Predicting Transfer in a Game-Based Adaptive Instructional System

Khanh-Phuong Thai[1], V. Elizabeth Owen[2], Ryan S. Baker[3]
[1]Age of Learning, [2]Learning Data Discovery, [3]University of Pennsylvania
kpthai@gmail.com, v.elizabeth.owen@gmail.com, ryanshaunbaker@gmail.com

ABSTRACT
Transfer, the application of knowledge to new problem-solving situations, and preparation for future learning are often considered among the primary goals of education, but it is difficult to achieve. This paper presents an initial exploration of young children's ability to transfer the specific knowledge and skills that they have learned from a game-based adaptive instructional system to the types of math problems they typically see in school. We describe a regression tree model predicting young children's far transfer performance of specific knowledge and skills following their performance in a game-based adaptive instructional system called *My Math Academy*. Findings suggest that *My Math Academy* gameplay empirically supports positive performance on a task that involves elements of both far transfer of learning and preparation for future learning.

Keywords
Transfer, Mathematics Learning, Game-Based Assessment, Preparation for Future Learning, Adaptive Instructional System, Classification and Regression Trees

1. INTRODUCTION
A key goal of an adaptive instructional system (AIS) – and of education in general – is to promote learning that is retained over time, can transfer to new situations, and prepares students to learn in the future [10, 42] . Transfer of learning and preparation for future learning, however, are often difficult to achieve [9, 11, 24]. AISs can consider what learners know, do not know, and are ready to learn to tailor instruction, feedback, and scaffolding to ensure mastery of specific learning objectives [39]. However, only a small proportion of studies attempt to directly measure whether knowledge learned from AISs will transfer or prepare students for future learning, particularly in game-based systems [1, 29]. Game-based AISs can leverage the many positive benefits of game-based learning [12, 13], but they also pose an additional challenge for learners to transfer the content and strategies outside of the game environment [20, 25]. How do we know whether learners can apply what they have learned to new learning or problem situations beyond the AISs?

This paper presents an initial exploration of children's ability to transfer the specific mathematics skills that they have learned from a game-based AIS to the type of problems they typically see in school. In this pilot, children were asked to use a game-based AIS called *My Math Academy* (*MMA*) as a supplement to their usual mathematics instruction. Upon their demonstration of mastery of each granular learning objective (i.e., skill, knowledge, fact, or ability) within *MMA*, children were asked to complete online worksheets consisting of exercises or assessment items designed to measure transfer of that learning objective. Previous research has demonstrated the effectiveness of *MMA* in accelerating mathematics learning outcomes in young children [4, 41]. The primary goal of the current pilot was to evaluate the degree to which *MMA* can provide the potential for the subsequent transfer of learning and preparation for future learning outside the game contexts. Insights from this pilot can be used to inform design improvements of *MMA* games. The data collected from this pilot enabled us to explore the relationship between gameplay and transfer performance. Here we describe a machine learning model predicting young children's performance on a set of tasks involving both transfer and a lesser degree of preparation for future learning, from performance of specific learning objectives from gameplay behaviors in *MMA*.

2. LITERATURE REVIEW
Transfer of learning is the application of previously acquired knowledge and skills in new problem-solving situations [5, 9, 31]. Achieving transfer has long been one of the biggest challenges of traditional education [42], and it is a particularly important challenge for game-based learning systems given the less-similar representations and contexts often used in game-based learning. Many have also argued that transfer itself is not as important for long-term student outcomes as preparation for future learning (PFL), where a student is better prepared to learn new content [9]. To achieve transfer and PFL, the similarities and analogies between current learning content and future actual context and processes of application often play a critical role. Learners can memorize facts and procedures, but often they do not know *when* to appropriately apply those facts and procedures. To see the applicability and relevance of learned facts and procedures in new contexts, learners must be able to actively recognize the meaningful problem structures underlying the source and transfer tasks amidst irrelevant details [8, 22]. In other words, transfer is a pattern recognition problem; learners must be able to see what matters and map it between learning and usage [16, 23]. The challenge lies in making sure that students grasp the underlying structure of the learned problems and content, and are able to do so fluently enough to apply what they have learned to novel problems.

