Zitiervorschlag: Kickmeier-Rust, M. D., Niggli, C., & Richter, K. (2024). An Introduction to Game-Based Competence Assessment Based on Cognitive Diagnostic Models. In P. Dondio, M. Rocha, A. Brennan, A. Schönbohm, F. de Rosa, A., & F. Bellotti (Eds), *Games and Learning Alliance: 12th International Conference, GALA 2023, Dublin, Ireland, November 29 – December 1, 2023, Proceedings* (Lecture Notes in Computer Science, Vol 14475, pp. 244-253). Springer. https://doi.org/10.1007/978-3-031-49065-1_24

An Introduction to Game-based Competence Assessment Based on Cognitive Diagnostic Models

Michael D. Kickmeier-Rust[1] [0000-0003-0957-0037], Corsin Niggli[1] and Katharina Richter[1]

[1] St.Gallen University of Teacher Education, St.Gallen, Switzerland
michael.kickmeier@phsg.ch
corsin.niggli@student.phsg.ch
katharina.richter@phsg.ch

Abstract. In Serious Games, in particular in learning and training games, the assessment of competencies and skills is crucial for monitoring learning progress, tailoring learning experiences, and providing individual formative feedback. A sound psychometric diagnostic of competencies is not trivial, however. Conventional scoring techniques have severe shortcomings in terms of accuracy and the degree to which actionable information can be drawn from them. In this paper we introduce Cognitive Diagnostic Models and in particular Competence-based Knowledge Space Theory as theoretical underpinnings of in-game competence assessments. We exemplify the approach by a gamified mathematics learning scenario named Mathiade and illustrate the steps of developing, implementing, and evaluating competence models.

Keywords: Competence Assessment, Cognitive Diagnostic Models, Competence-based Knowledge Space Theory, Mathematics

1 Introduction

Digitization is changing education; there is a strong movement toward stronger competence-orientation, formative and contextualized assessments, as well as evidence-based personalization of learning. Along with these developments, digital games are about to play an increasingly important role. Educational games, in a very natural way, focus on the learners' competencies and skills, they can provide clear goals and rules, a relevant learning context, an engaging storyline, immediate feedback, a high level of interactivity, challenge and competition, random elements of surprise, and rich and appealing learning environments. These factors not only determine motivation to play but are also considered important for successful and effective learning. Meta-reviews [1, 2, 3] revealed that digital games can significantly increase learning success as opposed to conventional learning media, even though the effect sizes are generally moderate. A substantial body of research reported that a key element of a serious game's success is a suitable personalization and balancing of learning and gaming experiences [4]. This endeavor, however, requires a solid understanding of individual learning processes, strengths, weaknesses, knowledge gaps, and learning dispositions by a game as an autonomous tutorial system. In other words, individualized learning and gaming require valid, reliable, and accurate assessments [5]. The approaches to in-game assessment and non-invasive adaptation of games have been refined significantly over the past decade [6]. State-of-the-art psychometric methods include the concept of stealth assessment [7], which is a method for embedding assessment seamlessly into games. A different approach to stealth assessment was introduced by [8]; it applies recurrent neural networks (long short-term memory networks) models for assessment. In addition, there exist structural, combinatorial models [9], cognitive classification models [10], Bayesian approaches [11], latent variable models [12] and methods from the fields of learning analytics research [13] and machine learning [14]. The likely most popular combinatorial methods are Cognitive Diagnostic Models (CDM) [15].

