To be cited as (≈APA):
Westera, W. (2022). Examining the productiveness of failure in serious gaming, International Journal of Technology Enhanced Learning, 14(1), 66–87. Doi: 10.1504/IJTEL.2022.120561

The productiveness of mistakes: on the value of failure in serious gaming

Abstract

This study presents a logistic model of knowledge growth to investigate the differences between performance and learning in serious games. In contrast with common performance assessment approaches the model takes into account the learning from failures. Monte-Carlo simulations of the model show that performance metrics systematically overestimate the player's actual learning at early stages in a game and underestimate these at the end. Three evaluation metrics describing the progression, efficacy and efficiency of learning show how these differences depend on the players' knowledge growth capacities and their success rates in the game. Results from the model when applied to a real serious game are consistent with those from Monte-Carlo simulations. The significance of the study goes beyond the particular details of this study in that it extends and complements the field of educational research with novel computational models and modelling methodologies.

Keywords

Serious game; assessment; learning; performance; simulation model

1. Introduction

Games have widely demonstrated their potential for learning, training and other serious purposes. A large number of empirical studies have shown significant benefits of game-based approaches over non-gaming alternatives (Clark, Tanner-Smith, and Killingsworth, 2015; Boyle et al., 2016). These so-called "serious games" (Abt, 1970) have gained considerable popularity among scholars and teachers, because of their dynamic, responsive and visualised nature, capable of promoting high motivations, strong user involvement and penetrating learning experiences (Westera, 2015). A player in a game doesn't simply digest informational clues but essentially becomes an influential actor in the course of events, which establishes active learning along with high engagement and a sense of control, empowerment and ownership. The powerful experiences that game can provide are considered the main source of learning. Already in the early 20th century John Dewey (1938) developed his experiential learning theory recognising the importance of powerful experiences for learning, stating that learning should be connected with some meaningful, real world context in order to allow the learner to relate symbolic content (e.g., concepts and principles) to real-world referents. Learning from experience refers to learning by active exploration and self-direction rather than learning from direct instruction. Many related ideas and concepts have been used ever since to indicate comparable approaches, such as discovery learning (Bruner, 1961), problem-based learning (Barrows and Tamblyn, 1980), inquiry learning (Papert,1980), experiential learning (Kolb, 1984), constructivism (Jonassen, 1991), situated learning (Lave and Wenger, 1991) and learning by doing (Schank, 1995; Aldrich 2005), respectively. All these models put the learner at the centre of the action and emphasise discovery, exercise, inquiry, problem solving, and authentic contextual knowledge, which activate learners and help them to acquire the tacit knowledge [Polanyi, 1966] that is intrinsically bound to the actions performed. Learning from experience is the predominant pedagogical paradigm for game-based learning: players in a game engage in a problem context, they are in charge of addressing the challenges posed and learn from the responses they obtain from the game world.

As a consequence of the action-oriented nature of games, the player's progression in a game is directly linked with performance: new problems and new levels will open up only when the present ones have been completed successfully. Accordingly, game scoring systems tend to be based on appropriate performance, which refers to doing the right actions and doing them in the right way. However, various studies have challenged the learning-by-doing argument (Kirschner, Sweller, and Clark, 2006; Mayer, 2004). Just doing things does not necessarily lead to deep cognitive processing and the associated insights and understandings. Consequently, the inherent dynamic nature of games may induce players to act rather than think. Studies into computer-assisted instruction and simulations have shown that learners often adopt a trial-and-error strategy, which involves a lot of doing, but lacks any thoughtful considerations (Vargas, 1986). In many games this gets even worse when time-constraints come into play. Essentially, performance in a game is not necessarily a good indicator for learning progress. Various authors (VandeWalle, Brown, Cron, and Slocum, 1999; Fisher and Ford, 1998) have pointed out the difference between performance and learning. Good performance is commonly linked with reaching milestones, the swift completion of tasks, meeting execution standards, avoiding mistakes, and reducing risks. In many respects, however, effective learning requires opposite conditions, such as having sufficient opportunities for reflection, revision, exploration, self-evaluation, and notably being prepared to make mistakes (Westera, 2015). Hence, the performance orientation of games may readily conflict with the objective of learning. Game score systems, mostly being performance-based, tend to neglect the productiveness of mistakes for learning or, even worse, disqualify mistakes as a source of learning by assigning penalty points.

This study investigates the fundamental differences between performance and learning in a game with a series of simulations. To this end it presents a computational model and explores how performance and learning are related in different conditions. The consequential inclusion of failure in assessment metrics will be investigated in detail. To study practical significance, the model is applied and evaluated with logged player data from a real game. In the next section, we first elaborate the theoretical background of the approach.

2. Theoretical background

2.1. Learning from errors and failure

Many studies have reported the productiveness of errors and failure for learning (e.g. Mory, 2003; Mathan and Koedinger, 2005; Tjosvold, Yu and Hui, 2004; Ivancic and Hesketh, 2000; Keith and Frese, 2008; Calhoun, Boone, Porter and Miller, 2014; Gardner, Abdelfattah, Wiersch, Ahmed, and Willis, 2015; Bridger and Mecklinger, 2014; Huelser and Metcalfe, 2012; Potts and Shanks, 2014; Radosavljević, 2015; Cattaneo and Boldrini, 2017). In a study on episodic memory Cyr and Anderson (2014) have demonstrated the positive effects of learning from mistakes in conceptual tasks as compared to retrieval tasks. They conclude that conceptual mistakes are remembered well and act as "stepping stones" for learning. An extensive study in nursing (Hegg Reime et al., 2016) showed that making mistakes during simulation-based team training improves the quality of patient care once the students returned to clinical practice as it made the students more vigilant. Reflecting on the errors made is crucial. Weinzimmer and Esken (2017) demonstrate the positive effects of mistake tolerance on organisational learning and performance. Simm (2005) poses the rhetorical question whether or not the performance demonstrated by students who are academically-strong, carry out a project effectively without major mistakes, produce a good report, but have little scope for self-critique, really outperform those students who are academically weak, initially struggle with the project resulting in poor 'products', but learn by their mistakes and produce a good self-critique.

2.2. Negative effects of performance-based score systems

Although games for learning are essentially protected spaces that allow for failure and mistakes, the included score systems are often performance-based, which stimulates learners to demonstrate high ability and to avoid poor performance. In such contexts failure becomes a threat to success and thereby it affects self-esteem, self-confidence, and motivation. The resulting self-defence reactions (Mory, 2003) may include discounting (Kelley, 1973), task avoidance, feigning boredom, and task-irrelevant actions to bolster self-image (Dweck and Legget, 1988), and learned helplessness (Seligman, Maier, and Geer, 1968). Also, many games include designed mechanisms for inducing stress, such as time pressure or time-dependent scores, which are likely to promote hurried, shallow or incomplete processing. Risk avoidance draws players toward activities that they are good at already and make them stay away from new approaches to avoid penalty points. Performance and learning are conflicting concepts that often require opposite operational conditions and opposite attitudes. A high performance score does not necessarily indicate high learning gains. To support a learning orientation, serious games should lower the price of failure (Gee, 2003; Westera, 2019). Time constraints and penalty points should ideally be avoided, while allowing players to make mistakes, to spend sufficient time and effort, to try and retry, to reflect on attainments and to decide upon their own strategies (autonomy). To exploit the full potential of learning by doing, games need to take into account the conflicts between performance and learning and promote deep processing, reflection and the consolidation of experiences.

2.3. Performance assessment in serious games

Performance assessment in serious games is generally based on the player's achievements. Whatever game mechanics or game scenarios are used, game play always involves active decision taking by the players, who are challenged to find solutions to posed problems. To decide upon the quality of performance the results or solutions can then be checked against some reference standard.