Game-based learning environments can provide the context for us to study game mechanics and scaffolds that promote transfer, and the player performance and behaviors within the game environment that promote transfer [38]. This pilot focuses on the latter. To evaluate transfer, we also need to have assessments that provide ecologically valid measures of transfer. This pilot incorporates the use of *MMA*, a game-based AIS, into classrooms and uses worksheets of problems typically found in teacher resources and standardized assessments to establish an ecologically valid classroom measure of transfer. The goals of the pilot were formative: (1) to measure the extent to which students could demonstrate transfer outside the game context to inform game design improvements, (2) to test and iterate transfer measures for future studies. In this paper, we explored the degree to which students' usage and performance within *MMA* could be used to predict their ability to succeed on a task involving transfer as well as a lesser degree of preparation for future learning.

Research has demonstrated that transfer can be defined along several contextual and content dimensions [5, 11]. For the purposes of this paper, the degree of transfer is based on task similarity, such that "near" and "far" transfer refer to the degree of similarity between the training and transfer situations (more on this below). Much of the evidence for transfer from game-based learning contexts has come from near-transfer items or responses taken directly from gameplay with a high degree of alignment with the game itself [33]. To evaluate how well students can transfer beyond the original context, we opted to predict performance only on far transfer items, given after the student had also had the opportunity to learn from near transfer items. Instances of far transfer, while rare, have been documented and prior research suggested that it may even be predictable once the relevant dimensions are specified [5]. Some prior work has investigated transfer through predictive modeling of academic skills learned from AISs (e.g., [3, 36]). There have been reports of transfer to executive functioning skills from game-based learning environments [32], but much remains unknown about transfer from games to academic learning outcomes.

Well-designed game-based learning systems leverage play to present authentic learning contexts for learners and provide formative assessment to foster learning and sustain engagement [15], translating game-based interactions to inferences about learning performance (e.g., [27]). In this data-rich context, the field of educational data mining (EDM) provides an ideal range of methods in which to make inferences about learning [29]. Specifically, EDM-based prediction using event-stream game data has offered insights into learning constructs such as strategic problem-solving (e.g., [43]), cognitive disengagement (e.g., [21]), and learning outcomes (e.g., [35]). Similarly, these EDM prediction and classification methods, already used to explore transfer in other contexts (e.g., [3]) are leveraged in our study for modeling transfer in a game-based learning environment.

3. RESEARCH AND DATA COLLECTION

3.1 *My Math Academy* (*MMA*)

My Math Academy is a game-based AIS designed to help young children build a strong foundation in number sense and operations. It includes 300 game-based activities, covering number sense and operations concepts and skills for pre-kindergarten through second grade. Topics range from counting to 10 to adding and subtracting three-digit numbers using the standard algorithm. The personalized mastery-learning system underlying *MMA* uses initial placement assessments (based on a simplified ECD design; e.g., [26]) to measure each child's prior knowledge and determine where they are placed within the system based on what they know and are ready to learn next [14].

Children play games to learn in *MMA*. Using an evidence-based design approach (e.g., [17]) to guide the work of generating in-game evidence, game-based assessment mechanics within *MMA* are designed to generate data on learning performance. Each game is associated with a learning objective, learning tasks, and evidence of learning (e.g., [30]). Learning objectives (LOs) are granular. Examples include count sequence 1-5, 1:1 correspondence and cardinality 6-10, and count out 11-15 objects from a group of up to 20 objectives. A knowledge map organizes the LOs and their prerequisite relationships [14].

Evidence of learning on each granular LO is collected as the student plays. As they progress in *MMA*, the adaptive system uses their performance to recommend learning for individualized difficulty, at the LO level (based on the knowledge map), and within the LO level (by calibrating difficulty of new activities based on prior activity performance). Each group of game-based activities that correspond to an LO is referred to here as an "LO unit". Each LO unit includes up to six learning activities at varying difficulty levels, including an in-game mastery assessment called the "boss" level. Students master the boss levels to demonstrate their skills and understanding for each LO, indicating that they are ready to move on to the next LO unit. Within individual activities, performance data are used to provide scaffolding, adjust difficulty, and offer formative feedback. Across the system, this adaptivity gives learners a customized pathway between skills based on prior performance. A student may pass, stay, or fail back within an LO unit (go to an easier activity level), and the various adaptivity and scaffolding mechanisms enable each student to have a highly personalized experience, tailored to his or her "ready to learn" math level and learning pace.