1.1 Cognitive Diagnostic Models and Knowledge Space Theory

Cognitive Diagnostic Models (CDMs) or Diagnostic Classification Models represent a psychometric framework for collecting, analyzing, and reporting diagnostic data. CDMs provide discrete multivariate fine-grained diagnostic feedback information about learners' strengths and weaknesses for developing targeted instruction and personalized support [16]. Knowledge Space Theory (KST) [17] can be considered a member of the CDM family. [18] for example, have elaborated on the relationships between CDM and KST. It represents a structural, combinatorial approach, which may serve as a counterpart to the statistical models. The starting point is the notion of a so-called knowledge domain, which is a set of problems taken from a certain area. For instance, additions, subtraction, multiplication, and division of positive integers are problems of the domain basic algebra. To provide an example, assume that the knowledge domain $Q = \{a, b, c, d, e, f\}$ consists of six problems a, b, c, d, e, and f. The performance state of a person is represented by the subset of problems from Q that the person can solve. However, if we look at the answer patterns a sufficiently large number of subjects exhibit on these six problems, then most certainly not all possible subsets (there are $2^{|Q|} = 64$ subsets in our example) will actually occur. A person who can solve a problem that requires multiplying two positive integers will also be capable of solving a problem that involves an addition of two positive integers. This means that from a correct solution to the first problem we can surmise a correct solution to the second problem. This kind of mutual dependency is captured in a so-called prerequisite relation. By this relation, the number of performance states is restricted: items cannot establish a performance state without their prerequisite items. The collection of the performance states corresponding to a prerequisite relation is called a (quasi-ordinal) performance space (or knowledge space). The original concept of KST only operated on the observable performance dimension. To account for the underlying latent cognitive processes, the approach has been extended, emphasizing the underlying latent competences, which are necessary to solve a problem. Such extensions came, for example, from [19, 20, 21]. We subsume the approaches, specifically that of [21], under the term Competence-based Knowledge Space Theory (CbKST). Competences (skills, knowledge, aptitude) can be defined as a fine-grained (often termed atomic) latent theoretical construct which determines a person's observable behavior respectively performance, i.e., if a person solves a test item/task or not. Competence and performance structures can be mapped to each other utilizing interpretation respectively representation functions. On this basis, a latent, structural competence model can be established, describing individual competence states and learning paths. A particular advantage is that learning processes – from the stage of holding none of the skills and competencies of a domain to possessing all – are not seen as a linear, unidimensional approach but a rather multidimensional processes with multiple, individual learning paths [5]. KST models have been advanced in various applied directions [17], and they have been applied in the field of adaptive serious games.

KST provides several standard validation methods. These methods investigate the consistency of empirical answer patterns and the patterns hypothesized by a knowledge (or competence) structure. Different numerical measures have been proposed, for example, the Discrepancy Index [22] or the Distance Agreement Coefficient [23] or the Minimal Symmetric Set Difference [24], which likely is the most common metric. An important quality indicator for the structural competence modelling approach KST is a size-fit trade-off; with increasing model size (number of competence/performance states) the fit (the number of explained empirical answer patterns) increases automatically. A validation method to account for this trade-off was presented by [25]. The principal idea is to compare all possible prerequisite relations for a given set of competences (or items) and select the one that offers the best fit to a set of empirically observed answer patterns. The method offers statistical data about the relations between the cumulative relative frequencies of partial order types, number of relations in a partial order, and a certain goodness-of-fit-criterion.

In this paper we demonstrate the potential of CDM in general and CbKST in particular, to conduct a fine-grained competence-oriented assessment. We exemplify the approach in high school mathematics in the domain of fractions. In a first step we explain the development of a competence model and in a second step we evaluate the model with the data from a gamified math app.

1.2 Modelling the Domain of Fraction Arithmetic

Fraction arithmetic presents students with complex challenges and requires students to have a solid understanding of the basic concepts in arithmetic. We started the generation of a CbKST-type competence model by analyzing curricula, for example, LehrplanPLUS (www.lehrplanplus.bayern.de) and the relevant literature in math didactics (e.g., [27]). A systematic content analysis revealed a set of key skills (cf. Table 1). The level of granularity of the model depends on the nature of tasks. In other words, with certain task types it is possible to identify certain skills and sets of skills. To avoid an unnecessary large model, the key skills have been grouped into nine distinct competencies (Table 1).

