Three Strikes and Who is Out? Individual Differences in Error-Induced Quitting

Annie M. Johansson[1], Alexander O. Savi[1], Han L. J. van der Maas[1], and Abe D. Hofman[1, 2]

[1]Department of Psychological Methods, University of Amsterdam
[2]Prowise Learn

June 6, 2024

Abstract

The biggest threat to learning is to not engage in it. Crucially, sequential errors have been found to be an important cause of quitting from learning. However, little is known about how students differ in their sensitivity to errors. Using intensive longitudinal practicing data from over $200,000$ primary-school students in a large-scale Online Learning Environment, we confirm previous findings 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 quitting following sequential errors. We validate these results in an independent dataset, and show that individual differences are stable across two arithmetic practice domains. Our results corroborate the theoretical notion that students differ in their tolerance to failure and pinpoint a need to individualize how computer-adaptive systems intervene after errors.

1 Introduction

The most effective way to harm learning is to disengage from it. A long-standing and crucial societal issue is keeping young children engaged in educational practice. Student engagement is a key contributor to academic success [1; 2; 3; 4] which, in turn, promotes development. This study seeks to identify what factors contribute to disengagement in educational practice. In doing so, it is possible to inform educational design to maximize learning potential for all students.

Parallel to promoting engagement, there is a growing need for education to adapt to an increasingly digitalized society. Technological developments in education, such as gamification and artificial intelligence, have opened several avenues in the pursuit of improving educational outcomes for students (see *e.g.*, [5; 6; 7; 8]). One example is adaptive education software: online learning environments (OLEs) designed with the goal to personalize the learning level for each student. Although these platforms enable flexibility and individualization in the learning process and in doing so can promote learning, they, too, can only be successful under the crucial condition that students keep engaging with the platform.

The ultimate form of disengagement from learning is quitting, deliberately removing oneself from the learning process before its intended end. While adult drop-out rates in Massive Open Online Courses (MOOCs) have been widely studied and are known to be high (*e.g.*, [9]), there is limited research investigating what factors influence quitting at the primary school level. Here, OLEs are intended to complement traditional classroom instruction rather than replace it, thus drop-out, in the way it is studied within MOOCs, does not exist. Therefore, quitting in this context is better understood as shorter-term disengagement, such as a lack of persistence.

Task persistence, and other related measures, such as perseverance of effort, self-efficacy, and conscientiousness [10; 11] are considered crucial for student engagement and academic achievement [12]. These traits have shown to improve with development. For instance, children's ability to stay persistent on a puzzle-solving task increases and becomes more consistent with age [13]. Apart from developmental mechanisms, persistence can fluctuate on a daily basis, both in adolescents [14], and younger children [13]. Such daily variability suggests that there may be immediate external or internal factors which affect children's tendencies to stay engaged in their educational practice.

One such factor is errors. While it is often argued that some error is necessary for effective learning, making repeated incorrect responses has been associated with less growth, and more disengagement, in OLEs [15; 16]. Closely related to quitting, many students skip problems when they require more effort [17; 18], such as after an error has been made. One study has modeled quitting behavior in Prowise Learn [19]. They estimated the probability of transitioning between three different states across time: (1) the student is playing (persisting state), (2) the student quits an exercise prematurely but stays within the learning environment (soft-quit state), or (3) the student quits an exercise prematurely and exits the learning environment entirely (hard-quit state). Importantly, in correspondence with previous research [15; 16], they found that the strongest predictor of both soft-quitting and hard-quitting was sequential errors. That is, making two, three, or more than three errors in a row largely increased students' risk of quitting both from specific games and the learning environment for the day. Ultimately, making sequential errors may be an important culprit in student quitting behavior.

Persevering in the face of consecutive failures requires that students possess a plethora of adaptive traits that help them resist quitting (for an overarching review, see [20]), including self-regulated learning (SRL; [21; 22]). Disengagement and effort-avoidance caused by errors, in this context, may be signaling a lack of self-regulation in the learner. Self-regulation is argued to be beneficial for success especially in the context of academic obstacles, as it requires a degree of meta-cognitive skills which allow for the child to monitor their performance while keeping their task goal in mind, and in doing so they can recognize when their performance does not match the task goal (i.e. the task has failed; [23; 24; 25]).