Theoretical foundations of learning sciences have been applied to inform *MMA*'s content, pedagogy, and design for learning and engagement (e.g., [7, 30]). Formative assessment mechanics require learners to actively retrieve new and previously learned information [34]. Children explore and practice with different examples of the same concept to ensure that they fully grasp the underlying concept across different examples [23]. In addition, *MMA* engages learners in real-world math with stories that contextualize math concepts within meaningful problem-solving situations [40]. For example, Figure 1 shows a game where the Shapeys (game characters) need help with counting swim rings.

Figure 1. Help the Lifeguard Shapey find the total number of swim rings needed by counting on from a number within 1-10

The architecture for event-stream data collection within *MMA* was based on data frameworks that can provide syntheses of assessment mechanics and resultant evidence while capturing a context-rich data stream of player interactions (e.g., [18, 37]). This structured data stream enables EDM investigations into emergent patterns of learner behaviors and outcomes, leveraging in-game logfile data and comprehensive feature engineering for exploration of play and learning [28].

3.2 Data Collection

In Spring 2021 during the COVID-19 pandemic, five classrooms of kindergarteners and first graders (N = 102 students) from two school districts virtually participated in this pilot. They used *MMA* as a supplement to their regular math instruction for an average total of 10.5 hours over 10-12 weeks. After each student demonstrated mastery on a LO in *MMA*, they received assignments of transfer items that addressed that LO. Transfer items were created or found in teacher workbooks or standardized assessments that addressed the granular LOs covered by *MMA*. Transfer was intentionally explored in this granular fashion to inform specific improvements in the design of *MMA* games.

Transfer items shared the same underlying problem structure as the source problems posed in *MMA*. They were classified as “near” or “far” transfer based on the extent to which they shared task features with the source problems. Task features included problem format, procedures, and cover story. Example transfer items are shown in Figure 2. All transfer items required a change in context from *MMA* games to worksheets in ClassDojo or Seesaw, interactive learning platforms that classrooms were already using for assigning and collecting homework. Each item was an editable worksheet delivered with replayable voiceover instruction, to ensure that children’s ability to read the instruction was not a deterrent to determining their math knowledge. Students could insert text and draw directly on the worksheet and submit audio or video recordings.

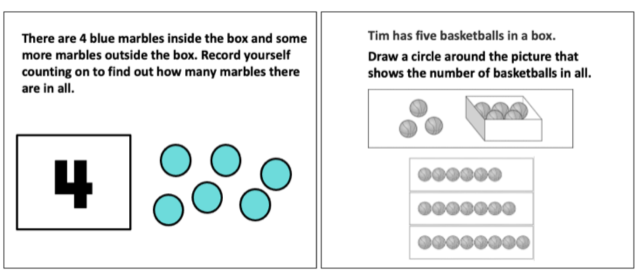

Figure 2. Example near transfer item (left) and far transfer item (right) for the Counting On 1-10 learning objective.

When a student “masters” an LO (i.e., passes the boss level) within *MMA*, near transfer items for that LO were assigned the next day. If they successfully demonstrated near transfer on at least one item, they received far transfer items for that LO the following day. Assignments were done manually by the research team. There was often a temporal delay of 1 to 5 days from when students demonstrated mastery within the games and when they responded to the given transfer items. Corrective feedback was provided for every transfer item, to support learning. Since students could have received feedback on near transfer items before receiving far transfer items, the far transfer measure captures some degree of preparation for future learning as well. However, since the far transfer items still involved different transfer than the near transfer items, the measure can be seen as primarily capturing far transfer.

Given the adaptive nature of the system and that students used *MMA* at home at varying time intervals, students varied in the amount of usage and mastery rate. Therefore, the number of transfer items seen by students varied. Students who completed skills at a faster rate received more total transfer items. Students were asked to complete the assignments independently without help, and students were limited to 1-5 assignments per student per week to allow room for other schoolwork. Additionally, due to the formative nature of the pilot, when certain transfer items did not function well (e.g., when the instructions were not clear to students), they were modified, or alternative transfer items were assigned in their place. Thus, while approximately 90% of the items remained the same during the pilot, some items were iteratively improved based on student data throughout the course of the pilot.

