265–282. https://doi.org/10.1080/08959285.2010.488463

Barzilai, S., & Blau, I. (2014). Scaffolding game-based learning: Impact on learning achievements, perceived learning, and game experiences. *Computers & Education*, 70, 65-79. https://doi.org/10.1016/j.compedu.2013.08.003

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 50(4), 1088–1101. https://doi.org/10.2307/2533446

Behrens, C. C., Dolmans, D. H., Gormley, G. J., & Driessen, E. W. (2019). Exploring undergraduate students' achievement emotions during ward round simulation: a mixed-method study. *BMC Medical Education*, 19(1), 1–7. https://doi.org/10.1186/s12909-019-1753-1.

Bellou, I., Papachristos, N. M., & Mikropoulos, T. A. (2018). Digital learning technologies in chemistry education: A review. In D. Sampson, D. Ifenthaler, J. M. Spector, & P. Isaías (Eds.), *Digital technologies: Sustainable innovations for improving teaching and learning* (pp. 57–80). Springer International Publishing.

Bennett, A. G., & Rebello, N. S. (2012). Retention and learning. In *Encyclopedia of the sciences of learning* (pp. 2856–2859). Springer. https://doi.org/10.1007/978-1-4419-1428-6_664

Bieg, M., Goetz, T., & Hubbard, K. (2013). Can I master it and does it matter? An intraindividual analysis on control–value antecedents of trait and state academic emotions. *Learning and Individual Differences, 28,* 102–108. https://doi.org/10.1016/j.lindif.2013.09.006

Bipp, T., & Van Dam, K. (2014). Extending hierarchical achievement motivation models: The role of motivational needs for achievement goals and academic performance. *Personality and Individual Differences*, 64(157-162). https://doi.org/10.1016/j.paid.2014.02.039

Bong, M. (2001). Between- and Within-Domain Relations of Academic Motivation Among Middle and High School Students: Self-Efficacy, Task-Value, and Achievement Goals. *Journal of Educational Psychology*, 93, 23-34. https://doi.org/10.1037/0022-0663.93.1.23

Bong, M. (2009). Age-Related Differences in Achievement Goal Differentiation. *Journal of Educational Psychology*, 101(4), 879-896. https://doi.org/10.1037/a0015945

Borenstein, M. (2019). *Common mistakes in meta-analysis and how to avoid them*. Biostat, Incorporated.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2013). *Comprehensive meta-analysis version 3*. Biostat, Incorporated.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons. https://doi.org/10.1002/9780470743386

Bourgonjon, J., De Grove, F., De Smet, C., Van Looy, J., Soetaert, R., & Valcke, M. (2013). Acceptance of game-based learning by secondary school teachers. *Computers & Education*, 67, 21-35. https://doi.org/10.1016/j.compedu.2013.02.010

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Ribeiro, C., & Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education*, 94, 178–192. https://doi.org/10.1016/j. compedu.2015.11.003

Brom, C., Starkova, T., & D'Mello, S. K. (2018). How effective is emotional design? A meta-analysis on facial anthropomorphisms and pleasant colors during multimedia learning. *Educational Research Review*, 25, 100–119. https://doi.org/10.1016/j.edurev.2018.09.004

Brown, T. L., Eugene LeMay, H., Bursten, B. E., Murphy, C. J., Woodward, P. M., Stoltzfus, M. W., & Lufaso, M. W. (2018). *Chemistry: The central science* (14th ed.). Pearson Education.

Bruder, M., Fischer, A., & Manstead, A. S. (2014). Social appraisal as a cause of collective emotions. In C. von Scheve & M. Salmela (Eds.), *Collective emotions: Perspectives from psychology, philosophy, and sociology* (pp. 141–155). Oxford University Press.

Burgess, L. G., Riddell, P. M., Fancourt, A., & Murayama, K. (2018). The influence of social contagion within education: A motivational perspective. *Mind, Brain, and Education*, 12(4), 164–174. https://doi.org/10.1111/mbe.12178

Byun, J., & Joung, E. (2018). Digital game-based learning for K–12 mathematics education: A meta-analysis. *School Science and Mathematics*, 118(3–4), 113–126. https://doi.org/10.1111/ssm.12271

*Cahyana, U., Paristiowati, M., Savitri, D. A., & Hasyrin, S. N. (2017). Developing and application of mobile game-based learning (M-GBL) for high school students performance in chemistry. *EURASIA Journal of Mathematics, Science and Technology Education*, 13(10), 7037–7047. https://doi.org/10.12973/ejmste/78728

Cai, Z., Mao, P., Wang, D., He, J., Chen, X., & Fan, X. (2022). Effects of scaffolding in digital game-based learning on student's achievement: A three-level meta-analysis. *Educational Psychology Review*, 34(2), 537–574. https://doi.org/10.1007/s10648-021-09655-0

Camacho-Morles, J., Slemp, G. R., Pekrun, R., Loderer, K., Hou, H., & Oades, L. G. (2021). Activity achievement emotions and academic performance: A meta-analysis. *Educational Psychology Review*, 33(3), 1051-1095. https://doi.org/10.1007/s10648-020-09585-3

Carver, C. (2003). Pleasure as a sign you can attend to something else: placing positive feelings within a general model of affect. *Cognition and Emotion*, 17(2), 241–261. https://doi.org/10.1080/02699930302294.

Cellar, D. F., Stuhlmacher, A. F., Young, S. K., Fisher, D. M., Adair, C. K., Haynes, S., et al. (2011). Trait goal orientation, self-regulation, and performance: A meta-analysis. *Journal of Business and Psychology*, 26(4), 467-483. http://dx.doi.org/10.1007/s10869-010-9201-6

*Cha, J., Kan, S. Y., Wahab, N. H. A., Aziz, A. N., & Chia, P. W. (2017). Incorporation of brainteaser game in basic organic chemistry course to enhance students' attitude and academic achievement. *Journal of the Korean Chemical Society*, 61(4), 218–222. https://doi.org/10.5012/jkcs.2017.61.4.218

Chan P., Van Gerven T., Dubois J.-L., Bernaerts K. Study of motivation and engagement for chemical laboratory safety training with VR serious game. Safety Science. (in review March 2023)

Charsky, D., & Ressler, W. (2011). "Games are made for fun": Lessons on the effects of concept maps in the classroom use of computer games. *Computers & Education*, 56(3), 604-615. https://doi.org/10.1016/j.compedu.2010.10.001

Checa, D., & Bustillo, A. (2019). A review of immersive virtual reality serious games to enhance learning and training. *Multimedia Tools and Applications*, 79, 1–27. https://doi.org/10.1007/s11042-019-08348-9

Chee, Y. S., & Tan, D. K. C. (2012). Becoming chemists through game-based inquiry learning: The case of legends of Alkhimia. *Electronic Journal of e-Learning*, 10(2), 185–198.

*Chen, M.-P., & Liao, B.-C. (2015). Augmented reality laboratory for high school electrochemistry course. In *2015 IEEE 15th International Conference on Advanced Learning Technologies* (pp. 132–136). IEEE. https:// doi.org/10.1109/ICALT.2015.105

Chen, C.-H., Shih, C.-C., & Law, V. (2020). The effects of competition in digital game-based learning (DGBL): A meta-analysis. *Educational Technology Research and Development*, 68(4), 1855–1873. https://doi.org/10.1007/ s11423-020-09794-1

Chen, M. H., Tseng, W. T., & Hsiao, T. Y. (2018). The effectiveness of digital game-based vocabulary learning: A framework-based view of meta-analysis. *British Journal of Educational Technology*, 49(1), 69–77. https://doi. org/10.1111/bjet.12526

Chen, J., Wang, M., Kirschner, P. A., & Tsai, C.-C. (2018). The role of collaboration, computer use, learning environments, and supporting strategies in CSCL: A meta-analysis. *Review of Educational Research*, 88(6), 799–843. https://doi.org/10.3102/0034654318791584

*Chen, M. P., Wong, Y. T., & Wang, L. C. (2014). Effects of type of exploratory strategy and prior knowledge on middle school students' learning of chemical formulas from a 3D role-playing game. *Educational Technology Research and Development*, 62(2), 163–185.

https://doi.org/10.1007/s11423-013-9324-3

Cheng, M. T., Chen, J. H., & Chu, S. J. (2015). The use of serious games in science education: a review of selected empirical research from 2002 to 2013. *Journal of Computer Education*, 2, 353–375. https://doi.org/10.1007/ s40692-015-0039-9

Chevrier, M., Muis, K. R., Trevors, G. J., Pekrun, R., & Sinatra, G. M. (2019). Exploring the antecedents and consequences of epistemic emotions. *Learning and Instruction*, 63, 101209. https://doi.org/10.1016/j. learninstruc.2019.05.006.

Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta-analyses: a structural equation modeling approach. *Psychological Methods*, 19(2), 211. https://doi.org/10.1037/a0032968

Cheung, M. W. L. (2019). A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychology Review*, 29(4), 387–396. https://doi.org/10.1007/s11065-019-09415-6

Cheung, A. C. K., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45(5), 283–292. https://doi.org/10.3102/0013189X16656615

*Chimeno, J. S., Wulfsberg, G. P., Sanger, M. J., & Melton, T. J. (2006). The rainbow wheel and rainbow matrix: Two effective tools for learning ionic nomenclature. *Journal of Chemical Education*, 83(4), 651. https://doi. org/10.1021/ed083p651

Chiu, M. H., & Wu, H. K. (2009). The roles of multimedia in the teaching and learning of the triplet relationship in chemistry. In J. K. Gilbert & D. Treagust (Eds.), *Multiple representations in chemical education* (pp. 251–283). Springer.

Clark, D., Nelson, B., Sengupta, P., & D'Angelo, C. (2009). Rethinking science learning through digital games and simulations: Genres, examples, and evidence. In *Proceedings of The National Academies Board on Science Education Workshop on Learning Science: Computer Games, Simulations, and Education* (pp. 36–41). National Academy of Sciences.

Clark, D., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research*, 86(1), 79–122. https://doi.org/10.3102/ 0034654315582065

Clark, R. E. (1982). Antagonism between achievement and enjoyment in ATI studies. *Educational Psychologist*, 17(2), 92–101. https://doi.org/10.1080/00461528209529247

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research*, 53(4), 445–459. https://doi.org/10.3102/00346543053004445

Clark, R. E. (1991). When researchers swim upstream: Reflections on an unpopular argument about learning from media. *Educational Technology*, 31(2), 34–40.

Clark, R. E. (1994a). Media and method. *Educational Technology Research and Development*, 42(3), 7–10. https:// doi.org/10.1007/BF02298090

Clark, R. E. (1994b). Media will never influence learning. *Educational Technology Research and Development*, 42(2), 21–29.

Clark, R. E. (2007). Learning from serious games? Arguments, evidence, and research suggestions.

Educational Technology, May-June(3), 56–59 https://www.jstor.org/stable/44429512

Clark, R. E., Bissell, A., Burrus, C. S., & Breck, J. (2008). What evidence would change your mind about the learning benefits of serious games? A reply to Parker, Becker, and Sawyer. *Educational Technology*, 48(3), 56–58 https://www.jstor.org/stable/44429581

Clark, R. E., & Feldon, D. F. (2014). Ten common but questionable principles of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 151–173). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.009

Clark, R. E., Howard, K., & Early, S. (2006). Motivational challenges experienced in highly complex learning environments. In J. Elen, & R. E. Clark (eds.), *Handling Complexity in Learning Environments: Theory and Research* (pp. 27–43). Elsevier.

Coe, R. (2002). It is the effect size, stupid—What effect size is and why it is important. In *Annual Conference of the British Education Research Association*. University of Exeter. http://www.leeds.ac.uk/educol/documents/ 00002182.htm

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education 8th edition*. Routledge.

Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, 59(2), 661–686. https://doi.org/10.1016/j.compedu.2012.03.004

Cook, D. A., & Artino, A. R. (2016). Motivation to learn: An overview of contemporary theories. Medical Education, 50(10), 997–1014. https://doi.org/10.1111/medu.13074

Cooper, H. (2015). *Research synthesis and meta-analysis: A step-by-step approach* (5th ed.). Sage https://us.sagepub.com/en-us/nam/research-synthesis-and-meta-analysis/book241775

Cooper, M. M., & Stowe, R. L. (2018). Chemistry education research—From personal empiricism to evidence, theory, and informed practice. *Chemical Reviews*, 118(12), 6053–6087. https://doi.org/10.1021/acs.chemrev.8b00020

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety: The experience of work and play in games*. Jossey Bass.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper and Row.

Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, 19(2), 272–309. https://doi.org/10.1080/15213269.2015.1015740

*da Silva Júnior, J. N., Nobre, D. J., do Nascimento, R. S., Torres, G. S., Leite, A. J. M., Monteiro, A. J., Alexandre, F. S. O., Rodríguez, M. T., & Rojo, M. J. (2018). Interactive computer game that engages students in reviewing organic compound nomenclature. *Journal of Chemical*

Education, 95(5), 899–902. https://doi.org/10.1021/acs.jchemed.7b00793

*Daubenfeld, T., & Zenker, D. (2015). A game-based approach to an entire physical chemistry course. *Journal of Chemical Education*, 92(2), 269–277. https://doi.org/10.1021/ed5001697

de Brabander, C. J., & Martens, R. L. (2014). Towards a unified theory of task-specific motivation. *Educational Research Review*, 11, 27–44. https://doi.org/10.1016/j.edurev.2013.11.001

De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science*, 38(2), 105-134. https://doi.org/10.1007/s11251-009-9110-0

de Jong, O., & Taber, K. S. (2007). Teaching and learning the many faces of chemistry. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 631–652). Routledge.

de Jong, K., Conijn, J. M., Gallagher, R. A., Reshetnikova, A. S., Heij, M., & Lutz, M. C. (2021). Using progress feedback to improve outcomes and reduce drop-out, treatment duration, and deterioration: A multilevel meta-analysis. *Clinical Psychology Review*, 85, 102002. https://doi.org/10.1016/j.cpr.2021.102002

Deci, R. M., & Ryan, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78 10.1037110003-066X.55.1.68.

Depaoli, S., & van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist. *Psychological Methods*, 22(2), 240–261. https://doi.org/10.1037/met0000065

Di Natale, A. F., Repetto, C., Riva, G., & Villani, D. (2020). Immersive virtual reality in K-12 and higher education: A 10-year systematic review of empirical research. *British Journal of Educational Technology*, 51(6), 2006–2033. https://doi.org/10.1111/bjet.13030

Dietrich, J., Schmiedek, F., & Moeller, J. (2022). Academic motivation and emotions are experienced in learning situations, so let's study them. *Introduction to the special issue. Learning and Instruction*, 101623. https://doi.org/10.1016/j.learninstruc.2022.101623

D'Mello, S., & Graesser, A. (2011). The half-life of cognitive-affective states during complex learning. *Cognition & Emotion*, 25(7), 1299-1308. https://doi.org/10.1080/02699931.2011.613668

Dominguez Alfaro, J. L., Gantois, S., Blattgerste, J., De Croon, R., Verbert, K., Pfeiffer, T., & Van Puyvelde, P. (2022). Mobile augmented reality laboratory for learning acid–base titration. *Journal of Chemical Education*, 99(2), 531-537. https://doi.org/10.1021/acs.jchemed.1c00894

Domínguez Alfaro, J. L., Udeozor, C., Solmaz, S. & Cermak-Sassenrath, D. (2022). Designing a mobile game for introducing learners to a soap making process. *Proceedings of the 16th European Conference on Games Based Learning (ECGBL)*, 21–27.

Docktor, J. L., & Mestre, J. P. (2014). Synthesis of discipline-based education research in physics. *Physical Review Special Topics-Physics Education Research*, 10(2), 020119. https://doi.org/10.1103/PhysRevSTPER.10.020119

Donker, M. H., van Vemde, L., Hessen, D. J., van Gog, T., & Mainhard, T. (2021). Observational, student, and teacher perspectives on interpersonal teacher behavior: Shared and unique associations with teacher and student emotions. *Learning and Instruction*, 73, 101414. https://doi.org/10.1016/j.learninstruc.2020.101414

Duit, R., Niedderer, H., & Schecker, H. (2007). Teaching physics. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 599–629). Routledge.