One such meta-cognitive process is error monitoring, the ability to keep track of the outcomes of one's actions and optimally adjust subsequent behavior. Previous research has found variability in the degree to which individuals are able to optimally monitor their errors [26], and lower error monitoring has been linked with lower levels of conscientiousness and a lack of perseverance [27]. Other research has focused on how individuals respond to errors. For example, adaptive post-error behavior, in the form of post-error slowing (PES), has been shown to be related to increased performance in Prowise Learn [28]. Moreover, lower levels of PES has been linked to attention deficit hyperactivity- [29] and anxiety disorders [30]. Other work, such as that on learned helplessness [31], fixed- and growth mindset [32; 33; 34] point to variability in the the degree of effort that students are willing to exert when faced with challenging tasks. Taken together, it is likely that there are variances in the ability to adaptively detect and respond to errors, giving rise to differences in the effect that sequential errors have on quitting across students.

1.1 The current study

The current study aims to replicate previous findings that sequential errors predict quitting in OLEs [19], and extend these findings by accounting for individual variability in the effect of sequential errors on quitting. We expect the probability of transitioning between persisting, soft-quit, and hard-quit states to be similar to the previous study, and expect the same pattern of results for the relative importance of 1, 2, 3, and more than 3 sequential errors on all state transitions.

Next, we estimate the degree of individual differences in the relationship between sequential errors and quitting. This is arguably the next step for improving the design of OLEs that promise an individualized learning experience. First, we examine whether the effect of sequential errors on quitting differs depending on age, difficulty level, speed of errors, and whether students play outside or inside school hours. Second, we look at how the tendency to quit following errors fluctuates across time and whether it depends on the amount of exposure to playing that a user has. Lastly, we fit a mixed-effects logistic regression and examine how the individual effect of sequential errors on quitting fluctuates between users. Based on previous research, we expect sequential errors to have an average effect on quitting, with significant between-subject variance. Some users may show a large tendency to quit following sequential errors, while others persist.

Besides between-subjects individual differences, developmental science concerns *within-subjects differences*: how do individuals' behaviors differ across time and situations [35]? Such questions are crucial for understanding the stability of individual differences across contexts. If we find meaningful individual differences in the effect of sequential errors on quitting in the addition game, our last goal is to examine how these individual differences relate to the subtraction domain. Is an individual's effect of sequential errors on quitting in the addition game similar in the subtraction game? Answering this question is a first step in determining the stability of error-induced quitting across learning contexts.

2 Results

We included different subsets of Prowise Learn gameplay data for our respective aims in this project. Detailed data selection criteria and analysis procedures are described in the methods section.

2.1 Sequential errors predict quitting

For the replication of [19], 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 ($n = 4,343$), four ($n = 4,816$), five ($n = 4,770$), six ($n = 4,304$), seven ($n = 3869$), and eight ($n = 2,787$).

Two different Multi-State Survival Models (MSSM) were fit to the data. First, we fit a constrained model, estimating the rates of transitioning between persisting, soft-quit, and hard-quit states without controlling for any extraneous variables. Second, we fit a covariate model, controlling for gender, grade, difficulty level, speed of error, playing within or outside school hours, and sequential errors. Including covariates in the model contributed to significantly better model fit, $\chi^2(44) = 1238246$, $p < .001$. Similarly to [19], 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, whereas the transition rate between a soft-quit state to a hard-quit state was 0.0284. 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. This is comparable to the results of [19], who reported a rate of 2. All transition intensities can be found in Table 1.

Table 1: Transition Intensities for the Covariate MSSM

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

Note. 95% confidence intervals are computed using normal approximation methods, assuming normality of the log effect. MSSM = Multi-State Survival Model; SE = Standard Error.

The proportional hazards model revealed that sequential errors predict transitions into both soft-quit and hard-quit states (Figure 1). The pattern of effects of 1, 2, 3, and >3 sequential errors was similar to [19], with particularly strong similarities for persisting to soft-quit transitions across both learning applications. For transitions into soft-quitting within the Math Garden, the effect of 1, 2, 3, and >3 sequential errors all fell within the previously reported confidence bounds. For Language Sea games, the effect of 1 and 3 sequential errors did, whereas the effect of 2 and >3 did not. For transitions into hard-quitting, the effects of 1 and 2 sequential errors was replicated across both learning applications, whereas the effects of 3 and >3 sequential errors were lower compared to the previous findings. The effects of 3 and >3 sequential errors in the persisting to hard-quit transition were similar to the effects of 1 and 2 sequential errors, which deviates from both the previous findings and the pattern of effects seen for the persisting to soft-quit transitions.