4. DATA SOURCES AND FEATURE ENGINEERING

Transfer item data was utilized to provide a salient outcome variable for predictive analysis. Since the study aimed to understand the breadth of transfer across early mathematics skills, we focused on the number of mathematics LOs with high-level transfer evidence (based on the passing of far transfer items used in the study). Thus the final outcome variable, named *number_of_LOs_with_far_transfer* in analysis, represented the total number of LOs with successful far transfer per student.

MMA event-stream data collection was grounded in recent educational game-based data frameworks (e.g., [18]), with feature engineering focusing on in-game progression and performance [28]. Progression features included the quantity of activities started per student in the entirety of their system use during the study, and well as their total activities canceled, and cancellation rate. Performance features included scores from an in-game pretesting system, as well as performance on activities and final LO assessments in the game. To measure a student’s degree of success rather than degree of completion (captured partially also by *activities_started*), in-game completion and learning measures were engineered as ratios and percents, representing performance per student at the activity level and at the LO level. Table 1 below details the features included in the main prediction analysis.

Table 1. Event-stream data features for input into prediction model

Feature Name	Description
activities_started	Number of activities started
activities_canceled	Number of activities canceled*
activity_cancellation_rate	Number of activities canceled / Number of activities started
pretests_passed	Total pretest score (prior knowledge proxy)
assessment_level_pass_rate	Pass rate of all assessment levels in the *MMA* system
dataLO_assess_pass_rate	Pass rate of fact-memorization LO assessment levels in the system**
masteryrate_finalboss-pass_to_activitiescomplete	Mastery rate: final boss levels passed / activities completed to get there (a learning efficiency ratio)
LO_completion_rate	Number of LO units passed / number of LO units started

*Canceling: a student leaves current activity, and returns to activity selection menu.
**”Data” LOs, or learning objectives focused on fact memorization (e.g., basic number recognition), were differentiated here because they have a timed “fluency” mechanic that is unique within the system (and thus assessed slightly differently); also, these LOs can be viewed as foundational to most of the early math skills in *MMA*.

5. METHODS: PREDICTING FAR TRANSFER

To effectively explore the relationship between *MMA* and transfer, our goal was to predict far transfer based on student interactions with the *MMA* personalized learning system. Thus, using the full feature list, a model was built using in-game features as input variables, and the number of far transfer LOs as the outcome variable. To control for formative evolution of items, we used a stable, universal subset of far transfer items to generate the final outcome variable. Interestingly, students had a mean of 8.4 LOs passed with far transfer (with a mean of 9.4 LO far transfer attempts), and an average of 1.25 hours of play per each successful far transfer LO. Before prediction analysis, the data set was filtered to include only students who completed at least one far transfer item (final n = 97) to ensure that the outcome variable only represented those with an opportunity for far transfer.

A set of algorithms appropriate for regression on a small data set were selected, including RepTREE, linear regression, Random Forest Regression, and M5'. M5' is an algorithm that fits a decision tree with linear regression equations at the leaves; it is particularly effective for regression tasks because it can fit piecewise regressions where the relationships between variables are locally regular but have discontinuities across which the relationships change. Built using the RWeka [19] implementations of these algorithms, models were evaluated under Leave One Out Cross Validation (LOOCV) at the student level (the overall level of analysis). A single final model was chosen based on the cross-validated Spearman's rank correlation, since the dependent variable was not normally distributed (confirmed by a Shapiro-Wilk test; W= .68604, p < .001).

6. RESULTS

Ultimately, M5' produced the best-performing model, yielding a cross-validated correlation of .539, comparable to similar transfer and game-based learning models (e.g., [2, 3, 28]). Interestingly, the model splits along assessment level pass rate, implying that students who had higher in-game performance had a different model of far transfer than those who did not. Each linear model, with detailed features, follows below:

assessment_level_pass_rate <= 0.344 : LM1 (39/28.504%)
assessment_level_pass_rate > 0.344 : LM2 (58/93.683%)

LM num: 1
number_of_LOs_with_far_transfer =
0.0215 activities_canceled - 0.0267* pretests_passed*
+ 5.1746 dataLO_assess_pass_rate*
- 4.0665 masteryrate_finalbosspass_to_activitiescomplete*
+ 15.0016 assessment_level_pass_rate*
+ 9.1112 activity_cancellation_rate - 5.5577*