Duval, S., & Tweedie, R. (2000a). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Duval, S., & Tweedie, R. (2000b). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95, 89–98. https://doi.org/10.1080/01621459.2000.10473905

Eastwood, J. L., & Sadler, T. D. (2013). Teachers' implementation of a game-based biotechnology curriculum. *Computers & Education*, 66, 11–24. https://doi.org/10.1016/j.compedu.2013.02.003

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist*, 34(3), 169-189. https://doi.org/10.1207/s15326985ep3403_3

Elliot, A. J., Dweck, C. S., & Yeager, D. S. (2017). *Handbook of competence and motivation: Theory and application* (2nd ed.). The Guilford Press.

Elliot, A. J., & Hulleman, C. S. (2017). Achievement goals. In A. J. Elliot, C. S. Dweck, & D. S. Yeager (Eds.), *Handbook of competence and motivation: Theory and application* (2nd ed., pp. 43–60). The Guilford Press.

Elliot, A. J., & McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology*, 80(3), 501–519. https://doi.org/10.1037/0022-3514.80.3.501

Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology, 100*(3), 613–628. https://doi.org/10.1037/0022-0663.100.3.613

Elliot, A. J., Murayama, K., & Pekrun, R. (2011). A 3 × 2 achievement goal model. *Journal of Educational Psychology*, 103(3), 632–648. https://doi.org/10.1037/a0023952

Ellis, P. D. (2010). The essential guide to effect sizes (Vol. 136). Cambridge University Press. https://doi.org/10.1017/CBO9780511761676

Erhel, S., & Jamet, E. (2013). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education*, 67, 156-167. http://dx.doi.org/10.1016/j.compedu.2013.02.019

Erhel, S., & Jamet, E. (2016). The effects of goal-oriented instructions in digital game-based learning. *Interactive Learning Environments*, 24(8), 1744-1757. https://doi.org/10.1080/10494820.2015.1041409

Erhel, S., & Jamet, E. (2019). Improving instructions in educational computer games: Exploring the relations between goal specificity, flow experience and learning outcomes. *Computers in Human Behavior*, 91, 106-114. https://doi.org/10.1016/j.chb.2018.09.020

European Commission. (2015a). *Encouraging STEM Studies for the Labour Market.* Publications Office of the EU. https://data.europa.eu/doi/10.2861/519030

European Commission. (2015b). *Science education for responsible citizenship*. Publications Office of the EU https://op.europa.eu/en/publication-detail/-/publication/a1d14fa0-8dbe-11e5-b8b7-01aa75ed71a1

*Fatokun, K. V. F., Egya, S. O., & Uzoechi, B. C. (2016). Effect of game instructional approach on chemistry students' achievement and retention in periodicity. *European Journal of Research and Reflection in Educational Sciences*, 4(7), 29–40. www.idpublications.org

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. https://doi.org/10.3758/BF03193146

Feldon, D. F., Callan, G., Juth, S., & Jeong, S. (2019). Cognitive load as motivational cost. *Educational Psychology Review*, 31(2), 319-337. https://doi.org/10.1007/s10648-019-09464-6

Feldon, D. F., Franco, J., Chao, J., Peugh, J., & Maahs-Fladung, C. (2018). Self-efficacy change associated with a cognitive load-based intervention in an undergraduate biology course. *Learning and Instruction*, *56*, 64-72. https://doi.org/10.1016/j.learninstruc.2018.04.007

Feldon, D. F., Jeong, S., & Clark, R. E. (2021). Fifteen common but questionable principles of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (3rd ed., pp. 25–40). Cambridge university press.

Fernández-Castilla, B., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2021). Detecting selection bias in meta-analyses with multiple outcomes: A simulation study. *Journal of Experimental Education*., 89(1), 125–144. https://doi.org/10.1080/00220973.2019.1582470

Fernández-Aguilar, L., Navarro-Bravo, B., Ricarte, J., Ros, L., & Latorre, J. M. (2019). How effective are films in inducing positive and negative emotional states? A meta-analysis. *PloS ONE*, 14(11), e0225040. https://doi.org/10.1371/journal.pone.0225040

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. SAGE Publications.

Field, A. P., & Wilcox, R. R. (2017). Robust statistical methods: A primer for clinical psychology and experimental psychopathology researchers. *Behaviour Research and Therapy*, 98, 19-38. http://dx.doi.org/10.1016/j.brat.2017.05.013

Fiedler, K. & Beier, S. (2014). Affect and cognitive processes in educational contexts. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 36–55). Routledge. https://doi.org/10.4324/9780203148211.ch3

Fiorella, L., & Mayer, R. E. (2012). Paper-based aids for learning with a computer-based game. *Journal of Educational Psychology*, *104*(4), 1074–1082. https://doi.org/10.1037/a0028088

Fiorella, L., van Gog, T., Hoogerheide, V., & Mayer, R. E. (2017). It's all a matter of perspective: Viewing first-person video modeling examples promotes learning of an assembly task. *Journal of Educational Psychology*, 109(5), 653–665. https://doi.org/10.1037/edu0000161

Fischer, A., & Hess, U. (2017). Mimicking emotions. *Current Opinion in Psychology*, 17, 151–155. https://doi.org/10.1016/j.copsyc.2017.07.008

Fitts, P. M., & Posner, M. I. (1967). *Human performance*. Brooks/Cole.

Fornós, S., Udeozor, C., Glassey, J., & Cermak-Sassenrath, D. (2022). The CHEM Jam-how to integrate a game creation event in curriculum-based engineering education. *Education for Chemical Engineers*, 40, 8-16. https://doi.org/10.1016/j.ece.2022.04.001

Forsthuber, B., Motiejunaite, A., de Almeida Coutinho, A. S., Baïdak, N., & Horvath, A. (2011). *Science education in Europe: National policies, practices and research. Education, Audiovisual and Culture Executive Agency*. European Commission.

Fredrickson, B. L. (2013). Positive emotions broaden and build. In P. Devine & A. Plant (Eds), *Advances in Experimental Social Psychology* (pp. 1–53). Academic Press.

Frenzel, A. C., Becker-Kurz, B., Pekrun, R., Goetz, T., & Lüdtke, O. (2018). Emotion transmission in the classroom revisited: A reciprocal effects model of teacher and student enjoyment. *Journal of Educational Psychology*, 110(5), 628–639. https://doi.org/10.1037/edu0000228

Friedman, R., Deci, E. L., Elliot, A. J., Moller, A. C., & Aarts, H. (2010). Motivational synchronicity: Priming motivational orientations with observations of others' behaviors. *Motivation and Emotion*, 34, 34–38. https://doi.org/10.1007/s11031-009-9151-3

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation and Gaming*, 33(4), 441–467. https://doi.org/10.1177/1046878102238607

Garzón, J., & Acevedo, J. (2019). Meta-analysis of the impact of augmented reality on students' learning gains. *Educational Research Review*, 27, 244–260. https://doi.org/10.1016/j.edurev.2019.04.001

Garzón, J., Pavón, J., & Baldiris, S. (2019). Systematic review and meta-analysis of augmented reality in educational settings. *Virtual Reality*, 23(4), 447–459. https://doi.org/10.1007/s10055-019-00379-9

Gasper, K., Spencer, L. A., & Hu, D. (2019). Does neutral affect exist? How challenging three beliefs about neutral affect can advance affective research. *Frontiers in Psychology*, 10, 2476. https://doi.org/10.3389/fpsyg.2019.02476

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis (3rd Edition)*. CRC.

Gerjets, P., & Scheiter, K. (2003). Goal configurations and processing strategies as moderators between instructional design and cognitive load: Evidence from hypertext-based instruction. *Educational Psychologist*, 38(1), 33-41. https://doi.org/10.1207/S15326985EP3801_5

Gibbons, R. E., & Raker, J. R. (2019). Self-beliefs in organic chemistry: Evaluation of a reciprocal causation, cross-lagged model. *Journal of Research in Science Teaching*, 56(5), 598–618. https://doi.org/10.1002/tea.21515

Gilbert, J. K., & Treagust, D. (2009). *Multiple representations in chemical education*. Springer. https://doi.org/10.1007/978-1-4020-8872-8

Girard, C., Ecalle, J., & Magnan, A. (2013). Serious games as new educational tools: How effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning,* 29(3), 207–219. https://doi.org/10.1111/j.1365-2729.2012.00489.x

Goetz, T., Frenzel, A. C., Stoeger, H., & Hall, N. C. (2010). Antecedents of everyday positive emotions: An experience sampling analysis. *Motivation and Emotion*, *34*(1), 49-62. https://doi.org/10.1007/s11031-009-9152-2

Goetz, T., Sticca, F., Pekrun, R., Murayama, K., & Elliot, A. J. (2016). Intraindividual relations between achievement goals and discrete achievement emotions: An experience sampling approach. *Learning and Instruction*, 41, 115–125. https://doi.org/10.1016/j.learninstruc.2015.10.007

Gogol, K., Brunner, M., Goetz, T., Martin, R., Ugen, S., Fischbach, A., et al. (2014). 'My questionnaire is too long!' the assessments of motivational-affective constructs with three-item and single-item measures. *Contemporary Educational Psychology*, 39, 188e205. http://dx.doi.org/10.1016/j.cedpsych.2014.04.002

Graesser, A. C. (2020). Emotions are the experiential glue of learning environments in the 21st century. *Learning and Instruction*, 70, 101212. https://doi.org/10.1016/j.learninstruc.2019.05.009

Graesser, A. C., D'Mello, S. K., & Strain, A. C. (2014). Emotions in advanced learning technologies. In R. Pekrun & L. Linnenbrink-Garcia (Eds.) *International handbook of emotions in education* (pp. 473–493). Taylor & Francis.

*Gupta, T. (2019). Game-based learning in chemistry: A game for chemical nomenclature. In *Technology integration in chemistry education and research* (TICER) (pp. 65–79). American Chemical Society. https://doi.org/10.1021/bk-2019-1318.ch005

Haddaway, N. R., Collins, A. M., Coughlin, D., & Kirk, S. (2015). The role of Google scholar in evidence reviews and its applicability to grey literature searching. *PLoS One*, 10(9), e0138237. https://doi.org/10.1371/journal. pone.0138237

Harley, J. M., Pekrun, R., Taxer, J. L., & Gross, J. J. (2019). Emotion regulation in achievement situations: an integrated model. *Educational Psychologist*, 54(2), 106–126. https://doi.org/10.1080/00461520.2019.1587297.

*Halpern, D. F., Millis, K., Graesser, A. C., Butler, H., Forsyth, C., & Cai, Z. (2012). Operation ARA: A computerized learning game that teaches critical thinking and scientific reasoning. *Thinking Skills and Creativity*, 7(2), 93–100. https://doi.org/10.1016/j.tsc.2012.03.006

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods,* 20(1), 102–116. https://doi.org/10.1037/a0038889

Hamaker, E. L., Asparouhov, T., & Muthén, B. (2021). Dynamic structural equation modeling as a combination of time series modeling, multilevel modeling, and structural equation modeling. To be published as Chapter 31. In R. H. Hoyle (Ed.), *The Handbook of Structural Equation modeling* (2nd ed.). Guilford Press

Harrer, M., Cuijpers, P., Furukawa, T. . A., & Ebert, D. D. (2019). D*oing meta-analysis in R: A* *hands-on guide*. Chapman and Hall/CRC. https://doi.org/10.5281/zenodo.2551803

Hatfield, E., Bensman, L., Thornton, P. D., & Rapson, R. L. (2014). New perspectives on emotional contagion: A review of classic and recent research on facial mimicry and contagion. *Interpersona*, 8(2), 159–179. https://doi.org/10.5964/ijpr.v8i2.162

Hatfield, E., Cacioppo, J. T., & Rapson, L. R. (1994). *Emotional Contagion*. Cambridge University Press.

Hattie, J. (2013). *Visible learning and the science of how we learn*. Routledge. https://doi.org/10.4324/ 9781315885025

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, 76(4), 408–420. https://doi.org/10.1080/03637750903310360

Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, 67(3), 451–470. https://doi.org/10.1111/bmsp.12028

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107–128.

Herrando, C., & Constantinides, E. (2021). Emotional contagion: a brief overview and future directions. *Frontiers in Psychology*, 12, 712606. https://doi.org/10.3389/fpsyg.2021.712606

Hidi, S., & Renninger, K. (2006). The four-phase model of interest development. *Educational Psychologist,* 41(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Hoch, E., Sidi, Y., Ackerman, R., Hoogerheide, V., & Scheiter, K. (2023). Comparing Mental Effort, Difficulty, and Confidence Appraisals in Problem-Solving: A Metacognitive Perspective. *Educational Psychology Review*, 35(2), 61. https://doi.org/10.1007/s10648-023-09779-5

*Hodges, G. W., Wang, L., Lee, J., Cohen, A., & Jang, Y. (2018). An exploratory study of blending the virtual world and the laboratory experience in secondary chemistry classrooms. *Computers & Education*, 122, 179–193. https://doi.org/10.1016/j.compedu.2018.03.003

Höfling, T. T. A., Gerdes, A. B., Föhl, U., & Alpers, G. W. (2020). Read my face: automatic facial coding versus psychophysiological indicators of emotional valence and arousal. *Frontiers in Psychology*, 11, 1388. https://doi.org/10.3389/fpsyg.2020.01388

Hofverberg, A., & Winberg, M. (2020). Achievement goals and classroom goal structures: Do they need to match?. *The Journal of Educational Research*, 113(2), 145-162. https://doi.org/10.1080/00220671.2020.1759495

Hoijtink, H. (2011). *Informative hypotheses: Theory and practice for behavioral and social scientists*. CRC Press.

Homer, B. D., Raffaele, C., & Henderson, H. (2020). Games as playful learning: Implications of developmental theory for game-based learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of Game-based Learning* (pp. 25–52). MIT Press.

Hou, H. T. (2022). Augmented reality board game with multidimensional scaffolding mechanism: A potential new trend for effective organizational strategic planning training. *Frontiers in*

Psychology, 13, 5515. https://doi.org/10.3389/fpsyg.2022.932328

Hu, Y., Gallagher, T., Wouters, P., van der Schaaf, M., & Kester, L. (2021). Game-based learning has good chemistry with chemistry education: A three-level meta-analysis. *Journal of Research in Science Teaching*, 1–45. https://doi.org/10.1002/tea.21765

Hu, Y., Wouters, P., van der Schaaf, M., & Kester, L. (2022a). Timing of information presentation matters: Higher motivation, more enjoyment, but less learning in game-based learning. *[Manuscript submitted for publication]*

Hu, Y., Wouters, P., van der Schaaf, M., & Kester, L. (2022b). The effects of achievement goal instructions in game-based learning on students' achievement goals, performance, and achievement emotions. *[Manuscript submitted for publication]*

Huang, C. (2011). Achievement goals and achievement emotions: A meta-analysis. *Educational Psychology Review*, 23(3), 359. https://doi.org/10.1007/s10648-011-9155-x

Huang, C. (2012). Discriminant and criterion-related validity of achievement goals in predicting academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104, 48–73. https://doi.org/10.1037/a0026223

Huang, C. (2016). Achievement goals and self-efficacy: A meta-analysis. *Educational Research Review*, 19, 119-137. https://doi.org/10.1016/j.edurev.2016.07.002

Hulleman, C. S., Schrager, S. M., Bodmann, S. M., & Harackiewicz, J. M. (2010). A meta-analytic review of achievement goal measures: Different labels for the same constructs or different constructs with similar labels? *Psychological Bulletin*, 136, 422–449. https://doi.org/10.1037/a0018947

Hung, H. T., Yang, J. C., Hwang, G. J., Chu, H. C., & Wang, C. C. (2018). A scoping review of research on digital game-based language learning. *Computers & Education*, 126, 89–104. https://doi.org/10.1016/j.compedu. 2018.07.001

Hwang, G.-J., & Wu, P.-H. (2012). Advancements and trends in digital game-based learning research: A review of publications in selected journals from 2001 to 2010. *British Journal of Educational Technology*, 43(1), E6–E10. https://doi.org/10.1111/j.1467-8535.2011.01242.x

*Jagodziński, P., & Wolski, R. (2015). Assessment of application technology of natural user interfaces in the crea- tion of a virtual chemical laboratory. *Journal of Science Education and Technology*, 24(1), 16–28. https://doi. org/10.1007/s10956-014-9517-5

Jaber, L. Z., & Hammer, D. (2016). Learning to feel like a scientist. *Science Education*, 100(2), 189–220. https:// doi.org/10.1002/sce.21202

Jansen, R. S., van Leeuwen, A., Janssen, J., Jak, S., & Kester, L. (2019). Self-regulated learning partially mediates the effect of self-regulated learning interventions on achievement in higher education: A meta-analysis. *Educational Research Review*, 28, 100292. https://doi.org/10.1016/j.edurev.2019.100292

*Joag, S. D. (2014). An effective method of introducing the periodic table as a crossword puzzle at the high school level. *Journal of Chemical Education*, 91(6), 864–867. https://doi.org/10.1021/ed400091w

Johnson, C. I., Bailey, S. K., & Van Buskirk, W. L. (2017). Designing effective feedback messages in serious games and simulations: A research review. In P. Wouters & H. van Oostendorp (Eds.), I*nstructional techniques to facilitate learning and motivation of serious games* (pp. 119–140). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-39298-1_7

*Johnson-Glenberg, M. C., Birchfield, D. A., Tolentino, L., & Koziupa, T. (2014). Collaborative embodied learning in mixed reality motion-capture environments: Two science studies. *Journal of Educational Psychology*, 106(1), 86–104. https://doi.org/10.1037/a0034008

Johnstone, A. H. (1991). Why is science difficult to learn? Things are seldom what they seem. Journal of Computer Assisted Learning, 7, 75–83.