In model-driven games, such as business simulations, player performance can be derived from the dependent variables of the model, for instance the realised business turnover, profit, deficits or customer satisfaction. Likewise, performance in a ballistic challenge (physics) may be expressed by using the distance from a full hit. If no parametric models are available, player performance need to be based on the compliancy with protocols and checklists. For instance, in a game about fire safety, player performance would be expressed as the degree of agreement with prescribed safety protocols. In a math game, performance could be simply expressed as the fraction of problems that were solved. Occasionally, time spent is included in the scoring system, which is disputable as such in the context of learning. In all cases, a quantitative score is defined, which is then compared with a reference to make a judgement. In its most simple form the judgement of a single game action is dichotomous: the result is either correct or incorrect, while numeric scores are assigned accordingly. Summation of the scores across all the game's challenges, eventually using different weights, yields the player’s overall performance score. By its focus on completion and achievements, in-game performance assessment is essentially outcome-oriented, neglecting any process information or player history. This focus on achievements, outcomes or results, is readily associated with “closures points”, which is a game design pattern that indicates the successful completion of a challenge, an episode, or a level (Björk and Holopaïnen, 2005). Hereby, performance assessment comes close to testing. Although it cannot be denied that testing may contribute to learning (the testing effect: Roediger and Karpicke, 2006), it may unnecessarily promote risk-avoiding behaviour as well as shallow and incomplete processing. Particularly in the rich and dynamic learning environments that games provide, learning from mistakes is a means too powerful to be neglected. To allow players to learn from mistakes, the history of failure should be taken into account. A failing performance may still contribute positively to the player’s mastery of knowledge, skill and competency. In current study, the consequential inclusion of failure in assessment metrics will be investigated in detail.

3. Model Starting points

3.1. A frugal representation of serious games

Serious games are complex, interactive systems composed of many interrelated game objects and their frequently changing attributes, all of which contribute to a combinatorial explosion of potential game states. Even a simple game such as tic-tac-toe (noughts and crosses) has a state space up to 3^9=19,683 different states (neglecting any symmetries) allowing for 9!=362,880 different trajectories through game state space. When accounting for symmetries and including games that end within 9 moves only, the number of trajectories is still 26,830 (Schaeffer, 2002). Frugality with respect to the wide range of variables that are available for describing the processes and conditions of serious gaming is dictated to avoid overfitting of an envisioned model. Rather than describing games by their numerous game states, which would account for every detailed player interaction (e.g. mouse clicks, keyboard strokes), a game is considered to provide the players with a coherent set of challenges, e.g., tasks, assignments, missions, scenes or levels that need to be passed through {Westera, 2017 }. Thereby a game challenge is conceived as a higher level aggregate of micro-actions, that constitute a well-defined chunk of the game scenario, it has a well-delineated scope, and its successful completion goes with a clear result or achievement that is supposed to contribute to raised mastery of knowledge or skills. Hence, a game challenge describes a part of the game rather than the whole, very similar to a learning task or learning activity in a lesson. Given the set of game activities, a serious game is represented as a network of challenges. Playing a serious game can effectively be interpreted as following a trajectory through the network of specified challenges (rather than through all possible game states).

3.2. Knowledge mastery versus performance

Serious games are designed to support the mastery of well-specified knowledge, skills or competences. For practical reasons we will use the term knowledge as a transcending, inclusive concept indicating the things to be learned. In education and training knowledge requirements are specified as learning objectives, which serve as the formalised benchmark for assessment and certification (e.g. Anderson and Krathwohl, 2001; Bloom, Engelhart, Furst, Hill, and Krathwohl, 1956; Heller, Steiner, Hockemeyer, and Albert, 2006; Shute, and Ventura, 2013). The learning objectives should not be confused with the specific goals that players are to pursue in the game (e.g. defeating enemies, collecting objects, or locating the treasure). Generally, learning objectives are represented as a hierarchical framework of interrelated knowledge elements, while child nodes in the hierarchy have a precedence relationship with their parent nodes. The learning objectives hierarchy is static by its nature of expressing the benchmark of required learning outcomes. In general, the relationship of game activities and learning objectives entails a many-to-many relationship: each game activity may address multiple learning objectives, while at the same time each learning objective may point to multiple activities. Observed player behaviours in the activity would provide evidence for the enhanced mastery of underlying knowledge, which is generally referred to as the evidence model (Mislevy, Steinberg, and Almond, 2003). Performance may well be covered by knowledge mastery, as it usually refers to the well-demonstrated mastery of certain skills or competencies. The confusion between knowledge mastery and performance arises from the fact that performance measurement is focusing on the correct execution of a task or often even the outcome only, while other dimensions of mastery (e.g. understanding, insight) are readily neglected. This inevitably translates into the performance-based scoring methods that are generally used in games. As a consequence, performance indicators do not fully capture the state of a player's knowledge mastery.

4. Model development

4.1. The performance assessment model

To award the player's achievements, most serious games include a performance-based scoring system. The scores may be either manifestly communicated to the player or they are left concealed and only used to drive the evolution of the game. In its most simple form, a scoring model assigns bonus points to favourable actions. Sometimes, the systems also account for penalty points when mistakes are made. Overall performance is thus expressed as the accumulation of the bonus points assigned for successes, reduced with the penalty points assigned for failures. The assignment of scores is a discrete process that is activated only upon completion of a challenge. Consequently, performance curves are represented as discrete, stepwise functions, even though the processes of learning may be assumed to be fully continuous.

Such performance score model can be expressed as follows. Let the game be represented as a set of well-designed game challenges labelled with a natural number *i* and let S_i be a binary success indicator for the *i*-th challenge, with

$$S_i = 0 \text{ for a failure, and } S_i = 1 \text{ for a success.} \tag{1}$$

Then the cumulative performance P_i after *i* challenges can be written as

$$P_i = \sum_i (S_i \cdot b_i - (1 - S_i) \cdot p_i) \tag{2}$$

with b_i and p_i the bonus and penalty assigned for challenge i, respectively.

Figure 1 displays a randomly generated performance curve (dots) in accordance with such model and average curves (solid) after 500 iterations (bonus points b_i=1, penalty points p_i=0, success rate S_i=0.6).

