Three Strikes and Who is Out? Individual Differences in Quitting Following Repeated Errors in Online Learning

Annie M. Johansson[a], Alexander O. Savi[a], Han L. J. van der Maas[a], and Abe D. Hofman[a, b]

[a]Department of Psychological Methods, University of Amsterdam, Nieuwe Achtergracht
129-B, 1018 WS Amsterdam, The Netherlands
[b]Prowise Learn, Luchthavenweg 1, 6021 PX Budel, The Netherlands

January 16, 2026

Abstract

No matter how well a learning system is designed, it becomes irrelevant when students quit; once learners disengage, learning ceases. It is therefore paramount to understand why students quit. To answer this question, we leverage intensive longitudinal practice data from over $200\,000$ primary-school students in a large-scale online learning environment. First, we show that sequential errors strongly increase the probability of quitting from learning. Second, we find large variability in this effect, ranging from no or small tendencies to quit to high sensitivities to quit following sequential errors. Third, the effect of errors on quitting differed across age, difficulty level, time of day, and response times. We validate these results in an independent dataset and show that individual differences are stable across two arithmetic practice domains. While the threat of disengagement is real, our results provide actionable insights that can both help identify undesirable sources of disengagement and ultimately help counter it.

Keywords student engagement, quitting, online learning, errors, Markov models, individual differences

1 Introduction

The topic of engagement, and its counterpart, disengagement, is central to educational psychology (Sinatra et al., 2015). The ultimate form of disengagement from learning is quitting, deliberately removing oneself from the learning process before its intended end. This may be in the form of dropout, as often seen in informal learning environments such as Massive Open Online Courses (MOOCs), where quitting is notably high (e.g., Borrella et al., 2022). In formal learning environments, particularly on a primary-school level, quitting less often takes the form of course dropout. Rather, it is more often characterized by a state of short-term disengagement from the current learning task. Conceptualizations of engagement and quitting at this finer grain size have become increasingly common in the literature, albeit under slightly different labels, including stopout (Botelho, Varatharaj, Inwegen, & Heffernan, 2019), level quitting (Karumbaiah et al., 2018), micro-persistence (Israel-Fishelson & Hershkovitz, 2020; Klein-Latucha & Hershkovitz, 2024), and momentary engagement (Symonds et al., 2024). In this paper, we refer to it as soft-quitting.

While soft-quitting does not directly imply withdrawal from schooling, it is nonetheless consequential. Educational resources—money and effort spent on instruction, curriculum materials, and digital learning platforms—can only generate their desired value when students enage with them sufficiently. Inefficient use of educational resources not only poses financial threats on society (Botelho, Varatharaj, Inwegen, & Heffernan, 2019; Brunello & Paola, 2014), they also imply consequences for individual learning outcomes (Fredricks et al., 2004), employability (Saqr & López-Pernas, 2021; Symonds et al., 2023), and well-being (Martins et al., 2022; Wong et al., 2024). On top of that, when quitting is systematic and unevenly distributed due to individual differences within and across schools, structural inefficiency in how educational resources are converted into learning is created, which risks widening educational inequalities in society at large (Bryan et al., 2021).

Soft-quitting can also be seen as an early warning sign for larger, more persistent, forms of quitting. Quitting in online learning has been found to dynamically change over time (López-Pernas & Saqr, 2024; Saqr & López-Pernas, 2021; Saqr et al., 2023; ten Broeke et al., 2022), and different students have been found to portray different patterns of soft-quitting behavior (Botelho, Varatharaj, Patikorn, et al., 2019; Karumbaiah et al., 2018; Klein-Latucha & Hershkovitz, 2024). Importantly, students who have made more soft-quits in the past are more likely to continue soft-quitting (Karumbaiah et al., 2018), and soft-quitting may predict longer-term quitting, such as staying disengaged for the rest of the day (ten Broeke et al., 2022, here referred to as hard-quitting). These findings suggest that quitting is not a one-off behavior but a dynamic process influenced by both contextual and individual factors.

A central question, therefore, is what experiences push learners toward disengaging rather than persisting when learning becomes difficult. In learning environments, one of the most immediate and informative signals of such difficulty is making an error. Errors provide feedback in learning, but may also function as motivational stressors that gives rise to quitting. In the following sections, we review theoretical and empirical work on the role of errors in learning and engagement, with particular attention to how learners differ in their responses to repeated failure, and how these differences may give rise to systematic variation in quitting behavior.

1.1 Previous Research on Errors and Quitting

There is a paradox to errors in learning. On the one hand, errors are necessary for learning. They serve as indicators of performance and signal potential flaws that can be analyzed and utilized for guiding learning (Metcalfe, 2017; Wilson et al., 2019). On the other hand, making errors while learning can be frustrating, and an adequate level of motivation is crucial to stay engaged (Brophy, 2013). For this reason, matching the complexity of the instructional material to the proficiency of the learner is a central goal in any learning environment.

However, educational environments and instructional materials are often developed with the implicit assumption that this tension, *i.e.,* the amount of error necessary for a student to maintain learning and simultaneously resist disengaging, is similar across learners. In this paper, we scrutinize this assumption by investigating how errors influence disengagement from learning, and, importantly, how this varies between learners. This is a crucial question for informing learning design; disregarding variability while it is present risks creating a mismatch between instructional challenge and learner engagement, and may widen existing inequalities between learners (Bryan et al., 2021).

Notably, errors have been shown to predict quitting in online learning environments. Repeated mistakes are associated with both reduced learning gains and an increased likelihood of disengagement (Hou, 2015; Kristensen et al., 2024). According to Hou (2015), maintaining an optimal balance between challenge and success is crucial for sustaining engagement—a state often described as “flow” (Csikszentmihalyi, 2000). Their study found that students experiencing lower levels of flow in a learning game were more likely to make repeated mistakes, suggesting that disengagement and errors may reinforce each other.

Supporting this view, Arztmann et al. (2025) found that students who quit a learning game prematurely tended to show recurring patterns of mistakes—more so than those who quit simply because they failed to reach the last level, or those who persisted throughout. Botelho, Varatharaj, Inwegen, and Heffernan (2019) identified distinct clusters of students who were more likely to soft-quit immediately following incorrect responses. Similarly, Karumbaiah et al. (2018) and Liu et al. (2024) found that students who quit more showed more random and irrelevant interactions in a learning game. Together, these findings point to patterns of errors as a key trigger of quitting behavior.

1.1.1 Sequential Errors and Dynamic Quitting Processes

Recent work suggests that the relationship between errors and quitting is not limited to isolated responses to failure, but is inherently temporal in nature. In a large-scale online learning context, ten Broeke et al. (2022) modeled quitting behavior as transitions between engagement states across time: (1) persisting, (2) quitting an exercise while remaining within the learning environment (soft-quitting), and (3) quitting an exercise and exiting the learning environment entirely (hard-quitting). Using Markov models to capture these state transitions, they found that the accumulation of sequential errors was the strongest predictor of both soft- and hard-quitting. Specifically, making two, three, or more consecutive errors substantially increased students’ likelihood of transitioning from persistence to quitting, and this relationship was subsequently shown to be causal.

A similar perspective on quitting is advanced by López-Pernas and Saqr (2024) and Saqr and López-Pernas (2021), who conceptualize engagement and disengagement as processes that develop across time and depend on learners’ prior states. Using longitudinal transition and Markov-based models, these studies demonstrate that engagement tends to show continuity, such that prior engagement or disengagement increases the likelihood of remaining in similar states in subsequent learning moments. Crucially, López-Pernas and Saqr (2024) found that students who experienced higher previous success were more likely to transition to more engaged states.

Taken together, it is likely that quitting can emerge through the accumulatation of micro-level disruptions that gradually weaken engagement and increase the likelihood of disengaging. From this perspective, repeated errors can be understood as local pertubations that increase the likelihood of disengagement, and eventually trigger quitting behavior. These studies strengthen the understanding of quitting as a state-dependent, dynamic process that unfolds over time, and demonstrate the use of transition-based modeling approaches to capture this process.

1.1.2 Individual Differences

Productively persevering after errors requires a range of adaptive traits, including self-regulation and meta-cognitive skills. Skinner et al. (2020) offer a comprehensive developmental and social-contextual framework for understanding motivational resilience—patterns of engagement and effort across episodes of success and failure. Students who rebound after difficulty exhibit adaptive motivational responses, whereas those who disengage, for example, by quitting a task prematurely, show signs of motivational vulnerability.

A central mechanism underlying motivational resilience is self-regulated learning (SRL), which can be described as the learner's ability to monitor performance, sustain goal-directed effort, and regulate emotions under challenge (Zimmerman, 2002). This involves meta-cognitive processes such as error monitoring, which helps students detect performance mismatches and adjust behavior accordingly (Dent & Koenka, 2016; Pintrich, 2000; Zimmerman, 2002). Importantly, lower error monitoring has been linked to reduced conscientiousness and lower persistence (Hill et al., 2016; Miller et al., 2012). Similarly, individual variation in post-error slowing (PES)—a cognitive control mechanism in which students momentarily slow down after errors to adjust their behavior—has been associated with better performance in adaptive learning environments (de Mooij et al., 2022).

A large body of work has examined how learners respond to failure. Dweck's Implicit Theories of Intelligence explain that people hold implicit beliefs about whether their intelligence is a fixed trait or a malleable quality which can be developed through learning and growth (Dweck, 1986). Learners with a fixed mindset about intelligence tend to interpret errors as signs of personal inadequacy, leading to avoidance and withdrawal, while students with a growth mindset view errors as learning opportunities, fostering persistence (Dweck, 2013; Moser et al., 2011; O'Rourke et al., 2014).

While Implicit Theories of Intelligence provide a framework for understanding the global traits that lead learners to appraise their errors differently, Tulis et al. (2016) define two momentary error-related reactions, affective-motivational and action-based, which can be adaptive or maladaptive. Having adaptive affective-motivational reactions, which involve maintaining confidence and regulating frustration, has been positively linked to persistence and engagement, while maladaptive reactions predict disengagement and reduced effort (Grassinger et al., 2018). These forms of adaptivity operate via different mechanisms and may be unequally distributed across students; learners with low affective-motivational adaptivity may quit tasks prematurely even if they possess the cognitive skills to correct errors, because they are unable to tolerate or recover emotionally from failure. Similarly to Dweck (2013), these momentary error reactions are guided by whether the learner believes they can learn from the error in the first place: holding positive beliefs about errors as learning opportunities has been shown to promote emotional stability, increased task engagement, and persistence in academic settings (Tulis et al., 2018).

Taken together, these theories and findings suggest that there are individual-level variances in the ability to adaptively detect and respond to errors, giving rise to differences in the effect that sequential errors have on quitting between students. In this study, we model each student's tendency to quit after making errors as a reflection of this variability.

1.1.3 Group Differences

Apart from individual-specific effects, there may be group-level factors that influence engagement. For example, the characteristics needed to successfully cope with negative feedback and stay persistent on a task increase with age (Oeri et al., 2024). Furthermore, success rate plays a role in engagement: experiencing early success predicts that students will stay motivated for a task (Leonard et al., 2023; Lomas et al., 2013; Toste et al., 2020; Vu et al., 2022), and theories such as Expectancy-Value Theory (Eccles & Wigfield, 2002) and the Expectancy-Value-Cost Model (Barron & Hulleman, 2015) highlight that students' motivation to complete a task is aided by their confidence in succeeding in it and hindered by emotional and cognitive stressors such as unexpected failure. Thus, the expected success rate, influenced by *e.g.* the difficulty of the learning material, may also influence error-induced quitting behavior. Another contextual factor is the setting in which the student practices the learning material. For example, students playing at home may rely more on intrinsic motivation and experience less influence from social factors, such as fear of negative evaluation after repeated errors which may occur in a classroom setting. Finally, patterns of quitting may shift over time, as repeated exposure to the learning material and its feedback mechanisms could build tolerance to it.

Recognizing these influences, we not only examine variability in error-induced quitting on an individual level, but also explore how these contextual and temporal factors shape the relationship between sequential errors and quitting.

1.2 Research Aims

The reviewed body of work suggests that understanding quitting requires modeling how repeated errors influence engagement over time, and how this process varies across learners and contexts. The present study aims to model the effect of sequential errors on the likelihood of quitting online learning, and extend this relationship by accounting for individual variability.