Johnstone, A. H. (2000). Teaching of chemistry-logical or psychological? *Chemistry Education Research and Practice*, 1(1), 9–15.

Joseph, D. L., Chan, M. Y., Heintzelman, S. J., Tay, L., Diener, E., & Scotney, V. S. (2020). The manipulation of affect: A meta-analysis of affect induction procedures. *Psychological Bulletin*, 146(4), 355–375. https://doi.org/10.1037/bul0000224

Kalyuga, S., & Plass, J. L. (2009). Evaluating and managing cognitive load in games. In R. E. Ferdig (Ed.), *Handbook of Research on Effective Electronic Gaming in Education* (Vol. 3, pp. 719–737). IGI Global Press.

Kapur, M. (2008). Productive failure. *Cognition and Instruction*, 26(3), 379-424. https://doi.org/10.1080/07370000802212669

Karakoç, B., Eryılmaz, K., Özpolat, E. T., & Yıldırım, I_. (2020). The effect of game-based learning on student achievement: A meta-analysis study. *Technology, Knowledge and Learning*, 1–16, 207–222. https://doi.org/10. 1007/s10758-020-09471-5

Kärkkäinen, K. and S. Vincent-Lancrin (2013), “Sparking Innovation in STEM Education with Technology and Collaboration: A Case Study of the HP Catalyst Initiative”, OECD Education Working Papers, No. 91, OECD Publishing. http://dx.doi.org/10.1787/5k480sj9k442-en

*Kavak, N. (2012). ChemOkey: A game to reinforce nomenclature. *Journal of Chemical Education*, 89(8), 1047– 1049. https://doi.org/10.1021/ed3000556

Ke, F. (2016). Designing and integrating purposeful learning in game play: A systematic review. *Educational Technology Research and Development*, 64(2), 219–244. https://doi.org/10.1007/s11423-015-9418-1

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, 44(3), 486–507. https://doi.org/10.1177/00491241145432

Kester, L., Kirschner, P. A., & Van Merriënboer, J. J. (2004a). Timing of information presentation in learning statistics. *Instructional Science*, 32(3), 233-252. https://doi.org/10.1023/B:TRUC.0000024191.27560.e3

Kester, L., Kirschner, P. A., & van Merriënboer, J. J. G. (2004b). Information presentation and troubleshooting in electrical circuits. *International Journal of Science Education*, 26, 239–256. https://doi.org/10.1080/69032000072809

Kester, L., Kirschner, P. A., & van Merriënboer, J. J. G. (2006). Just-in-time information presentation: Improving learning a troubleshooting skill. *Contemporary Educational Psychology*, 31, 167–185. https://doi.org/10.1016/j.cedpsych.2005.04.002

Kester, L., Lehnen, C., Van Gerven, P. W., & Kirschner, P. A. (2006). Just-in-time, schematic supportive information presentation during cognitive skill acquisition. *Computers in Human Behavior*, 22(1), 93-112. https://doi.org/10.1016/j.chb.2005.01.008

Kester, L. & Van Merriënboer, J. J. G. (2021). Implications of the Four Component Instructional Design Model for Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (3rd ed., pp. 222–262). Cambridge University Press.

King, R. B., & Mendoza, N. B. (2020). Achievement goal contagion: Mastery and performance-based goals spread among classmates. *Social Psychology of Education*, 23(3), 795-814. https://doi.org/10.1007/s11218-020-09559-x

King, D., Ritchie, S. M., Sandhu, M., Henderson, S., & Boland, B. (2017). Temporality of emotion: Antecedent and successive variants of frustration when learning chemistry. *Science Education*, 101(4), 639–672. https:// doi.org/10.1002/sce.21277

Kim, E. S., & Willson, V. L. (2010). Evaluating pretest effects in pre–post studies. *Educational and psychological measurement*, 70(5), 744-759. https://doi.org/10.1177/0013164410366687

Kirschner, F., Paas, F., & Kirschner, P. A. (2009). A cognitive load approach to collaborative learning: United brains for complex tasks. *Educational Psychology Review*, 21(1), 31–42. https://doi.org/10.1007/s10648-008- 9095-2

Kirschner, F., Paas, F., & Kirschner, P. A. (2011). Task complexity as a driver for collaborative learning efficiency: The collective working-memory effect. *Applied Cognitive Psychology*, 25(4), 615–624. https://doi.org/ 10.1002/acp.1730

Kirschner, P. A., & Hendrick, C. (2020). How learning happens (1st ed.). Routledge.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Klopfer, E., & Thompson, M. (2020). Game-based learning in science, technology, engineering, and mathematics. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 387–408). MIT Press.

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine*, 22(17), 2693–2710. https://doi.org/10.1002/sim.1482

Knapp, F., Viechtbauer, W., Leonhart, R., Nitschke, K., & Kaller, C. P. (2017). Planning performance in schizophrenia patients: A meta-analysis of the influence of task difficulty and clinical and sociodemographic variables. *Psychological Medicine*, 47(11), 2002–2016. https://doi.org/10.1017/S0033291717000459

Knörzer, L., Brünken, R., & Park, B. (2016). Facilitators or suppressors: Effects of experimentally induced emotions on multimedia learning. *Learning and Instruction*, 44, 97–107. https://doi.org/10.1016/j.learninstruc.2016.04.002

Korbach, A., Brünken, R., & Park, B. (2018). Differentiating different types of cognitive load: A comparison of different measures. *Educational Psychology Review*, 30(2), 503–529.

https://doi.org/10.1007/s10648-017-9404-8

Kozma, R. B. (1991). Learning with media. *Review of Educational Research*, 61(2), 179–211.

Kozma, R. B. (1994a). A reply: Media and methods. E*ducational Technology Research and Development*, 42(3), 11–14. https://doi.org/10.1007/BF02298091

Kozma, R. B. (1994b). Will media influence learning? Reframing the debate. *Educational Technology Research and Development*, 42(2), 7–19. https://doi.org/10.1007/BF02299087

Krieglstein, F., Beege, M., Rey, G. D., Ginns, P., Krell, M., & Schneider, S. (2022). A Systematic Meta-analysis of the Reliability and Validity of Subjective Cognitive Load Questionnaires in Experimental Multimedia Learning Research. *Educational Psychology Review*, 1-57. https://doi.org/10.1007/s10648-022-09683-4

Kromrey, J., & Rendina-Gobioff, G. (2006). On knowing what we do not know: An empirical comparison of methods to detect publication bias in meta-analysis. *Educational and Psychological Measurement*, 66(3), 357–373. https://doi.org/10.1177/0013164405278585

Kuiper, R. (2022). AIC-type Theory-Based Model Selection for Structural Equation Models, *Structural Equation Modeling: A Multidisciplinary Journal*, 29(1), 151–158. https://doi.org/10.1080/10705511.2020.1836967

Kuiper, R. M., Hoijtink, H., & Silvapulle, M. J. (2011). An Akaike-type information criterion for model selection under inequality constraints. *Biometrika*, 98, 495–501. https://doi.org/10.1093/biomet/asr002

Kumar, J. A., Muniandy, B., & Wan Yahaya, W. A. J. (2019). Exploring the effects of emotional design and emotional intelligence in multimedia-based learning: an engineering educational perspective. *New Review of Hypermedia and Multimedia*, 25(1-2), 57–86. https://doi.org/10.1080/13614568.2019.1596169

Laffey, J. M., Sadler, T. D., Goggins, S. P., Griffin, J., & Babiuch, R. N. (2019). Mission HydroSci: Distance learning through game-based 3D virtual learning environments. In *Virtual Reality in education: Breakthroughs in research and practice* (pp. 623–643). IGI Global.

Lakens, D., & Caldwell, A. R. (2021). Simulation-based power analysis for factorial analysis of variance designs. *Advances in Methods and Practices in Psychological Science*, 4(1), 2515245920951503. https://doi.org/10.1177/2515245920951503

Lamb, R. L., Annetta, L., Firestone, J., & Etopio, E. (2018). A meta-analysis with examination of moderators of student cognition, affect, and learning outcomes while using serious educational games, serious games, and simulations. *Computers in Human Behavior*, 80, 158–167. https://doi.org/10.1016/j.chb.2017.10.040

Law, W., Elliot, A. J., & Murayama, K. (2012). Perceived competence moderates the relation between performance-approach and performance-avoidance goals. *Journal of Educational Psychology*, *104*(3), 806-819. https://doi.org/10.1037/a0027179

Lawson, A. P., & Mayer, R. E. (2022). Does the Emotional Stance of Human and Virtual Teachers in Instructional Videos Affect Learning Processes and Outcomes?. *Contemporary Educational Psychology*, 102080. https://doi.org/10.1016/j.cedpsych.2022.102080

*Lay, A.-N., & Osman, K. (2018). Developing 21st century chemistry learning through designing digital games. J*ournal of Education in Science, Environment and Health (JESEH)*, 4(1), 81–92. https://doi.org/10.21891/jeseh.387499

Lazowski, R. A., & Hulleman, C. S. (2016). Motivation interventions in education: A meta-analytic review. *Review of Educational Research*, 86(2), 602–640. https://doi.org/10.3102/0034654315617832

*le Maire, N. V., Verpoorten, D. P., Fauconnier, M.-L. S., & Colaux-Castillo, C. G. (2018). Clash of chemists: A gamified blog to master the concept of limiting reagent stoichiometry. *Journal of Chemical Education*, 95(3), 410–415. https://doi.org/10.1021/acs.jchemed.7b00256

Lee, K. M. (2004). Presence, explicated. Communication Theory, 14(1), 27–50. https://doi.org/10.1111/j.1468-2885.2004.tb00302.x

Lench, H. C., Flores, S. A., & Bench, S. W. (2011). Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: A meta-analysis of experimental emotion elicitations. *Psychological Bulletin*, 137(5), 834–855. https://doi.org/10.1037/a0024244

Leutner, D. (1993). Guided discovery learning with computer-based simulation games: Effects of adaptive and non-adaptive instructional support. *Learning and Instruction, 3(2)*, 113-132. https://doi.org/10.1016/0959-4752(93)90011-N

Levy, D. A., & Nail, P. R. (1993). Contagion: A theoretical and empirical review and reconceptualization. *Genetic, Social, and General Psychology Monographs*, 119(2), 233–284.

Li, M. C., & Tsai, C. C. (2013). Game-based learning in science education: A review of relevant research. *Journal of Science Education and Technology*, 22(6), 877–898. https://doi.org/10.1007/s10956-013-9436-x

Limelight Networks (2020). The state of online gaming – 2020 [White paper]. https://lg-static.techrepublic.com/direct/whitepapers/The_SoOG_2020_en.pdf

Linnenbrink, E. A. (2005). The Dilemma of Performance-Approach Goals: The Use of Multiple Goal Contexts to Promote Students' Motivation and Learning. *Journal of Educational Psychology*, 97(2), 197–213. https://doi.org/10.1037/0022-0663.97.2.197

Linnenbrink-Garcia, L., Patall, E. A., & Pekrun, R. (2016). Adaptive motivation and emotion in education: research and principles for instructional design. *Policy Insights From the Behavioral and Brain Sciences*, 3(2), 228–236. https://doi.org/10.1177/2372732216644450

Linnenbrink, E. A., & Pintrich, P. R. (2002). Achievement goal theory and affect: An asymmetrical bidirectional model. *Educational Psychologist*, 37(2), 69–78. https://doi.org/10.1207/S15326985EP3702_2

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. National Center for Special Education Research (NSCER 2013–3000).

Locke, E. A., & Latham, G. P. (2019). The development of goal setting theory: A half century retrospective. *Motivation Science*, 5(2), 93–105. https://doi.org/10.1037/mot0000127

Loderer, K., Pekrun, R., & Lester, J. C. (2018). Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments. *Learning and Instruction*, 70, 101162. https://doi.org/10.1016/j.learninstruc.2018.08.002

Loderer, K., Pekrun, R., & Plass, J. L. (2020). Emotional foundations of game-based learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 111–151). MIT Press.

López-López, J. A., Marín-Martínez, F., Sánchez-Meca, J., Van den Noortgate, W., & Viechtbauer, W. (2014). Estimation of the predictive power of the model in mixed-effects meta-regression: A simulation study. *British Journal of Mathematical and Statistical Psychology*, 67(1), 30–48. https://doi.org/10.1111/bmsp.12002

López-López, J. A., Page, M. J., Lipsey, M. W., & Higgins, J. P. (2018). Dealing with effect size multiplicity in systematic reviews and meta-analyses. *Research Synthesis Methods*, 9(3), 336–351. https://doi.org/10.1002/jrsm. 1310

Lou, A. J., & Jaeggi, S. M. (2020). Reducing the prior-knowledge achievement gap by using technology-assisted guided learning in an undergraduate chemistry course. *Journal of Research in Science Teaching*, 57(3), 368–392. https://doi.org/10.1002/tea.21596

*Low, K. S. (2010). Effect of the card game SynTactic on the learning of organic chemistry reactions (Master's the- sis). http://hdl.handle.net/10497/4594

Macaskill, P., Walter, S. D., & Irwig, L. (2001). A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine*, 20(4), 641–654. https://doi.org/10.1002/sim.698

Maechler, M., Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., ... & di Palma, M. A. (2021). Package 'robustbase'. Basic Robust Statistics. R package version (version 0.93-9). https://cran.r-project.org/web/packages/robustbase/

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science*, 5(4), 333–369. https://doi.org/10.1016/S0364-0213(81)80017-1

Maria, F., Dos Santos, T., & Mortimer, E. F. (2003). How emotions shape the relationship between a chemistry teacher and her high school students. *International Journal of Science Education*, 25, 1095–1110. https://doi. org/10.1080/0950069032000052216

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103, 391–410. doi:10.1037/0033-2909.103.3.391

*Martin, M. W., & Shen, Y. (2014). The effects of game design on learning outcomes. Computers in the Schools, 31(1–2), 23–42. https://doi.org/10.1080/07380569.2014.879684

Martinez-Garza, M., Clark, D., & Nelson, B. C. (2013). Digital games and the US National Research Council's science proficiency goals. *Studies in Science Education*, 49(2), 170–208. https://doi.org/10.1080/03057267.2013.839372

*Martinez-Hernandez, K. J. (2010). *Development and assessment of a chemistry-based computer video game as a learning tool* (Doctoral dissertation). ProQuest Dissertations Publishing.