In a next step, we developed the CbKST-type surmise relation among the competencies. The surmise relation states whether we can assume from mastering one task the mastery of another. This relation is shown in Fig. 1, left panel. This graphical representation is identical with the Q matrix of CDMs. The graph reads from bottom to top; lower competencies are considered prerequisites of higher ones (e.g., *c1* is a

Table 1. Competencies in the domain of fraction arithmetic.

Level	Description	Skill	Competency
1.	Basic mathematical concept	a_1	$\mathbf{c_1}$
2.	Whole number arithmetic	b_1	$\mathbf{c_2}$
	Mathematical symbols	b_2	
	Non-symbolic representations	b_3	
	Non-symbolic proportional reasoning	b_4	
	Number line estimation	b_5	
	Equivalence of ratios	b_6	
3.	Multiplication with whole number	c_1	$\mathbf{c_3}$
	Division with whole number	c_2	
	Addition with whole number	c_3	
	Subtraction with whole number	c_4	
4.	Recognizing equivalent fractions	d_1	$\mathbf{c_4}$
	Ordering fractions	d_2	
	Describing fractions in different form	d_3	
	Knowledge of fraction magnitudes	d_4	
5.	Multiplication with fractions	e_1	$\mathbf{c_5}$
	Division with fractions	e_2	$\mathbf{c_6}$
	Subtraction with fractions	e_3	$\mathbf{c_7}$
	Addition with fractions	e_4	$\mathbf{c_8}$
6.	Reduce fractions	f_1	$\mathbf{c_9}$

prerequisite of *c3*). Competence *c9* is considered independent from the others, which are built upon each other. From the surmise (or prerequisite) relation we can derive the competence structure by set inclusion (Fig. 1, right panel). The competence structure (depending on its mathematical properties, it is called competence space or learning space [17]) is the set of all admissible states – i.e., the combination of competencies – in which a learner can be if the assumed surmise relation is true. The competence structure includes the empty and the full set of competencies. The edges of the graph indicate admissible learning paths.

1.3 *Mathiade*: A gamified math competition

In educational settings, gamification is utilized primarily to facilitate learners' motivation, engagement, concentration, probably decrease frustration and thus – as the ultimate consequence – improve (learning) performance [28]. There is a broad range of studies investigating the effects of various gamification applications in education. *Mathiade* is a gamification scenario for an entire lesson in a classroom setting that combines competition and cooperation in mathematics learning, both are supposed to be

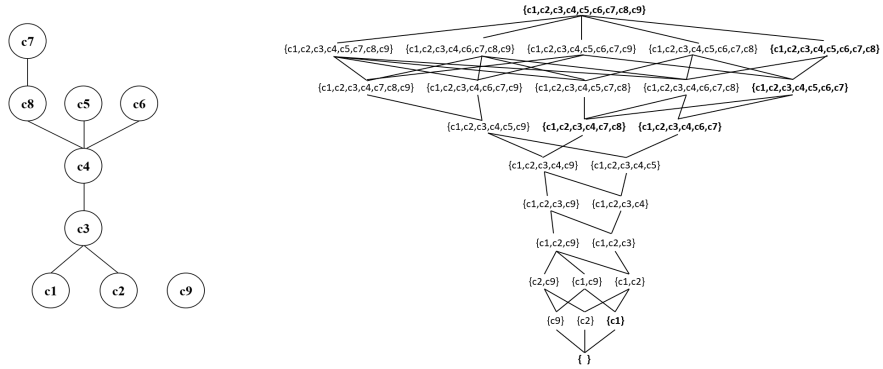

Fig 1. The left panel shows the surmise relation for the domain described in Table 1; the right panel shows the resulting competence structure.