We do this by analyzing learning data from Prowise Learn, a large-scale adaptive learning system for Dutch primary school students to practice language and mathematics (Brinkhuis et al., 2018; Klinkenberg et al., 2011; Vermeiren et al., 2025). Computerized adaptive learning platforms are increasingly being used in all levels of education (NCES, 2021; OECD, 2025), and offer promising tools to tailor instruction and track the engagement of learners in real time (Roschelle et al., 2020; Toomla et al., 2025). Crucially, the scale of the data provided by this online learning platform offers a unique opportunity to study heterogeneity across a wide sample of learners interacting with educational material in a real-world context. Obtaining a representative sample is a crucial condition to properly assess deviations from mean-level effects (Bryan et al., 2021), making our data ideal for this goal. Also, the scale of the data enables us to take an exploratory approach and validate those explorations in an independent dataset. As such, we make no specific predictions on the size or direction of effects, but adopt a training/testing approach, elaborated in Section 3.

Our analyses proceed in three steps:

1. We assess whether the effect of sequential errors on quitting is moderated by age, difficulty,

error speed, and context (in vs. out of school). Following from the reviewed literature, we expect: that younger learners to quit more following errors; when playing on a more difficult level increases quitting following errors; playing in school, compared to at home, increases the risk of quitting following errors; and making fast, compared to slow, errors is associated with a higher probability of quitting.

2. We investigate how quitting patterns change over time, and expect that, with repeated exposure to the learning environment and errors therein, learners are less likely to quit after errors.
3. Using mixed-effects modeling, we estimate individual-level effects of sequential errors on quitting. We also assess their stability across two domains (addition and subtraction). That is, what is the association between individual learners' post-error quit probabilities in addition and subtraction? We expect stable individual differences, where individual learners range from low to high likelihoods of quitting after making repeated errors.

By combining large-scale behavioral data with a focus on within- and between-subject variability, this study aims to illuminate how students differ in their motivational responses to failure—a key consideration for both adaptive system design and theories of individual learning differences.

2 Methods

2.1 Online Learning Environment

We use online learning data from the application Prowise Learn (Klinkenberg et al., 2011). In this OLE, children can choose to enter the Math Garden (for mathematics practice), Language Sea (for language practice), or Words and Birds (for English language learning). Figure 1 shows a screenshot of the addition game in the math garden environment. The program utilizes computer-adaptive practice (CAP), which means that it selects the appropriate difficulty level of the item given each player's estimated ability level. The algorithm is a combination of the Elo algorithm (for more information, see Vermeiren et al. (2025)) and the explicit scoring rule proposed by Maris and Van der Maas (2012). Importantly, the games are time-limited and ratings increase when a student performs above expectation (typically accurate and fast), and decrease if the student performs below expectation (typically incorrect and/or slow). The use of CAP allows us to interpret the findings in the context of students practicing within their optimal learning level.

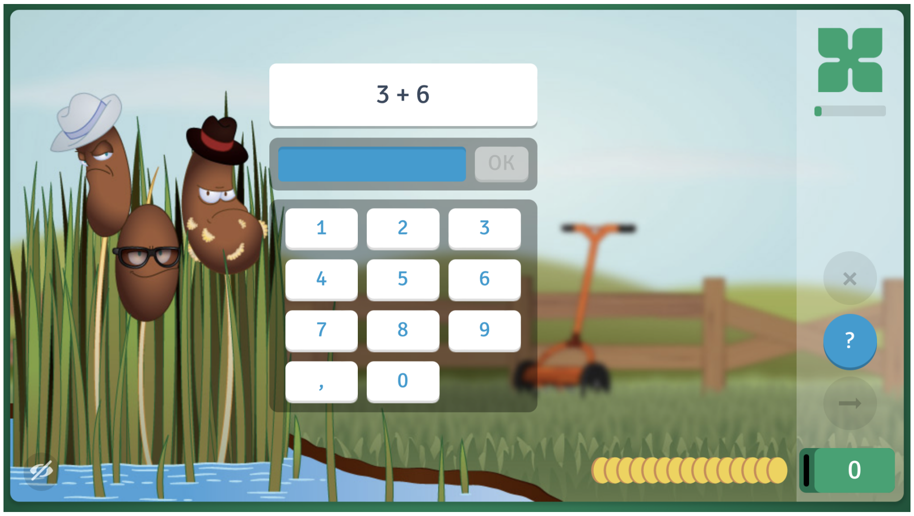

Figure 1: A screenshot of the addition game in Prowise Learn.

Each game that the user can choose to start consists of 10 items. The time given to respond to an item is shown on the screen in the form of coins, which decrease by one coin per second, until the time is up or a response is submitted. If the item is answered correctly, the user receives the number of coins that are left. If the user gives an incorrect response, a fixed number of coins is subtracted from their total. The next item is selected according to the expected probability that the user will answer it correctly, which is based on the current item difficulty and user ability rating. This probability can be manipulated by the difficulty level that the user plays on, and can be 0.90 (easy), 0.75 (medium), or 0.60 (difficult). The difficulty level is set at medium as a default unless the player chooses otherwise. Apart from difficulty selection, the player can manipulate whether they can see the disappearing coins at each item, and they can retrieve the answer for the item by clicking a question mark. In the latter case, the player receives a null score, and the item is skipped. Lastly, the player can continue to the next item or quit the game by clicking an exit button between the presentation of items. If no choice is made, the game moves automatically to the next item.

2.2 Data

2.2.1 Ethics

This work was approved by the Ethics Review Board of the Psychological Methods department at the University of Amsterdam (reference number 2022-PML-15260), and is in accordance with the

Declaration of Helsinki. Schools or families signing up for Prowise Learn provide informed consent for their data's scientific use. Agreements between Prowise Learn and individual schools ensure that parents are informed about data usage and that participation is voluntary. Data from children without parental consent were not included, and all data were anonymized before researchers accessed them.

2.2.2 Inclusion criteria

Users who show non-deliberate gameplay were excluded from all datasets, according to the following criteria: (1) sessions where the game was started but exited immediately; (2) sessions with sequences of only incorrect responses; or (3) sessions with three fast incorrect responses in a row (as detected by the system, resulting in an automatic ending of a game). We also excluded users in grades 1 and 2, as most games are not suited for children in this age; thus, they provide too little data.

2.2.3 Statistical Analyses

All data analyses were conducted using R Statistical Software (R Core Team, 2024, version 4.3.2). Fitting procedures for the Markov Models were carried out using the `msm` package (Jackson, 2011, version 1.7). These models were chosen as they allow for longitudinal modeling of transitions into multiple states, which fits into the nature of the learning data at hand. The significance of these results was assessed by deriving confidence intervals using bootstrap methods provided in the `msm` package. Confidence intervals for the transition intensity matrix and its covariate values were computed from $10\,000$ bootstrap samples, assuming a symmetric distribution on the log scale. The 2-State Markov Model was fitted on $105\,864$ observations in the training dataset, and $105\,733$ observations in the testing dataset.

Mixed-effects logistic regressions were fitted using the `glmer` function from the `lme4` package (Bates et al., 2015, version 1.1.35.1). This method is suited for the analysis of variability in the effect of sequential errors on quitting, as it allows us to adhere to a nested structure in the data. In this way, each individual is treated as a separate level in the analysis and thus their individual effect of sequential errors on quitting can be estimated. We fit this model on 3998 users in the training data, and 4091 users in the testing data. All tests were two-tailed with an alpha level of .05.

2.3 Measures

Quitting A soft-quit is defined as when the learner exits a game before it's intended end (*i.e.,* before completing 10 items), within the current learning session. A learning session ends if the learner exits the learning environment and is inactive for at least 30 minutes.

Sequential Errors This variable denotes how many errors a student has made in a row, directly before the current item. When the user makes a correct response, the variable is reset to 0. For the 2-state Markov Models, the variable is discretized and can take on the values 0, 1, 2, 3, or >3. For the mixed-effect models, the variable is continuous and can take on any value between 0 (no error made) and 10, which is the maximum amount of sequential errors possible before a game is complete.

Difficulty setting We measure whether the student chooses to play on the easy, medium, or difficult setting.

Speed of responding We include a binary response time variable, where a response is considered fast if the response time is faster than the median response time, and slow if it is slower than the median response time.

Playing inside or outside school hours Additionally, we measured whether students play during school hours or not, as a proxy for the context in which they were practicing. School hours were considered to fall between 07:00 and 15:00 on weekdays. This is a binary variable and takes on the values 0 (in school) or 1 (outside school).

User ability To estimate user ability, we extracted the last user rating estimate (determined by the Elo Rating System and Explicit Scoring Rule, described above) for each user within the relevant domain and data collection period.

3 Results

In this section we first show that, in general, making sequential errors substantially increases the risk of quitting from learning. We then demonstrate wide individual differences in this effect; while some students have a very high risk of quitting after errors, some are resilient. We showcase these results within two popular arithmetic games in Prowise Learn: addition and subtraction. Prior to narrowing down the analyses to these domains, we examined the dynamics of quitting across all games on the platform. In addition to estimating transition probabilities between persisting and soft-quit states, this allowed us to include a hard-quit state (*i.e.,* the student leaves the platform entirely), as has been done previously by ten Broeke et al. (2022), and served as a baseline measurement of quitting behavior across all domains in the current data. Analysis procedures and all estimated transition rates can be found in the Supplementary Material, Section 1.

For the subsequent analyses, we used log data from users playing the addition and subtraction games within the Math Garden environment in Prowise Learn. Both games consist of arithmetic items suited for students with a wide range of addition and subtraction ability. We chose data from the addition game spanning a three-year period between 2020-09-01 and 2023-07-31, resulting in more than 25 million responses from $255\,568$ unique learners. We followed the same procedures for data from the subtraction game (n = $196\,895$). Here, we report results from the addition game; findings from the subtraction game were comparable and are reported in the Supplementary Material, Section 2.

Furthermore, given the size of our data, we took a data-driven approach, exploring model parameters in a training dataset (n = $105\,864$ in the addition game), and later validating our models on an untouched testing dataset (n = $105\,733$). Grade, gender, ability, and choice of difficulty level were equally distributed across both datasets (Supplementary Materials Section 3.2). Here, we report results from the training dataset; results from the testing data set are reported in the Supplementary Material, Sections 2 and 3.

3.1 Sequential Errors Increase the Risk of Quitting from Learning

Across training and testing datasets, about 31% of sessions ended prematurely. To estimate the general probability of switching between a persisting to a quitting state, three 2-state Simple Markov Models were defined. First, a model estimating baseline transition rates without the influence of any covariates. Second, we model the likelihood of a state transition, while controlling for the following covariates: gender (male vs. female), age (3–4 vs. 7–8 and 5–6 vs. 7–8 years old), difficulty level (easy vs. medium, hard vs. medium), playing outside vs. inside school hours, and sequential errors (1, 2, 3, and $>$3, each level compared to making no error). Lastly, we fit a model including an interaction term between sequential errors and all other covariates. That is, we estimated how the effect of quitting following sequential errors differs across grade, difficulty level, response time, and playing during or outside school hours. These three models are hereafter referred to as the constrained, covariate, and interaction models, respectively.

The interaction model provided the best fit to the data, according to AIC ($AIC_{baseline} = 2835833$; $AIC_{covariate} = 2367165$; $AIC_{interaction} = 2359080$) and likelihood ratio test ($\chi^2(35) = 8133, p < .001.$) measures. Results of this model are displayed in Figure 2. In this model, users transitioned between a persisting to quitting state at an instantaneous rate of 0.018. Hazard ratios demonstrated that 1, 2, 3, and $>$3 sequential errors had a considerable effect on the likelihood of transitioning from a persisting to quitting state: learners were 6.45 to 14.62 times more likely to quit when encountering one or more errors in a row while practicing addition (Figure 2, left).

Main effects of difficulty level, speed of errors, and playing inside or outside school hours were small, but significant (Figure 2, left). When compared to the default medium difficulty level, learners were slightly more likely to quit when playing on an easy (5% increased risk) and hard (8% increased risk) difficulties. Similarly, fast responding slightly reduces (around 5%) the quitting risk, compared to slow responding. Younger students have a progressively higher risk of quitting compared to older age-groups (there is 15% increased risk of quitting for 5–6 year old children, and 44% for 3–4 year old children, compared to 7–8 year old children). Lastly, practicing during school hours is associated with a 38% decrease in the risk of quitting, compared to practicing outside school hours.