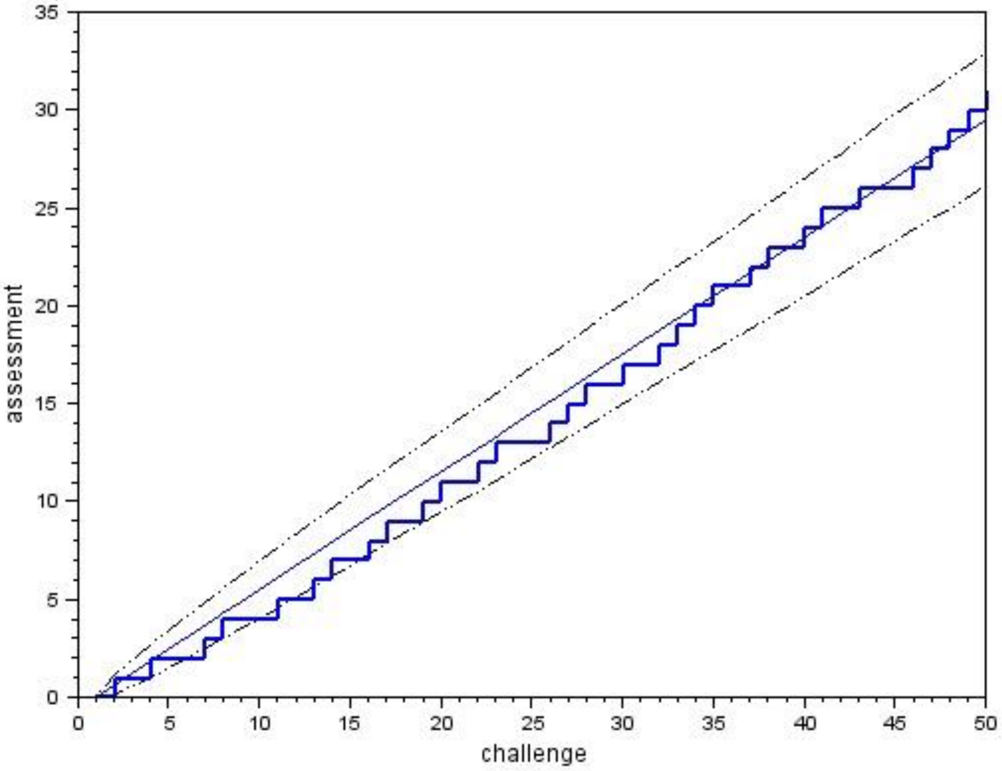

Figure 1. A common player performance example based on a randomised sequence of 50 random successes or failures (staircase line), and the mean result of 500 iterations (thin solid line) and its variability range (dashed-dotted).

The average performance is proportional with the number of challenges, while the standard deviation is typically 7%. Changing the bonuses, penalties or success rates preserves the overall linear pattern.

A similar curve can be drawn to represent the performance score as a function of time. If D_i denotes the duration of time required to complete the i-th challenge, then the timestamp for the completion of challenge i is given by:

$$t_i = \sum_{j \leq i} D_i \tag{3}$$

When D_i is constant (each challenge requiring the same amount of time), the overall linear shape of the performance curve is preserved. Indeed, in many games, the performance score of a player will grow more or less linearly with time spent. Likewise, time may be readily replaced with effort, which would yield a similar pattern, provided that the player preserves the same intensity of effort during the game.

The linearity of these performance systems is problematic and highly unrealistic, because in practice it will become increasingly difficult to approach the level of perfection. To account for such asymptotic saturation effect we turn to an assessment model based on logistic growth.

4.2. The logistic assessment model for knowledge mastery

In many respects, the process of individual learning is all about change, growth and development. Therefore, the interplay of the individual learner and a serious game can be considered a dynamic system, which internal structure and state continually change over time. The process of learning displays many similarities with system dynamics, which generally describes growth as a function of restricted available resources. The dynamic systems approach incorporates complex feedback mechanisms which are essentially nonlinear in nature. Systems dynamics is able to account for chaotic behaviours, instabilities, strange attractors, phase transitions and other phenomena that are associated with complexity. The complexity of learning and instruction processes certainly argues in favour of such exploration of nonlinearities, as it may yield insights in the conditions for optimum knowledge growth to occur. Both internal and external variables put limitations to the growth of knowledge, which suggests that the learning of an individual readily qualifies as a system's constrained growth process. In particular, logistic functions could be used to cover the constrained growth process of learning.

The starting point for the logistic growth of the player's knowledge mastery is the following recurrent difference equation, expressed as a function of time t:

$$M_{t+\Delta t} - M_t = g \cdot M_t \cdot (M_{max} - M_t) \cdot \Delta t \tag{4}$$

where

t is time

Δt is the increment of time

M_t is the knowledge mastery level at time t

M_{max} is the level of ultimate perfection, and

g the unconstrained growth rate.

The logistic model complies with the proverb that "every beginning is hard", because initial growth of knowledge mastery is proportional to the actual mastery level M_t: low mastery would lead to small growth, whereas high mastery means large growth. But the growth of mastery is also proportional to the remaining knowledge mastery gap (M_{max}-M_t), which accounts for the concept of "full mastery" or "perfection" (M_{max}) and the phenomenon that striving for perfection becomes increasingly hard when one approaches this ultimate level.

Again, it is assumed that the learning environment is composed of a set of separate challenges. Given the duration D_i of the i-th challenge (cf. equation(3)), the time interval associated with the challenge is given by

$$\sum_{j \leq i-1} D_i < t \leq \sum_{j \leq i} D_i \tag{5}$$

The main starting point is that the engagement of a player in a game challenge inherently leads to enhanced mastery, be it that the growth rate *g* may be different for successes and failures. To this end, we use S_i as the binary success indicator for the i-th challenge, as introduced in equation(1). Let *α* be the unconstrained growth rate in case of success and *β* the unconstrained growth rate in case of failure, then the growth rate *g* in equation (2) can be rewritten as g_i

$$g_i = \alpha \cdot S_i + \beta \cdot (1 - S_i) \tag{6}$$

Although g_i is constant during the challenge it may vary between different challenges (either equal to α or β), making it time dependent.

Equation(4) representing the knowledge mastery growth model can now be rewritten as

$$M_{t+1} - M_t = \left(\alpha \cdot S_i + \beta \cdot (1 - S_i)\right) \cdot M_t \cdot (M_{max} - M_t) \cdot \Delta t \tag{7}$$

This difference equation generally can be solved analytically, be it constrained to each interval (fixed growth rate). Also, solutions can be generated through Monte-Carlo simulations, which will be presented in the next sections. The resulting mastery curves will be demonstrated to comply with the general S-shape of constrained growth given by the logistic principles: starting from a certain prior knowledge level, the knowledge mastery first growths exponentially, but gradually tends to saturate and asymptotically approach the level of perfection, as it becomes increasingly hard to improve any further.

5. Model exploration

5.1. Technical implementation

To allow for Monte-Carlo Simulations for investigating knowledge growth behaviours described by equation(7) under various conditions, the model was technically implemented in a series of Scilab programmes (www.scilab.org), while using its math functions and graphics libraries.

5.2. Baseline parameters

After preliminary exploration and testing the following parameters were set to define a baseline model with realistic growth (Table 1):

Table 1. Baseline model parameters and explanations.

Parameter	Value	Explanation
Initial knowledge mastery level M_1	0.01	This is an arbitrary low prior knowledge mastery level. This value does not fundamentally influence the shape of the resulting growth curves.
Success growth rate α	0.005	A growth rate as small as this requires a substantial number of iterations to approach high mastery
Failure growth rate β	0.001	The failure growth rate was chosen to be considerably lower that the success growth rate.
Challenge success rate S	0.60	The odds of successes are 6 to 4.
Time increment Δt	1	Time is no more than a scaling factor. Each time step covers the completion of one game challenge out of an unlimited pool of challenges.
Challenge durations D_i	50	Expressed as the number of time increments Δt
Number of challenges n	50	Similar to the linear case (cf. Figure 1)
Total number of time steps N	2500	This equal to D_i times n

5.3. Logistic learning curves

Figure 2 displays a randomly generated growth curve and the average growth curve after 500 iterations for the baseline case. It also shows the range of the standard deviations and the two extreme cases of all successes (upper dashed line) and all failures (lower dashed line).