Mayer, R. E. (2014a). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.005

Mayer, R. E. (2014b). *Computer games for learning: An evidence-based approach*. MIT Press.

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

Mayer, R. E. (2020). Cognitive foundations of game-based learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 83–110). MIT Press.

Mayer, R. E. (2021). *The Cambridge Handbook of Multimedia Learning*. Cambridge university press.

Mayer, R. E., & Fiorella, L. (2021). Principles for managing essential processing in multimedia learning: segmenting, pre-training, and modality principles. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (3rd ed., pp. 243–360). Cambridge university press.

Mayer, R. E., Mautone, P., & Prothero, W. (2002). Pictorial aids for learning by doing in a multimedia geology simulation game. *Journal of Educational Psychology*, 94(1), 171. https://doi.org/10.1037/0022-0663.94.1.171

McNeish, D. M., & Stapleton, L. M. (2016). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review*, 28(2), 295-314. https://doi.org/10.1007/s10648-014-9287-x

*Meesuk, K., & Srisawasdi, N. (2014). Implementation of student-associated game-based open inquiry in chemistry education: Results on students' perception and motivation. In *Proceedings of the 22nd International Conference on Computers in Education* (pp. 219–226). Asia-Pacific Society for Computers in Education.

Meinhardt, J., & Pekrun, R. (2003). Attentional resource allocation to emotional events: An ERP study. *Cognition and Emotion*, 17(3), 477-500. https://doi.org/10.1080/02699930244000039

Melo, C., Madariaga, L., Nussbaum, M., Heller, R., Bennett, S., Tsai, C. C., & van Braak, J. (2020). Editorial: Educational technology and addictions. *Computers & Education*, 145, 103730. https://doi.org/10.1016/j.compedu.2019.103730

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education*, 70, 29–40. https://doi.org/10.1016/j.compedu.2013.07.033

*Merchant, Z., Goetz, E. T., Keeney-Kennicutt, W., Cifuentes, L., Kwok, O., & Davis, T. J. (2013). Exploring 3-D virtual reality technology for spatial ability and chemistry achievement. *Journal of Computer Assisted Learning*, 29(6), 579–590. https://doi.org/10.1111/jcal.12018

Merkle, E. C., & Rosseel, Y. (2018). blavaan: Bayesian structural equation models via parameter expansion. *Journal of Statistical Software*, 85(4), 1–30. https://doi.org/10.48550/arXiv.1511.05604

Merrill, D. M. (2012). First principles of instruction. John Wiley & Sons.

Midgley, C., Kaplan, A., & Middleton, M. J. (2001). Performance-approach goals: Good for what, for whom, under what circumstances, and at what cost? *Journal of Educational Psychology*, 93, 77–86. https://doi.org/10.1037/0022-0663.93.1.77

Miller, C. S., Lehman, J. F., & Koedinger, K. R. (1999). Goals and learning in microworlds. *Cognitive Science,* 23(3), 305-336. https://doi.org/10.1016/S0364-0213(99)00007-5

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group* (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*, 151(4), 264–269.

Moreno, R., & Mayer, R. E. (2002). Learning science in virtual reality multimedia environments: Role of methods and media. *Journal of Educational Psychology*, 94(3), 598–610. https://doi.org/10.1037//0022-0663.94.3.598

Moreno, R., & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology*, 96(1), 165–173. https://doi.org/10.1037/0022-0663.96.1.165

Moreno, R., & Mayer, R. E. (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of Educational Psychology*, 97(1), 117–128. https://doi.org/10.1037/0022-0663.97.1.117

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review,* 19(3), 309–326. https://doi.org/10.1007/s10648-007-9047-2

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods*, 11(2), 364–386. https://doi.org/10.1177/1094428106291059

Mulder, J. D., & Hamaker, E. L. (2021). Three extensions of the random intercept cross-lagged panel model. *Structural Equation Modeling: A Multidisciplinary Journal*, 1-11, 638–648. https://doi.org/10.1080/10705511. 2020.1784738

Muis, K. R., Ranellucci, J., Trevors, G., & Duffy, M. C. (2015). The effects of technology-mediated immediate feedback on kindergarten students' attitudes, emotions, engagement and learning outcomes during literacy skills development. *Learning and Instruction*, 38, 1–13. https://doi.org/10.1016/j.learninstruc.2015.02.001

Murayama, K., & Elliot, A. J. (2009). The joint influence of personal achievement goals and classroom goal structures on achievement-relevant outcomes. *Journal of Educational Psychology*, 101(2), 432–447. https://doi.org/10.1037/a0014221

Muthén, L. K., & Muthén, B. (2017). *Mplus user's guide* (8^{th} Edition). Muthén & Muthén.

Muthén, B., Muthén, L. & Asparouhov, T. (2016). *Regression and mediation analysis using Mplus*. Muthén & Muthén.

National Academies of Science, National Research Council (NRC, 2012). Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century. *National Academies Press*. https://doi.org/10.17226/13398. ISBN 978-0-309-25649-0.

National Research Council [NRC]. (2009). A new biology for the 21st century. *The National Academic Press*. https://doi.org/10.17226/12764

National Research Council [NRC] (2011a). In M. A. Honey & M. L. Hilton (Eds.), *Learning science through computer games and simulations*. The National Academies Press.

National Research Council [NRC]. (2011b). *Successful K-12 STEM education: Identifying* *effective approaches in science, technology, engineering, and mathematics*. The National Academies Press. https://doi.org/10.17226/ 13158

National Research Council [NRC]. (2012a). *Discipline-based education research: understanding and improving learning in undergraduate science and engineering*. The National Academies Press. https://doi.org/10.17226/13362

National Research Council [NRC]. (2012b). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. The National Academies Press. https://doi.org/10.17226/13165

National Research Council [NRC]. (2012c). *Challenges in chemistry graduate education: A workshop summary*. The National Academies Press. https://doi.org/10.17226/13407

National Research Council [NRC]. (2014). *Undergraduate Chemistry Education: A Workshop Summary*. The National Academies Press. https://doi.org/10.17226/18555

National Research Council [NRC]. (2015). *Reaching students: What research says about effective instruction in undergraduate science and engineering*. The National Academies Press. https://doi.org/10.17226/18687

National Research Council [NRC]. (2016). *Science Literacy: Concepts, Contexts, and Consequences*. The National Academies Press. https://doi.org/10.17226/23595

Navarro, M., & Siegel, J.T. (2018). Solomon Four-Group Design. In B.B. Frey (Ed.), *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. SAGE Publications.

Nebel, S., Schneider, S., Schledjewski, J., & Rey, G. D. (2017). Goal-setting in educational video games: Comparing goal-setting theory and the goal-free effect. *Simulation & Gaming*, 48(1), 98-130. https://doi.org/10.1177/1046878116680869

Neelen, M., & Kirschner, P. A. (2020). *Evidence-informed learning design: Creating training to improve performance*. Kogan Page Publishers.

Niemivirta, M. (2002). Motivation and performance in context: The influence of goal orientations and instructional setting on situational appraisals and task performance. *Psychologia: An International Journal of Psychology in the Orient*, 45(4), 250–270. https://doi.org/10.2117/psysoc.2002.250

Noordzij, G., Giel, L., & van Mierlo, H. (2021). A meta-analysis of induced achievement goals: The moderating effects of goal standard and goal framing. *Social Psychology of Education,* 24(1), 195-245. https://doi.org/10.1007/s11218-021-09606-1

Noroozi, O., Busstra, M. C., Mulder, M., Biemans, H. J., Tobi, H., Geelen, A., ... & Chizari, M. (2012). Online discussion compensates for suboptimal timing of supportive information presentation in a digitally supported learning environment. *Educational Technology Research and Development*, 60(2), 193-221. https://doi.org/10.1007/s11423-011-9217-2

Office of Science and Technology Policy (OSTP) (2018). *Charting a Course for Success: America's Strategy for STEM Education.* https://files.eric.ed.gov/fulltext/ED590474.pdf

*Okonkwo, I. G. A. (2012). *Effects of concept mapping and simulation-game teaching strategies on students' achievement and interest in environmental concepts in chemistry* (Doctoral dissertation). University of Nigeria.

http://www.unn.edu.ng/publications/files/images/MRSOKONKWOFINALPROJECTPh.D.pdf

O'Laughlin, K. D., Martin, M. J. & Ferrer, E. (2018). Cross-sectional analysis of longitudinal mediation processes. *Multivariate Behavioral Research*, 53(3), 375–402. http://doi.org/10.1080/00273171.2018.1454822

Opfermann, M., Schmeck, A., & Fischer, H. E. (2017). Multiple representations in physics and science education–why should we use them? In D. F. Treagust, R. Duit, & H. F. Fischer (Eds.), *Multiple representations in physics education* (pp. 1–22). Springer.

Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, 84(4), 429. https://doi.org/10.1037/0022-0663.84.4.429

Paas, F., & Sweller, J. (2012). An evolutionary upgrade of cognitive load theory: Using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educational Psychology Review*, 24(1), 27–45. https://doi.org/10.1007/s10648-011-9179-2

Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 27–42). Cambridge University Press. https:// doi.org/10.1017/CBO9781139547369.004

Paas, F., Tuovinen, J. E., Van Merrienboer, J. J., & Darabi, A. A. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educational Technology Research and Development*, 53(3), 25-34. https://doi.org/10.1007/BF02504795

Pahljina-Reinić, R., & Kolić-Vehovec, S. (2017). Average personal goal pursuit profile and contextual achievement goals: Effects on students' motivation, achievement emotions, and achievement. *Learning and Individual Differences*, 56, 167–174. https://doi.org/10.1016/j.lindif.2017.01.020

Park, B., Knörzer, L., Plass, J. L., & Brünken, R. (2015). Emotional design and positive emotions in multimedia learning: An eyetracking study on the use of anthropomorphisms. *Computers & Education*, 86, 30–42. https://doi.org/10.1016/j.compedu.2015.02.016

Parker, J., Becker, K., & Sawyer, B. (2008). Re-reconsidering research on learning from media: Comments on Richard E. Clark's point of view column on serious games. *Educational Technology: The Magazine for Managers of Change in Education*, 48(1), 39–43 https://www.jstor.org/stable/44429544

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. Journal of *Educational Psychology*, 110(6), 785–797. https://doi.org/10.1037/edu0000241

Parong, J., & Mayer, R. E. (2021). Learning about history in immersive virtual reality: Does immersion facilitate learning? *Educational Technology Research and Development*, 1-19, 1433–1451. https://doi.org/10.1007/ s11423-021-09999-y

Paas, F., Tuovinen, J. E., Van Merrienboer, J. J., & Darabi, A. A. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educational Technology Research and Development*, 53(3), 25-34. https://doi.org/10.1007/BF02504795

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18(4), 315-341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R. (2020). Commentary: Self-report is indispensible to assess students' learning. *Frontline Learning Research*, 8(3), 185–193. https://doi.org/10.14786/flr.v8i3.637

Pekrun, R., Cusack, A., Murayama, K., Elliot, A. J., & Thomas, K. (2014). The power of anticipated feedback: Effects on students' achievement goals and achievement emotions. *Learning and Instruction*, 29, 115–124. https://doi.org/10.1016/j.learninstruc.2013.09.002

Pekrun, R., Elliot, A. J., & Maier, M. A. (2006). Achievement goals and discrete achievement emotions: A theoretical model and prospective test. *Journal of Educational Psychology*, 98(3), 583-597. https://doi.org/10.1037/0022-0663.98.3.583

Pekrun, R., Elliot, A. J., & Maier, M. A. (2009). Achievement goals and achievement emotions: Testing a model of their joint relations with academic performance. *Journal of Educational Psychology*, 101(1), 115–135. https://doi.org/10.1037/a0013383

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, 36(1), 36-48. https://doi.org/10.1016/j.cedpsych.2010.10.002

Pekrun, R., Hall, N. C., Goetz, T., & Perry, R. P. (2014). Boredom and academic achievement: testing a model of reciprocal causation. *Journal of Educational Psychology*, 106(3), 696–710. https://doi.org/10.1037/ a0036006

Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement emotions and academic performance: longitudinal models of reciprocal effects. *Child Development*, 88(5), 1653–1670. https://doi.org/10.1111/cdev.12704

Pekrun, R., & Linnenbrink-Garcia, L. (2014). International Handbook of Emotions in Education. Routledge.

Pellas, N., Fotaris, P., Kazanidis, I., & Wells, D. (2018). Augmenting the learning experience in primary and secondary school education: A systematic review of recent trends in augmented reality game-based learning. *Virtual Reality*, 23(4), 329–346. https://doi.org/10.1007/s10055-018-0347-2

Pilegard, C., & Mayer, R. E. (2016). Improving academic learning from computer-based narrative games. *Contemporary Educational Psychology*, 44, 12–20. http://dx.doi.org/10.1016/j.cedpsych.2015.12.002

Pilegard, C., & Mayer, R. E. (2018). Game over for Tetris as a platform for cognitive skill training. *Contemporary Educational Psychology*, 54, 29–41. https://doi.org/10.1016/j.cedpsych.2018.04.003

Plass, J. L., Heidig, S., Hayward, E. O., Homer, B. D., & Um, E. (2014). Emotional design in multimedia learning: Effects of shape and color on affect and learning. *Learning and Instruction*, 29, 128–140. https://doi.org/10. 1016/j.learninstruc.2013.02.006

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning.

Educational Psychologist, 50(4), 258–283. https://doi.org/10.1080/00461520.2015.1122533

Plass, J. L., Homer, B. D., MacNamara, A., Ober, T., Rose, M. C., Pawar, S., ... & Olsen, A. (2020). Emotional design for digital games for learning: The effect of expression, color, shape, and dimensionality on the affective quality of game characters. *Learning and Instruction*, 70, 101194. https://doi.org/10.1016/j.learninstruc.2019.01.005

Plass, J. L., & Kalyuga, S. (2019). Four ways of considering emotion in cognitive load theory. *Educational Psychology Review*, 31(2), 339-359. https://doi.org/10.1007/s10648-019-09473-5

Plass, J. L., & Kaplan, U. (2015). Emotional design in digital media for learning. In S. Y. Tettegah & M. Gartmeier (Eds.), *Emotions, technology, design, and learning* (pp. 131–161). Academic Press. https://doi.org/10.1016/ b978-0-12-801856-9.00007-4

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

Plass, J. L., Mayer, R. E., & Homer, B. D. (2020). Theoretical foundations of game-based learning and playful learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 3–24). MIT Press.

Prensky, M. (2001). *Digital game-based learning*. McGraw-Hill.

Putwain, D. W., Schmitz, E. A., Wood, P., & Pekrun, R. (2021). The role of achievement emotions in primary school mathematics: Control–value antecedents and achievement outcomes. *British Journal of Educational Psychology*, *91*(1), 347-367. https://doi.org/10.1111/bjep.12367

Raaijmakers, S. F., Baars, M., Schaap, L., Paas, F., & Van Gog, T. (2017). Effects of performance feedback valence on perceptions of invested mental effort. *Learning and Instruction*, 51, 36–46. https://doi.org/10.1016/j.learninstruc.2016.12.002

Rahimi, S., Shute, V. J., Fulwider, C., Bainbridge, K., Kuba, R., Yang, X., ... & D'Mello, S. K. (2022). Timing of learning supports in educational games can impact students' outcomes. *Computers & Education*, 190, 104600. https://doi.org/10.1016/j.compedu.2022.104600

Rahman, T., & Lewis, S. E. (2020). Evaluating the evidence base for evidence-based instructional practices in chemistry through meta-analysis. *Journal of Research in Science Teaching*, 57(5), 765–793. https://doi.org/10.1002/tea.21610

Raker, J. R., Gibbons, R. E., & Cruz-Ramírez de Arellano, D. (2019). Development and evaluation of the organic chemistry-specific achievement emotions questionnaire (AEQ-OCHEM). *Journal of Research in Science Teaching*, 56(2), 163–183. https://doi.org/10.1002/tea.21474

*Rastegarpour, H., & Marashi, P. (2012). The effect of card games and computer games on learning of chemistry concepts. *Procedia— Social and Behavioral Sciences*, 31, 597–601. https://doi.org/10.1016/j.sbspro.2011.12.111

Rawsthorne, L. J., & Elliot, A. J. (1999). Achievement goals and intrinsic motivation: A meta-analytic review. *Personality and Social Psychology Review*, 3(4), 326–344. https://doi.org/10.1207/s15327957pspr0304_3

*Renner, J. C. (2014). *Does augmented reality affect high school student's learning outcomes in chemistry?* (Doctoral dissertation). ProQuest LLC.