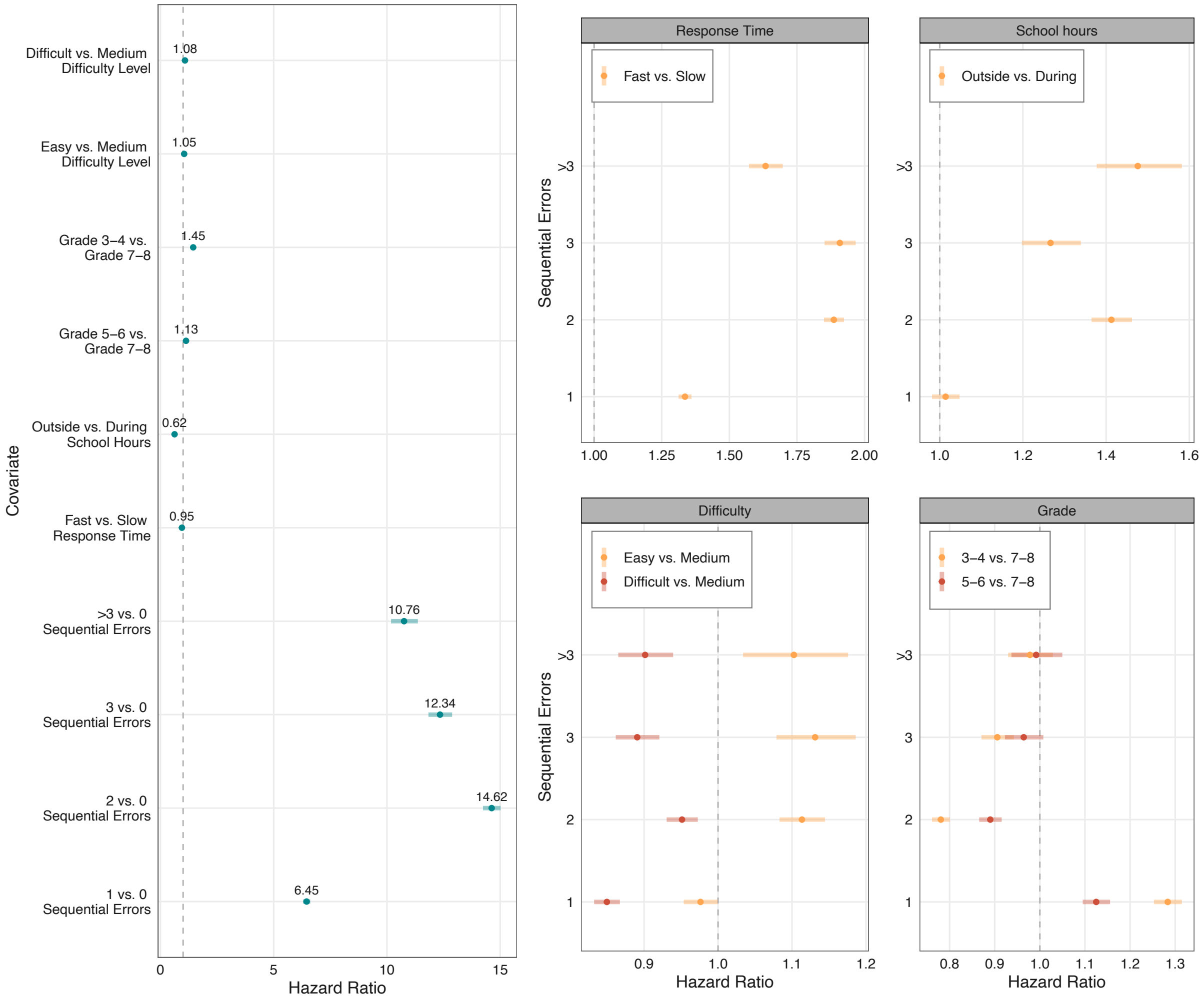

Figure 2: Hazard ratios for the main (left) and interaction (right) effects derived from the 2-state markov model on the addition game. Main effects are on a log scale. Each hazard ratio for the interaction effects represents the relative increase or decrease in the likelihood of transitioning into a quitting state for each covariate across each level of sequential error. For example, the transitioning into a quitting state when making more than 3 sequential errors is approximately 1.35 times more likely if the response was fast, compared to slow. Lines represent the 95% confidence interval.

3.2 Error-induced Quitting Differs Across Age, Response Time, Difficulty, and Time of Day

All interaction effects were significant (Figure 2, right). The interaction effect between sequential errors and response time revealed that users were less likely to quit after making 1 error with a fast response time, but more likely to quit after 2, 3, and >3 sequential errors when they have a fast, compared to a slow, response time. Further, users playing on the easy difficulty level were more likely to quit following 2, 3, and $>$3 sequential errors, while users playing on the difficult level were less likely to quit following all levels of sequential errors. Effects of grade reveal that younger users were less likely to quit after 1 error, but more likely to quit after 2 sequential errors, compared to older users, with no difference in quitting for 3 or > 3 sequential errors across grades. Lastly, playing during school hours led to less quitting following 2, 3, and > 3 sequential errors, compared to playing outside school hours.

3.3 Error-Induced Quitting Reduces over Time

Within Prowise Learn, there is a large variance in the amount of sessions that players have played. This also means that some users encounter errors more often than others. To explore whether the amount of sessions a user has played affects their tendency to quit, we looked at the average quitting rate across sessions. Specifically, we examined how the probability of quitting after an incorrect versus a correct response changes as a function of the number of sessions a user has played. We separated these analyses by difficulty level and prior experience with the platform. Prior experience was necessary to control for, as some learners played their first ever session within the collected data (labeled new user) while some already had experience with the platform before data collection (labeled existing user). This exploratory analysis revealed two important trends in our data: (1) post-error quitting decreases with more playing experience, while post-correct quitting does not, (2) post-error quitting is more likely while playing easy and difficult levels, but not post-correct quitting. These longitudinal effects are displayed in Figure 3.

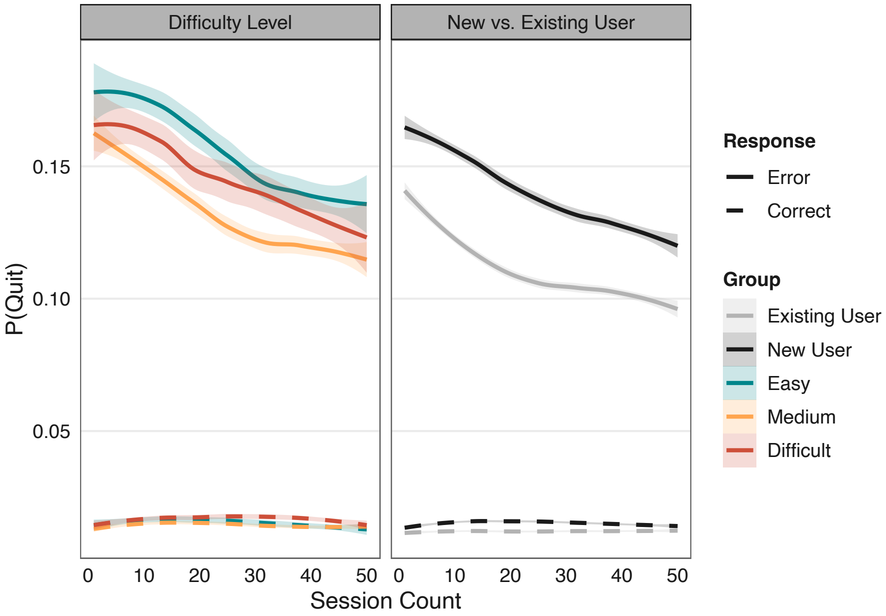

Figure 3: Post-error vs. post-correct quit probabilities, separated across difficulty levels and new vs. existing users. A user was classified as a new user if their first experience in the addition game occurred during the period of data collection, and existing user if they had played the addition game before the start of data collection. Ribbons around the estimates represent a 95% confidence interval.

3.4 Error-induced Quitting Differs Across Individuals

Finally, to examine the variability in the effect of sequential errors on quitting across users in the OLE, we performed a mixed-effects logistic regression. Importantly, in order to have enough data to estimate individual effects, we excluded users who had played less than 50 sessions and made less than 10 quits in total. This resulted in a sample of 3998 users. [1]

We fit three separate mixed-effects models. In all models, we used a student identifier as the grouping variable and modeled the fixed effect of sequential errors on quitting. In this case, sequential errors was treated as a continuous variable, ranging from 0 to 10. The first model estimated quitting predicted by sequential errors and included a random intercept term allowing for the baseline quit rates

---

[1] Due to the relatively low base rate of quitting in our data, these stringent selection criteria were necessary for proper convergence of our logistic models. To investigate potential bias due to these selection criteria, we fitted `glmer` models where we only removed extreme cases from our sample (students who either never quit or always quit), as these cases cause the convergence issues. This resulted in a sample size of $77\,873$, from which we randomly sampled 2000 students and performed the same mixed-effect model (including a random intercept and slope). For robustness, we reiterated this process 10 times. The effect size of sequential errors on quitting ranged from 1.14 (SE = 0.021) to 1.18 (SE = 0.022), which is larger than the originally reported 0.84 (SE = 0.007). This ensured that our selection criteria did not bias the results in a positive direction.

(at 0 sequential errors) to vary across users. Second, we modeled quitting predicted by sequential errors, including a random intercept and a random slope, allowing the effect of sequential errors on quitting to vary across individual users. Lastly, we fit a model including a random intercept, a random slope, and fixed effects of user ability and grade, thus allowing these two variables to be controlled for. [2]

The third model provided the best fit to the data (Table 1). There was a significant fixed effect of sequential errors ($\beta = 0.81$; SE = 0.001; $p < .001$) on quitting. This is equivalent to an odds ratio (OR) of about 2.25, meaning that for every sequential error committed, the risk of quitting increases by about 2.25. There was a significant effect of user ability ($\beta = -0.10; SE = 0.01; p < .001; OR = 0.90$) and grade ($\beta = -0.05; SE = 0.004; p < .001; OR = 0.95$) , meaning that the risk of quitting is higher for users with lower ability ratings and in lower grades. Correlation estimates between fixed effects can be found in Supplementary Table S10.

Adding random person variance to the fixed effect of sequential errors on quitting resulted in a substantially better model fit (Table 1). The variance estimate for the intercept was 0.39 ($\sigma = 0.62$) and for the effect of sequential errors was 0.14 ($\sigma = 0.38$), signaling a large variance around the fixed effect of sequential errors on quitting. This can also be demonstrated by the variance in odds ratios: a student who lies 1 standard deviation below the average effect is 1.53 times more likely to quit per sequential error, whereas a student who lies 1 standard deviation above the average effect is 3.28 times more likely to quit per sequential error that they have committed. There was a weak negative correlation between random intercept and slope estimates ($r = -.10$), implying that users who quit less after a correct response are slightly more likely to quit when faced with errors. All `glmer` model estimates are reported in Table 2.

Results on Test Data The testing dataset consisted of 4091 users. Despite random sampling to training and testing datasets, and identical data selection across both datasets, model fitting procedures on the testing data had poorer convergence rates compared to the training data. Nevertheless, BIC values and a likelihood ratio test provided evidence that the model including fixed effects of

---

[2] In light of the previous results that tendency to quit differs across difficulty levels, we aimed to fit a model also including difficulty level as a covariate, but due to our stringent data selection criteria, each difficulty group was too small for the model to fit adequately.