strong facilitators of performance. In educational settings, leaderboards represent rankings of learners according to certain measurable variables, including performance, achievements, diligence, and perseverance. [29]. Leaderboards compare these variables of individuals with several opponents and, thereby, induce a competitive environment. Competition, in turn, is considered a key element of game and play, that is supposed to encourage learners to persistently carry out certain tasks. As [30] pointed out, the board leaders may provide others with a (more or less desirable) goal to reach, that is, keeping up with the leaders. This goal could improve performance by guiding attention, improving motivation and persistence, and promoting the use of goal-relevant strategies. The second gamification element in our scenario is cooperation. Cooperation can be seen as a behavior that is facilitated by gamification, but it also can be considered a game mechanic on its own [31]. Cooperation supports a mutual exchange of ideas, knowledge, and skills on the one hand, and it influences psychological variables and sentiment (e.g., group norms and social identity), which in turn raise motivation, engagement, and perseverance [32].

First, the students were arranged in groups of three or four students by the teacher. The groups evenly included stronger and weaker students as well as high and low communicative students. The groups competed for points by solving mathematics tasks in an online interface (Fig. 2). Each student worked separately, however, in the interface the students always saw the live results (i.e., the tasks completed and the points) of other group members. Assuming that each member (especially the high performers and students with highly competitive characteristics) of a group wanted to win the competition, collaboration and mutual help were possible. High performers were intended to support other group members when they encountered those who underperformed. The way the students communicated and collaborated was open. On a center screen in the classroom, all students saw the live rankings of the groups (Fig. 2) – without information about the performance of individuals. By having a public ranking via the group leaderboard, low performers were less exposed and perhaps stigmatized due to possibly poor performance. This approach was intended to reduce the perception of stress and other socially and emotionally detrimental effects for individuals. Technically, the

Fig. 2. Screenshots of the *Mathiade* app.

interface has been realized in HTML 5 as a browser-based app. Dynamic functions (i.e., display of tasks, scoring, display of live results, data storage) have been realized with PHP and JavaScript. The backend was a mySQL database. The application consisted of an item pool, the student interface, the group leaderboard, and additional administrative functions.

2 Study Implementation

To exemplify the CbKST assessment procedure, we implemented Mathiade for the 10th grade in a secondary school in Switzerland. We investigated general performance measures, and we used the students' response data to evaluate and revise the competence model. The research question is whether the gamification scenario is superior to a conventional, non-gamified setting.

2.1 Participants

In total, 43 students participated in the study. We recruited four classes of a 10th grade of a secondary school in Liechtenstein with 10 to 12 students in each class. The average age of the students was 17 years. For the study the class teacher formed heterogenous groups (in terms of performance and openness) of 2 to 4 students. For practical reasons we assigned the students of the entire classes to either control or experimental group.

2.2 Procedure and Materials

All classroom activities were administered by the regular class teacher. Beforehand, the teachers were instructed about the procedures and specifically about keeping systematic logs about collaborative events during the lesson. First, we made a baseline survey on student motivation. A week later, we carried out the mathematics competition scenario in two groups, one with and one without gamification. Another week later we carried out the scenario again and changed control and experimental group. For the mathematics competition students had exactly 30 minutes. In both groups, communicating and working together in groups was allowed but not explicitly required. Directly after the session, we issued the motivation questionnaire again. During the sessions, teachers kept logs about any events, specifically collaborative events. For the mathematics competition we developed a set of 120 test items in the domain of fraction arithmetic. The items cover the basic arithmetic operators addition, subtraction, multiplication and division. Devisor as well as dividend were either integers or fractions. Examples are 5 + 3/8, 7/9 - 8y, and 6/12 * 7/24. The results had to be entered in text fields, either as integer or as a fraction (Fig. 2). In the control group, the students worked on the tasks in the form of a paper-and-pencil test. The students were instructed to complete as many items as possible within 30 minutes.