LM num: 2
number_of_LOs_with_far_transfer =
0.0752 activities_started - 0.1011* activities_canceled*
- 0.0198 pretests_passed + 12.3236* dataLO_assess_pass_rate*
- 3.0081 masteryrate_finalbosspass_to_activitiescomplete*
+ 16.729 activity_cancellation_rate - 6.8657*

In both branches, assessment level performance boosted transfer (with a direct relationship to data LO assessment pass rates in both LMs, and higher general assessment level pass rates within LM1). In addition, the more practice activities students completed in order to reach a unit's final assessment level, the better their far transfer. This is represented in the mastery rate feature, which divides final boss assessments passed by the number of activities completed to get there – a kind of efficiency ratio, in which a higher value implies a shorter trajectory between unit start and final assessment pass. Its negative coefficient in both LMs suggests that the most direct path to the final assessment wasn't the best for transfer; in fact, it appears to better foster transfer when students have more practice along the way, consistent with cognition and memory research supporting benefits of retrieval practice and repetition over time [34]. Pretests passed, a proxy for prior knowledge, also had a negative coefficient in both equations (albeit very small). This implies that lower prior knowledge contributes to a greater number of LOs with far transfer. One possible explanation is a ceiling effect, i.e., that some students with higher prior knowledge coming into the system – thus initially placed further along in the system – may have simply run out of new LOs to learn about (*MMA* covers up to 2nd grade math). Indeed, two students did reach the end of all material in the app (and other students came close). Lastly, cancellation rate of activities is associated with more successful transfer, suggesting that the decision to cancel activities somehow involves engaging with the mathematics material – possibly when evaluating the underlying math skill enough to decide whether to continue with an activity.

Thus, the model's far transfer outcome was boosted with a lot of preceding practice, slightly lower prior knowledge, and higher cancellation rate. One key difference between the branches (split based on assessment pass rate) was that the group that performed more poorly on assessments (LM1) showed a positive relationship between the number of canceled activities and transfer outcomes, while the higher performing group (LM2) yielded a negative coefficient for that feature. Since players in the LM1 group had lower prior knowledge (with average pretest scores 31% lower than LM2 students) and potentially lower confidence in math skills, one possibility is that cancellations may have represented a conscious, engaged selection of activities with careful thought about what each student did and didn't feel ready for. In the LM2 group, with higher prior knowledge and higher assessment performance, it's possible that cancellations represented more of a spontaneous decision (since they were more likely to have the skills needed to confidently tackle each unit), with less thoughtful evaluation of underlying math content. This suggests that different metacognitive processes may be involved with cancellations in different contexts, and that cancellations can occur for multiple reasons. Further research is needed to understand these associations more deeply, but it's an interesting finding that activity cancellations – which one might automatically assume have a negative impact – have varied relationships with performance and transfer.

7. DISCUSSION AND LIMITATIONS

As an initial exploration of mathematics transfer based on *MMA*, this study found consistent relationships between system play and performance on the study's transfer items, succeeding in fitting a model that could predict a measure involving far transfer and a lesser degree of preparation for future learning. These pilot findings show promise for future research in establishing transfer based on *MMA* play, and can be used to refine both the system and the transfer items themselves. In the quest to have positive impact on students' early math skills through engaged, playful learning, addressing transfer to more formal, school-based representations is essential. Unearthing findings that suggest *MMA* helps kids develop knowledge that transfers to school learning is a positive step in the right direction – fueling more work in this area to further improve student experience and results from *MMA*.

As a pilot study, there were many formative elements and limitations to consider, which we now have the opportunity to refine in future research. A main challenge was that we wanted to measure transfer, but there wasn't a previously established set of items that built a specific bridge from *MMA* play to far transfer in early mathematics. Indeed, part of our goal was to create and refine items for greatest relevance and usability in a formal learning environment (in the vein of ecological validity). In cases when the transfer items did not function well (e.g., when instructions were not clear), they were modified for higher clarity. In cases when some near transfer items proved to be too difficult, alternative items were assigned. As a result, while most transfer items remained the same during the pilot, some items were iteratively improved based on student data throughout the course of the study. To control for this in the main analysis, we used a stable, universal subset of far transfer items to generate our final outcome variable. Another limitation to our study was that the learning support (feedback) provided for the near transfer items, while increasing the number of students who could be investigated in terms of far transfer (and supporting overall student learning), introduced a degree of preparation for future learning into our far transfer measure, making our findings somewhat less interpretable. However, since far transfer and preparation for future learning are related, and both important, this limitation was – in our view – less important than ensuring that a maximal number of students learned to transfer their knowledge. Overall, in future studies, this set of pilot items can serve as a foundation for ongoing refinement of a broad base of relevant transfer assessment items.