Riopel, M., Nenciovici, L., Potvin, P., Chastenay, P., Charland, P., Sarrasin, J. B., & Masson, S. (2020). Impact of serious games on science learning achievement compared with more conventional instruction: An overview and a meta-analysis. *Studies in Science Education*, 55(2), 169–214. https://doi.org/10.1080/03057267.2019. 1722420

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, 130(2), 261–288. https://doi.org/10. 1037/0033-2909.130.2.261

Rollins, L., Bertero, E., & Hunter, L. (2021). Developmental differences in the visual processing of emotionally ambiguous neutral faces based on perceived valence. *PloS One*, 16(8), e0256109. https://doi.org/10.1371/journal.pone.0256109

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin,* 86(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

RStudio Team (2022). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/

Rutten, N., van Joolingen, W. R., & van der Veen, J. T. (2012). The learning effects of computer simulations in science education. *Computers & Education*, 58, 136–153. https://doi.org/10.1016/j.compedu.2011.07.017

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Press.

Ryan, R. M., & Rigby, C. S. (2020). Motivational foundations of game-based learning. In J. L. Plass, R. E. Mayer & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 25–52). MIT Press.

Sabourin, J. L., & Lester, J. C. (2014). Affect and engagement in game-based learning environments. *IEEE Transactions on Affective Computing*, 5(1), 45–56. https://doi.org/10.1109/T-AFFC.2013.27

Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review*, 32(1), 77-112. https://doi.org/10.1007/s10648-019-09498-w

Salen, K., & Zimmerman, E. (2004). Rules of play: Game design fundamentals. MIT Press.

Salomon, G. (1984). Television is “easy” and print is “tough”: The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology*, 76(4), 647–658. https://doi.org/10.1037/0022-0663.76.4.647

Sarason, I. G., Sarason, B. R., Keefe, D. E., Hayes, B. E., & Shearin, E. N. (1986). Cognitive interference: Situational determinants and traitlike characteristics. *Journal of Personality and social Psychology*, 51(1), 215. https://doi.org/10.1037/ 0022-3514.51.1.215

Sato, W., Fujimura, T., & Suzuki, N. (2008). Enhanced facial EMG activity in response to dynamic facial expressions. *International Journal of Psychophysiology*, 70, 70–74. https://doi.org/10.1016/j.ijpsycho.2008.06.001

Schiefele, U. (2009). Situational and individual interest. In K. R. Wentzel & D. B. Miele (Eds.), *Handbook of motivation at school* (1st ed., pp. 197–222). Routledge.

Schneider, S., Beege, M., Nebel, S., Schnaubert, L., & Rey, G. D. (2022). The cognitive-affective-social theory of learning in digital environments (CASTLE). *Educational Psychology Review*, 34(1), 1–38. https://doi.org/10.1007/s10648-021-09626-5

Schnotz, W., Fries, S., & Horz, H. (2009). Motivational aspects of cognitive load theory. In M. Wosnitza, S. A. Karabenick, A. Efklides, & P. Nenniger (Eds.), *Contemporary motivation research: From global to local perspectives* (pp. 69–96). Hogrefe & Huber.

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, 19(4), 469-508. https://doi.org/10.1007/s10648-007-9053-4

Schola Europaea. (2019). Chemistry Syllabus – S4-S5. https://www.eursc.eu/Syllabuses/2019-01-D-46-en-2.pdf Selig, J. P., & Little, T. D. (2012). Autoregressive and cross-lagged panel analysis for longitudinal data. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 265–278). Psycnet. https://psycnet.apa.org/record/2012-07988-016

Schrader, C., Kalyuga, S., & Plass, J. L. (2021). Motivation and affect in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (3rd ed., pp. 121–131). Cambridge university press.

Senko, C., & Dawson, B. (2017). Performance-Approach Goal Effects Depend on How They Are Defined: Meta-Analytic Evidence From Multiple Educational Outcomes. *Journal of Educational Psychology*, 109(4), 574-598. https://doi.org/10.1037/edu0000160

Senko, C., & Hulleman, C. S. (2013). The role of goal attainment expectancies in achievement goal pursuit. *Journal of Educational Psychology,* 105(2), 504–521. https://doi.org/10.1037/a0031136

Senko, C., Hulleman, C. S., & Harackiewicz, J. M. (2011). Achievement goal theory at the crossroads: Old controversies, current challenges, and new directions. *Educational Psychologist*, 46(1), 26–47. https://doi.org/10.1080/00461520.2011.538646

Senko, C., & Miles, K. M. (2008). Pursuing their own learning agenda: How mastery-oriented students jeopardize their class performance. *Contemporary Educational Psychology*, 33(4), 561-583. https://doi.org/10.1016/j.cedpsych.2007.12.001

Setiawan, H., & Phillipson, S. (2019). The effectiveness of game-based science learning (GBSL) to improve students' academic achievement: A meta-analysis of current research from 2010 to 2017. *REiD (Research and Evaluation in Education)*, 5(2), 152–168. https://doi.org/10.21831/reid.v5i2.28073

Shute, V., Ke, F., & Wang, L. (2017). Assessment and adaptation in games. In P. Wouters & H. van Oostendorp (Eds.), I*nstructional techniques to facilitate learning and motivation of serious games* (pp. 59–78). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-39298-1_4

Shute, V. J., & Ke, F. (2012). Games, learning, and assessment. In D. Ifenthaler, et al. (Eds.), *Assessment in game-based learning: foundations, innovations, and perspectives* (pp. 43–58). Springer. https://doi.org/10.1007/978- 1-4614-3546-4_4

Shute, V. J., & Rahimi, S. (2021). Stealth assessment of creativity in a physics video game. *Computers in Human Behavior*, 116, 106647. https://doi.org/10.1016/j.chb.2020.106647

Sicart, M. (2014). Play matters. MIT Press.

Simpson, A. (2017). The misdirection of public policy: Comparing and combining standardised effect sizes. *Journal of Education Policy*, 32(4), 450–466. https://doi.org/10.1080/02680939.2017.1280183

Sinatra, G. M., Broughton, S. H., & Lombardi, D. O. U. G. (2014). Emotions in science education. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International Handbook of Emotions in Education* (pp. 425–446). Routledge.

Sinha, T., & Kapur, M. (2021). When problem solving followed by instruction works: Evidence for productive failure. *Review of Educational Research*, 91(5), 761-798. https://doi.org/10.3102/00346543211019105

Sirhan, G. (2007). Learning difficulties in chemistry: An overview. *Journal of Turkish science education*, 4(2), 2.

Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology*, 64(2), 489–528. https://doi.org/10.1111/j.1744-6570.2011.01190.x

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. Appleton-Century-Crofts.

Skulmowski, A., & Xu, K. M. (2022). Understanding cognitive load in digital and online learning: A new perspective on extraneous cognitive load. *Educational Psychology Review*, 34(1), 171–196. https://doi.org/10.1007/s10648-021-09624-7

Slavin, R. E. (2008). Perspectives on evidence-based research in education—What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, 37(1), 5–14. https://doi.org/10.3102/0013189X08314117

Slavin, R., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational Evaluation and Policy Analysis*, 31(4), 500–506. https://doi.org/10.3102/0162373709352369

Solmaz, S., & Van Gerven, T. (2022). Interactive CFD simulations with virtual reality to support learning in mixing. *Computers & Chemical Engineering*, 156, 107570. https://doi.org/10.1016/j.compchemeng.2021.107570

Sommet, N., Elliot, A. J., & Sheldon, K. M. (2021). Achievement goal complexes: Integrating the "what" and the "why" of achievement motivation. In O. P. John & R. W. Robins (Eds.), *Handbook of personality: Theory and research* (p. 104–121). The Guilford Press.

*Sousa Lima, M. A., Monteiro, Á. C., Melo Leite Junior, A. J., de Andrade Matos, I. S., Alexandre, F. S. O., Nobre, D. J., Monteiro, A. J., & da Silva Júnior, J. N. (2019). Game-based application for helping students review chemical nomenclature in a fun way. *Journal of Chemical Education*, 96(4), 801–805. https://doi.org/ 10.1021/acs.jchemed.8b00540

*Srisawasdi, N., & Panjaburee, P. (2019). Implementation of game-transformed inquiry-based learning to promote the understanding of and motivation to learn chemistry. *Journal of Science Education and Technology*, 28(2), 152–164. https://doi.org/10.1007/s10956-018-9754-0

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, 5, 60–78. https://doi.org/10.1002/jrsm.1095

Stark, L., Brünken, R., & Park, B. (2018). Emotional text design in multimedia learning: A mixed-methods study using eye tracking. *Computers & Education*, 120, 185–196. https://doi.org/10.1016/j.compedu.2018.02.003

Steegh, A., Höffler, T., Höft, L., & Parchmann, I. (2021). First steps toward gender equity in the chemistry Olympiad: Understanding the role of implicit gender-science stereotypes. *Journal of Research in Science Teaching*, 58(1), 40–68. https://doi.org/10.1002/tea.21645

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rucker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, 343, d4002. https://doi.org/10. 1136/bmj.d4002

Stewart, J., Bleumers, L., Van Looy, J., Mariën, I., All, A., Schurmans, D., ... & Misuraca, G. (2013). *The potential of digital games for empowerment and social inclusion of groups at risk of social and economic exclusion: evidence and opportunity for policy*. Joint Research Centre, European Commission.

*Stringfield, T. W., & Kramer, E. F. (2014). Benefits of a game-based review module in chemistry courses for non-majors. *Journal of Chemical Education*, 91(1), 56–58. https://doi.org/10.1021/ed300678f

Strunk, K. K., Lester, W. S., Lane, F. C., Hoover, P. D., & Betties, J. S. (2020). Testing the mastery-avoidance construct in achievement goal theory: A meta-analytic confirmatory factor analysis (MA-CFA) of two achievement goals scales. *Educational Psychology*, 1-18. https://doi.org/10.1080/01443410.2020.1824268

*Su, C.-H., & Cheng, T.-W. (2019). A sustainability innovation experiential learning model for virtual reality chemistry laboratory: An empirical study with PLS-SEM and IPMA. *Sustainability*, 11(4), 1027. https://doi. org/10.3390/su11041027

*Sugiyarto, K. H., Ikhsan, J., & Lukman, I. R. (2018). The use of an android–based-game in the team assisted individualization to improve students' creativity and cognitive achievement in chemistry. *Journal of Physics: Conference Series*, 1022(1), 012037. https://doi.org/10.1088/1742-6596/1022/1/012037

Sullivan, G. M., & Artino Jr, A. R. (2013). Analyzing and interpreting data from Likert-type scales. *Journal of graduate medical education*, 5(4), 541-542. https://doi.org/10.4300/JGME-5-4-18

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, 22(2), 123-138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296. https://doi.org/10.1023/A:1022193728205

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. Educational Psychology Review, 31(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Taber, K. S. (2009). Learning at the symbolic level. In J. K. Gilbert & D. Treagust (Eds.), *Multiple* *representations in chemical education* (pp. 75–105). Springer. https://doi.org/10.1007/978-1-4020-8872-85

Taber, K. S. (2013). Revisiting the chemistry triplet: Drawing upon the nature of chemical knowledge and the psychology of learning to inform chemistry education. *Chemistry Education Research and Practice*, 14(2), 156–168. https://doi.org/10.1039/c3rp00012e

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48, 1273-1296. https://doi.org/10.1007/s11165-016-9602-2

Talanquer, V. (2011). Macro, submicro, and symbolic: The many faces of the chemistry "triplet". *International Journal of Science Education*, 33(2), 179–195. https://doi.org/10.1080/09500690903386435

Talsma, K., Schüz, B., Schwarzer, R., & Norris, K. (2018). I believe, therefore I achieve (and vice versa): A meta-analytic cross-lagged panel analysis of self-efficacy and academic performance. *Learning and Individual Differences*, 61, 136–150. https://doi.org/10.1016/j.lindif.2017.11.015

The European Chemical Industry Council (Cefic, 2023). Facts and Figures 2023. Retrieved June 19, 2023, from https://cefic.org/app/uploads/2023/03/2023-Facts-and-Figures.pdf

Thompson, C. G., & von Gillern, S. (2020). Video-game based instruction for vocabulary acquisition with English language learners: A Bayesian meta-analysis. *Educational Research Review*, 30, 100332. https://doi.org/10. 1016/j.edurev.2020.100332

Tobias, S., & Fletcher, J. D. (2012). Reflections on "a review of trends in serious gaming". *Review of Educational Research*, 82(2), 233–237. https://doi.org/10.3102/0034654312450190

Tobias, S., Fletcher, J. D., Dai, D. Y., & Wind, A. P. (2011). Review of research on computer games. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 127–222). Information Age.

Tokac, U., Novak, E., & Thompson, C. G. (2019). Effects of game-based learning on students' mathematics achievement: A meta-analysis. *Journal of Computer Assisted Learning*, 35(3), 407–420. https://doi.org/10. 1111/jcal.12347

Towns, M., & Kraft, A. (2011). Review and synthesis of research in chemical education from 2000–2010. In Paper presented at the Second Committee Meeting on the Status, Contributions, and Future Directions of Discipline-Based Education Research. National Academies https://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_072594.pdf

Tsai, Y. L., & Tsai, C. C. (2018). Digital game-based second-language vocabulary learning and conditions of research designs: A meta-analysis study. *Computers & Education*, 125, 345–357. https://doi.org/10.1016/j. compedu.2018.06.020

Tsai, Y. L., & Tsai, C. C. (2020). A meta-analysis of research on digital game-based science learning. *Journal of Computer Assisted Learning*, 36(3), 280–294. https://doi.org/10.1111/jcal.12430

Tsai, M. J., Wu, A. H., & Wang, C. Y. (2022). Pre-training and cueing effects on students' visual behavior and task outcomes in game-based learning. *Computers in Human Behavior Reports*, 6, 100188. https://doi.org/10.1016/j.chbr.2022.100188

Tsui, C. Y., & Treagust, D. F. (2013). Introduction to multiple representations: Their importance in biology and biological education. In D. F. Treagust & C. Y. Tsui (Eds.), *Multiple representations in biological education* (pp. 3–18). Springer.

Tze, V. M., Daniels, L. M., & Klassen, R. M. (2016). Evaluating the relationship between boredom and academic outcomes: a meta-analysis. *Educational Psychology Review*, 28(1), 119–144. https://doi.org/10.1007/s10648-015-9301-y.