2.2 Individual differences in Error-Induced Quitting

For the analysis of individual differences, we used log data from users playing the addition game within the Math Garden environment in Prowise Learn. This game exists of arithmetic items suited for students with a large range of addition ability. The data spanned a three-year period between 2021-09-01 and 2023-07-01,

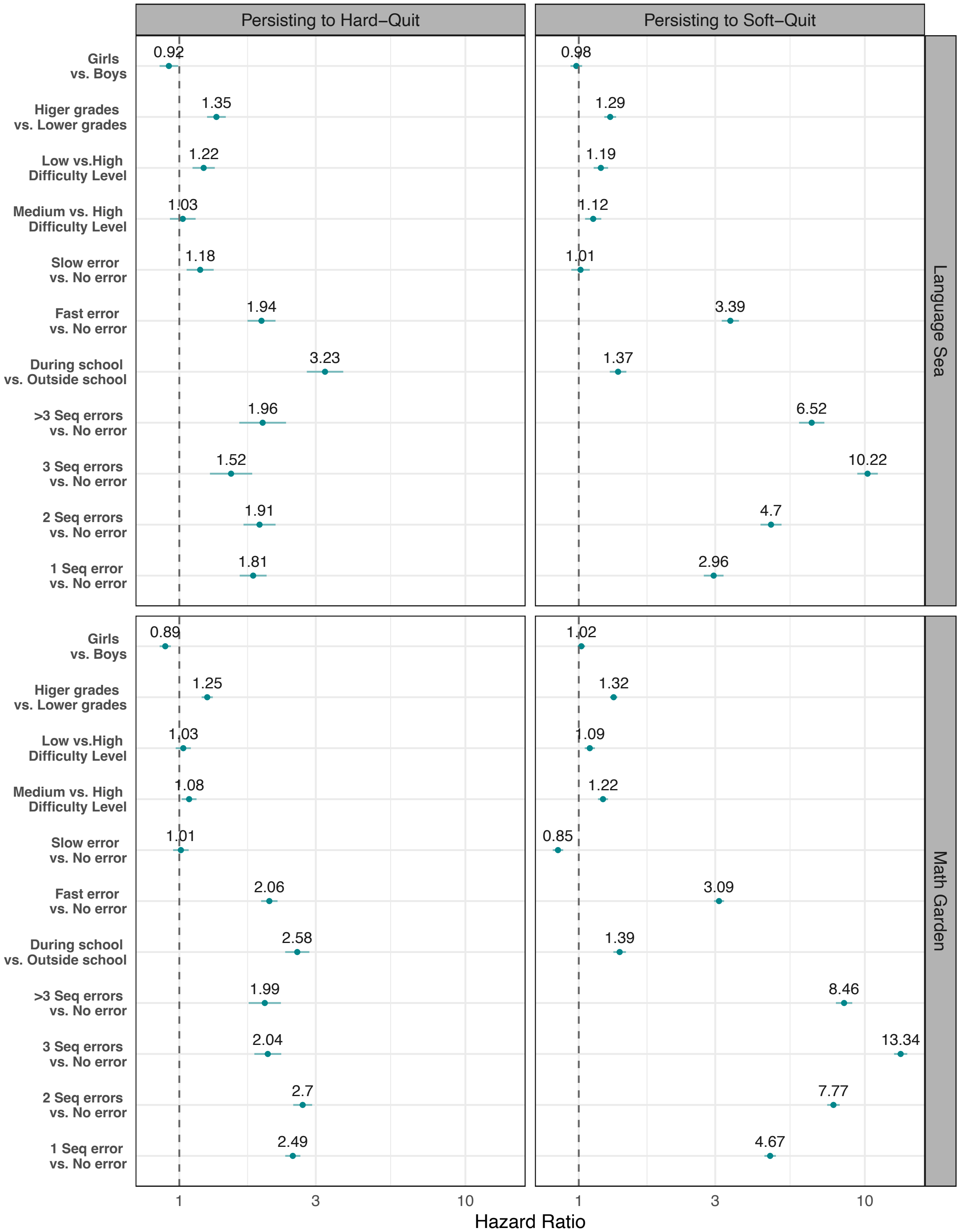

Figure 1: Hazard ratios derived from the Multi-State Survival Model, presented on a logarithmic scale. Each hazard ratio represents the relative increase or decrease in quits for each covariate value compared to its reference category. Lines represent the 95% confidence interval. MSSM = Multi-State Survival Model.

resulting in more than 25 million responses from 255,568 unique users. Given the size of our data, we took a data driven approach, exploring model parameters in a training dataset ($n = 105,864$), and later validating our models on an untouched testing dataset ($n = 105,733$). Grade, gender, ability, and choice of difficulty level were equal distributed across both datasets (Supplementary Materials C1).