Posing additional challenges, the study took place in the middle of brick-and-mortar school closures due to COVID-19, and the delivery of the transfer items shifted to digital delivery (instead of in-person classroom assignment). This required some adaptation on the researchers' part, changing item format slightly to fit the digital platform requirements, and becoming more dependent on parents to check in with the platform and accept the assignment. Given the adaptive nature of the system and that students used the system at home at varying time intervals, students varied greatly in the amount of usage and they demonstrated mastery at different rates. With in-person classroom instruction resuming, this element could be better controlled for in the future with structured in-class playtime and assignments (as originally planned).

8. CONCLUSION AND FUTURE WORK

With AISs – and specifically game-based AIS, designed for engagement as well as rigor – transfer of learning is a critical element in which there's a large opportunity for new and impactful research. For education to succeed, learning must apply well beyond the learning context – far transfer and PFL are required. Instances of far transfer and PFL, while rare, have been documented and predicted [3, 5]. This work suggests that game-based event stream data can be used to reliably predict robust learning of this nature, building on prior work in other contexts [3, 5, 42].

In terms of future research and development, the insights gained in this study show promising patterns around learning transfer based on *MMA* usage. In subsequent studies, based on item analysis and review of individual student and teacher responses, there is now a basis for establishing a more refined, seasoned set of *MMA* transfer items. As a fully developed item bank is set, it can be used in future studies to more formally assess item reliability and validity. In turn, this can be used for more formal measurement of transfer in the context of *MMA* usage, potentially yielding more prediction models across broader contexts (e.g., experimental studies with multiple classrooms, districts, and students; [41]). Ideally, *MMA* could then develop a research-informed automated transfer system for the classroom, in which students receive transfer and PFL support materials and assessment items at an individualized pace to help them reach far transfer of their playfully-learned math skills. Transfer prediction could then be used as a partial substitute for more formal assessment of transfer *and* to support better learning outcomes. Given the volume of data generated by *MMA*, A/B tests can also be run to support continuous improvement of the learning experience.

Results of this particular prediction model suggest areas for future research. Based on the role of prior knowledge in transfer and PFL, expansion of the system to include higher level math skills may be helpful for younger students who are ready for more of a challenge. In addition, these results highlight an intriguing relationship between transfer and activity cancellation. Additional research is needed to better understand why activity cancellations have a positive impact on some students and not for others. Key questions might include when is cancellation a meaningless whim, when does it reflect deep consideration of the underlying material, and what is its relationship to a child's enjoyment, cognitive engagement with math, and ability to retain focus? This is particularly relevant given the variance in executive function in very early elementary school [6]. Answering this would give designers and researchers more information about how to interpret cancellation stats to inform design and learning outcomes, particularly for these younger students.

This study provides a first look into the relationship into how *MMA* can support and measure robust learning, yielding promising patterns suggesting that gameplay can empirically support transfer and PFL. Findings can help inform future research and development for more engaging, effective playful learning in early math education.

9. ACKNOWLEDGMENTS

We gratefully acknowledge the support of Sunil Gunderia, Hee Jin Bang, Kirsten Collins, Emily Persons, Eric Keylor, Sarah Buchan, Amanda Kates, Elana Blinder, Jessica Alexandre, Katy Sinclair, and our participating teachers, students, and parents.

10. REFERENCES

[1] Bainbridge, K., Shute, V., Rahimi, S., Liu, Z., Slater, S., Baker, R.S. and D'Mello, S.K. 2022. Does embedding learning supports enhance transfer during game-based learning? *Learning and Instruction*. 77, (Feb. 2022), 101547.

[2] Baker, R.S. and Clarke-Midura, J. 2013. Predicting Successful Inquiry Learning in a Virtual Performance Assessment for Science. *Proc. 21st Int'l Conf on User Modeling, Adaptation, and Personalization* (New York, NY), 203–214.