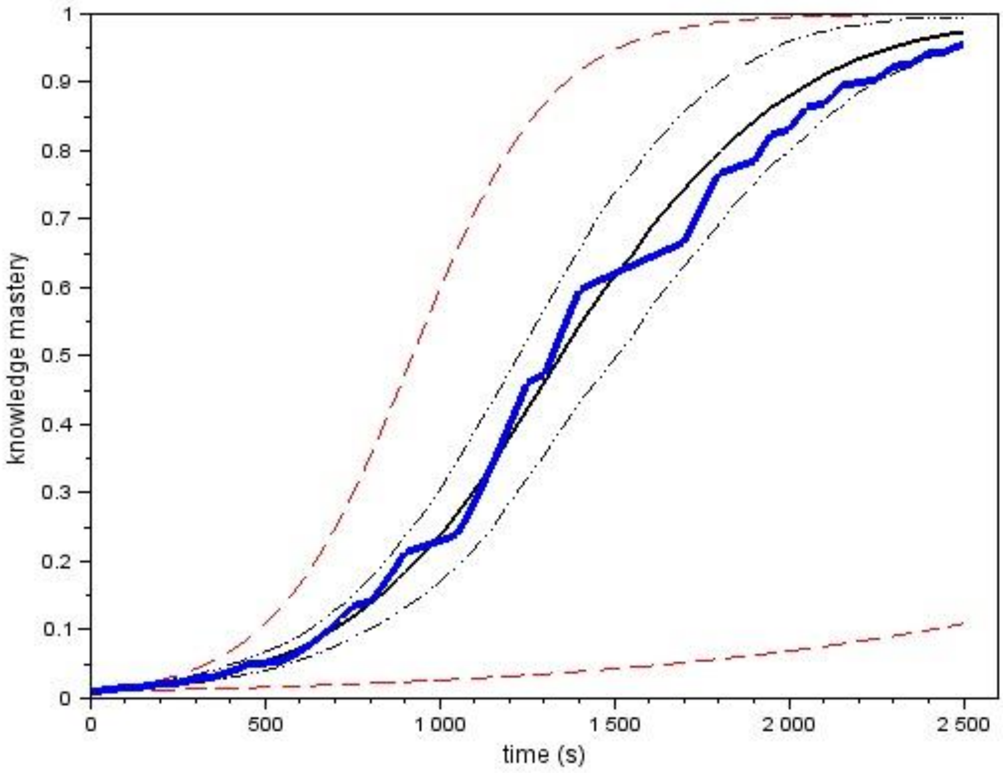

Figure 2. An exemplary growth curve of the baseline logistic model (bold solid), the mean baseline growth curve (thin solid) after 500 runs, the variability range (dash-dotted lines), and full success curve and full failure curve, respectively (dashed lines).

The mastery growth curve for a single run (bold solid line in figure 2) roughly follows the logistic pattern, while it is composed of different logistic curve fragments defined by either the growth rate for successes α or for failures β.

5.4. Comparing linear and logistic assessments

The different assessment models should be compared for the very same game session. The (arbitrary) performance scale is aligned with the logistic mastery scale by setting the final linear performance score P_n (after n challenges) and the attained (logistic) mastery level M_N at the end of the game to coincide. By definition, the final mastery level is constrained to the normalised scale and any realistic level of mastery will inevitably be well below one (perfection). For a balanced comparison, three distinct assessment metrics should be evaluated: 1) the progression metric, and 2) the efficacy metric, and 3) the efficiency metric, respectively, all expressed as a normalised scalar.

5.4.1. The progression ratio

The progression ratio for logistic mastery as a function of time t is defined as

$$RM_{progression,t} = \frac{M_t}{M_N} \tag{7}$$

which indicates the player’s progression at each step *i* toward the final (target) mastery level. Likewise, the progression ratio for the linear performance at challenge i is written as

$$RP_{progression,i} = \frac{P_i}{P_n} \tag{8}$$

Figure 3 shows the progression ratios for logistic mastery and linear performance, respectively, as a function of time.

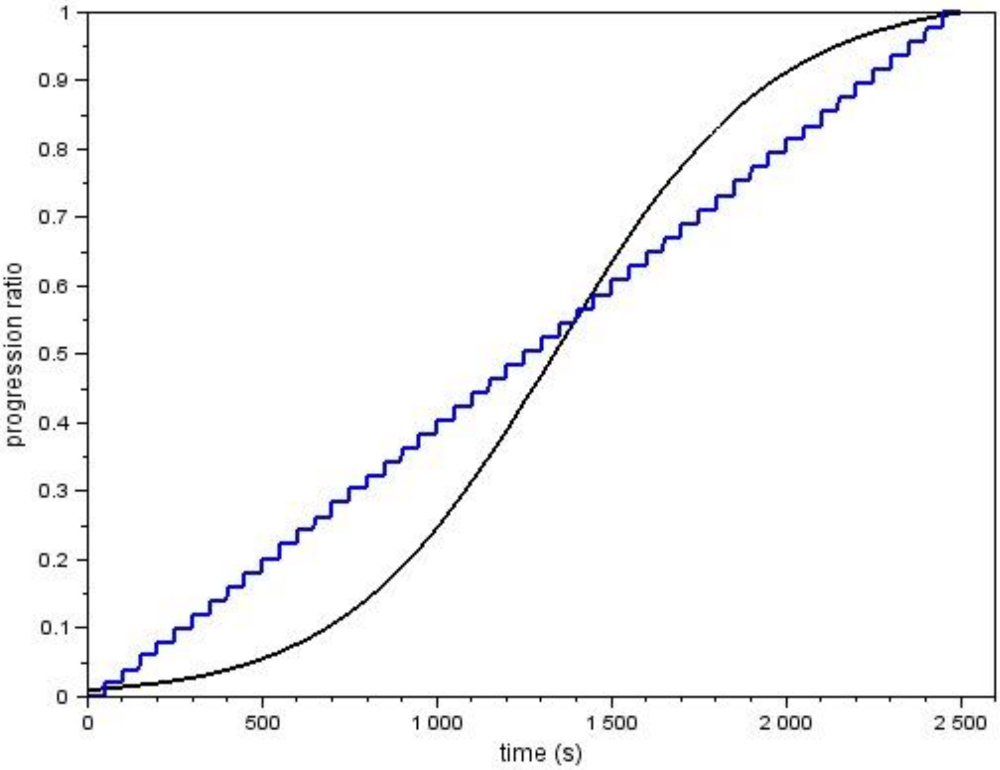

Figure 3. Progression ratios with respect to the final achievement level for logistic mastery (fluent curve) and linear performance (staircase line), respectively, based on 500 simulated game runs for the baseline case (cf. Table 1).

A low value of the progression ratio signals that the player still has a long way to go. The staged performance curve reflects the discrete nature of assigning scores only after completion of each challenge. The figure shows that linear performance initially tends to overrate the player's achievements, which is then compensated for during the final stages of the game, when levels of substantial mastery are reached. The progression ratios only reflect the distance to the final level, but do not necessarily reflect the quality of work so far. Hence an additional metric is needed to cover this.

5.4.2. The efficacy ratio

The quality of achievement so far is expressed by the efficacy ratio, which is defined as follows for the mastery model

$$\mathrm{RM}_{efficacy,t} = \mathrm{M}_t / \mathrm{M}_{max,t} \tag{9}$$

The progression ratio relates the player's achievements at time *t* to the best possible achievement level that the player could have reached at time *t*, rather than to the final target level M_N. It is the fair ratio of the actual achievement M_t and the optimum achievement $M_{max,t}$ with M_{max} being the curve of successes only. The analogous formula for the linear performance efficacy ratio after challenge i reads

$$\mathrm{RP}_{efficacy,i} = \mathrm{P}_i/\mathrm{P}_{max,i} \tag{10}$$

where *Pmax,i* is the curve of maximal performance (all successes). Figure 4 shows the efficacies for both models.