Table 1: GLMER Model Fit Indices

		Training Data			Testing Data		
Model	Params (N)	AIC	BIC	Log-Likelihood	AIC	BIC	Log-Likelihood
1.	3	1047301	1047340	-523647	1041610	1041649	-520802
2.	5	1032731	1032796	-516360	1027581	1027646	-513785
3.	7	1013950	1014041	-506968	1027433	1027525	-513710

Note. All models include a fixed intercept, and fixed effect of sequential errors on probability of quitting. Random effects are modeled across individual users. Models 1, 2, and 3 are ordered by increasing complexity: Model 1 = Model with random intercept; Model 2 = Model with random intercept and random slope; Model 3 = Model with random intercept, random slopes, and covariates grade and user ability ratings.

sequential errors, user ability, and grade, a random intercept and random effect of sequential errors on quit probability fit the data best, similar to the training data (Table 1). BIC values were similar across datasets. Full model comparison results and estimated model parameters in the testing data are reported in Table 2, as well as Supplementary Tables S8–S11.

Table 2: GLMER Effect Estimates

	Training Data						Testing Data		
	Addition			Subtraction			Addition		
Fixed effects	β	SE	*p*	β	SE	*p*	β	SE	*p*
(Intercept)	-3.48	0.023	$<$.001	-3.59	0.025	$<$.001	-3.57	0.024	$<$.001
Sequential errors	0.81	0.007	$<$.001	0.74	0.007	$<$.001	0.81	0.006	$<$.001
Ability rating	-0.10	0.011	$<$.001	-0.07	0.011	$<$.001	-0.09	0.011	$<$.001
Grade	-0.05	0.005	$<$.001	-0.07	0.004	$<$.001	-0.03	0.004	$<$.001
Random effects	σ^2	SD	*cor*	σ^2	SD	*cor*	σ^2	SD	*cor*
(User) intercept	0.39	0.62		0.39	0.62		0.41	0.64	
(User) sequential errors	0.14	0.38	-0.10	0.12	0.34	-0.09	0.13	0.36	-0.10

Note. cor denotes the correlation between the random intercept and random effect of sequential errors. SE = Standard Error; SD = Standard Deviation.

3.5 Individual Differences are Stable Across Two Arithmetic Domains

The aforementioned findings reveal individual differences in the extent to which quitting rates of users playing the addition game within Prowise Learn are affected by sequential errors. In this final step of the analysis, we sought to examine the robustness of these findings by extracting individual effects of sequential errors on quitting from the same users, but in a different domain, namely, subtraction.

Here, we included users who played at least 50 sessions and made a minimum of 10 quits in both the addition and subtraction game within the given data collection period; $n = 1765$.

Similar to the addition game, there was a significant main effect of sequential errors ($\beta = 0.74$; SE = 0.007; $p < .001$), user ability ($\beta = -0.07$; SE = 0.011; $p < .001$), and grade ($\beta = -0.07$; SE = 0.004; $p < .001$; Table 2). The random variance estimate for the intercept was 0.39 ($\sigma = 0.62$) and for the effect of sequential errors was 0.12 ($\sigma = 0.34$; Table 2). Importantly, to estimate the stability of individual differences across the two domains, we computed correlations between the random intercepts and random slopes extracted from users' data for both the subtraction and addition games. Figure 4 (left) shows individual effect estimates in both domains from 300 randomly sampled users. There was a strong correlation between users' individual effects of sequential errors on quitting (random slopes) in the addition and the subtraction game ($r = .64$; $p < .001$). Likewise, there was a strong correlation between users' baseline quitting rates in both domains ($r = .80$; $p < .001$; Figure 4, right).

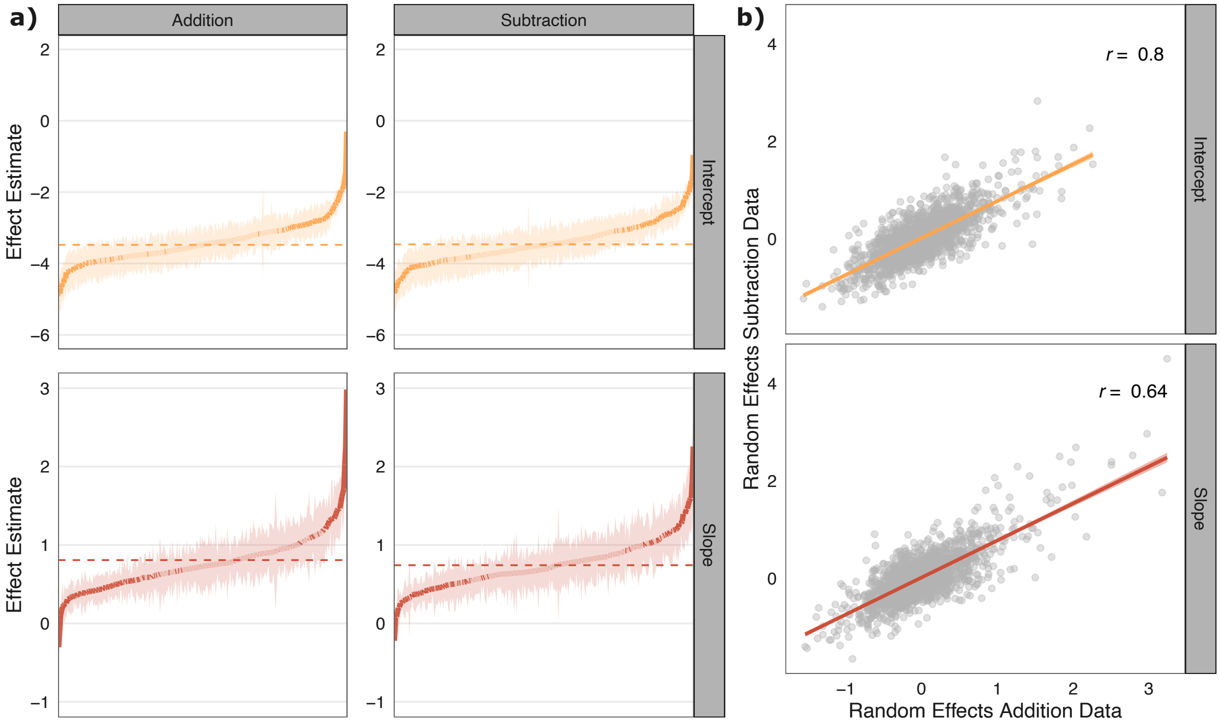

Figure 4: Left: Baseline quitting rates (intercept) and effects of sequential errors on quitting (slope) for 300 randomly sampled users, across both the addition and subtraction domain. Datapoints are ordered from lowest to highest effect estimate in both graphs. Horizontal lines denote the fixed effect. Vertical lines represent the 95% confidence interval of each user's effect, estimated from repeated resampling (n = 200) from the posterior distribution of the random effects. The *REsim* function from the package merTools (v. 0.6.2; Knowles et al. (2016)) in R was used to achieve this. The points of users whose effect estimate is not significantly different from the average main effect has a more transparent color. Right: Scatterplots representing the correlation between random effects in the addition and subtraction game. The shaded region represents the 95% confidence interval.

4 Discussion

Disengagement is detrimental for learning. While a central goal of all learning environments is to match the complexity of the learning material to the proficiency of the learner, there are substantial individual differences in how learners deal with the frustrations of failure. The increasing popularity of online adaptive learning software, coupled with a growing awareness that one size does not fit all in learning, demands research that pinpoints what elements make OLEs effective, and for whom that is the case. In this study, we have combined both urgencies by investigating predictors of quitting, and the variability thereof, in a large-scale OLE. Results support previous findings that sequential errors predict quitting and add to these findings that errors do not affect quitting equally across students.

Our results demonstrate that, when faced with repeated mistakes, students are highly likely to quit the learning session. We have demonstrated this effect using data from diffrent time periods, learning domains, and statistical methods. Moreover, the effect is in line with previous research highlighting repeated mistakes as a factor disrupting flow (Hou, 2015) and increasing the likelihood of disengagement in learning (Arztmann et al., 2025; Lomas et al., 2013; ten Broeke et al., 2022; Vu et al., 2022). These findings emphasize errors as a key trigger in momentary quitting behavior.

Crucially, we extend from this effect and find considerable inter-individual variability: students differ both in their general likelihood of quitting a learning session, and in their tendency to quit following errors. These findings suggest that students vary in how they adaptively regulate behavior in response to failure. From the perspective of motivational resilience (Skinner et al., 2020), such variability may reflect differences in learners' ability to sustain goal-directed effort, manage emotions, and monitor errors (Dent & Koenka, 2016; Hill et al., 2016; Miller et al., 2012). It is likely that there are students who are better able to tolerate frustration and maintain confidence after errors, displaying adaptive motivational responses (Grassinger & Dresel, 2017; Grassinger et al., 2018; Tulis et al., 2016) and thus higher persistence, whereas those who experience stronger negative affect in the face of failure are quicker to disengage. Likewise, learners who are more likely to perceive errors as learning opportunities as opposed to a reflection of low ability may be better able to persist after errors (Dweck, 1986, 2013; Moser et al., 2011; Tulis et al., 2018). Overall, our findings are consistent with frameworks that view engagement as the outcome of dynamic interactions between cognitive, emotional, and motivational regulation processes and underscore the importance of considering individual differences in adaptive regulation when examining persistence in digital learning environments.

Furthermore, we identified group-level variables which impact the tendency to quit after errors. First, compared to those playing on the default medium level, users on an easy level were more likely to quit, while those on a difficult level were less likely to quit following errors. This pattern aligns with expectancy-value theory (Eccles & Wigfield, 2002) and the expectancy-value-cost model (Barron & Hulleman, 2015). Players on an easier level may have a higher expectancy of success, making errors more surprising and emotionally impactful, increasing the likelihood of quitting. Conversely, players on a difficult level may anticipate errors, reducing their emotional impact and fostering persistence. This explanation is particularly strengthened by the fact that difficulty level, on its own, did not influence quitting. Rather, errors played an especially important role for engagement in the respective difficulty levels. The ability to measure reactions to errors in varying degrees of difficulty presents a unique potential to further explore how students' expectancy beliefs and the structure of their environment impact their persistence.

Second, while the previously discussed research indicates that children's tolerance to errors should increase with age, our results are less straightforward. Younger children (6–7 years) were more likely to quit after a single error compared to older children (8–9 and 10–11 years). This aligns with research indicating that younger children have less developed emotional regulation and metacognitive strategies, making them more reactive to failure (Bjorklund, 2022; Cole et al., 2004; Oeri et al., 2024). Younger students may also be less able to stay engaged on a task in general, due to *e.g.,* lower sustained attention (Betts et al., 2006), which is also reflected in small, but significant, main effects of age on quitting. However, after two errors, this pattern reverses: younger children who persist past the initial failure are less likely to quit than older children. This suggests that these students may be a more resilient subset. In contrast, older children may initially tolerate failure but reach a breaking point after repeated errors. After three or more errors, quitting rates stabilize and are likely better explained by general individual differences. In fact, developmental research points to increasingly complex and individualized motivational profiles as children advance through primary school (Guay et al., 2010; Hornstra et al., 2023; Skinner et al., 1998; Wan et al., 2021)

Third, in line with the intuition that when students engage with learning material in their free time, compared to during school hours, they are more motivated to learn and thus less likely to quit, we see less baseline quitting when students practice mathematics outside school hours. However, the interaction effects show that error-induced quitting is more likely outside school hours, compared to during school time. These effects signal that while the home environment might be generally protective against disengagement, students may feel more prone to frustration following errors, as they are less likely to feel the social evaluation from peers taking place in the classroom. Another account may be a more positive error climate induced by teachers in the classroom (Steuer et al., 2013), or a social contagion effect where motivation is passed on through peer relationships (Burgess et al., 2018). In any respect, these accounts are speculative and demand further research.

Lastly, looking at error-induced quitting over time revealed that the average probability of quitting reduces with more sessions played. Thus, it may be fruitful to intervene on error-induced quitting early in the playing process, and decrease the strength of intervention as users gain more experience. These results suggest that there is an experience-related build-up of tolerance to errors, which should be studied in more detail in the future.

4.1 Limitations

There are several limitations and potentials for future research. First, the addition and subtraction domains are very close in relation to each other. Recent research has shown that motivation likely differs across broader learning domains, such as between mathematics and language (Gaspard et al., 2018). Thus, it is possible that the stability of individual differences in error-induced quitting does not generalize across broader learning domains. Investigating how stable individual error-induced quitting effects are over domains with varying degrees of difference would have large implications for whether engagement in online learning is domain- or user-specific.