Um, E. R., Plass, J. L., Hayward, E. O., & Homer, B. D. (2012). Emotional design in multimedia learning. *Journal of Educational Psychology*, 104(2), 485–498. https://doi.org/10.1037/a0026609

U.S. Department of Education, Office of Educational Technology (OET) (2017). *Reimagining the Role of Technology in Education: 2017 National Education Technology Plan Update*. https://tech.ed.gov/files/2017/01/NETP17.pdf

Utman, C. H. (1997). Performance effects of motivational state: A meta-analysis. *Personality and Social Psychology Review*, 1, 170-182. doi: 10.1207/sl5327957pspiO102_4

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, 39(2), 111–133. https://doi.org/10.1207/ s15326985ep3902_3

van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2019). Effects of flipping the classroom on learning outcomes and satisfaction: A meta-analysis. *Educational Research Review*, 28, 100281. https://doi.org/10.1016/j. edurev.2019.05.003

van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review*, 22, 271-296. https://doi.org/10.1007/s10648-010-9127-6

van Harsel, M., Hoogerheide, V., Verkoeijen, P., & van Gog, T. (2020). Examples, practice problems, or both? Effects on motivation and learning in shorter and longer sequences. *Applied Cognitive Psychology*, *34*(4), 793-812. https://doi.org/10.1002/acp.3649

van Merriënboer, J. J. G., & Kester, L. (2014). The four-component instructional design model: Multimedia principles in environments for complex learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 104–148). Cambridge University Press. https://doi.org/10.1017/ CBO9781139547369.007

van Merriënboer, J. J., & Kirschner, P. A. (2018). *Ten steps to complex learning: A systematic approach to four component instructional design*. Routledge.

van Yperen, N. W., Blaga, M., & Postmes, T. (2014). A meta-analysis of self-reported achievement goals and nonself-report performance across three achievement domains (work, sports, and education). *PloS one*, 9(4), e93594. https://doi.org/10.1371/journal.pone.0093594

van Yperen, N. , Blaga, M. , & Postmes, T. (2015). A meta-analysis of the impact of situationally induced achievement goals on task performance. *Human Performance*, 28 (2), 165-182. https://doi.org/10.1080/08959285.2015.1006772

Vandercruysse, S., Desmet, E., Vandewaetere, M., & Elen, J. (2015). Integration in the curriculum as a factor in math-game effectiveness. In J. Torbeyns et al. (eds.), *Describing and Studying Domain-Specific Serious Games* (pp. 133-153). Springer.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Viechtbauer, W. (2017). *Metafor package: bias and sensitivity diagnostics*. Retrieved, June 6, 2019, from https://stats.stackexchange.com/questions/155693/metafor-package-bias-and-sensitivity-diagnostics

Viechtbauer, W., López-López, J. A., Sánchez-Meca, J., & Marín-Martínez, F. (2015). A comparison of procedures to test for moderators in mixed-effects meta-regression models. *Psychological Methods*, 20(3), 360–374. https://doi.org/10.1037/met0000023

Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research*, 34(3), 229–243. https://doi.org/10.2190/FLHV-K4WA-WPVQ-H0YM

Vygotsky, L. S. (1978). Mind in society: The development of higher mental processes. Harvard University Press. Wandersee, J. H., Fischer, K. M., & Moody, D. E. (2000). The Nature of Biology Knowledge. In K. M. Fisher, J. H. Wandersee, & D. E. Moody (Eds.), *Mapping biology knowledge* (Vol. 11, pp. 25–38). Springer.

Warfa, A.-R. M. (2016). Using cooperative learning to teach chemistry: A meta-analytic review. *Journal of Chemical Education*, 93, 248–255. https://doi.org/10.1021/acs.jchemed.5b00608

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond “p < 0.05”. *American Statistician*, 73, 1–19. https://doi.org/10.1080/00031305.2019.1583913

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, 92(4), 548–573. https://doi.org/10.1037/0033-295X.92.4.548

*Weng, H. J., Chang, D. F., & Shyu, H. Y. (2015). Testing the effect of mnemonic strategy embedded in digital game. *ICIC Express Letters*, 9(3), 827–833.

What Works Clearinghouse [WWC]. (2019). Standards handbook (Version 4.1). Institute of Education Sciences. Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Willson, V. L., & Putnam, R. R. (1982). A meta-analysis of pretest sensitization effects in experimental design. *American Educational Research Journal*, 19(2), 249-258.

Wirthwein, L., Sparfeldt, J. R., Pinquart, M., Wegerer, J., & Steinmayr, R. (2013). Achievement goals and academic achievement: A closer look at moderating factors. *Educational Research Review*, 10, 66-89. https://doi.org/10.1016/j.edurev.2013.07.001

Wolters, C. A. (2004). Advancing achievement goal theory: Using goal structures and goal orientations to predict students’ motivation, cognition, and achievement. *Journal of Educational Psychology*, 96(2), 236–250. https://doi.org/10.1037/0022-0663.96.2.236

Wong, R. M., & Adesope, O. O. (2021). Meta-analysis of emotional designs in multimedia learning: A replication and extension study. *Educational Psychology Review*, 33(2), 357-385. https://doi.org/10.1007/s10648-020-09545-x

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100. https://doi.org/10.1111/j.14697610.1976.tb00381.x

*Wood, J., & Donnelly-Hermosillo, D. F. (2019). Learning chemistry nomenclature: Comparing the use of an electronic game versus a study guide approach. *Computers & Education*, 141, 103615. https://doi.org/10.1016/j.compedu.2019.103615

Wu, H. K., & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education*, 88(3), 465–492. https://doi.org/10.1002/sce.10126

Wouters, P., Paas, F., & van Merriënboer, J. J. G. (2008). How to optimize learning from animated models: A review of guidelines based on cognitive load. *Review of Educational Research*, 78(3), 645–675. https://doi.org/10.3102/0034654308320320

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, 105(2), 249–265. https://doi.org/10.1037/a0031311

Wouters, P., & van Oostendorp, H. (2013). A meta-analytic review of the role of instructional support in game-based learning. *Computers & Education*, 60(1), 412–425. https://doi.org/10.1016/j.compedu.2012.07.018

Wouters, P., & van Oostendorp, H. (2017). Overview of instructional techniques to facilitate learning and motivation of serious games. In P. Wouters & H. van Oostendorp (Eds.), *Instructional techniques to facilitate learning and motivation of serious games* (pp. 1–16). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-39298-1_1

Yang, J. C., Chung, C. J., & Chen, M. S. (2022). Effects of performance goal orientations on learning performance and in-game performance in digital game-based learning. *Journal of Computer Assisted Learning*, 38(2), 422-439. https://doi.org/10.1111/jcal.12622

Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., Simeoni, Z., Tran, M., & Yukhymenko, M. (2012). Our princess is in another castle: A review of trends in serious gaming for education. *Review of Educational Research*, 82(1), 61–89. https://doi.org/10.3102/0034654312436980

Zhao, X., Lynch Jr, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, 37(2), 197-206. https://doi.org/10.1086/651257

Zhao, F., & Mayer, R. E. (2023). Benefits of turning the illustrations in a narrated slideshow into cartoons: An extension of the positivity principle. *Learning and Instruction*, 86, 101779. https://doi.org/10.1016/j.learninstruc.2023.101779

Supplementary materials

Chapter 2 Supplementary Materials

Table S2.1

Moderator information for individual study

Author	ALCG	AI	NGS	UG	R	SS	PS	LO
Akkuzu & Uyulgan (2016)	Passive	Yes	One	Multiple	QED	62	Published	Cognition
Cahyana et al. (2017)	Active	Yes	One	One	QED	40	Published	Cognition
Cha et al. (2017)	Active	Yes	Multiple	Multiple	QED	198	Published	Retention
Chee & Tan (2012)	Passive	Yes	Multiple	Multiple	QED	77	Published	Retention
Chimeno et al. (2006)	Passive	Yes	One	One	QED	40	Published	Retention
da Silva Júnior et al. (2018)	Active	Yes	Multiple	One	QED	246	Published	Retention
Daubenfeld & Zenker (2015)	Unknown	Yes	One	Multiple	QED	46	Published	Retention
Fatokun et al. (2016)	Unknown	Unknown	One	Multiple	QED	96	Published	Both
Gupta (2019)	Active	No	One	One	RCT	67	Published	Both
Halpern et al. (2012)	Unknown	No	Multiple	One	RCT	136	Published	Retention
Hodges et al. (2018)	Passive	Yes	One	Multiple	QED	351	Published	Retention
Jagodziński & Wolski (2015)	Passive	Yes	One	One	QED	100	Published	Both
Joag (2014)	Active	Yes	One	One	QED	104	Published	Cognition
Johnson-Glenberg et al. (2014)	Passive	Yes	One	Multiple	QED	51	Published	Cognition
Kavak (2012)	Passive	Yes	One	Multiple	QED	49	Published	Cognition
Lay & Osman (2018)	Active	Yes	One	Multiple	QED	131	Published	Retention
le Maire et al. (2018)	Unknown	Yes	Multiple	One	QED	210	Published	Retention
Low (2010)	Active	Yes	One	Multiple	QED	75	Gray literature	Retention
Martin & Shen (2014)	Unknown	No	Multiple	One	RCT	70	Published	Retention
Martinez-Hernandez (2010)	Active	No	One	One	QED	40	Gray literature	Retention
Merchant et al. (2013)	Passive	Yes	Multiple	One	QED	382	Published	Retention
Okonkwo (2012)-1	Active	Yes	One	Multiple	QED	234	Gray literature	Retention
Rastegarpour & Marashi (2012)	Unknown	Unknown	One	One	RCT	105	Gray literature	Cognition
Renner (2014)	Active	Yes	One	One	QED	78	Gray literature	Retention
Sousa Lima et al. (2019)	Active	Yes	One	One	QED	144	Published	Retention
Srisawasdi & Panjaburee (2019)	Passive	Yes	One	One	QED	62	Published	Cognition
Stringfield & Kramer (2014)	Passive	Yes	One	Multiple	QED	120	Published	Retention

Author	ALCG	AI	NGS	UG	R	SS	PS	LO
Su & Cheng (2019)	Passive	Unknown	One	One	RCT	72	Published	Cognition
Sugiyarto et al. (2018)	Passive	Unknown	One	Multiple	QED	64	Gray literature	Retention
Weng et al. (2015)	Active	Yes	Multiple	One	RCT	135	Published	Cognition
Wood & Donnelly-Hermosillo (2019)	Passive	Yes	One	Multiple	QED	470	Published	Cognition

Note. ALCG = activity level of control group; AI = additional instruction; NGS = No. of game sessions; UG = user grouping; R = randomization; SS = sample size; PS = publication source; LO = leaning outcome.

Table S2.2

An example of GBL activities for each game genre

Game genre	Example	GBL activities
Puzzle game	Martin & Shen (2014)	In the game Element Solitaire, players place element “cards” on a periodic table based on basic information about the element. With each correct placement, they get points based on speed and accuracy. The goal is to place all elements on the periodic table as fast and accurate as possible.
Action game	Gupta (2019)	The game Molebots teaches chemical nomenclature. Players use a map to find molecules. The molebots are molecules with wrong chemical names and need to be shot by a laser gun. When pointing the molecules, players see their 3D structures. The goal is to shoot as many molebots as possible. With each correct shoot, they get points and an energy boost. When the energy meter is full, players levels up. To win the game, players must complete all four levels.
Adventure game	Daubenfeld & Zenker (2015)	The game has a 3D landscape with different locations referring to seven levels or seven chapters of phase equilibria. In each level, players get a background story, a picture of the location, and further learning content such as lectures and its notes, weblinks and online self-assessment tests. Players play the game in groups. At certain location, players need to solve a puzzle. At the end of each level, one random player of the team gets a level examination which asks the player to solve a chemistry problem. If the chosen player passes the level examination, the whole team get immediate feedback on the performance, bonus points, and the password to next level. In the final in-game examination, the player plays three different games individually (i.e., Chemory, three-component game and Labyrinth) and get bonus points.
Strategy game	Low (2010)	SynTacic© is a turn-based card game, which teaches organic synthesis. Players get a pool of four different cards: Reactant, Product, Reagent Action, and Tactic Action cards. Reagent and Product card give structure formula and IUPAC names of compounds. Reagent Action cards give the reagents, names, conditions, and mechanism of the chemical reactions. Tactic Action cards allow players to do things beyond the rules, such as steal a card from another player or force another player to skip a turn. At the beginning, players randomly draw one Reagent card and one Product card to start the synthesis and five action cards. During the players’ turn, they take one Main action (i.e., lay out, exchange, or organize cards) and may also first use one Tactic Action card and then draw one Action card. They need to lay out a Reagent card to complete the synthesis. Players think multiple syntheses and come up with the synthesis that need least number of cards. The first player who completes the synthesis win the game.
Role-playing game	Chen et al. (2014)	In the game Alchemist’s Fort, the player plays a scientist and defeats two monsters and one villain. The game has three stages with seven missions. To go to the next stage, they need to collect items and use them to solve the problems in a certain time limit. Players learn and apply knowledge on chemical reactions to accomplish the missions.
Simulation game	Hodges et al. (2018)	The game Blended Reality Environment blends interactive simulation game with a hand-on experiment to teach redox reaction. In the beginning, player get worksheets with background knowledge and the procedures of “gold rush” lab. Then, they get an interactive simulation game on “laboratory manual”, in which they learn more complex chemistry concepts, review difficult concepts, answer questions, and get real-time feedback on their solutions. After that, they propose a hypothesis, conduct the experiment, and indicate in the game whether the experiment is successful or not. If not, they play a video of the successful reaction and review the 3D simulation of the reaction. Finally, players write a report on the investigation.

Figure S2.1

Funnel plot for cognition

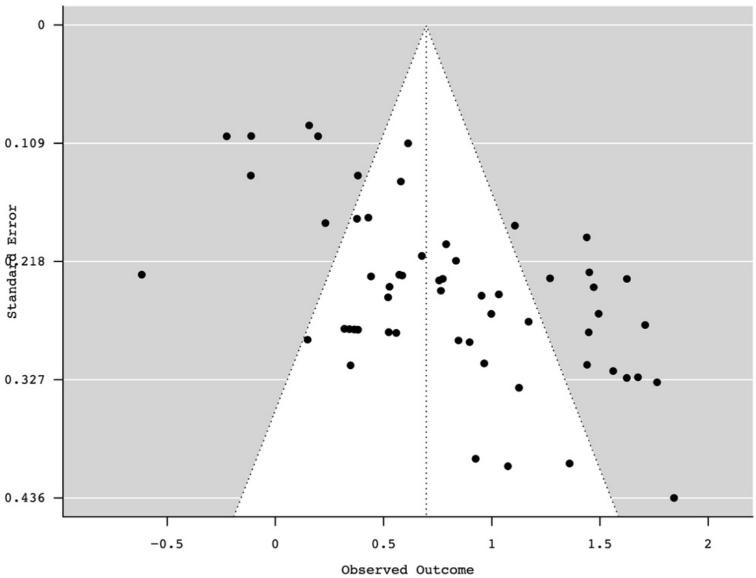
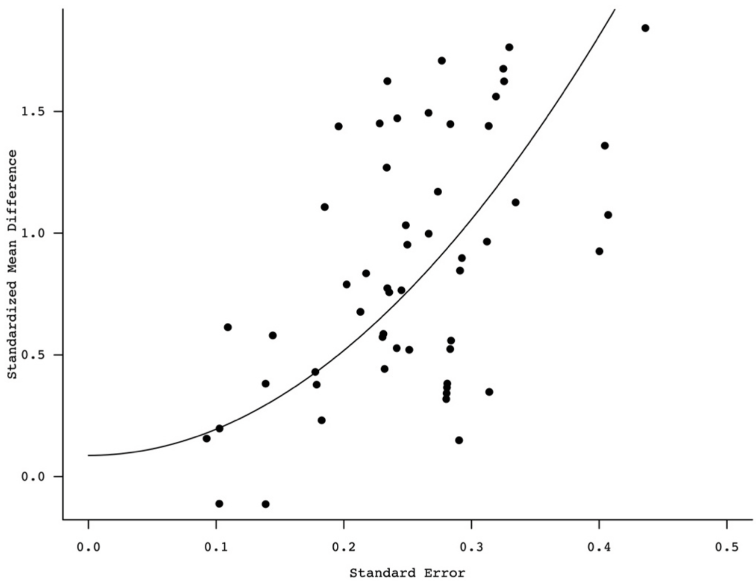


Figure S2.2

Funnel plot with the standard errors against the observed standardized mean differences



Chapter 3 Supplementary Materials

Table S3.1

Learning objectives (based on the four-component instructional design model), game levels, characters, and scenes of CosmiClean