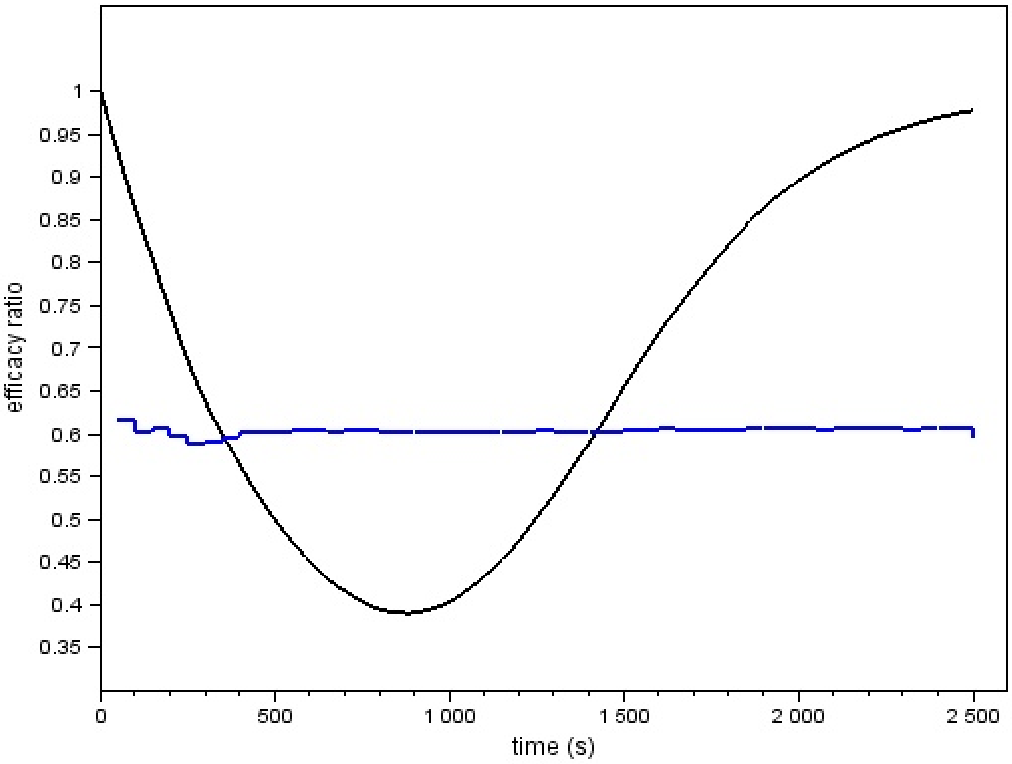

Figure 4. Efficacy ratios for logistic mastery (fluent line) and linear performance (staircase line), respectively, based on 500 simulated game runs of the baseline case (cf. Table 1).

While the efficacy for the linear model gravitates around a stable level (given by the challenge success rate S=0.60), the logistic model show a profound dependency on time. This is the direct consequence of the exponential growth occurring at low mastery levels: every early failure disproportionally reduces the cumulative growth as compared to the optimal achievement. In the course of time, however, when the optimal achievement curve starts to saturate, the arrears are being overtaken. This metric is thus informative about the quality of work so far as compared to what could have been achieved at this stage. It does not tell anything about the remaining distance to the final level M_N, which is covered by the progression ratio as explained before.

5.4.3. The efficiency ratio

The efficiency of achievements is given by the time spent so far compared to the minimum time (or number of steps i_{min}) required for having successes only, which can be expressed as follows:

$$\mathrm{RM}_{efficiency,i} = \frac{t_{min}}{t} = \mathrm{M}_{max,t}^{-1}(M_t)/t \tag{11}$$

and similarly for linear performance after challenge i

$$\mathrm{RP}_{efficiency,i} = \frac{i_{min}}{i} = \mathrm{P}_{max,i}^{-1}(P_i)/i \qquad (12)$$

Figure 5 shows the two efficiency ratios.

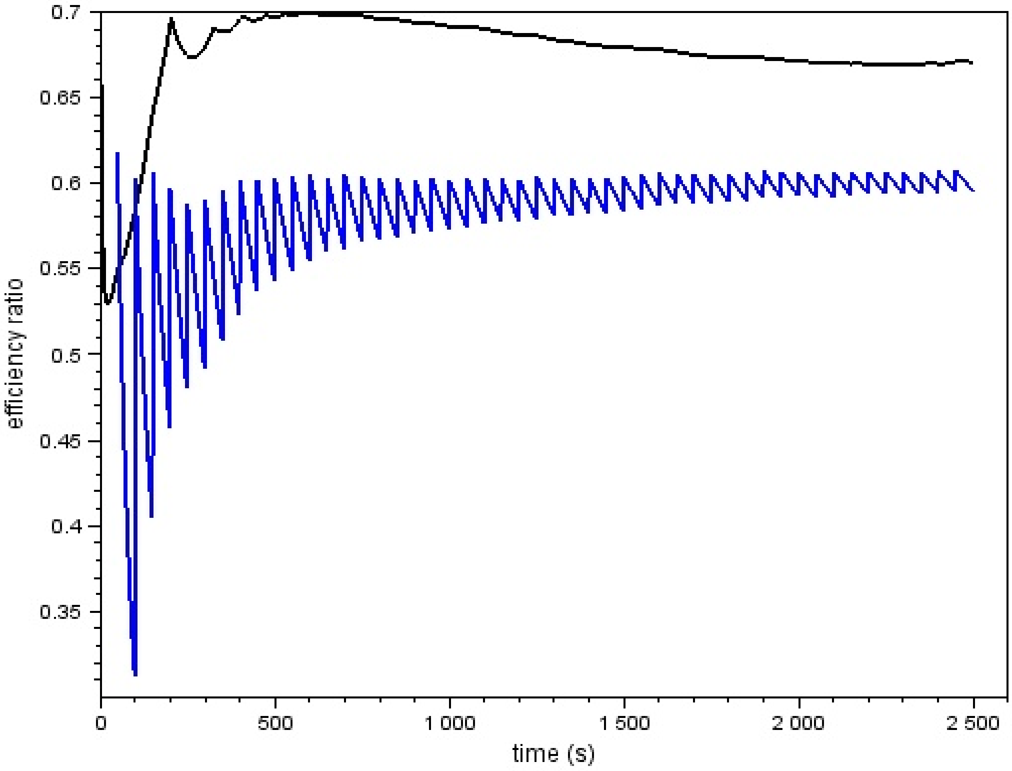

Figure 5. Efficiency ratios for logistic mastery (fluent curve) and linear performance (sawtooth), respectively, based on 500 simulated game runs for the baseline case (cf. Table 1).

The efficiency ratio of the average performance repetitively touches on the success rate level (S=0.60), but decreases hyperbolically during the dead times of challenge execution, since performance scores are assigned only after completion of the challenge. The nonlinear skew of the logistic model causes the efficiency ratio to strongly deviate from the efficacy ratio (cf. Figure 4). The efficiency ratio of the logistic model tends to be systematically higher than the ratio for the linear mode, which can be explained by the average growth factor (cf. equation (6)), which predicts a final efficiency level for the baseline case of 0.68. Overall, the linear model systematically underestimates the player’s efficiency.

5.5. Influence of growth rates α, β and success rate S

As the effective growth rate of the logistic model is given by equation(6), both α, β en S affect the shape of the resulting growth curves. When β approaches α, which means that learning from mistakes approaches the learning from successes, the logistic growth will be faster (cf. equation(6)), whereby the logistic progression ratio (cf. figure 3) will start to exceed the performance progression ratio at an earlier stage. The opposite happens when β gets smaller. For the logistic efficacy ratio the same holds: when β approaches α, the efficacy goes up, which leads to a radically less pronounced dip in the curve (cf. figure 4). Likewise, the logistic efficiency ratio goes up, while the linear performance efficiency ratio remains unaffected.