Next, our inferences are arguably limited to a population of children that are naturally more persistent. This in an inherent issue when researching quitting behavior; children who begin interacting in the system but quit very early are left unstudied due to their lack of data. With regard to disengagement and drop-out, the behavioral patterns of these students, and whether they differ from students who stay in the system for a longer time, may be particularly interesting to investigate. Therefore, future research should find ways to model this quitting behavior, for example, by defining quitting on a longer-term basis, as opposed to a session level. However, for the purpose of the current research, defining quitting on a session level is beneficial as it elucidates finer-grained behavioral dynamics.

Relatedly, this study has focused on the effects of errors on exiting a session, but has not studied whether it affects the time taken before starting a new session. Such findings would further elucidate the dynamics of student quitting behavior.

While we unveil quitting as an important trigger for quitting, it is important to consider whether this form of momentary quitting has consequences on learning: is quitting after errors necessarily maladptive, and should it always be counteracted? Our findings suggest that, albeit small, there is a subgroup of students who are *more* prone to persisting in the face of errors. While this may be considered very perseverant, research on wheel-spinning suggests that perseverance is not always productive for learning. Rather, when a student wheel-spins, they show a form of unproductive persistence: spending too much effort in persisting with a task despite a lack of progress (Beck & Gong, 2013). Likewise, some of these students may already be cognitively disengaged before quitting—employing strategies to game the system (Baker et al., 2004; Baker et al., 2010; Nagy et al., 2023). While the goal of the current work was not to delineate students who are productively and unproductively quitting after errors, a possible next step in this research is to relate error-induced quitting behaviors to learning gains. This way, it could be possible to differentiate between different types of productive and unproductive persistence. This would also open research questions about whether there is a "sweet-spot" in the error rate for tasks, which promotes productive persistence while simultaneously leading to optimal learning gains (Wilson et al., 2019).

4.2 Implications

As demonstrated in the current work, errors can serve as motivational triggers that produce momentary lapses in engagement. This sensitivity to errors implies that the design of errors in learning systems is not neutral: repeated failure systematically increases disengagement risk. Thus, errors should be understood not only as signals of cognitive ability, but also as patterns of failure which shape learners' motivation to persist. Ignoring this dual role risks overlooking a key mechanism through which difficulty translates into disengagement. Crucially, our findings do not challenge the pedagogical value of errors for learning, but highlight that their motivational consequences must be managed in instructional design. This can be done, for example, by regulating task difficulty, varying feedback after errors, or providing adequate room for recovery following unsuccessful attempts.

On top of that, the fact that learners respond differently to repeated errors has significant implications for both adaptive learning systems and educational design more broadly. It is crucial to recognize that students do not respond uniformly to errors; instead, if the goal is to optimize motivation and retain students in the learning process, their individual tolerances for error must be considered.

Adaptive learning systems can achieve this by dynamically adjusting the difficulty level of future items based on individual students' quitting patterns. Similarly, because post-error quitting patterns are strongly correlated between addition and subtraction games, learning environments can use individual patterns of behavior in one domain to design individualized interventions in the other.

If individual differences in error-induced quitting are ignored, learning environments risk confounding ability with motivation and may put students who are better able to tolerate errors at an unfair advantage. From the perspectives of online learning environments, students who are more able to cope with a higher error rate are more likely to be retained in the system for longer. This leads to more precise ability estimates, and give the student more opportunities to learn at their optimal level. Similarly, in classroom contexts, such dynamics may contribute to self-reinforcing motivation–achievement cycles: students who cope effectively with errors are more likely to persist, leading to greater learning gains and enhanced self-efficacy, whereas those who are more sensitive to failure may disengage earlier, constraining both achievement and motivation (Vu et al., 2022). Over time, these processes can amplify initial individual differences, creating cumulative disparities in performance and engagement. Consequently, it is essential that both adaptive learning systems and educational practices explicitly account for individual variability in error tolerance. Although a complete disentanglement of motivation and performance is likely not realistic, doing so can help to reduce the conflation between ability and perseverance and ensure that learners are evaluated and supported on a more equitable basis.

4.3 Conclusion

Effective learning relies on engaged learners. Behavioral patterns in giving up from learning, and the role of errors therein, provide a novel way to study children's motivational variability in an increasingly digital world. This study elucidates individual differences in perseverance and paves the way for future research on interventions to keep children motivated to learn.

5 Data Availability

The data that support the findings of this study are available from Prowise Learn but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are, however, available from the authors upon reasonable request and with the permission of Prowise Learn.

References

Arztmann, M., Domínguez Alfaro, J. L., Hornstra, L., Wong, J., Jeuring, J., & Kester, L. (2025). Game over? Investigating students' working memory, situational interest, and behavioral patterns as predictors of dropout in an educational game. *International Journal of Child-Computer Interaction*, *43*, 100721. https://doi.org/10.1016/j.ijcci.2025.100721

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-Task behavior in the cognitive tutor classroom: When students" game the system". *Proceedings of the SIGCHI conference on Human factors in computing systems*, 383–390.

Baker, R. S., D'Mello, S. K., Rodrigo, M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, *68*(4), 223–241. https://doi.org/10.1016/j.ijhcs.2009.12.003

Barron, K. E., & Hulleman, C. S. (2015). Expectancy-Value-Cost model of motivation. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 503–509). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.26099

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beck, J. E., & Gong, Y. (2013). Wheel-spinning: Students who fail to master a skill. *International conference on artificial intelligence in education*, 431–440. https://doi.org/10.1007/978-3-642-39112-5_44

Betts, J., Mckay, J., Maruff, P., & Anderson, V. (2006). The development of sustained attention in children: The effect of age and task load. *Child Neuropsychology*, *12*(3), 205–221. https://doi.org/10.1080/09297040500488522

Bjorklund, D. F. (2022). *Children's thinking: Cognitive development and individual differences*. Sage publications.

Borrella, I., Caballero-Caballero, S., & Ponce-Cueto, E. (2022). Taking action to reduce dropout in MOOCs: Tested interventions. *Computers & Education*, *179*, 104412. https://doi.org/10.1016/j.compedu.2021.104412

Botelho, A. F., Varatharaj, A., Inwegen, E. G. V., & Heffernan, N. T. (2019). Refusing to try: Characterizing early stopout on student assignments. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 391–400. https://doi.org/10.1145/3303772.3303806

Botelho, A. F., Varatharaj, A., Patikorn, T., Doherty, D., Adjei, S. A., & Beck, J. E. (2019). Developing early detectors of student attrition and wheel spinning using deep learning. *IEEE Transactions on Learning Technologies*, *12*(2), 158–170. https://doi.org/10.1109/TLT.2019.2912162

Brinkhuis, M. J. S., Savi, A. O., Hofman, A. D., Coomans, F., Van Der Maas, H. L. J., & Maris, G. (2018). Learning as it happens: A decade of analyzing and shaping a large-scale online learning system. *Journal of Learning Analytics*, *5*(2), 29–46. https://doi.org/10.18608/jla.2018.52.3

Brophy, J. E. (2013). *Motivating students to learn.* Routledge.

Brunello, G., & Paola, M. d. (2014). The costs of early school leaving in europe. *IZA Journal of Labor Policy*, *3*(1), 22. https://doi.org/10.1186/2193-9004-3-22

Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behavior*, *5*(8), 980–989. https://doi.org/10.1038/s41562-021-01143-3

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

Cole, P. M., Martin, S. E., & Dennis, T. A. (2004). Emotion regulation as a scientific construct: Methodological challenges and directions for child development research. *Child Development*, *75*(2), 317–333. https://doi.org/10.1111/j.1467-8624.2004.00673.x

Csikszentmihalyi, M. (2000). *Beyond boredom and anxiety.* Jossey-Bass.

de Mooij, S. M., Dumontheil, I., Kirkham, N. Z., Raijmakers, M. E., & Van Der Maas, H. L. (2022). Post-error slowing: Large scale study in an online learning environment for practising mathematics and language. *Developmental Science*, *25*(2), e13174. https://doi.org/10.1111/desc.13174

Dent, A. L., & Koenka, A. C. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis. *Educational Psychology Review*, *28*(3), 425–474. https://doi.org/10.1007/s10648-015-9320-8

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, *41*(10), 1040. https://doi.org/10.1037/0003-066X.41.10.1040

Dweck, C. S. (2013). *Self-theories: Their role in motivation, personality, and development*. Psychology press.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*(1), 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of educational research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Gaspard, H., Wigfield, A., Jiang, Y., Nagengast, B., Trautwein, U., & Marsh, H. W. (2018). Dimensional comparisons: How academic track students' achievements are related to their expectancy and value beliefs across multiple domains [10.1016/j.cedpsych.2017.10.003]. *Contemporary Educational Psychology*, *52*, 1–14. https://doi.org/10.1016/j.cedpsych.2017.10.003

Grassinger, R., & Dresel, M. (2017). Who learns from errors on a class test? Antecedents and profiles of adaptive reactions to errors in a failure situation. *Learning and Individual Differences*, *53*, 61–68. https://doi.org/10.1016/j.lindif.2016.11.009

Grassinger, R., Scheunpflug, A., Zeinz, H., & Dresel, M. (2018). Smart is who makes lots of errors? The relevance of adaptive reactions to errors and a positive error climate for academic achievement. *High Ability Studies*, *29*(1), 37–49. https://doi.org/10.1080/13598139.2018.1459294

Guay, F., Chanal, J., Ratelle, C. F., Marsh, H. W., Larose, S., & Boivin, M. (2010). Intrinsic, identified, and controlled types of motivation for school subjects in young elementary school children. *British Journal of Educational Psychology*, *80*(4), 711–735. https://doi.org/10.1348/000709910X499084

Hill, K. E., Samuel, D. B., & Foti, D. (2016). Contextualizing individual differences in error monitoring: Links with impulsivity, negative affect, and conscientiousness. *Psychophysiology*, *53*(8), 1143–1153. https://doi.org/10.1111/psyp.12671

Hornstra, L., Mathijssen, A. S., Denissen, J. J., & Bakx, A. (2023). Academic motivation of intellectually gifted students and their classmates in regular primary school classes: A multidimensional, longitudinal, person- and variable-centered approach. *Learning and Individual Differences*, *107*, 102345. https://doi.org/10.1016/j.lindif.2023.102345

Hou, H.-T. (2015). Integrating cluster and sequential analysis to explore learners' flow and behavioral patterns in a simulation game with situated-learning context for science courses: A video-based process exploration. *Computers in Human Behavior*, *48*, 424–435. https://doi.org/10.1016/j.chb.2015.02.010

Israel-Fishelson, R., & Hershkovitz, A. (2020). Persistence in a game-based learning environment: The case of elementary school students learning computational thinking. *Journal of Educational Computing Research*, *58*(5), 891–918. https://doi.org/10.1177/0735633119887187

Jackson, C. H. (2011). Multi-State models for panel data: The msm package for R. *Journal of Statistical Software*, *38*(8), 1–28. https://doi.org/10.18637/jss.v038.i08

Karumbaiah, S., Baker, R. S., & Shute, V. (2018). Predicting quitting in students playing a learning game. *International Educational Data Mining Society*. https://eric.ed.gov/?id=ED593115

Klein-Latucha, O., & Hershkovitz, A. (2024). When leaving is persisting: Studying patterns of persistence in an online game-based learning environment for mathematics. *Journal of Learning Analytics*, *11*(2), 42–51. https://doi.org/10.18608/jla.2023.8219

Klinkenberg, S., Straatemeier, M., & van der Maas, H. L. (2011). Computer adaptive practice of maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education*, *57*(2), 1813–1824. https://doi.org/10.1016/j.compedu.2011.02.003

Knowles, J. E., Frederick, C., Whitworth, A., & Knowles, M. J. E. (2016). *Package 'mertools'* [R package version 0.5]. https://cran.r-project.org/web/packages/merTools/merTools.pdf

Kristensen, J. K., Torkildsen, J. V. K., & Andersson, B. (2024). Repeated mistakes in app-based language learning: Persistence and relation to learning gains. *Computers & Education*, *210*, 104966. https://doi.org/10.1016/j.compedu.2023.104966

Leonard, J. A., Cordrey, S. R., Liu, H. Z., & Mackey, A. P. (2023). Young children calibrate effort based on the trajectory of their performance. *Developmental Psychology*, *59*(3), 609. https://doi.org/10.1037/dev0001467

Liu, X., Slater, S., Swanson, L., & Baker, R. (2024). Identifying when and why students choose to quit jobs in a science exploration game. *International Conference on Educational Data Mining (EDM)*.