	Details
Learning objectives	After playing the game, learners will be able to recycle resources by 1) investigating the difference between the properties of common materials, such as particle size, melting point, magnetic metal, non-ferrous metal, phase, boiling point, solubility, solvent and density of plastics, woods, iron, bricks, copper, sand, fuel, and gold 2) selecting the appropriate processors such as sieve, melter, magnet, shredder, non-ferrous separator, stream separator, boiler, dissolver, and centrifuge 3) formulating the best sequence of separations based on cost, safety, and accuracy 4) performing the recycling within the game
Game levels	The game has 57 levels: 1) Level 1-35 introduce new processors and materials 2) Level 36-57 involve recycling materials in daily life, such as IKEA furniture, party stuff and construction waste
Characters	The game includes three characters: 1) AI: an artificial intelligence who has access to all areas and systems of the spaceship and is required to restore the spaceship 2) Bob: the pilot who serves as the narrative hub introducing new challenges and problems 3) MEMO: a robot who follows the orders from AI to repair and maintain the spaceship
Scenes	The game has six scenes: 1) Cargo Bay: a large, open area where the player places processors 2) Bridge: the central control room of the spaceship 3) Inventory Hold: Displaying all the resources the player collected 4) Omnifacturing Lab: 3D-printing new Processors, MEMO units or replacement components for the ship 5) AI Core: Server room where the player changes game settings such as language, the volume of background music 6) Personnel Quarters: Bob’s main residence

Table S3.2

Standardized total effect and indirect effects of mediation

	Timing of domain-specific info presentation		Timing of game-specific info presentation	
	β	95% CI	*β*	95% CI
Test performance				
Total effect	.17*	[.02, .33]	.08	[-.07, .23]
Indirect effect via Mental effort	-.02	[-.07, .02]	.04	[-.01, .09]
Time-on-task				
Total effect	-.01	[-.17, .15]	-.01	[-.17, .14]
Indirect effect via Mental effort	.02	[-.03, .07]	-.04	[-.09, .01]
Mastery-approach goal				
Total effect	-.17*	[-.30, -.04]	.06	[-.06, .18]
Total indirect effect	-.03	[-.07, .01]	.03	[-.01, .07]
Specific indirect effect via Mental effort	-.001	[-.01, .01]	.001	[-.02, .02]
Specific indirect effect via Perceived competence	-.03	[-.07, .01]	.03	[-.01, .07]
Specific indirect effect via Mental effort + Perceived competence	0	[-.003, .002]	.001	[-.01, .01]
Performance-approach goal				
Total effect	-.21*	[-.32, -.09]	.06	[-.05, .18]
Total indirect effect	-.02	[-.05, .01]	.02	[-.01, .02]
Specific indirect effect via Mental effort	-.002	[-.01, .01]	.004	[-.01, .02]
Specific indirect effect via Perceived competence	-.02	[-.05, .01]	.02	[-.01, .04]
Specific indirect effect via Mental effort + Perceived competence	0	[-.002, .001]	0	[-.003, .004]
Mastery-avoidance goal				
Total effect	-.17	[-.29, -.05]	-.03	[-.14, .09]
Total indirect effect	-.01	[-.04, .02]	.02	[-.02, .05]
Specific indirect effect via Mental effort	-.003	[-.02, .01]	.01	[-.01, .03]
Specific indirect effect via Perceived competence	-.01	[-.04, .01]	.01	[-.01, .03]
Specific indirect effect via Mental effort + Perceived competence	0	[-.001, .001]	0	[-.002, .002]
Performance-avoidance goal				
Total effect	-.14	[-.26. -.02]	.06	[-.06, .18]
Total indirect effect	-.01	[-.04, .02]	.02	[-.01, .04]
Specific indirect effect via Mental effort	-.01	[-.03, .01]	.01	[-.01, .03]
Specific indirect effect via Perceived competence	-.004	[-.02, .01]	.004	[-.01, .02]

	Timing of domain-specific info presentation		Timing of game-specific info presentation	
	β	95% CI	β	95% CI
Specific indirect effect via Mental effort + Perceived competence	0	[0, 0]	0	[-.001, .001]
Enjoyment				
Total effect	-.14	[-.31, .02]	.01	[-.09, .12]
Total indirect effect	-.10	[-.20, .01]	.01	[-.09, .12]
Specific indirect effect via Mental effort	.004	[-.01, .02]	-.01	[-.03, .02]
Specific indirect effect via Perceived control	-.03	[-.07, .02]	.03	[-.02, .07]
Specific indirect effect via Perceived value	-.07	[-.15, .01]	-.02	[-.14, .10]
Specific indirect effect via Mental effort + Perceived control	-.004	[-.02, .01]	.01	[-.01, .03]
Specific indirect effect via Mental effort + Perceived value	.001	[-.01, .01]	-.002	[-.02, .01]
Anger				
Total effect	.04	[-.13, .20]	-.08	[-.18, .03]
Total indirect effect	.09	[-.01, .19]	-.08	[-.18, .03]
Specific indirect effect via Mental effort	.01	[-.03, .05]	-.03	[-.07, .01]
Specific indirect effect via Perceived control	.03	[-.01, .07]	-.04	[-.11, .03]
Specific indirect effect via Perceived value	.06	[-.01, .13]	.01	[-.03, .05]
Specific indirect effect via Mental effort + Perceived control	.01	[-.02, .03]	-.02	[-.04, .01]
Specific indirect effect via Mental effort + Perceived value	0	[-.003, .002]	.001	[-.01, .01]
Boredom				
Total effect	.13	[-.06, .31]	-.02	[-.13, .09]
Total indirect effect	.10	[-.002, .21]	-.02	[-.13, .09]
Specific indirect effect via Mental effort	.003	[-.01, .02]	-.01	[-.03, .01]
Specific indirect effect via Perceived control	.02	[-.02, .06]	-.02	[-.05, .02]
Specific indirect effect via Perceived value	.08	[-.01, .17]	.02	[-.08, .11]
Specific indirect effect via Mental effort + Perceived control	.003	[-.01, .01]	-.01	[-.02, .01]
Specific indirect effect via Mental effort + Perceived value	-.001	[-.01, .01]	.002	[-.01, .02]
Shame				
Total effect	-.02	[-.20, .17]	-.10	[-.20, .01]
Total indirect effect	.06	[-.04, .16]	-.10	[-.20, .01]
Specific indirect effect via Mental effort	.01	[-.02, .03]	-.01	[-.04, .02]
Specific indirect effect via Perceived control	.06	[-.04, .15]	-.06	[-.15, .04]
Specific indirect effect via Perceived value	-.01	[-.04, .02]	-.002	[-.01, .01]
Specific indirect effect via Mental effort + Perceived control	.01	[-.02, .04]	-.02	[-.06, .01]
Specific indirect effect via Mental effort + Perceived value	0	[-.001, .001]	0	[-.002, .002]

Note. N = 145; β = regression coefficient; CI = confidence interval (if 95% CI contains zero, it means β is not statistically significant); Total effect = the effect of independent variable on dependent variable including any indirect effects through a mediator (the sum of the direct effect and [total] indirect effect); Direct effect = the effect of independent variable on dependent variable excluding any indirect effects through a mediator; Indirect effect = a sequence of the effect of independent variable on dependent variable through a mediator (calculated by multiplying the paths coefficients that constitute the effect); Specific indirect effect = the indirect effect of independent variable on dependent variable through a given mediator controlling for all other included mediators; Total indirect effect = the effect of independent variable on dependent variable through multiple mediators (the sum of specific indirect effects).

Figure S3.1

The scene of level 16: the particles to be recycled – separate plastics, concrete, and iron from plastics and concrete with iron fillings (top) and the processors to be chosen (bottom)

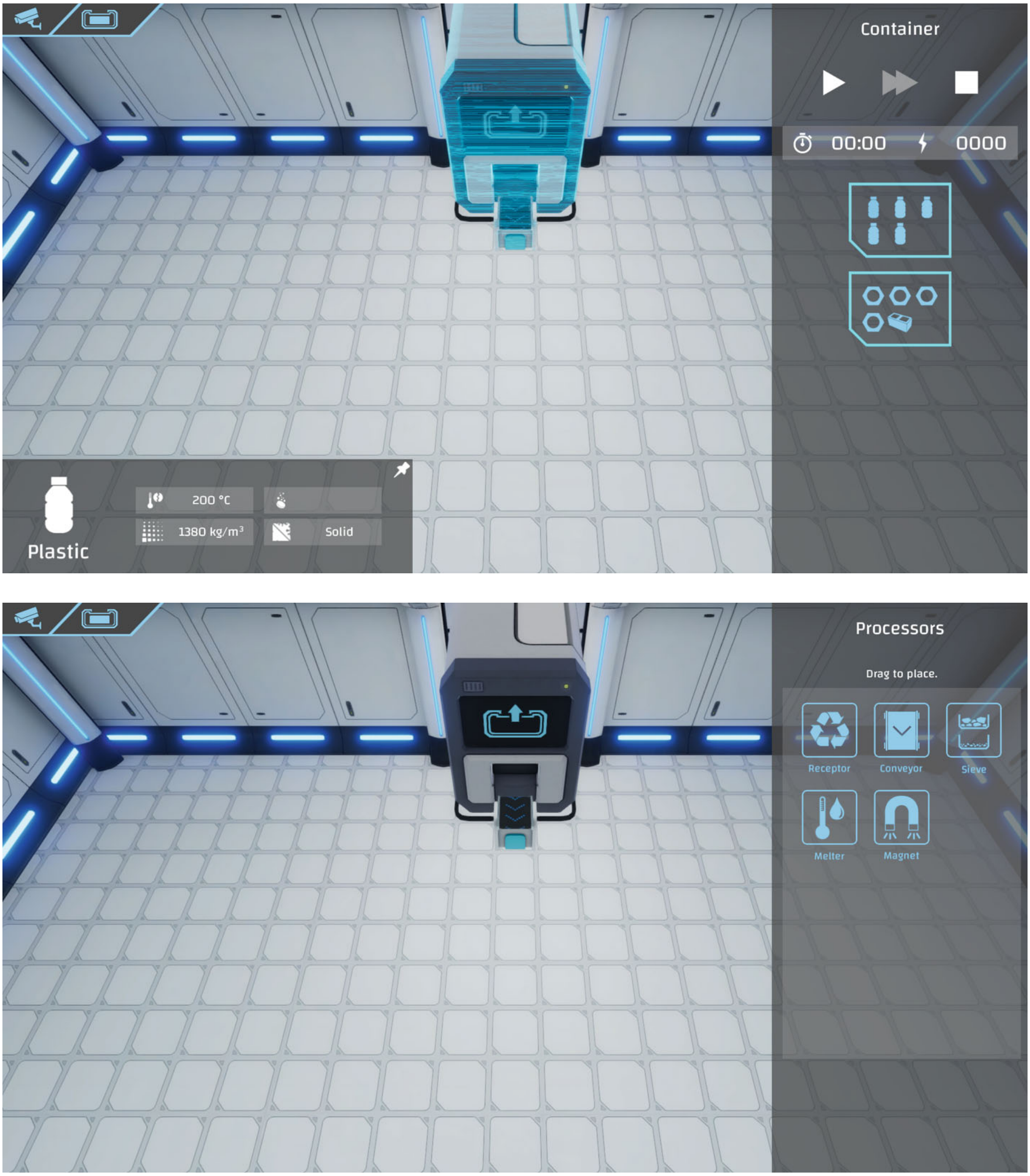
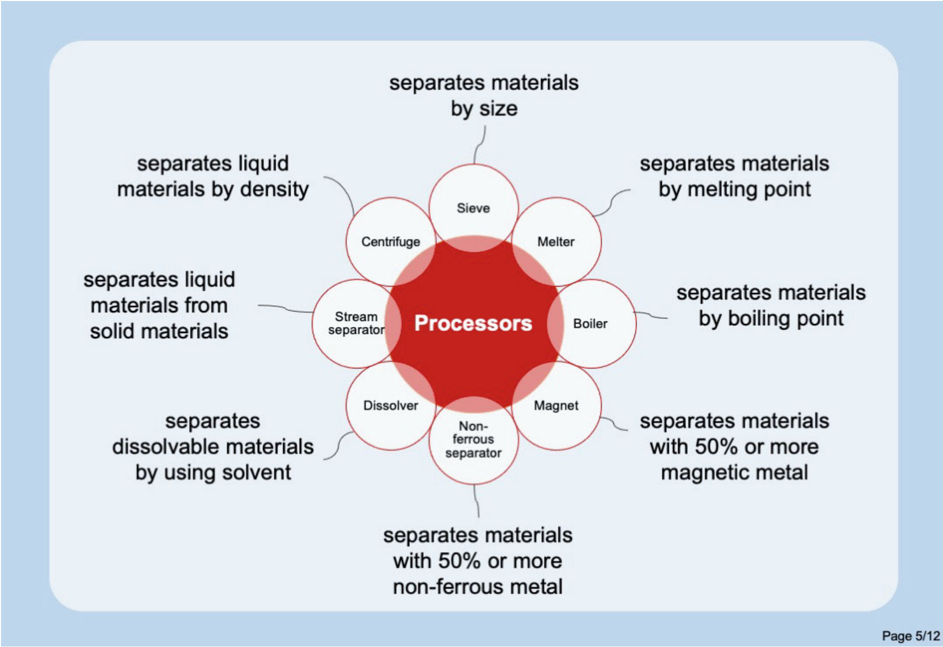


Figure S3.2

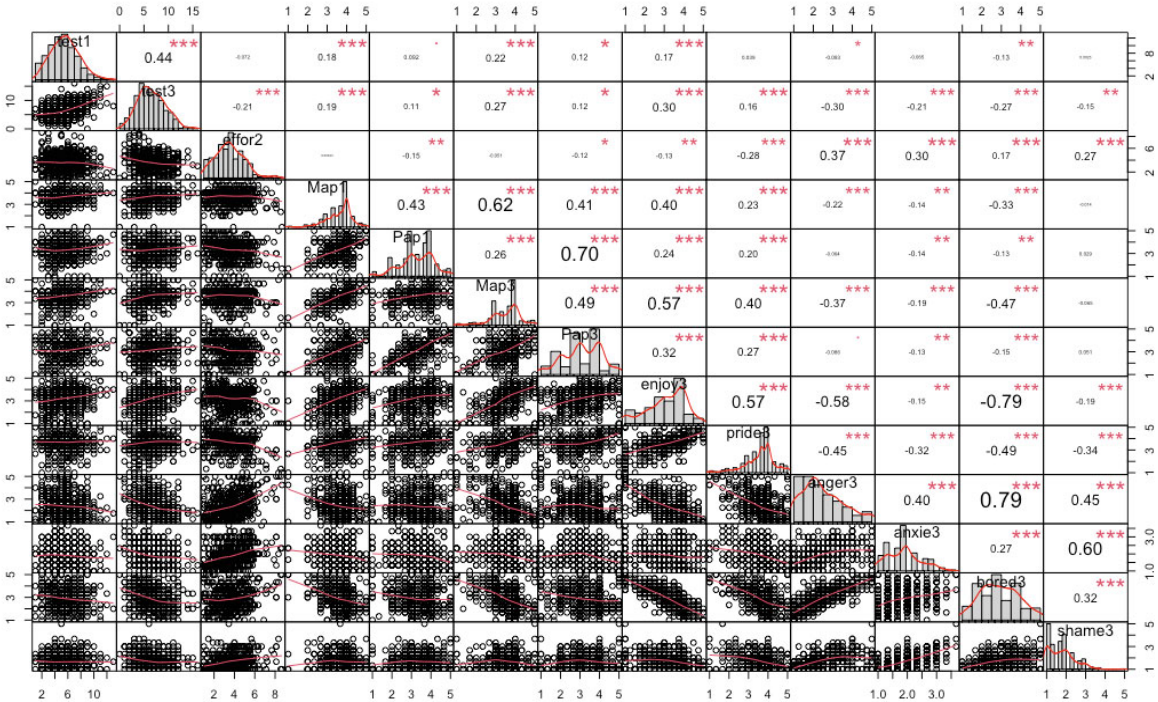
An example manual of domain-specific information (top) and game-specific information (bottom)



Chapter 4 Supplementary Materials

Figure S4.1

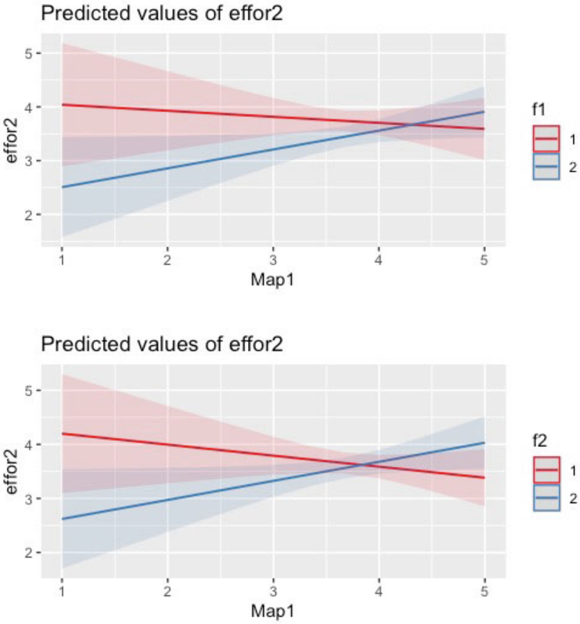
Pearson correlation between variables



Note. test1 = pretest performance; test3 = posttest performance; effor2 = mental effort; Map1 = prior mastery-approach goals; Pap1= prior performance-approach goals; Map3 = mastery-approach goals; Pap3 = performance-approach goals; enjoy3 = enjoyment; pride3 = pride; anger3 = anger; anxie3 = anxiety; bored3 = boredom; shame3 = shame; *$p < .05$; **$p < .01$; ***$p < .0$

Figure S4.2

Interaction plots of moderation



Note. f1 = mastery-approach goal instruction; f2 = performance-approach goal instruction; Map1 = prior mastery-approach goals; 1 = with; 2 = without; effor2 = mental effort (1-9).