Similarly, a raised success rate S leads to higher growth rates, procuring the same effects, be it, however, that now the performance efficacy ratios and efficiency ratios no longer remain the unaffected, but increase as well. The linear progression ratio remains unaffected as a result of the enforced calibration of final performance to the final mastery level. Overall, the exploration with different values for α, β en S show consistent outcomes that entail either less-pronounced or more-pronounced differences between logistic curves and linear performance curves. We wish to emphasise that we have only described here tendencies of behaviours that are averaged over a large number of simulated game runs. Individual play sessions could deviate considerably from the average behaviours (cf. the edgy curve in figure 2).

6. Case study: the Playground Game

To further explore the model's behaviours with real-world data, log file data obtained from the Playground game were used (Westera, Slootmaker, and Kurvers , 2014). The Playground game deals with statistics, a domain infamous among university students because of its inherent complexity (Beurze, Donders, Zielhuis, Vegt, and Verbeek, 2013; Griffith, Adams, Gu, Hart, and Nichols-Whitehead, 2012). The game offers students a situated problem to get acquainted with statistical concepts and their practical significance. The player's task in the game is to decide upon the most suitable location for laying out a children's playground in a fictitious town, while taking into account a variety of factors and data. The starting point of the game is a research report written by a "consultant". This report, however, contains deliberate flaws and fallacies, some of which are manifest and some of which are obscured or subtle. The player's task is to judge the correctness of the approach and the validity of the outcomes by interrogating the consultant who is the author of the report and a contra-expert who criticises the report - both represented in interactive videos. The game is structured along eight problem areas surfacing in the consultant's report, which the players can address in arbitrary order. For each problem area a set of challenges is offered, each requiring well-considered decision taking. The game's score system assigns bonus points for each correct decision. The Playground Game is a web-based game, a demo-version of which (in Dutch) is available at http://goo.gl/mwH9YL.

6.1. Data collection and processing

The game was administered to psychology students from the bachelor programme of Leuven University. It was combined with a post-questionnaire composed of 28 items concerning the players' appreciations, judgements about the game, but importantly also including 1) a self-assessment, 2) a set of questions about familiarity with statistical concepts, and 3) a set of five methodological test questions, which were combined into an aggregated metric reflecting the post-game mastery level. Out of 125 subscriptions, 117 participants completed the game and questionnaire. The anonymised log files of students were filtered to identify the key events, in particular, the timestamps of each game challenge and the awarded scores. This way, the game data represented a chain of subsequent challenges and the associated successes and failures, which is in agreement with the frugal model.

The session times were typically 1 hour (standard deviation 53 minutes), showing large variability at the high end. Performances were based on the scoring model (assigning 1 point for a success) and rescaling these by enforcing the final performance to equal the post-game test outcome. Model computations were carried out with time steps Δt of 1 second, the initial mastery level set to 0.01 and the ratio of failure growth rate and success rate β/α set to 0.2 (in accordance with the baseline model). Equating the final mastery level that a student attained with the test metric from the questionnaire allowed to estimate the growth rates α and β for each student. This was done by first estimating seed values of α and β from the alignment of an analytical solution with the test score, and then subsequently refining the numerical solution with a difference reduction metric (tolerance < 1%).

6.2. Case results

For each participant the log file data were used to calculate the performance curve and the knowledge mastery curve. Figure 6 shows both curves as a function of time, averaged over the population. Although the resulting curves seem to confirm the general pattern obtained from the simulations that the performance scores most of the time exceed the mastery scores (cf. figure 3), the curves are deceptive because they do not take into account the different session times of participants: at each point in time different players will be at different stages in the game. Consequently, it means that that pool of participants gradually decreases with time because of more and more players will have completed and exited their session. The time axis in figure 6 is cut off at 5733 seconds, the ultimate point in time that a minimum of 16 players is preserved in the game.

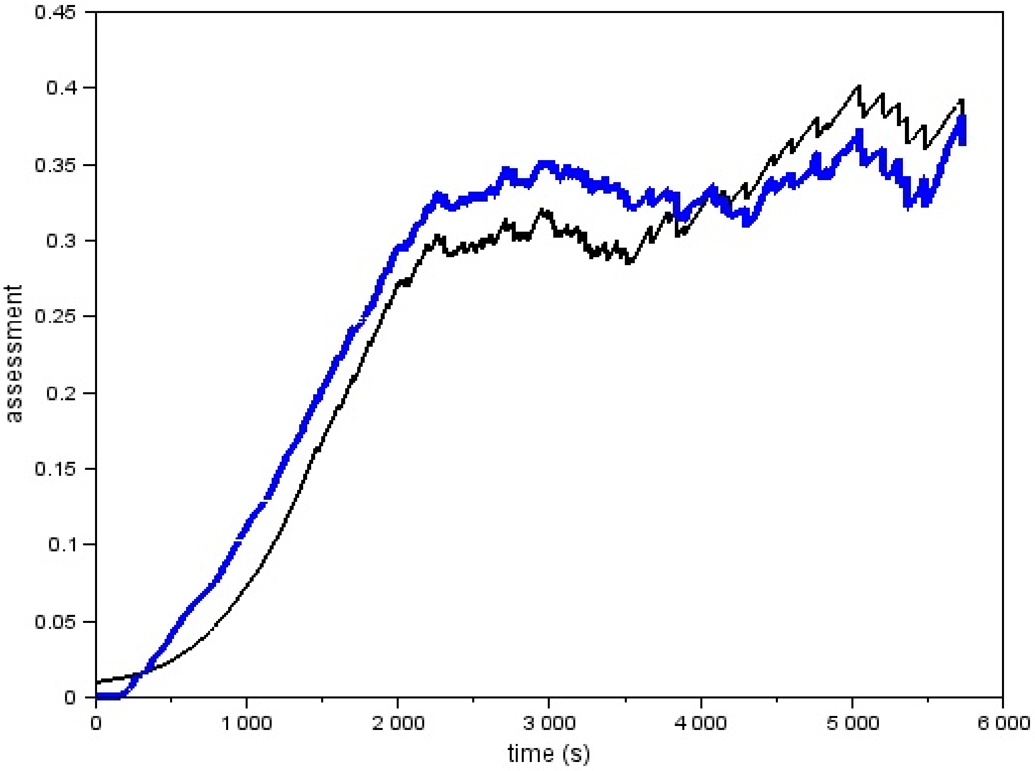

Figure 6. Average performance (bold solid line) and knowledge mastery (thin line) for the Playground game participants as a function of time.

To allow a more appropriate comparison across the pool of participants the different time scales were rescaled to match a fix standard covering each participant's full session. Figure 7 shows the average performance curve and knowledge mastery curve mapped onto a reference scale of 10,000 units: it represents the average patterns of progression toward completion of the game, irrespective of true time.