Because this analysis looks only at data from one game in the system, it is not possible to differentiate between hard- and soft-quitting. Thus, the outcome variable quitting used hereafter is defined as when a user stops the addition game prematurely (i.e., before completing 10 items), regardless of whether they continued interacting with the application or not. Across both datasets, about 31% of sessions ended prematurely.

2.2.1 Error-induced quitting differs across age, response time, difficulty, and time of day

To estimate the general probability of switching between a persisting to a quitting state in the addition game, 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 state transitions, while controlling for the same covariates as done in the MSSM model. 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 the levels of each covariate (grade, difficulty level, response time, playing during or outside school hours). These three models are hereafter referred to as the constrained, covariate, and interaction model, respectively.

The interaction model provided the best fit to the data ($AIC_{baseline} = 29044256$; $AIC_{covariate} = 2371495$; $AIC_{interaction} = 2364586$). Results of this model are displayed in figure 2. A likelihood ratio test showed best model fit for the interaction model ($\chi^2(35) = 6956, p < .001.$). In this model, users transitioned between a persisting to quitting state at an instantaneous rate of 0.018. 1, 2, 3, and >3 sequential errors had a considerable effect on the likelihood of transitioning from a persisting to quitting state, and there were small effects of difficulty, fast vs. slow errors, and playing outside vs. inside school hours (figure 2, left).

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.

The Markov models were subsequently fit to the testing data to validate the aforementioned results. Results in the testing dataset were highly similar to those found in the training dataset, with the interaction model showing best model fit, $AIC = 2375501$. Fixed and interaction effects were also similar. All Markov

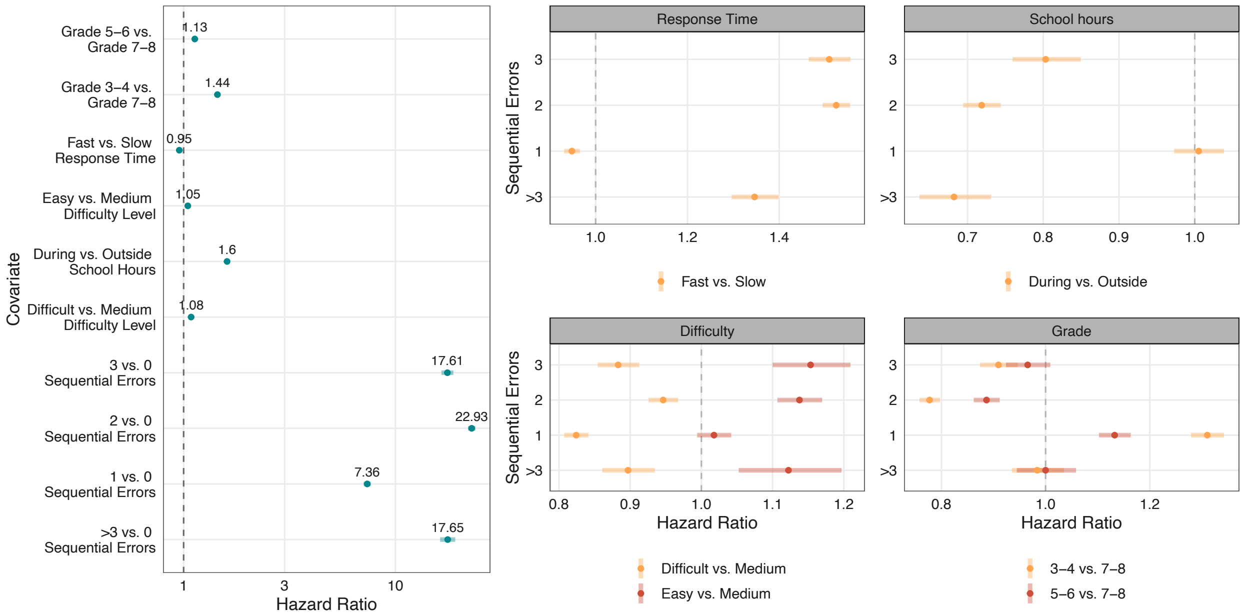

Figure 2: Hazard ratios for the main (left) and interaction (right) effects derived from the 2-state markov model on the addition game. Each hazard ratio for the interaction effects represents the relative increase or decrease in the likelihood of transitioning into a quitting state for each covariate 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.

modeling results on the testing data are reported in Supplementary Materials B.