[3] Baker, R.S., Gowda, S. and Corbett, A. 2011. Towards predicting future transfer of learning. *Artificial intelligence in education* (2011), 23–30.

[4] Bang, H.J., Li, L. and Flynn, K. 2022. Efficacy of an Adaptive Game-Based Math Learning App to Support Personalized Learning and Improve Early Elementary School Students' Learning. *Early Childhood Education Journal*.

[5] Barnett, S.M. and Ceci, S.J. 2002. When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*. 127, 4 (2002), 612–637.

[6] Best, J.R., Miller, P.H. 2010. A developmental perspective on executive function. *Child development*. 81, 6, 1641–1660.

[7] Betts, A. 2019. Mastery learning in early childhood mathematics through adaptive technologies. *IAFOR International Conference on Education.*

[8] Bransford, J., Brown, A.L., Cocking, R.R., National Research Council (U.S.), and Committee on Developments in the Science of Learning 1999. *How people learn: brain, mind, experience, and school*. National Academy Press.

[9] Bransford, J.D. and Schwartz, D.L. 1999. Rethinking transfer: A simple proposal with multiple implications. *Review of research in education*, 61–100.

[10] Brown, A.L. 1989. Analogical learning and transfer: What develops? *Similarity and analogical reasoning*. 369–412.

[11] Carraher, D. and Schliemann, A. 2002. The Transfer Dilemma. *Journal of the Learning Sciences*. 11, 1, 1–24.

[12] Clark, D.B., Tanner-Smith, E., Killingsworth, S.S., 2016. Digital Games, Design, and Learning: A Systematic Review and Meta-Analysis. *Review of Educational Research, 86* (1), 79-122.

[13] Craighead, J. 2008. Distributed, game-based, intelligent tutoring systems — The next step in computer based training? *2008 International Symposium on Collaborative Technologies and Systems*, 247–256.

[14] Dohring, D., Hendry, D., Gunderia, S., Hughes, D., Owen, V.E., Jacobs, D., Betts, A. and Salak, W. System and method for dynamically editing online interactive elements architecture. US Patent 11,151,187.

[15] Gee, J.P. 2005. Learning by design: Good video games as learning machines. *E-Learning and Digital Media*. 2, 1 (2005), 5–16.

[16] Gick, M.L. and Holyoak, K.J. 1983. Schema induction and analogical transfer. *Cognitive Psychology*. 15, 1 , 1–38.

[17] Groff, J., Clarke-Midura, J., Owen, V.E., Rosenheck, L. and Beall, M. 2015. Better Learning in Games: A Balanced Design Lens for a New Generation of Learning Games [white paper]. Cambridge, MA: MIT Education Arcade and Learning Games Network..

[18] Halverson, R. and Owen, V.E. 2014. Game Based Assessment: An Integrated Model for Capturing Evidence of Learning in Play. *Int'l Journal of Learning Technology [Special Issue: Game-Based Learning]*. 9, 2, 111–138.

[19] Hornik, K., Buchta, C. and Zeileis, A. 2009. Open-source machine learning: R meets Weka. *Computational Statistics*. 24, 2, 225–232.

[20] Hung, W. and Van Eck, R. 2010. Aligning Problem Solving and Gameplay: A Model for Future Research and Design. *Interdisciplinary Models and Tools for Serious Games: Emerging Concepts and Future Directions*, 227–263.

[21] Karumbaiah, S., Baker, R.S. and Shute, V. 2018. Predicting Quitting in Students Playing a Learning Game. *Proc. of the 11th International Conference on Educational Data Mining*.

[22] Kellman, P.J. 2002. Perceptual learning. *Steven's handbook of experimental psychology: Learning, motivation, and emotion, Vol. 3, 3rd ed*. John Wiley & Sons Inc. 259–299.

[23] Kellman, P.J. and Massey, C.M. 2013. Perceptual Learning, Cognition, and Expertise. *Psychology of Learning and Motivation*. B.H. Ross, ed. Academic Press. 117–165.

[24] Long, Y. and Aleven, V. 2013. Skill Diaries: Improve Student Learning in an Intelligent Tutoring System with Periodic Self-Assessment. *Proc. AIED.*

[25] Mayer, R.E. 2011. Multimedia learning and games. *Computer games and instruction*. Information Age Publishing. 281–305.