Lomas, D., Patel, K., Forlizzi, J. L., & Koedinger, K. R. (2013). Optimizing challenge in an educational game using large-scale design experiments. *Proceedings of the SIGCHI conference on human factors in computing systems*, 89–98. https://doi.org/10.1145/2470654.2470668

López-Pernas, S., & Saqr, M. (2024). How the dynamics of engagement explain the momentum of achievement and the inertia of disengagement: A complex systems theory approach. *Computers in Human Behavior*, *153*, 108126. https://doi.org/10.1016/j.chb.2023.108126

Maris, G., & Van der Maas, H. (2012). Speed-accuracy response models: Scoring rules based on response time and accuracy. *Psychometrika*, *77*(4), 615–633. https://doi.org/10.1007/s11336-012-9288-y

Martins, J., Cunha, J., Lopes, S., Moreira, T., & Rosário, P. (2022). School engagement in elementary school: A systematic review of 35 years of research. *Educational psychology review*, *34*(2), 793–849. https://doi.org/10.1007/s10648-021-09642-5

Metcalfe, J. (2017). Learning from errors. *Annual Review of Psychology*, *68*(1), 465–489. https://doi.org/10.1146/annurev-psych-010416-044022

Miller, A. E., Watson, J. M., & Strayer, D. L. (2012). Individual differences in working memory capacity predict action monitoring and the error-related negativity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(3), 757. https://doi.org/10.1037/a0026595

Moser, J. S., Schroder, H. S., Heeter, C., Moran, T. P., & Lee, Y.-H. (2011). Mind your errors: Evidence for a neural mechanism linking growth mind-set to adaptive posterror adjustments. *Psychological Science*, *22*(12), 1484–1489. https://doi.org/10.1177/0956797611419520

Nagy, G., Ulitzsch, E., & Lindner, M. A. (2023). The role of rapid guessing and test-taking persistence in modelling test-taking engagement. *Journal of Computer Assisted Learning*, *39*(3), 751–766. https://doi.org/10.1111/jcal.12719

NCES. (2021). Condition of education [Accessed: 2025-07-17]. *U.S. Department of Education, Institute of Education Sciences*. https://nces.ed.gov/programs/coe/

OECD. (2025). Trends shaping education 2025 [Accessed: 2025-07-17]. *OECD Publishing*. https://doi.org/10.1787/ee6587fd-en

Oeri, N., Kunz, N. T., & Kälin, S. (2024). Task persistence through a dynamic lens: Understanding temporal-behavioral dynamics in kindergarten children. *Journal of Applied Developmental Psychology*, *92*, 101642. https://doi.org/10.1016/j.appdev.2024.101642

O'Rourke, E., Haimovitz, K., Ballweber, C., Dweck, C., & Popović, Z. (2014). Brain points: A growth mindset incentive structure boosts persistence in an educational game. *Proceedings of the SIGCHI conference on human factors in computing systems*, 3339–3348. https://doi.org/10.1145/2556288.2557157

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In *Handbook of self-regulation* (pp. 451–502). Elsevier.

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Roschelle, J., Lester, J., & Fusco, J. (2020). AI and the future of learning: Expert panel report. https://circls.org/reports/ai-report

Saqr, M., & López-Pernas, S. (2021). The longitudinal trajectories of online engagement over a full program. *Computers & Education*, *175*, 104325. https://doi.org/10.1016/j.compedu.2021.104325

Saqr, M., López-Pernas, S., & Vogelsmeier, L. V. (2023). When, how and for whom changes in engagement happen: A transition analysis of instructional variables. *Computers & Education*, *207*, 104934. https://doi.org/10.1016/j.compedu.2023.104934

Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The challenges of defining and measuring student engagement in science. *Educational Psychologist*, *50*(1), 1–13. https://doi.org/10.1080/00461520.2014.1002924

Skinner, E. A., Graham, J. P., Brule, H., Rickert, N., & Kindermann, T. A. (2020). "I get knocked down but I get up again": Integrative frameworks for studying the development of motivational resilience in school. *International Journal of Behavioral Development*, *44*(4), 290–300. https://doi.org/10.1177/0165025420924122

Skinner, E. A., Zimmer-Gembeck, M. J., Connell, J. P., Eccles, J. S., & Wellborn, J. G. (1998). Individual differences and the development of perceived control. *Monographs of the Society for Research in Child Development*, *63*(2/3), i–231. https://doi.org/10.2307/1166220

Steuer, G., Rosentritt-Brunn, G., & Dresel, M. (2013). Dealing with errors in mathematics classrooms: Structure and relevance of perceived error climate. *Contemporary Educational Psychology*, *38*(3), 196–210. https://doi.org/10.1016/j.cedpsych.2013.03.002

Symonds, J. E., D'Urso, G., & Schoon, I. (2023). The long-term benefits of adolescent school engagement for adult educational and employment outcomes. *Developmental Psychology*, *59*(3), 503. https://doi.org/10.1037/dev0001458

Symonds, J. E., Kaplan, A., Upadyaya, K., Aro, K. S., Torsney, B. M., Skinner, E., & Eccles, J. S. (2024). Momentary student engagement as a dynamic developmental system. *Journal of Theoretical and Philosophical Psychology*. https://doi.org/10.1037/teo0000288

ten Broeke, N., Hofman, A., Kruis, J., de Mooij, S., & van der Maas, H. (2022). Predicting and reducing quitting in online learning [Preprint]. https://doi.org/10.31219/osf.io/htzvm

Toomla, K., Weng, X., Kikas, E., Malleus-Kotšegarov, E., Aus, K., Azevedo, R., & Hooshyar, D. (2025). Measuring (meta) emotion,(meta) motivation, and (meta) cognition using digital trace data: A systematic review of k-12 self-regulated learning. *Proceedings of the 40th ACM/SIGAPP Symposium on Applied Computing*, 91–100. https://doi.org/10.1145/3672608.3707961

Toste, J. R., Didion, L., Peng, P., Filderman, M. J., & McClelland, A. M. (2020). A meta-analytic review of the relations between motivation and reading achievement for K–12 students. *Review of Educational Research*, *90*(3), 420–456. https://doi.org/10.3102/0034654320919352

Tulis, M., Steuer, G., & Dresel, M. (2016). Learning from errors: A model of individual processes. *Frontline Learning Research*, *4*(2), 12–26. https://doi.org/10.14786/flr.v4i2.168

Tulis, M., Steuer, G., & Dresel, M. (2018). Positive beliefs about errors as an important element of adaptive individual dealing with errors during academic learning. *Educational Psychology*, *38*(2), 139–158. https://doi.org/10.1080/01443410.2017.1384536

Vermeiren, H., Kruis, J., Bolsinova, M., Van Der Maas, H. L., & Hofman, A. D. (2025). Psychometrics of an Elo-based large-scale online learning system. *Computers and Education: Artificial Intelligence*, *8*, 100376. https://doi.org/10.1016/j.caeai.2025.100376

Vu, T., Magis-Weinberg, L., Jansen, B. R. J., Van Atteveldt, N., Janssen, T. W. P., Lee, N. C., Van Der Maas, H. L. J., Raijmakers, M. E. J., Sachisthal, M. S. M., & Meeter, M. (2022). Motivation-Achievement cycles in learning: A literature review and research agenda. *Educational Psychology Review*, *34*(1), 39–71. https://doi.org/10.1007/s10648-021-09616-7

Wan, S., Lauermann, F., Bailey, D. H., & Eccles, J. S. (2021). When do students begin to think that one has to be either a "math person" or a "language person"? A meta-analytic review. *Psychological Bulletin*, *147*(9), 867–889. https://doi.org/10.1037/bul0000340

Wilson, R. C., Shenhav, A., Straccia, M., & Cohen, J. D. (2019). The eighty five percent rule for optimal learning. *Nature Communications*, *10*(1), 4646. https://doi.org/10.1038/s41467-019-12552-4

Wong, Z. Y., Liem, G. A. D., Chan, M., & Datu, J. A. D. (2024). Student engagement and its association with academic achievement and subjective well-being: A systematic review and meta-analysis.

Journal of Educational Psychology, *116*(1), 48. https://psycnet.apa.org/doi/10.1037/edu0000833

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, *41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Supplementary Materials

Three Strikes and Who is Out? Individual Differences in Quitting Following Repeated Errors in Online Learning

Contents

1 Multi-State Survival Model

1.1 Analysis

Fitting procedures for the Multi-State Markov Models were carried out using the msm package (Jackson, 2011, version 1.7). These models were chosen as they allow for continuous-time modeling of transitions into multiple states, which fits into the nature of the learning data at hand. The significance of these results was assessed by deriving confidence intervals using bootstrap methods provided in the msm package. Confidence intervals for the transition intensity matrix and its covariate values were computed from 10,000 bootstrap samples, assuming symmetric distribution on the log scale. The replication analysis was performed on a dataset with $n = 23,459$ observations.

The procedures outlined below match those specified in ten Broeke et al. (2022). See Figure S1 for a schematic representation of the MSSM. Survival models were originally developed for the modeling of disease progression in three states: health, illness, and death. Here, we apply it to data from Prowise Learn and define three quitting states: persisting (user is playing), soft-quitting, and hard-quitting. Hard-quitting is considered an absorbing state because it cannot be left once entering (the student has logged out of the learning environment). The model allows us to estimate the probability of transitioning between one state, h, at time t into another state, j, at time $t + \Delta$t.

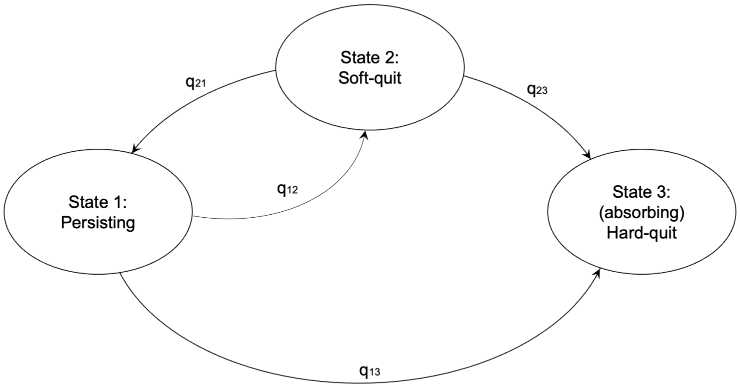

Figure S1: A schematic representation of the continuous Multi-State Survival Model. The model depicts the transitions between each possible state in continuous time: the student is playing and then soft-quits (q_{12}); the student completes a full session after soft-quitting (q_{21}); the student exits the OLE entirely following a soft-quit (q_{23}) or a completed session (q_{13}).

The MSSM assumes that the probability of transitioning from one state to another is dependent on the current state and time, regardless of the history of the system, and in doing so fulfills the Markov property. This allows the probability of transitioning between states, $P(t)$, to be defined by a state transition intensity matrix, Q, in the form:

$$\mathbf{Q} = \begin{bmatrix} -(q_{12} + q_{13} & q_{12} & q_{13} \\ q_{21} & -(q_{21} + q_{23}) & q_{23} \\ 0 & 0 & 0 \end{bmatrix}$$

Another important characteristic of the model is that it allows the effects of covariates, $z(t)$, on transition intensities q_{hj} to be estimated. This is done by calculating hazard ratios, which denote how much each covariate increases or decreases the likelihood of a state transition. Hazard ratios are computed with the proportional hazards model:

$$q_{hj}(z(t)) = q_{hj}^{0} \exp(\beta_{(z(t))}^{\top})$$

in which the baseline transition rate is modified by the exponentiated effect of each covariate's influence, represented by the product of the covariate's corresponding β parameter and its value.

The MSSM is fitted to the data using a maximum-likelihood algorithm. The individual likelihood of transitioning from one state, S_{tj}, to another state, S_{tj+1}, for student i, is:

$$L_{i,j} = p_{S(t_j)S(t_{j+1})}(t_{j+1} - t_j)$$

This represents the entry of the transition matrix at the $S(t_j)$th row and $S(t_{j+1})$th column at $t = t_{j+1} - t_j$. Consequently, the full likelihood is the product of all L_{ij} terms over all students and all transitions.