2.2.2 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 average quitting rate across sessions. Specifically, we compared the probability of quitting after a correct compared to an incorrect response, across session count (how many sessions in the addition game that the user has played at that point in time), separated across difficulty levels and experience (denoting whether a user had played the game before the start of data collection or not). 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 difficulty levels, but not post-correct quitting. These longitudinal effects are displayed in Figure 3.

2.2.3 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

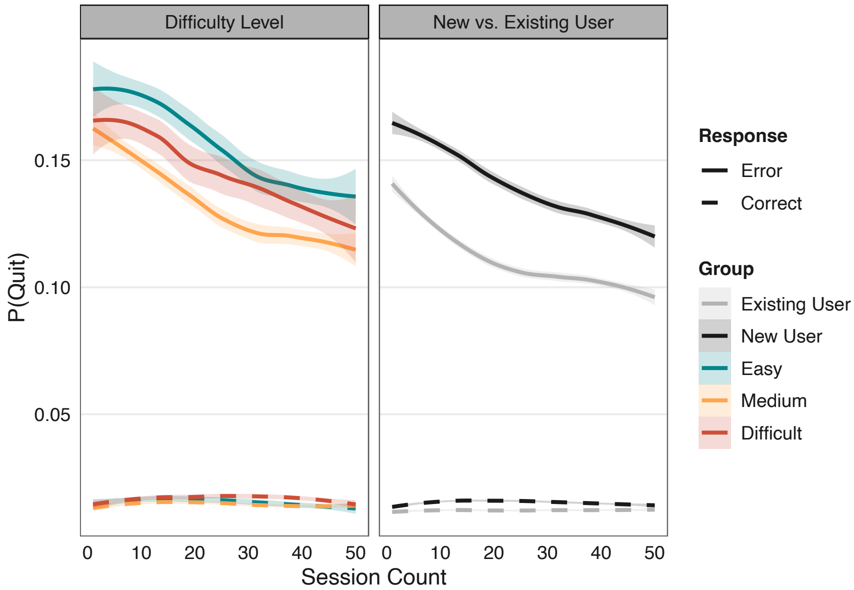

Figure 3: Post-error vs. post-correct quit probabilities, separated across difficulty levels and new vs. existing users. A user was classified as a new user if their first experience in the addition game occurred during the period of data collection, and existing user if they had played the addition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

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 3,998 users.

Fixed effects estimates We fit three separate mixed-effects models. In all models, we used subjects 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 (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 intercepts and individual users. Lastly, we fit a model including a random intercept, random slope, and the fixed effects of user ability and grade, thus allowing these two variables to be controlled for. [1]

The third model provided the best fit to the data (Table 2). 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

[1]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.

Supplementary Materials C; Table C3.

Random effects estimates Adding random variance in the effect of sequential errors on quitting resulted in better model fit. Additionally, intra-class correlations (ICC) demonstrated that this model accounted for about 68% of between-subject variance in baseline quitting rates, and 56% of between-subject variance in the effect of sequential errors on quitting. The variance estimate for the intercept was 0.39 ($\sigma = 0.62$) and for the effect of sequential errors was 0.14 ($\sigma = 0.37$). There was a weak negative correlation between random intercept and slope estimates ($r = -.10$), implying that users who quit less in the absence of errors, are slightly more likely to quit when faced with errors.

Results on Test Data The testing dataset consisted of 4,091 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 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 2). This model accounted for 69% of inter-individual variance in baseline quitting rates and 56% of inter-individual variance in effects of sequential errors on quitting. BIC values were also similar across datasets. Full model comparison results and estimated model parameters in the testing data are reported in Supplementary Materials C.