3 Results

In terms of performance (i.e., correctly solved tasks) we did not find a significant difference in the gamification condition (GC) as opposed to the control (non-gamification) group (CG). In the GC the number of correctly solved tasks was somewhat lower ($M = 41.89$, $SD = 15.74$) than in the CG (61.00, $SD = 22.77$). The session date (gamification in the first session, no gamification in the second and vice versa) had an effect as well; in the first session the results were higher ($M = 48.87$, $SD = 29.56$) in comparison to the second session ($M = 32.58$, $SD = 19.89$). Given the study setup, where the cooperation of students was the focus, absolute performance plays a subordinate role, however. More important is the loss in performance from session 1 to session 2, which is an indicator of the students' motivation. In the gamification condition (GC) the loss in performance (22%) is clearly lower than in the CG (69%). Which is a remarkable difference. A repeated measures ANOVA yielded a non-significant main effect for condition ($p =.157$), however, a significant main effect of time ($F(1,23) = 17.59$, $p = .001$) and a significant interaction ($F(1,23) = 13.64$, $p = .003$). This result provides some evidence that gamification can indeed sustain motivation in repeated tasks. The domain modelling (section 3) resulted in a surmise relation (Fig. 1, left panel), from which we derived the latent competence structure (Fig. 1, right panel). This structure denotes all admissible combinations of competences a learner may hold. This is a latent model, meaning that the competencies cannot be observed directly. Therefore, we link the competencies to the tasks. In terms of CbKST this linkage is named representation function. It specifies all the competencies that are necessary to master a certain task (cf. Table 2). This linkage results in the set of possible test results, that is, the set of mastered and not mastered tasks. If a learner possesses all the competencies that are necessary to master a task, this student should in fact master that task. Given that the real answer patterns of students may include errors such as slips and lucky guesses the identification of competence states is not unique. Therefore, we applied the minimal symmetric set distance (MSSD) [24] to analyze the data. This metric represents the difference between a given empirical response pattern and the closest competence state. The corresponding algorithm is described by [33]. For the hypothesized model (Fig. 1), the mean MSSD was 9.03 ($SD = 3.69$); the maximum set difference for 120 tasks is $N/2 = 60$. The number of identified competence states was 6, including the full and the empty set. The identified states are bolded in Fig. 1. The average number of tasks presented to students was 75.83 ($SD = 37.02$) a mean MSSD of 9.03 is satisfying in terms of identifying the competencies of learners; it means – on average – that 9 of 75 tasks results (12%) deviated from the hypothesized model.

Table 2. Representation function – linkage of competencies and tasks

		Competencies								
		c1	c2	c3	c4	c5	c6	c7	c8	c9
Tasks	1	1	1	1	1				1	
	2	1	1	1	1				1	
	…					…				
	119	1	1	1	1		1			1
	120	1	1	1	1		1		1	1

The evaluation of a model’s goodness of fit, typically arises from the comparison of several models, given that there is no ground truth. Such comparisons are beyond of the scope of this paper.

4 Discussion

In the present paper, we have demonstrated the use of CbKST, a family member of CDM, to conduct assessment in situations with a large degree of freedom, as it usually occurs in serious games. As opposed to a conventional approach to performance or competence assessment, the approach does not result in a single value, e.g., the percentage of correctly solved tasks, but in a set of probabilities with which a student may hold each of the competencies of a domain. In the presented *Mathiade* setting, the advantage is obvious. In the data, we found rather heterogenous and often contradictory results, where equivalent tasks were solved and not solved by the same student. When, however, computing the solution frequencies we obtain a mean of 0.49 with a SD of 0.19. By the CbKST-based assessment we could identify the competencies the learners hold, ranging from all to none of the competencies. The biggest advantage of the approach is that it uncovers actionable information about learners. From the identified competence states tailored interventions (e.g., providing individual learners with concrete learning content) or highly specific feedback can be derived. One can argue that the certainty of assessments may be weak. However, the same goes for computing solution frequencies, as discussed above. Other psychometric approaches, for example, Item Response Theory, are much more demanding, they require, for example, standardization studies of tasks. In serious games, this is most likely not possible for cost reasons. And, more importantly, in serious games all sorts of actions can be linked to certain available and lacking competencies. This concept has been described under the term micro adaptivity [34]. The technical implementation, in turn, is more challenging since the CbKST approach requires (a) a detailed modelling of the domain and (b) the implementation of assessment algorithms. There do exist, however, comprehensive introductions including *R* Shiny demonstrations (cf. tquant.eu). The modelling procedures described in this paper follow the framework of Evidence Centered Design (ECD) [35], which serves as a scaffolding for the realization of CbKST based in-game assessments.