1.2 Transition Intensities

To investigate the baseline transitions between persisting, soft-quitting, and hard-quitting, and the effects of sequential errors thereof, we fit two different Multi-State Survival Models (MSSM) to the data. This was a direct replication of the analyses of ten Broeke et al. (2022). We included data from all domains in the Learning Sea and Math Garden environments spanning the 2-month period between 2023-05-29 and 2023-07-30 from which users had played a total of at least 5 games. This resulted in a sample size of 24,859 users, across grades three (n = 4343), four (n = 4816), five (n = 4770), six (n = 4304), seven (n = 3869), and eight (n = 2787). The first model was a constrained model, which estimated the rates of transitioning between persisting, soft-quit, and hard-quit states without controlling for any extraneous variables. The second model, the proportional hazards model, controlled for the effects of 6 covariates: gender, grade, difficulty level, speed of error, playing within or outside school hours, and sequential errors.

The proportional hazards model contributed to a significantly better model fit compared to the constrained model, χ^2(44) = $1\,238\,246$, $p < .001$. This model provided evidence for the existence of quitting states across time. The rate of transitioning from a persisting state to a hard-quit state was 0.0227 (95% CI [0.022; 0.023]), whereas the transition rate between a soft-quit state to a hard-quit state was 0.0384 (95% CI [0.037; 0.040]). Thus, users are approximately 1.7 times more likely to transition into a hard-quit state when they are in a soft-quit state compared to a persisting state. All transition intensities can be found in Table S1.

1.3 The Proportional Hazards Model

The proportional hazards model revealed that sequential errors predict transitions to both soft-quit and hard-quit states (Figure S2; Table S2). The pattern of effects of 1, 2, 3, and $>$3 sequential errors was similar to ten Broeke et al. (2022), with particularly strong similarities for persisting to soft-quit transitions across both learning applications. These findings reflect strong effects of all levels of sequential errors on the risk of transitioning from persisting to soft- and hard-quitting states, respectively. For example, when practicing within Language Sea, a student is 10 times more likely to transition into a soft-quit state after having made three errors in a row, compared to no error.

Table S1: Transition Intensities for the Covariate MSSM

	Estimate	SE	95% CI	Probability
Persisting → Persisting	-0.057	$< .001$	$[-0.058; -0.056]$	0.948
Persisting → Soft-Quit	0.034	$< .001$	$[0.034; 0.035]$	0.029
Persisting → Hard-Quit	0.023	$< .001$	$[0.022; 0.023]$	0.023
Soft-Quit → Persisting	0.241	0.002	$[0.238; 0.245]$	0.204
Soft-Quit → Soft-Quit	-0.280	0.002	$[-0.283; -0.276]$	0.759
Soft-Quit → Hard-Quit	0.038	0.001	$[0.037; 0.040]$	0.036

Note. 95% confidence intervals are computed using normal approximation methods, assuming normality of the log effect. Transitions from the hard-quit state are not included as this is an absorbing state, thus all estimated transitions from this state are zero, and the probability of staying in the hard quit state is 1. MSSM = Multi-State Survival Model; SE = Standard Error.

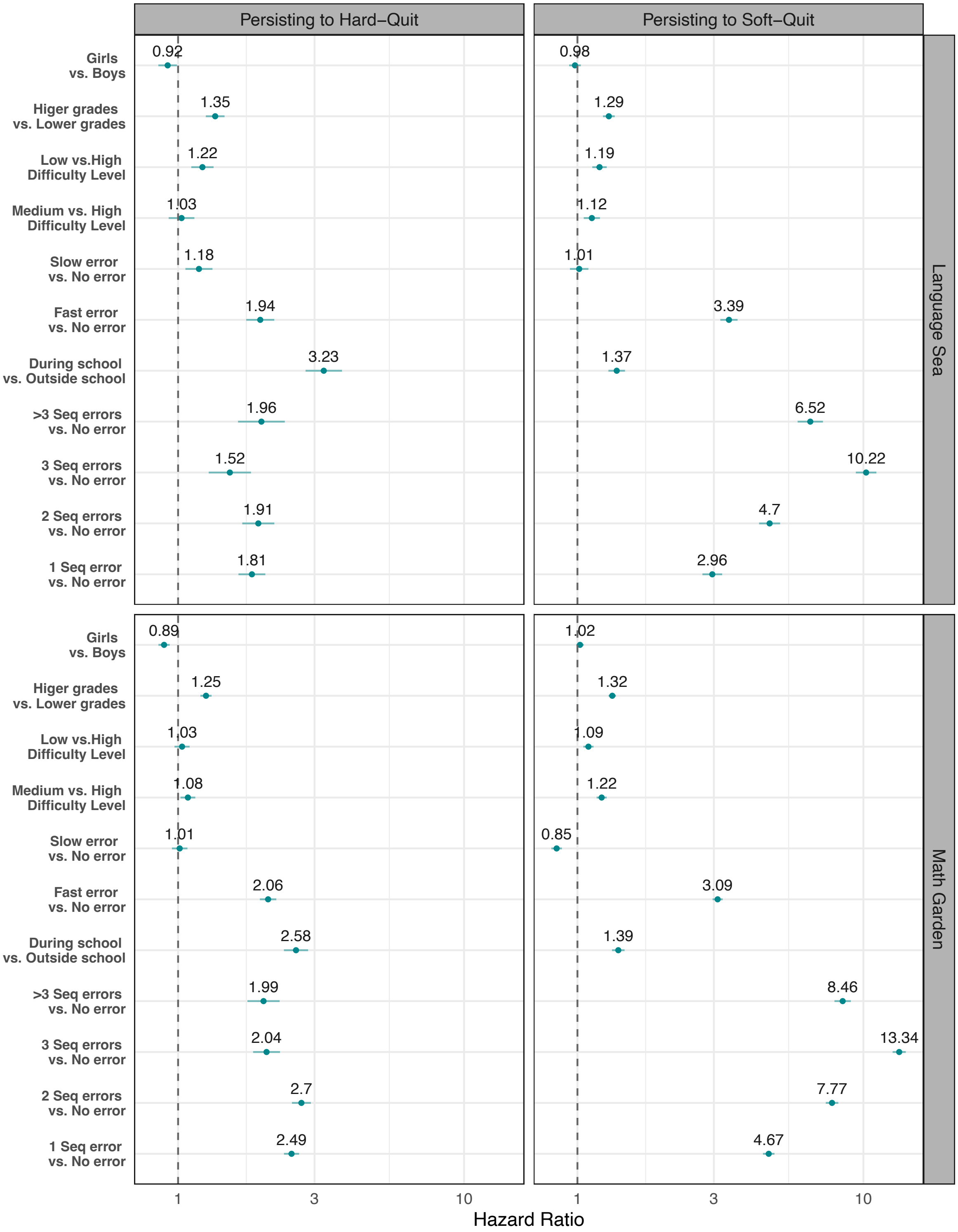

Figure S2: Hazard Ratios derived from the Multi-State Survival Model. Each hazard ratio represents the relative increase or decrease in risk of a state transition for each covariate value compared to its reference category. Hazard ratios are presented on a logarithmic scale. Lines represent the 95% confidence interval.

Table S2. Hazard Ratios and 95% Confidence Intervals for Language Sea and Math Garden

	Transition Type	Covariate	Hazard Ratio	95% CI
Language Sea	P → HQ	1 Seq error vs. No error	1.812	[1.626, 2.019]
		2 Seq errors vs. No error	1.907	[1.676, 2.170]
		3 Seq errors vs. No error	1.517	[1.279, 1.798]
		>3 Seq errors vs. No error	1.956	[1.620, 2.361]
		Fast error vs. No error	1.937	[1.730, 2.169]
		Slow error vs. No error	1.183	[1.061, 1.319]
		During school vs. Outside school	3.231	[2.790, 3.741]
		High diff vs. Medium diff	1.216	[1.112, 1.330]
		High diff vs. Low diff	1.028	[0.927, 1.140]
		Higher grades vs. Lower grades	1.348	[1.250, 1.453]
		Girls vs. Boys	0.919	[0.853, 0.991]
	P → S-Q	1 Seq error vs. No error	2.962	[2.736, 3.207]
		2 Seq errors vs. No error	4.699	[4.318, 5.115]
		3 Seq errors vs. No error	10.221	[9.410, 11.101]
		> 3 Seq errors vs. No error	6.52	[5.888, 7.220]
		Fast error vs. No error	3.387	[3.160, 3.631]
		Slow error vs. No error	1.014	[0.941, 1.092]
		During school vs. Outside school	1.372	[1.283, 1.467]
		High diff vs. Medium diff	1.195	[1.127, 1.266]
		High diff vs. Low diff	1.123	[1.052, 1.199]
		Higer grades vs. Lower grades	1.288	[1.228, 1.350]
		Girls vs. Boys	0.98	[0.935, 1.027]
Math Garden	P → H-Q	1 Seq error vs. No error	2.493	[2.345, 2.649]
		2 Seq errors vs. No error	2.699	[2.501, 2.913]
		3 Seq errors vs. No error	2.038	[1.829, 2.271]
		> 3 Seq errors vs. No error	1.989	[1.748, 2.265]
		Fast error vs. No error	2.064	[1.932, 2.205]
		Slow error vs. No error	1.013	[0.951, 1.078]
		During school vs. Outside school	2.584	[2.344, 2.848]
		High diff vs. Medium diff	1.032	[0.971, 1.097]
		High diff vs. Low diff	1.082	[1.019, 1.149]
		Higer grades vs. Lower grades	1.252	[1.197, 1.309]
		Girls vs. Boys	0.893	[0.854, 0.934]
	P → S-Q	1 Seq error vs. No error	4.666	[4.455, 4.887]
		2 Seq errors vs. No error	7.768	[7.386, 8.171]
		3 Seq errors vs. No error	13.341	[12.652, 14.068]
		> 3 Seq errors vs. No error	8.462	[7.923, 9.039]
		Fast error vs. No error	3.092	[2.969, 3.220]
		Slow error vs. No error	0.846	[0.811, 0.882]
		During school vs. Outside school	1.39	[1.321, 1.462]
		High diff vs. Medium diff	1.093	[1.049, 1.138]
		High diff vs. Low diff	1.215	[1.167, 1.265]
		Higer grades vs. Lower grades	1.324	[1.285, 1.364]
		Girls vs. Boys	1.021	[0.990, 1.052]

2 Model Statistics from the 2-State Markov Model

2.1 Main effects.

Main effects of all covariates from the 2-State Markov Model, for training and testing data. Estimates represent the associated risk of transitioning from a persisting to quitting state, and their associated 95% confidence interval.

Table S3: 2-State Markov Model: Main effects.

	Training Data		Testing Data	
	HR	95% CI	HR	95% CI
1 vs. 0 Sequential Errors	6.45	[6.29 ; 6.61]	6.98	[6.81 ; 7.16]
2 vs. 0 Sequential Errors	14.62	[14.24 ; 15.02]	16.08	[15.64 ; 16.51]
3 vs. 0 Sequential Errors	12.34	[11.83 ; 12.88]	13.46	[12.90 ; 14.04]
$>$3 vs. 0 Sequential Errors	10.76	[10.18 ; 11.37]	11.90	[11.26 ; 12.56]
Fast vs. Slow Response Time	0.95	[0.93 ; 0.96]	0.93	[0.91 ; 0.94]
During vs. Outside School Hours	0.62	[0.61 ; 0.63]	0.62	[0.61 ; 0.64]
Grade 5-6 vs. Grade 7-8	1.13	[1.11 ; 1.15]	1.24	[1.23 ; 1.28]
Grade 3-4 vs. Grade 7-8	1.44	[1.42 ; 1.47]	1.55	[1.53 ; 1.58]
Easy vs. Medium Difficulty Level	1.05	[1.03 ; 1.07]	1.02	[1.00 ; 1.04]
Difficult vs. Medium Difficulty Level	1.08	[1.07 ; 1.10]	1.10	[1.08 ; 1.11]

Note. 95% confidence intervals are computed using normal approximation methods, assuming normality of the log effect. HR = Hazard Ratio.