Table 2: GLMER Model Fit Indices

	Training Data			Testing Data		
Model	AIC	BIC	Log-Likelihood	AIC	BIC	Log-Likelihood
1.	1047301	1047340	-523647	1041610	1041649	-520802
2.	1032731	1032796	-516360	1027581	1027646	-513785
3.	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.

2.2.4 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 the both the addition and subtraction game within the given data collection period; $n = 1765$.

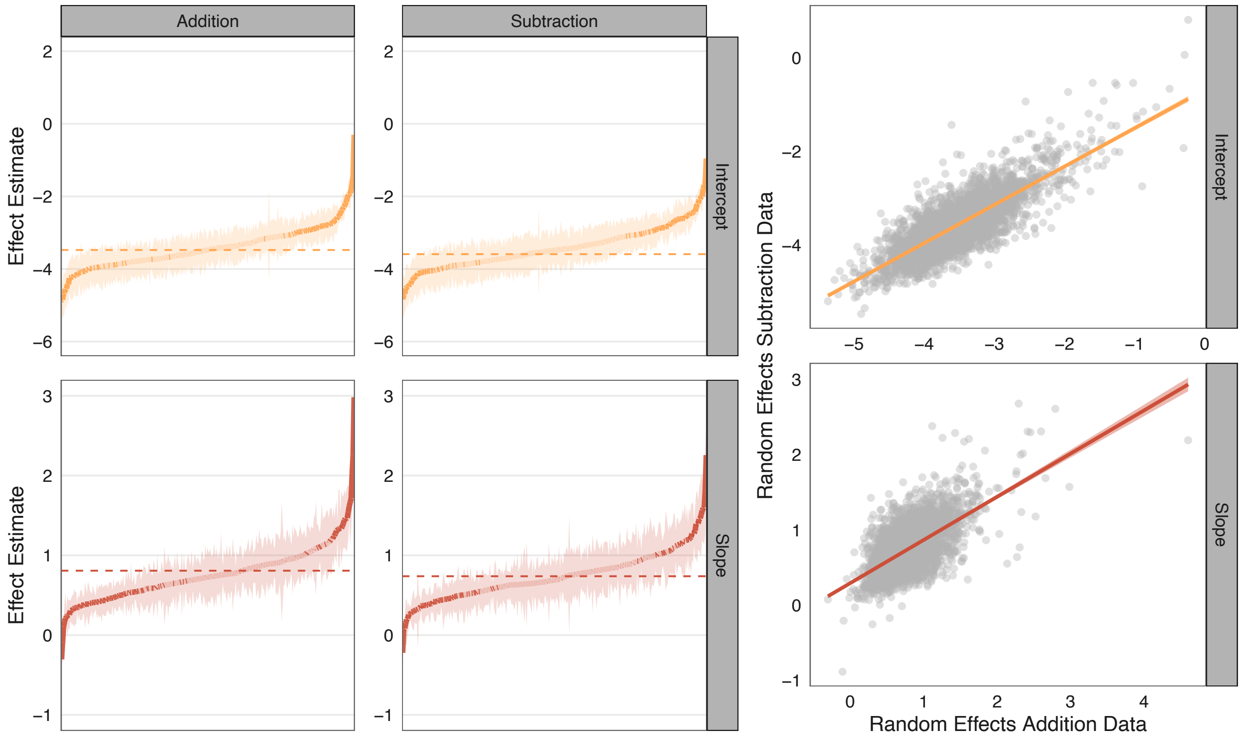

Figure 4: Left: Baseline quitting rates (intercept) and effects of sequential errors on quitting (slope) for 300 randomly sampled users, across both the addition and subtraction domain. Datapoints are ordered from lowest to highest effect estimate in both graphs. Horizontal lines denote the fixed effect. 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, [36]) 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.

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; $p < .001$) . The random variance estimate for the intercept was 0.39 ($\sigma = 0.62$) and for the effect of sequential errors was 0.12 ($\sigma = 0.35$). 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).

3 Discussion

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.

The replication was performed by fitting a continuous time MSSM to estimate the likelihood of transitioning between persisting, soft-quit and hard-quit states in the arithmetic- and language-learning environments. Similarly to [19], the associated risk of transitioning to a hard-quit state is higher when in a soft-quit state compared to a persisting state. This reinforces the previous notion of quitting as a state-dependent process. Importantly, we investigated how the associated risk of quitting changes given the presence of 1, 2, 3, or more than 3 sequential errors. The majority of these effects fall within the confidence bounds of the previous paper. As expected, we also see a clear resemblance in the pattern of effects of all covariates on soft-quitting when comparing with the results of [19].