[26] Mislevy, R., Oranje, A., Bauer, M., von Davier, A., Hao, J., Corrigan, S., Hoffman, E., DiCerbo, K. and John, M. 2014. *Psychometric Considerations in Game-based Assessment*. GlassLab.

[27] Mislevy, R.J., Almond, R.G. and Lukas, J.F. 2003. A brief introduction to evidence-centered design. *RESEARCH REPORT-EDUCATIONAL TESTING SERVICE PRINCETON RR*. 16.

[28] Owen, V.E. and Baker, R.S. 2020. Fueling Prediction of Player Decisions: Foundations of Feature Engineering for Optimized Behavior Modeling in Serious Games. *Technology, Knowledge and Learning*. 24, (2020), 1–26..

[29] Owen, V.E. and Baker, R.S. 2019. Learning Analytics for Games. *Handbook of Game-Based Learning*. J.L. Plass, R. Meyer, and B.D. Homer, eds. MIT Press. 513–535.

[30] Owen, V.E. and Hughes, D. 2019. Bridging Two Worlds: Principled Game Based Assessment in Industry for Playful Learning at Scale. *Game-Based Assessment Revisited*. D. Ifenthaler and Y.J. Kim, eds. Springer. 229–256.

[31] Pan, S.C. and Rickard, T.C. 2018. Transfer of test-enhanced learning: Meta-analytic review and synthesis. *Psychological Bulletin*. 144, 7 (2018), 710–756.

[32] Parong, J., Mayer, R.E., Fiorella, L., MacNamara, A., Homer, B.D. and Plass, J.L. 2017. Learning executive function skills by playing focused video games. *Contemporary Educational Psychology*. 51, 141–151.

[33] Pilegard, C. and Mayer, R.E. 2016. Improving academic learning from computer-based narrative games. *Contemporary Educational Psychology*. 44–45, 12–20.

[34] Roediger III, H.L. and Karpicke, J.D. 2006. Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science*. 17, 3 (2006), 249–255.

[35] Rowe, E., Baker, R.S. and Asbell-Clarke, J. 2015. Strategic game moves mediate implicit science learning. *Proc. of the 8th Int'l Conference on Educational Data Mining*, 432–435.

[36] Ruipérez-Valiente, J.A., Muñoz-Merino, P.J. and Delgado Kloos, C. 2018. Improving the prediction of learning outcomes in educational platforms including higher level interaction indicators. *Expert Systems*. 35, 6 (2018), 1–6.

[37] Salmon, A., Goodin, I., Stenerson, M. and Ramirez, D. 2013. *ADAGE: Assessment Data Aggregator for Game Environments (API and Client Resources)*. Games+Learning+Society.

[38] Sinatra, A.M., Oiknine, A.H., Patton, D., Ericson, M., Passaro, A.D., Files, B.T., Dalangin, B., Khooshabeh, P. and Pollard, K.A. 2019. Development of cognitive transfer tasks for virtual environments and applications for adaptive instructional systems. *International Conference on Human-Computer Interaction*, 340–351.

[39] Sottilare, R.A. 2018. A Comprehensive Review of Design Goals and Emerging Solutions for Adaptive Instructional Systems. *Technology, Instruction, Cognition & Learning*. 11, 1 (2018).

[40] Sullivan, P., Zevenbergen, R. and Mousley, J. 2003. The contexts of mathematics tasks and the context of the classroom: Are we including all students? *Mathematics Education Research Journal*. 15, 2 (2003), 107–121.

[41] Thai, K.-P., Bang, H.J. and Li, L. 2022. Accelerating Early Math Learning with Research-Based Personalized Learning Games: A Cluster Randomized Controlled Trial. *J. of Research on Educational Effectiveness*. 15, 1, 28–51.

[42] Thorndike, E.L. and Woodworth, R.S. 1901. The influence of improvement in one mental function upon the efficiency of other functions. II. The estimation of magnitudes. *Psychological Review*. 8, 4 (1901), 384–395.

[43] Weber, B.G. and Mateas, M. 2009. A data mining approach to strategy prediction. *2009 IEEE Symposium on Computational Intelligence and Games* (Piscataway, NJ), 140–147.