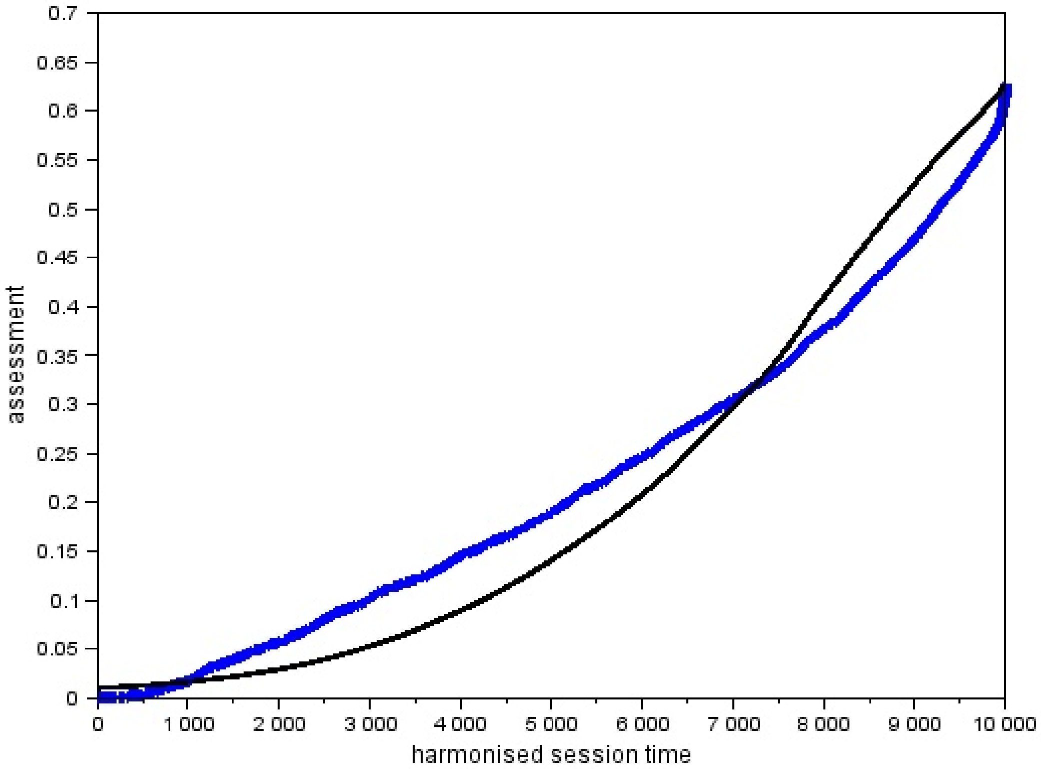

Figure 7. Average performance (bold solid line) and knowledge mastery (thin solid line) of the Playground population aligned to a reference time scale of 10,000 units.

These harmonised curves show quite good agreement with the pattern displayed in figure 3: performance levels tend to exceed the mastery levels for the larger part from the start, which is then compensated for toward the end. A slight deviation from the simulation-based pattern may be observed in the increasing rather than constant slope of the performance curve, which suggests that the time to successfully complete a challenge gets shorter toward the end of the game. This might be explained by the players becoming more knowledgeable about the game's content during the session, which procures faster achievement of successes. Such a mechanism is not assumed in the proposed computational model, but may be easily added if challenge complexity data were available.

The average growth rate for successes (α) was found to be 0.0028 per second with a large standard deviation of 0.0013 s^{-1}. Given the pre-set value of α/β=0.2, the corresponding rate for failures is 0.0014 s^{-1} (standard deviation 0.0007 s^{-1}). Logarithms of the individual growth rate α and game session times showed substantial, significant correlation (Spearman's ρ=-0.88). This is consistent with the fact the session time was used for deriving an estimated seed value for α. From linear regression the following hyperbolic relationship was derived, using α as a predictor for session time T in this game:

$$T = \frac{8.67}{\alpha} \tag{13}$$

Further correlations were found between players' average challenge duration and session time (ρ=0.79, slope=56.2, intercept=616 (s)), between average challenge duration and growth factor α (log-log ρ=-0.66, log slope=-0.66 (s^{-2}), log intercept=-3.44 (s^{-1})) and, be it less pronounced, between the fraction of time spent to successes and growth factor α (ρ =-0.48, slope=-0.0035 (s^{-2}), intercept=0.0047 (s^{-1})).

7. Discussion and conclusion

Model exploration has demonstrated that the differences between common performance metrics and logistic models of knowledge mastery are substantial. Linear performance metrics, which are used in many serious games, tend to overestimate the player's logistic knowledge mastery at early stages and underestimate it at the end. Although the study was inspired by the idea of learning from mistakes, the differences remain when this factor is excluded (β=0), be it less pronounced. Evaluation of three assessment metrics describing progression, efficacy and efficiency, respectively, revealed structural differences between linear performance and logistic mastery. Empirical testing with real-world game data showed results that are largely consistent with the results from Monte-Carlo simulation. Various correlations were identified, most notably the hyperbolic relationship between session time and growth rate α. Since the latter may be regarded as a stable personal characteristic (within a given domain), it may act as a personal predictor of session time (speed).
This study, which is positioned as an early exploration of the topic, is not without limitations. The presented model provides an unmistakably simplified representation of serious game environments. First, it treats knowledge as a single scalar, while in practice multidimensional competency constructs may be needed. If the superposition principle applies, however, this multidimensionality need not be a fundamental issue. Second, the model uses the dichotomy of successes and failures. It may need extension to more detailed scales. Third, the approach so far assumes that the game challenges are exchangeable and independent. In practise, game challenges may be of different complexity and they are likely to be conditionally dependent requiring a preferential sequence order possibly enforced by narrative. Although accounting for this would require considerable model extension, it would not be a principle problem. Fourth, the model assumes uninterrupted learning of stable intensity at every stage of the game by neglecting any periods of concentration flaws, persistent failure, mood changes, pauses, interruptions or any unproductive action such as navigation. Finally, the assumption that the processes of learning can be adequately described by a logistic growth model would need further investigation, including a well-controlled set-up that allows for frequent, monitoring of knowledge mastery.

Overall, this investigation has revealed substantial differences between linear performance assessment and the logistic assessment of knowledge mastery. Linear performance models may be well suited to assess correctness and fluency of task execution and the associated operational skills, but partially neglect the underlying factors of understanding, deep processing and reasoning. The ideal student completing the game quickly without errors may have learned less than the slow, cautious and reflective learner that frequently fails but sturdy reconsiders and gradually recovers. Linear performance metrics are simply based on milestones, but disregard the concept of growth. The logistic model fully embodies the concept of growth and thereby it possesses a more robust theoretical grounding, which goes beyond the pragmatics of performance assessment. The relevance of the work goes beyond the particular results of this study in that it extends and complements the field of educational research with new computational modelling methodologies, which have proven highly successful in various other domains by recreating observed phenomena and manipulating these by varying contextual parameters and constraints (De Marchi, 2005). Computational modelling methodologies are likely to positively contribute to enhanced learning theory formation and testing, enhancing the explanatory and predictive power of educational research, and to anticipating the ever growing opportunities for data driven research methodologies in the digital era.

References

Abt, C. (1970). *Serious games*. New York: Viking Press.

Aldrich, C. (2005). *Learning by Doing: the Essential Guide to Simulations, Computer Games, and Pedagogy E-Learning and other Educational Experiences*. San Francisco, CA: John Wiley & Sons.

Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives.* New York: Longman.

Barrows, H. S., & Tamblyn, R. M. (1980). *Problem-based learning: An approach to medical education.* New York: Springer.

Beurze, S, Donders, R., Zielhuis, G. , Vegt, F., & Verbeek, A. (2013). Statistics Anxiety: A Barrier for Education in Research Methodology for Medical Students?. *Medical Science Educator, 34(3), 377-384*.
http://dx.doi.org/10.1007/BF03341649

Björk, S., & Holopainen, J. (2005). *Patterns in Game Design.* Hingham: Charles River Media.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, H. W., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The cognitive domain*. New York: Longman.