Interestingly, some of the current effects are lower than those presented in the previous study. This is particularly the case for the associated risks of transitioning between persisting and hard-quit states following sequential errors. It could be that differences in the time period from which the current data was collected had an effect on the estimated effect sizes. The current data collection period took place in June and July, whereas the previous one took place in March and April. In the Netherlands, primary school students have a summer break from school starting at the beginning of July. Consequently, we saw a large drop in data around this time. It is possible that the lower amount of students that practice within the OLE in July are qualitatively different from the average player, which may have deflated our effect estimates. Notwithstanding, our pattern of results closely match those of the previous study and contribute to our understanding of predictors of quitting in online learning.

The effect of sequential errors on quitting is further strengthened by our results in the larger addition and subtraction datasets. The 2-state Markov model demonstrated large increases in the likelihood of transitioning into a quitting state following 1, 2, 3, and >3 sequential errors. Similarly, the fixed effect of quitting following sequential errors in the mixed-effect logistic regression indicated that the likelihood of quitting more than doubles for each sequential error a student makes. Taken together, this is robust evidence supporting the role of sequential errors on students' average propensity to quit, demonstrated across different time periods, learning domains, and statistical methods. These findings emphasize the need for continued use of interventions to reduce quitting in OLEs, such as that designed by [19]. Intervention research will also help in establishing causality in the relationship between errors and quitting.

Despite the fact, our findings make clear that such interventions might not be equally effective for all students. By allowing for effects to differ between users, we find considerable inter-individual variability, both in average quitting rates and in the tendency to quit following errors. These effects are stable across addition and subtraction domains. This is an important first finding highlighting the need to take individual differences in online learning engagement into account. In particular, there are two major implications of these results to the design of interventions on errors in OLEs. First, interventions should be adapted to the individual player. It is possible that users that are less likely to quit following errors will be less impacted by interventions which, for example, reduce the difficulty of an item following an error. Such a persistent user might in contrast be motivated to continue trying on a similar level following a mistake and therefore be discouraged at the presentation of an easier item. On the other hand, a user who is highly reactive to making multiple errors may benefit from a stronger intervention. Second, our findings that individual differences are stable across both addition and subtraction games imply that a user's individual effect of sequential errors on quitting in the addition game can be used to tailor their individual error intervention in the subtraction game. Future research should investigate the effectiveness of such cross-domain interventions.

Our findings indirectly support previous work demonstrating that children vary in their ability to monitor and react to their errors. We see that some children show very large tendencies to quit after encountering errors, which may reflect poor error monitoring, fixed mindset, or a helplessness approach to learning, and vice versa. One main limitation, however, is that the mechanisms underlying individual differences in error-induced quitting are left unstudied. Thus, no clear conclusions can be drawn for the relation between differences in quitting behavior and either cognitive nor behavioral motivational differences. Future work should investigate potential moderators in the relationship between sequential errors and quitting, such as PES or mindset.

The current study has also identified other variables which impact the tendency to quit after errors. Compared to users playing on the default medium level, users playing on a difficult level were less likely to quit, and users playing on an easy level were more likely to quit, following errors. In line with the expectancy-value theory of motivation [37; 38; 39], users playing on an easier level may have a higher expectancy of success, and thus have a higher level of surprise when making several mistakes in a row. This leads to a stronger emotional reaction when encountering the error, in turn leading to a greater likelihood of quitting the task. In contrast, the value component of the expectancy-value theory posits that individuals are more motivated when a task is perceived as valuable or important. This may reflect users that choose to play on a more difficult level, who are quitting less despite encountering errors. The ability to measure reactions to errors in varying degrees of difficulty presents a unique potential to further explore how users' beliefs about their own competency, and the structure of their environment, impacts their persistence. This opens up for a wide array of research questions revolved around whether there exists a "sweet-spot" in difficulty in which a task is neither too easy nor too difficult which allows the learner to stay engaged and motivated, and whether this differs between learners.

Second, looking at error-induced quitting over sessions 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.