Chapter 5 Supplementary Materials

Table S5.1

Fit indexes of bivariate random-intercept cross-lagged models

	CFI	TLI	RMSEA	SRMR	df	χ^2	AIC	BIC
Enjoyment and mental effort model								
RI-CLPM	.930	.913	.088	.113	261	525	8731	9064
CLPM	.854	.820	.126	.186	264	816	9015	9340
RI-CLPM with constraints over time	.902	.906	.092	.107	338	709	8761	8873
RI-CLPM with predictors and outcomes	.920	.904	.077	.101	410	728	10904	11321
Enjoyment and mastery-approach goals model								
Mastery-approach goals								
RI-CLPM	.943	.929	.083	.169	261	498	6199	6533
CLPM	.881	.854	.120	.176	264	760	6455	6780
RI-CLPM with constraints over time	.919	.922	.088	.160	338	677	6225	6337
RI-CLPM with predictors and outcomes	.938	.924	.075	.148	358	625	6786	7165
Enjoyment and performance-approach goals model								
RI-CLPM	.945	.932	.081	.131	261	487	6452	6785
CLPM	.866	.835	.126	.173	264	817	6776	7100
RI-CLPM with constraints over time	.918	.921	.087	.128	338	675	6485	6598
RI-CLPM with predictors and outcomes	.929	.914	.081	.108	358	662	7118	7497
Enjoyment and mastery-avoidance goals model								
RI-CLPM	.925	.906	.095	.180	261	567	6429	6762
CLPM	.839	.802	.138	.181	264	918	6774	7099
RI-CLPM with constraints over time	.896	.900	.098	.127	338	762	6470	6582
RI-CLPM with predictors and outcomes	.910	.890	.090	.145	358	740	7083	7462
Enjoyment and performance-avoidance goals model								
RI-CLPM	.946	.932	.081	.098	261	485	6427	6761
CLPM	.882	.854	.119	.168	264	753	6689	7014
RI-CLPM with constraints over time	.920	.923	.087	.113	338	670	6458	6571
RI-CLPM with predictors and outcomes	.939	.926	.075	.086	358	621	7056	7435
Relaxation and mental effort model								
RI-CLPM	.933	.916	.084	.117	261	503	8361	8694
CLPM	.838	.800	.130	.186	264	850	8702	9027
RI-CLPM with constraints over time	.900	.903	.090	.081	338	700	8404	8516
RI-CLPM with predictors and outcomes	.919	.903	.076	.108	410	719	10551	10968
Relaxation and mastery-approach goals model								
RI-CLPM	.933	.916	.088	.152	261	526	5932	6265
CLPM	.862	.830	.126	.182	264	809	6208	6533
RI-CLPM with constraints over time	.904	.908	.092	.140	338	716	5967	6079
RI-CLPM with predictors and outcomes	.925	.909	.081	.130	358	664	6518	6898
Relaxation and performance-approach goals model								
RI-CLPM	.937	.922	.085	.141	261	508	6143	6477
CLPM	.855	.822	.128	.174	264	832	6461	6786

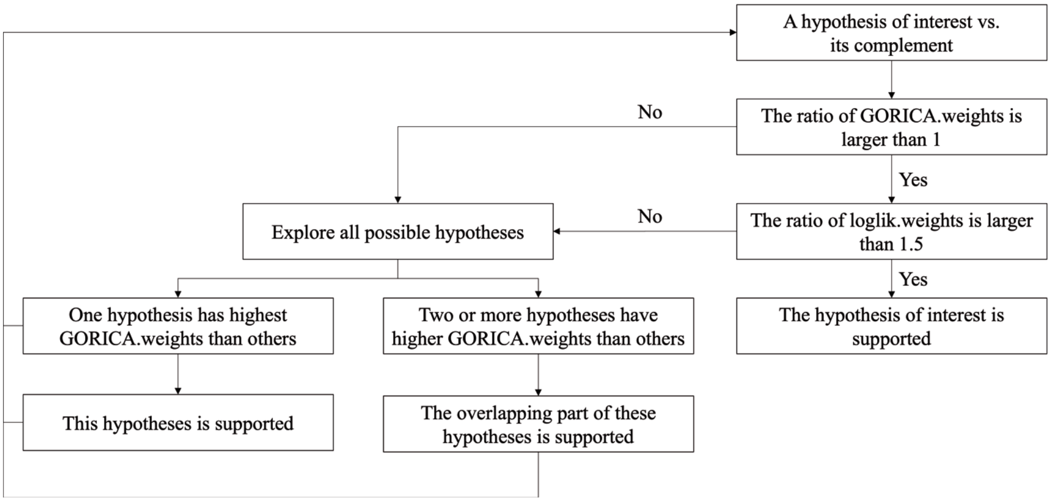
	CFI	TLI	RMSEA	SRMR	df	χ^2	AIC	BIC
RI-CLPM with constraints over time	.902	.906	.093	.110	338	722	6203	6315
RI-CLPM with predictors and outcomes	.917	.900	.085	.107	358	700	6813	7192
Relaxation and mastery-avoidance goals model								
RI-CLPM	.922	.903	.094	.157	261	563	6111	6445
CLPM	.830	.790	.138	.184	264	922	6465	6790
RI-CLPM with constraints over time	.889	.893	.099	.112	338	768	6163	6275
RI-CLPM with predictors and outcomes	.903	.883	.091	.132	358	749	6780	7160
Relaxation and performance-avoidance goals model								
RI-CLPM	.948	.935	.077	.087	261	466	6086	6419
CLPM	.878	.850	.118	.165	264	743	6356	6681
RI-CLPM with constraints over time	.920	.923	.084	.095	338	653	6119	6231
RI-CLPM with predictors and outcomes	.936	.922	.075	.077	358	623	6715	7095
Frustration and mental effort model								
RI-CLPM	.908	.886	.091	.113	261	546	9490	9824
CLPM	.825	.785	.125	.177	264	808	9746	10071
RI-CLPM with constraints over time	.881	.885	.092	.110	338	710	9500	9612
RI-CLPM with predictors and outcomes	.901	.880	.078	.100	408	736	11716	12139
Frustration and mastery-approach goals model								
RI-CLPM	.927	.910	.084	.171	261	503	7117	7450
CLPM	.851	.817	.120	.172	264	759	7367	7692
RI-CLPM with constrains over time	.909	.912	.083	.153	338	642	7102	7214
RI-CLPM with predictors and outcomes	.917	.899	.078	.154	358	643	7790	8170
Frustration and performance-approach goals model								
RI-CLPM	.928	.910	.083	.144	261	498	7364	7697
CLPM	.843	.807	.122	.167	264	780	7640	7965
RI-CLPM with constraints over time	.898	.902	.087	.120	338	672	7384	7496
RI-CLPM with predictors and outcomes	.907	.888	.082	.123	358	677	8109	8489
Frustration and mastery-avoidance goals model								
RI-CLPM	.897	.871	.100	.161	261	600	7332	7665
CLPM	.805	.760	.136	.174	264	904	7630	7954
RI-CLPM with constraints over time	.874	.879	.096	.124	338	749	7327	7440
RI-CLPM with predictors and outcomes	.903	.879	.085	.144	351	687	7960	8359
Frustration and performance-avoidance goals model								
RI-CLPM	.937	.922	.078	.117	261	467	7317	7651
CLPM	.871	.841	.111	.159	264	690	7534	7859
RI-CLPM with constraints over time	.920	.923	.077	.114	338	602	7298	7410
RI-CLPM with predictors and outcomes	.928	.913	.072	.109	358	604	8026	8406
Anger and mental effort model								
RI-CLPM	.918	.898	.095	.168	261	567	8450	8784
CLPM	.851	.817	.127	.175	264	819	8696	9021
RI-CLPM with constraints over time	.894	.898	.095	.149	338	734	8463	8575
RI-CLPM with predictors and outcomes	.900	.880	.086	.125	410	805	10611	11028
Anger and mastery-approach goals model								
RI-CLPM	.926	.908	.094	.213	261	560	6026	6360
CLPM	.873	.843	.122	.173	264	779	6239	6564
RI-CLPM with constraints over time	.905	.909	.093	.183	338	722	6034	6146

	CFI	TLI	RMSEA	SRMR	df	χ^2	AIC	BIC
RI-CLPM with predictors and outcomes	.914	.896	.087	.195	358	716	6591	6971
Anger and performance-approach goals model								
RI-CLPM	.917	.897	.099	.165	261	596	6256	6589
CLPM	.856	.823	.130	.167	264	848	6501	6826
RI-CLPM with constraints over time	.888	.892	.101	.156	338	791	6297	6409
RI-CLPM with predictors and outcomes	.903	.882	.094	.148	357	769	6879	7261
Anger and mastery-avoidance goals model								
RI-CLPM	.903	.879	.107	.191	261	650	6204	6537
CLPM	.830	.790	.141	.173	264	947	6495	6820
RI-CLPM with constraints over time	.874	.879	.107	.154	338	842	6242	6354
RI-CLPM with predictors and outcomes	.903	.880	.094	.143	354	760	6770	7161
Anger and performance-avoidance goals model								
RI-CLPM	.926	.908	.094	.220	261	561	6220	6554
CLPM	.877	.848	.120	.158	264	763	6416	6741
RI-CLPM with constraints over time	.910	.914	.091	.146	338	702	6207	6319
RI-CLPM with predictors and outcomes	.920	.903	.085	.110	358	695	6826	7206
Boredom and mental effort model								
RI-CLPM	.954	.942	.073	.127	261	442	8698	9031
CLPM	.885	.858	.114	.171	264	715	8964	9289
RI-CLPM with constraints over time	.935	.938	.076	.097	338	592	8693	8805
RI-CLPM with predictors and outcomes	.947	.936	.064	.103	410	628	10893	11310
Boredom and mastery-approach goals model								
RI-CLPM	.954	.943	.076	.206	261	460	6176	6510
CLPM	.894	.870	.115	.157	264	723	6433	6758
RI-CLPM with constraints over time	.936	.939	.079	.149	338	615	6177	6289
RI-CLPM with predictors and outcomes	.942	.930	.074	.207	358	618	6765	7144
Boredom and performance-approach goals model								
RI-CLPM	.948	.935	.081	.145	261	483	6457	6791
CLPM	.881	.854	.121	.161	264	773	6741	7066
RI-CLPM with constraints over time	.924	.926	.086	.120	338	666	6486	6598
RI-CLPM with predictors and outcomes	.932	.917	.081	.126	358	663	7128	7507
Boredom and mastery-avoidance goals model								
RI-CLPM	.936	.920	.089	.196	261	534	6393	6727
CLPM	.858	.825	.132	.170	264	871	6724	7049
RI-CLPM with constraints over time	.915	.918	.090	.118	338	699	6405	6517
RI-CLPM with predictors and outcomes	.915	.897	.090	.151	358	734	7071	7450
Boredom and performance-avoidance goals model								
RI-CLPM	.956	.946	.074	.087	261	447	6436	6770
CLPM	.906	.885	.108	.153	264	664	6647	6972
RI-CLPM with constraints over time	.941	.944	.075	.104	338	589	6423	6535
RI-CLPM with predictors and outcomes	.951	.941	.068	.078	358	572	7076	7455

Note. CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean residual; df = degree of freedom; χ^2 = chi-square; AIC = Akaike information criterion; BIC = Bayesian information criterion. All χ^2 values are significant at $p < .001$.

Figure S5.1

Decision tree of interpreting GORICA output



List of publications

Hu, Y., Wouters, P., van der Schaaf, M., Elliot, A. J., Pekrun, R., & Kester, L. (2023). Effects on peers' achievement emotions on students' achievement emotions, achievement goals, and performance in game-based learning. *Manuscript submitted for publication*. Department of Education, Utrecht University

Hu, Y., Wouters, P., van der Schaaf, M., & Kester, L. (2022). The effects of achievement goal instructions in game-based learning on students' achievement goals, performance, and achievement emotions. *Manuscript submitted for publication*. Department of Education, Utrecht University

Hu, Y., Wouters, P., van der Schaaf, M., & Kester, L. (2022). Timing of information presentation matters: Higher motivation, more enjoyment, but less learning in game-based learning. *Manuscript submitted for publication*. Department of Education, Utrecht University

Hu, Y., Gallagher, T., Wouters, P., van der Schaaf, M., & Kester, L. (2021). Game-based learning has good chemistry with chemistry education: A three-level meta-analysis. *Journal of Research in Science Teaching*, 1–45. https://doi.org/10.1002/tea.21765

Garcia Fracaro, S., Hu, Y., Gallagher, T., van Loenen, S., Solmaz, S., Cermak-Sassenrath, D., & Van Gerven, T. (2021). Immersive tools for teaching and training in a science and technology environment. *Policy brief*. https://charming-etn.eu/wp-content/uploads/2021/04/D5.4_CHARMING_Policy_Brief.pdf

胡媛媛 (2018). 不同价态硫物质的转换. 张立云(主编), *中学化学实验教程* (p151–154). 广东教育出版社

胡媛媛 (2018). 同族、同周期元素性质的递变. 张立云(主编), *中学化学实验教程* (p159–162). 广东教育出版社

Hu, Y., Ma, B., Zhang, Y., & Wang, M. (2014). Small molecule—folic acid modification on nanopatterned PDMS and investigation on its surface property. *Biomedical microdevices*, 16(3), 487-497. https://doi.org/10.1007/s10544-014-9851-7

Background

Game-based learning (GBL) uses a game as the medium for learning. Well-designed GBL should and can promote cognitive processes, motivation to learn, and positive emotions, all of which contribute to learning. However, from a societal perspective, it is still unclear whether teachers and students should use GBL, and which design features improve GBL. From a scientific perspective, evidence for the effects of GBL in comparison to non-GBL was inconsistent and research on instructional design features that improve GBL was limited.

Aims

This thesis investigated the effects of GBL in comparison to non-GBL (Chapter 2) and investigated not only the effects of three instructional design features that manipulated students' cognition (i.e., timing of information presentation; Chapter 3), motivation (i.e., achievement goals; Chapter 4), and emotion (i.e., peers' achievement emotions; Chapter 5) on three learning processes and outcomes, namely, cognition (i.e., mental effort and performance), motivation (i.e., achievement goals), and emotion (i.e., achievement emotions), but also their interconnection in GBL in chemistry education.

Methods

This thesis included a meta-analysis on 34 empirical studies in GBL in chemistry (Chapter 2) and three experimental studies with 1257 students from secondary and higher education (Chapter 3, 4, and 5).

Conclusions

GBL enhances chemistry learning more than non-GBL, and instructional design features differentially affect cognition, motivation, and emotion in GBL in chemistry education. Specifically, students felt more enjoyed and motivated in the learning environment from which they learned less.