References

1. Clark, D. B., 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.
2. Wouters, P.J.M., & van Oostendorp, H. (2013). A meta-analytic review of the role of instructional support in game-based learning. Computers & Education, 60, 412-425.
3. Wouters, P.J.M., 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, 249–265.
4. Wouters P.J.M., & van Oostendorp H. (2017) Overview of Instructional Techniques to Facilitate Learning and Motivation of Serious Games. In: Wouters P., van Oostendorp H. (Eds), Instructional Techniques to Facilitate Learning and Motivation of Serious Games. Advances in Game-Based Learning. Cham: Springer.
5. Kickmeier-Rust, M.D., & Albert, D. (2012). Educationally adaptive: Balancing serious games. International Journal of Computer Science in Sport, 11(1), 15-28.
6. Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P, & Berta, R. (2013). Assessment in and of serious games: an overview. Advances in Human-Computer Interaction, 2013, 1-11.
7. Shute, V., Ke, F., & Wang, L. (2016). Assessment and adaptation in games. In P. Wouters and H. van Oostendorp (Eds.), Techniques to Improve the Effectiveness of Serious Games, Advances in Game-Based Learning. Cham: Springer.
8. Akram, B., Wookhee M., Wiebe, E., Mott, B., Boyer, K.E., & Lester, J. (2018). Improving stealth assessment in game-based learning with LSTM-based analytics. In Proceedings of the Eleventh International Conference on Educational Data Mining, Buffalo, New York.
9. Nyamsuren, E., van der Maas, H.L.J., & Maurer, M. (2018). Set-theoretical and combinatorial instruments for problem space analysis in adaptive serious games. International Journal of Serious Games, 5(1), 5-18.
10. Heller, J., Stefanutti, L., Anselmi, P., & Robusto, E. (2015). On the link between Cognitive Diagnostic Models and Knowledge Space Theory. *Psychometrika*, 80, 995-1019.
11. Käser, T., Klingler, S., Schwing, A.G., & Gross, M. (2017). Dynamic Bayesian Networks for student modeling. IEEE Transactions on Learning Technologies, 10(4), 450-462.
12. Mislevy, R.J., Oranje, A., Bauer, M., von Davier, A.A., Hao, J., et al. (2014). Psychometric considerations in game-based assessment. Redwood City, CA: GlassLab.
13. Kickmeier-Rust, M.D. (Ed.) (2014). Learning Analytics for an in serious games. Proceedings of the Joint workshop of the GALA Network of Excellence and the LEA's BOX project at EC-TEL 2014, September 17, 2014, Graz, Austria.
14. Rowe, J., & Lester. J. (2015). Improving student problem solving in narrative-centered learning environments: A Modular Reinforcement Learning Framework. In Proceedings of the Seventeenth International Conference on Artificial Intelligence in Education, Madrid, Spain.
15. de la Torre, J., Carmona, G., Kieftenbeld, V., Tjoe, H., & Lima, C. (2016). Diagnostic classification models and mathematics education research: Opportunities and Challenges. In A. Izsák, J. T. Remillard, & J. Templin (Eds.), Psychometric methods in mathematics education: Opportunities, challenges, and interdisciplinary collaborations (pp. 53–72). Reston, VA: National Council of Teachers of Mathematics.
16. Xin, T., Wang, C., Chen, P., & Liu, Y. (2021). Editorial: Cognitive Diagnostic Models: Methods for Practical Applications. Frontiers in Psychology, 13.