Boyle, E. A., Hainey, T, Connolly, T. M., Graya, 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

Bridger, E. K., & Mecklinger, A. (2014). Errorful and errorless learning: The impact of cue-target constraint in learning from errors. *Memory & Cognition, 42, 898–911.*
http://dx.doi.org/10.3758/s13421-014-0408-z

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review, 31(1), 21–32.*

Calhoun, A. W., Boone, M. C., Porter, M. B., & Miller, K. H. (2014). Using simulation to address hierarchy-related errors in medical practice. *The Permanente Journal, 18(2), 14-20.*
http://dx.doi.org/10.7812/TPP/13-124

Cattaneo, A., & Boldrini, E. (2017). You Learn by your Mistakes. Effective Training Strategies Based on the Analysis of Video-Recorded Worked-out Examples. *Vocations and Learning, 10(1), 1-26.*
http://dx.doi.org/10.1007/s12186-016-9157-4

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.*
https://doi.org/10.3102/0034654315582065

Cyr, A.-A. , & Anderson, N. D. (2015). Mistakes as stepping stones: Effects of errors on episodic memory among younger and older adults. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 41(3), 841 -850.*

De Marchi, S. (2005). *Computational and Mathematical Modelling in the Social Sciences*. New York: Cambridge University Press.

Dewey, J. (1938). *Experience and Education.* New York: The Macmillan Publishing Company.

Dweck, C.S., & Legget, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychology Review, 95(2), 256–273.*

Fisher, S .L., & Ford, J. K. (1998). Differential effects of learner effort and goal orientation on two learning outcomes. *Personnel Psychology, 51(2), 397–420.* https://doi.org/10.1111/j.1744-6570.1998.tb00731.x

Gardner, A. K., Abdelfattah, K., Wiersch, J., Ahmed, R. A., & Willis, R. E. (2015). Embracing errors in simulation-based training: the effect of error training on retention and transfer of central venous catheter skills. *Journal of Surgical Education, 72 (6), e158-e162.* http://dx.doi.org/10.1016/j.jsurg.2015.08.002.

Gee, J. P. (2003). *What Video Games Have to Teach Us About Learning and Literacy.* New York: Palgrave/Macmillan.

Griffith, J. D., Adams, L. T., Gu, L. L, Hart, C. L., & Nichols-Whitehead, P. (2012). Students' attitudes toward statistics across the disciplines: A mixed-methods approach. *Statistics Education Research Journal Archives, 11(2), 45-56.* http://iase-web.org/documents /SERJ/SERJ11(2)_Griffith.pdf Accessed 23 July 2019.

Hegg Reime, M., Johnsgaard, T., Kvam, F. I., Aarflot, M., Breivik, M., Engeberg, J. M., & Brattebø, G. (2016). Simulated settings; powerful arenas for learning patient safety practices and facilitating transference to clinical practice. A mixed method study. *Nurse Education in Practice, 21, 75-82.* https://doi.org/10.1016/j.nepr.2016.10.003.

Heller, J., Steiner, C., Hockemeyer, C., & Albert, D. (2006). Competence-based knowledge structures for personalised learning. *International Journal on E-learning, 5(1), 75-88.*

Huelser, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & Cognition, 40, 14–527.*

http://dx.doi.org/10.3758/s13421-011-0167-z

Ivancic, B., & Hesketh, K. (2000). Learning from error in a driving simulation: Effects on driving skill and self-confidence. *Ergonomics, 43, 1966-1984.*

Jonassen, D. (1991). Objectivism versus constructivism*. Educational Technology Research and Development, 39(3), 5–14.*

Keith, N., & Frese, M. (2008). Effectiveness of error management training: A meta-analysis. *Journal of Applied Psychology, 93, 59-69.*

Kelley, H. H. (1973). The Processes of causal attribution. *American Psychologist, 28(2), 107–128.*

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An Analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41(2), 75–86.*

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development.* Englewood Cliffs, NJ: Prentice-Hall.

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge, MA: Cambridge University Press.

Mathan, S. A., & Koedinger, K. R. (2005). Fostering the intelligent novice: learning from errors with meta-cognitive tutoring. *Educational Psychology, 40(4) 257–265.*

Mayer, R. (2004). Should there be a three-strikes rule against pure discovery learning? The Case for guided methods of instruction. *American Psychologist, 59(1), 14–19.*

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1(1), 3-62.* https://doi.org/10.1207/S15366359MEA0101_02

Mory, E. H. (2003). Feedback Research Revisited. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp.745-783). New York: MacMillan Library Reference.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. New York: Basic Books.

Polanyi, M. (1966). *The Tacit Dimension.* Chicago: University of Chicago Press.

Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General, 143, 644–667.* http://dx.doi.org/10.1037/a0033194

Radosavljević, M. (2015). Strategy Of Learning From Mistakes*. International Journal of Law & Economics, 13, 125 -131.*

Roediger H. L., & Karpicke J. D. (2006). Test-enhanced learning: taking memory tests improves long-term retention. *Psychological Science, 17, 249–255.* https://doi.org/10.1111/j.1467-9280.2006.01693.x

Schaeffer, S. (2002). Tic-Tac-Toe (Naughts and Crosses, Cheese and Crackers, etc.). Mathematical Recreations Website. http://www.mathrec.org/old/2002jan/solutions.html Accessed 23 July 2019.

Schank, R. C. (1995). *Engines for Education.* New York: Lawrence Erlbaum.

Seligman, M. E., Maier, S. F., & Geer, J. H. (1968). Alleviation of learned helplessness in the dog. *Journal of Abnormal Psychology, 73(3), 256–262.*

Shute, V. J., & Ventura, M. (2013). *Measuring and supporting learning in games: Stealth assessment.* Cambridge, MA: The MIT Press.

Simm, D. (2005). Experiential learning: assessing process and product. *Planet, 15(1), 16-19.* https://doi.org/10.11120/plan.2005.00150016

Tjosvold, D., Yu, Z., & Hui, C. (2004). Team learning from mistakes: The contribution of cooperative goals and problem-solving. *Journal of Management Studies, 41, 1223-1245.*

VandeWalle, D., Brown, S. P., Cron, W. L., & Slocum, L. W. (1999). The influence of goal orientation and self-regulation tactics on sales performance: A longitudinal field test. *Journal of Applied Psychology, 84(2), 249–259.*

Vargas, J. S. (1986). Instructional Design Flaws in Computer-Assisted Instruction. *The Phi Delta Kappan, 67(10), 738-744.* http://www.jstor.org/stable/20403230 Accessed 23 July 2019.

Weinzimmer, L. G., & Esken, C. A. (2017). Learning from mistakes: How mistake tolerance positively affects organizational learning and performance*. Journal of Applied Behavioral Science, 53(3), 322-*

348.
http://dx.doi.org/10.1177/0021886316688658

Westera, W. (2015). Games are motivating, aren´t they? Disputing the arguments for digital game-based learning. *International Journal of Serious Games 2(2).*
http://journal.seriousgamessociety.org/index.php?journal=IJSG&page=article&op=view&path%5B%5D=58 Accessed 23 July 2019.

Westera, W. (2017). How people learn while playing serious games: A computational modelling approach. *Journal of Computational Science, 18(1), 32-45.*
http://dx.doi.org/10.1016/j.jocs.2016.12.002.

Westera, W. (2019). Why and How Serious Games can Become Far More Effective: Accommodating Productive Learning Experiences, Learner Motivation and the Monitoring of Learning Gains. *Educational Technology & Society, 22 (1), 113–123.*

Westera, W., Slootmaker, A.,& Kurvers, H. (2014). The Playground Game: Inquiry-Based Learning About Research Methods and Statistics. In C. Busch (ed.), *Proceedings of the 8th European Conference on Games Based Learning* (pp. 620-627). Sonning Common (UK): ACPI.