2.2 Interaction effects.

Interaction effects of all covariates from the 2-State Markov Model, for training and testing data. Estimates represent the change in associated risk of transitioning from a persisting to quitting state for levels of sequential errors, given the relevant covariate, and their associated 95% confidence interval.

Table S4: 2-State Markov Model: Interaction effects.

	Training Data		Testing Data	
	HR	95% CI	HR	95% CI
1 SE * Fast RT	1.34	[1.31 ; 1.36]	1.36	[1.34 ; 1.39]
2 SE * Fast RT	1.89	[1.85 ; 1.92]	1.87	[1.83 ; 1.91]
3 SE * Fast RT	1.91	[1.85 ; 1.97]	1.90	[1.84 ; 1.96]
>3 SE * Fast RT	1.63	[1.57 ; 1.70]	1.60	[1.54 ; 1.66]
1 SE * Easy Level	0.98	[0.95 ; 1.00]	0.97	[0.95 ; 0.91]
2 SE * Easy Level	1.11	[1.08 ; 1.14]	1.18	[1.15 ; 1.21]
3 SE * Easy Level	1.13	[1.08 ; 1.19]	1.18	[1.13 ; 1.24]
>3 SE * Easy Level	1.10	[1.03 ; 1.18]	1.21	[1.14 ; 1.30]
1 SE* Difficult Level	0.85	[0.83 ; 0.87]	0.83	[0.82 ; 0.85]
2 SE * Difficult Level	0.95	[0.93 ; 0.97]	0.91	[0.89 ; 0.93]
3 SE * Difficult Level	0.89	[0.86 ; 0.92]	0.90	[0.87 ; 0.93]
>3 SE * Difficult Level	0.90	[0.87 ; 0.94]	0.94	[0.90 ; 0.97]
1 SE* Grade 3-4	1.28	[1.25 ; 1.31]	1.18	[1.15 ; 1.21]
2 SE * Grade 3-4	0.78	[0.76 ; 0.80]	0.71	[0.70 ; 0.73]
3 SE * Grade 3-4	0.91	[0.87 ; 0.94]	0.82	[0.78 ; 0.85]
>3 SE * Grade 3-4	0.98	[0.93 ; 1.03]	0.87	[0.82 ; 0.91]
1 SE * Grade 5-6	1.12	[1.10 ; 1.16]	1.02	[1.00 ; 1.05]
2 SE * Grade 5-6	0.89	[0.87 ; 0.92]	0.80	[0.78 ; 0.82]
3 SE * Grade 5-6	0.96	[0.92 ; 1.01]	0.84	[0.81 ; 0.88]
>3 SE * Grade 5-6	0.99	[0.94 ; 1.05]	0.89	[0.84 ; 0.95]
1 SE * During School Hours	1.01	[0.98 ; 1.05]	1.04	[1.01 ; 1.08]
2 SE * During School Hours	1.41	[1.37 ; 1.46]	1.33	[1.28 ; 1.37]
3 SE * During School Hours	1.27	[1.20 ; 1.34]	1.32	[1.25 ; 1.40]
>3 SE * During School Hours	1.48	[1.38 ; 1.58]	1.39	[1.29 ; 1.49]

Note. 95% confidence intervals are computed using normal approximation methods, assuming normality of the log effect. HR = Hazard Ratio; SE = Sequential Error; RT = Response Time.

2.3 Fit Estimates.

AIC and Log Likelihood estimates across all three Markov Models, for training and testing data.

Table S5: 2-State Markov Model: Fit Estimates.

	Training Data		Testing Data	
Model	AIC	-2Log-Likelihood	AIC	-2Log-Likelihood
Baseline	2835833	2835831	2848723	2848721
Covariate	2367165	2367143	2378413	2378391
Interaction	2359080	2359010	2370467	2370397

3 Model Statistics from the Mixed Effects Logistic Regression

All model estimates related to the mixed-effects logistic regression fitted on the addition and subtraction data.

3.1 Data Selection Effects

Due to the relatively low base rate of quitting in our sample (approximately 0.03), the level 1 sample sizes need to be larger to ensure reliable estimated individual quitting probability. This is also reflected in convergence issues in the glmer models if we loosen the selection criteria. However, to investigate potential bias due to these selection criteria, we performed glmer models where we only removed extreme cases from our sample (students who either never quit, or always quit), as these cases cause the convergence issues. Following this approach, we did not constrain a minimum amount of sessions completed. This resulted in a sample size of $77\,873$, compared to 3998 following the previous criteria. To speed up computation, we randomly sampled 2000 students ten times and performed the same glmer model (including a random intercept and slope) on each subsample. The effect size of sequential errors on quitting ranged from 1.14 (SE = 0.021) to 1.18 (SE = 0.022), which is larger than the reported effect of the original model (0.81 (SE = 0.007)). Figure S3 displays the difference in the distribution of effects estimated by the model with the original data selection criteria and the model with only extreme cases removed. All estimated fixed (Table S6) and random (Table S7) effects from one randomly sampled subset are reported below.

Figure S3: Distribution of random slope estimates, comparing the glmer model with the original data selection criteria (pink), and the glmer model were only extreme cases were removed (purple). Veritcal lines display the fixed effect of sequential errors on quitting in each model, with 95% confidence bounds.

Table S6: GLMER Fixed effect estimates on data selection subset

	Estimate	SE	*z* value	*p* value
Intercept	-3.71	0.021	-170.90	¡.001
Sequential Errors	1.16	0.022	52.40	¡.001
Rating	-0.26	0.025	-10.34	¡.001
Grade	0.03	0.014	2.46	.014

Note. Fixed effects estimated in one of 10 subsets (N = 2000) of the addition data, where only extreme cases were removed from the data. SE = Standard Error.

Table S7: GLMER variance estimates of random parameters on data selection subset

	Estimate	Std. Deviation
Intercept	0.29	0.54
Sequential Error	0.37	0.60

Note. Variance estimates of the random parameters, estimated in one of 10 subsets (N = 2000) of the addition data, where only extreme cases were removed from the data.

3.2 Descriptive statistics across training and testing data.

Histograms for the training data is shown in orange, testing data in purple, and their overlap in pink. A: Distribution of grades. B: Distribution of sequential errors. C: Distribution of ability (user ratings). D: Distribution of accuracy (binary variable). E: Distribution of chosen difficulty setting (categorical variable with levels 1 (easy), 2 (medium), 3 (hard).

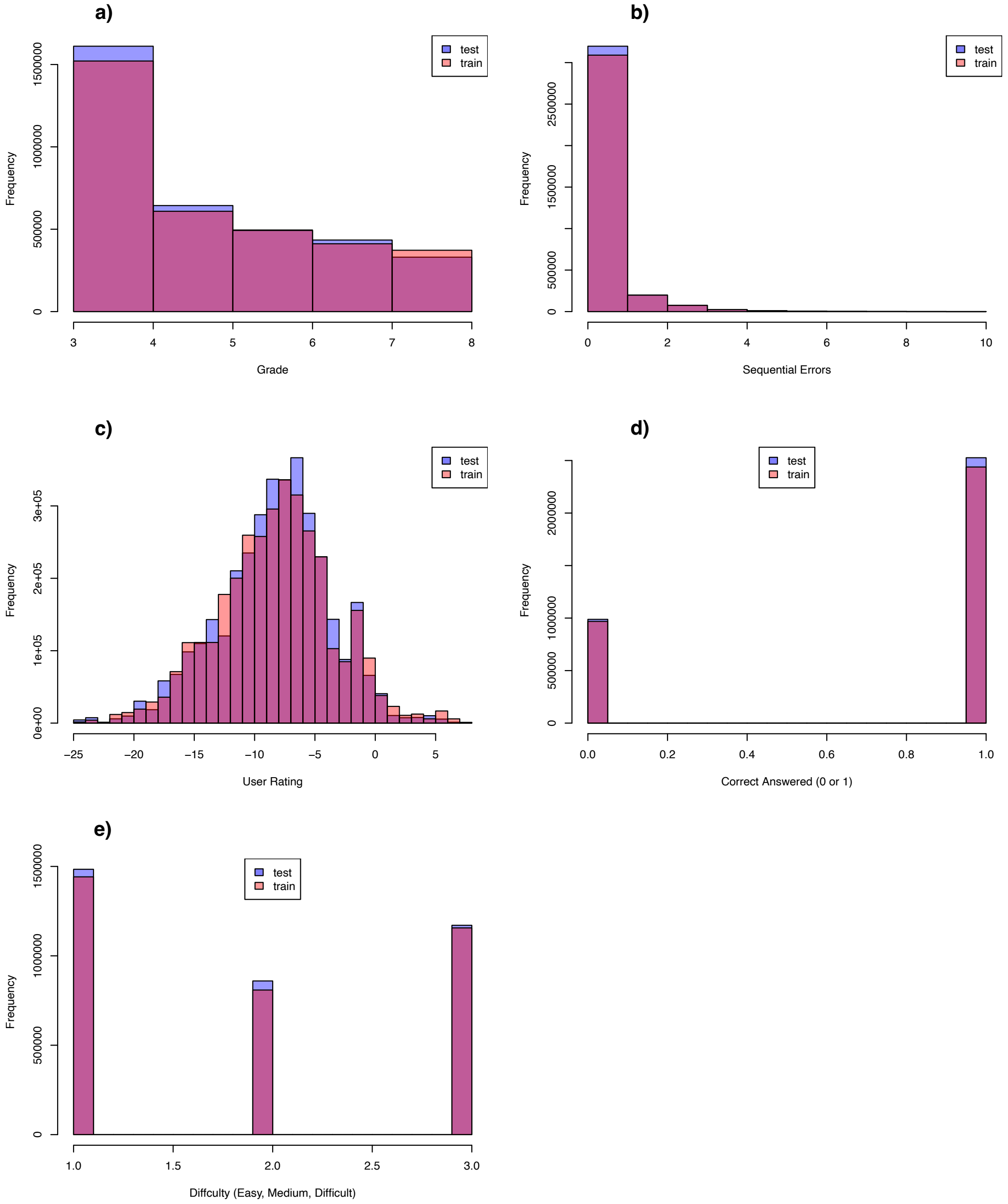

Figure S4: Descriptive statistics across training and testing data.

3.3 GLMER model estimates on the addition data.

Table S8: Fixed effects on the addition data.

	Training Data				Testing Data			
	Estimate	SE	*z* value	*p* value	Estimate	SE	*z* value	*p* value
Intercept	-3.48	0.02	-152.11	<0.001	-3.57	0.02	-149.97	<0.001
Sequential Error	0.81	0.01	119.57	<0.001	0.81	0.01	125.46	<0.001
Rating	-0.10	0.01	-8.83	<0.001	-0.09	0.01	-8.09	<0.001
Grade	-0.05	0.00	-12.07	<0.001	-0.03	0.00	-7.30	<0.001

Note. SE = Standard Error.

Table S9: Random Effects on the Addition Data

	Training Data		Testing Data	
	Variance	Std. Deviation	Variance	Std. Deviation
Intercept	0.39	0.62	0.41	0.64
Sequential Error	0.14	0.38	0.13	0.36

Table S10: Correlation between fixed effects

	Training Data				Testing Data			
	1.	2.	3.	4.	1.	2.	3.	4.
1. Intercept	1.000	-0.073	0.281	-0.886	1.000	-0.095	0.263	-0.891
2. Sequential Errors	-0.073	1.000	0.007	-0.018	-0.095	1.000	-0.005	0.006
3. User Rating	0.281	0.007	1.000	-0.286	0.263	-0.005	1.000	-0.269
4. Grade	-0.886	-0.018	-0.286	1.000	-0.891	0.006	-0.269	1.000

Table S11: Correlation between random effects.

Dataset	Variance-Covariance	Std. Correlation	*p* value
Training	-0.02	-0.10	$<$0.001
Testing	-0.02	-0.10	$<$0.001

References

Jackson, C. H. (2011). Multi-State models for panel data: The msm package for R. *Journal of Statistical Software*, *38*(8), 1–28. https://doi.org/10.18637/jss.v038.i08

ten Broeke, N., Hofman, A., Kruis, J., de Mooij, S., & van der Maas, H. (2022). Predicting and reducing quitting in online learning [Preprint]. https://doi.org/10.31219/osf.io/htzvm