17. Falmagne, J.-C. (2015). Thirty years of knowledge space theory: the beginning, the core ideas, and the assessment spaces. Paper presented at the EMPG Meeting 2015 at Padua, Italy, 01-03 September.
18. Heller, J., Stefanutti, L., Anselmi, P., & Robusto, E. (2016). Erratum to: On the link between Cognitive Diagnostic Models and Knowledge Space Theory. Psychometrika, 81, 250-251.
19. Doignon, J. (1994). Probabilistic assessment of knowledge. In D. Albert (Ed.), Knowledge Structures (pp. 1–56). New York: Springer.
20. Düntsch, I., & Gediga, G. (1995). Skills and Knowledge Structures. British Journal of Mathematical and Statistical Psychology, 48, 9–27.
21. Korossy, K. (1999). Modelling Knowledge as Competence and Performance. In D. Albert & J. Lukas (Eds.), Knowledge Spaces: Theories, Empirical Research Applications (pp. 103–132). Mahwah, NJ: Lawrence Erlbaum Associates.
22. Kambouri, M., Koppen, M., Villano, M. & Falmagne, J.-C. (1994). Knowledge assessment: tapping human expertise by the QUERY routine. International Journal of Human-Computer Studies, 40, 119-151.
23. Schrepp, M., Held, T., & Albert, D. (1999). Component–based Construction of Surmise Relations for Chess Problems. In D. Albert & J. Lukas (Eds.), Knowledge Spaces: Theories, Empirical Research, and Applications (pp. 41–66). Mahwah, NJ: Lawrence Erlbaum Associates.
24. Schrepp, M. (2001). A Method for Comparing Knowledge Structures Concerning Their Adequacy. Journal of Mathematical Psychology, 45, 480–496.
25. Albert, D., Kickmeier-Rust, M.D., & Matsuda, F. (2008). A formal framework for modelling the developmental course of competence and performance in the distance, speed, and time domain. Developmental Review, 28, 401-420.
26. Obersteiner, A., Dresler, T., Bieck, S.M., Moeller, K. (2019). Understanding Fractions: Integrating Results from Mathematics Education, Cognitive Psychology, and Neuroscience. In: Norton, A., Alibali, M.W. (eds) Constructing Number. Research in Mathematics Education. Springer, Cham.
27. Padberg, F. (2009). Didaktik der Bruchrechnung für Lehrerausbildung und Lehrerfortbildung. 4th Ed. Heidelberg: Spektrum.
28. Oliveira, W., Hamari, J., Shi, L., et al. (2023). Tailored gamification in education: A literature review and future agenda. Education Information Technology, 28, 373–406.
29. Nebel, S., Beege, M., Schneider, S., & Rey, G. D. (2016). The higher the score, the higher the learning outcome? Heterogeneous impacts of leaderboards and choice within educational videogames. Computers in Human Behavior,65, 391–401.
30. Landers, R. N., Bauer, K. N., & Callan, R. C. (2017). Gamification of task performance with leaderboards: A goal setting experiment. Computers in Human Behavior, 71, 508–515.
31. Erickson, L.V., & Sammons-Lohse, D. (2021). Learning through video games: The impacts of competition and cooperation. E-Learning and Digital Media, 18(1) 1–17.
32. Reza Keramati, M., & Gillies, R.M. (2022). Teaching cooperative learning through cooperative learning environment: a qualitative follow-up of an experimental study. Interactive Learning Environments, 1-13.
33. Hockemeyer, C. (2000). Documentation of the libsrbi Library. Available online at https://kst.cord-hockemeyer.info/techreports/libsrbi_TechRep_FWF00.pdf
34. Kickmeier-Rust, M.D., & Albert, D. (Eds.) (2012). An Alien's guide to multi-adaptive educational games. Santa Rosa, CA: Informing Science Press.
35. Kim, Y.J., Almond, R.J., & Shute, V. (2016). Applying Evidence-Centered Design for the development of game-based assessments in Physics Playground. *International Journal of Testing*, 0, 1-22.