"The value of within-person approaches in educational psychology cannot be emphasised enough" ([35], p. 83). In this study, we were able to pinpoint within-person effects of errors on quitting across two arithmetic domains. Previous research has also brought forward within-subjects differences in children and adolescents' fluctuations in task persistence over time [14; 40]. In addition, [41] demonstrate moment-to-moment variability in children's performance in an online math practice environment, as well as reliable inter-individual differences in these intra-individual variabilities. Notably, students with higher degrees of moment-to-moment fluctuations in performance had lower average math performance. These findings bring to light a new path within individual differences research, wherein longitudinal within-subjects fluctuations in tolerance to errors, and their relations to overall cognitive performance, should be established.

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 [42], which our replication results also allude to. 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 persistence in online learning is domain- or user-specific.

Next, our stringent selection of users limits our inferences to a population of children that are inherently more persistent, and children that begin interacting in the system but quit before having played 50 sessions are left unstudied. In regards to disengagement and drop-out, the behavioral patterns of these students, and whether they differ from students who stay in the system on 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. Relatedly, our study has only 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. Nevertheless, we found reliable individual differences in a narrow sub-population of users in the OLE, a strong finding in and of itself.

The efficacy of online learning platforms hinge on their ability to retain their users. Moreover, 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 way for future research on interventions to keep children motivated to learn.

4 Methods

4.1 Online Learning Environment

We use online learning data from the application Prowise Learn, developed by [43]. 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 5 shows a screenshot of the addition game in the math garden environment. The program utilizes computer-adaptive practice (CAP), meaning that it selects the appropriate difficulty level of the item given each player's estimated ability rating. The algorithm is a combination of the Elo algorithm (for further information see [5]) and the explicit scoring rule proposed by [44]. Importantly, the games are time-limited and rating increases when a student is accurate and fast, and decreases if the student performs incorrectly or slowly. The use of CAP allows us to interpret findings in the context of students that are practicing within their optimal learning level.

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 an answer is made. 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 are subtracted from their total. The next item is selected according to the expected probability that the user will answer it correctly, 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.

4.2 Data

This work was approved by the Ethics Review Board of the Psychological Methods department at the University of Amsterdam (approval number 2022-PML-15260). Schools or families signing up for Prowise Learn provide 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.

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 long sequences of 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

Figure 5: A screenshot of the addition game in Prowise Learn.

most games are not suited for children in this age, thus they provide too little data.

Analysis. All data analyses were conducted using R Statistical Software (v4.3.2; [45]). Fitting procedures for the Multi-State and Simple Markov Models were carried out using the msm package (v1.7, [46]). Mixed-effects models were fitted using the *glmer* function from the lme4 package (v.1.1.35.1, [47]).

4.3 The Multi-State Survival Model

The procedures outlined below match those specified in [19]. See Figure 6 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 $t + \Delta\text{t}$.

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

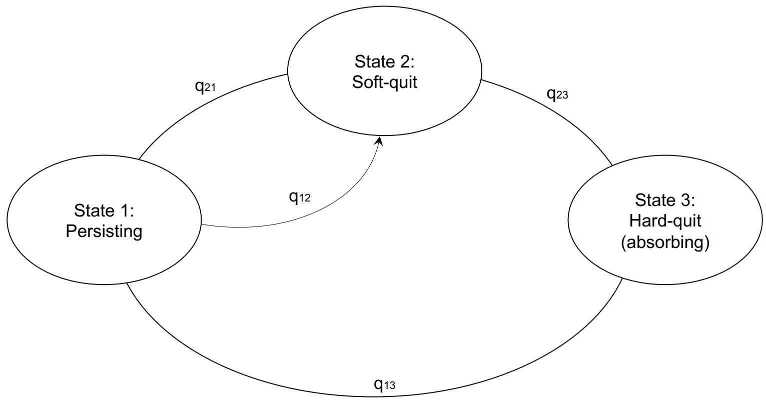

Figure 6: Schematic representation of the continuous Multi-State Survival Model. The model depicts the transitions between each possible state in continuous time: the student is playing and then soft-quits (q_{12}); the student completes a full session after soft-quitting (q_{21}); the student exits the OLE entirely following a soft-quit (q_{23}) or a completed session (q_{13}).

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:

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

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.

4.4 Measures

Quitting Quitting is operationalized in two ways. A soft-quit occurs when the user exits a game before the session has ended (i.e., before completing 10 items), but stays within the learning environment. Second, a hard-quit occurs when the user exits the game prematurely and leaves the application completely. In the replication analysis, both of these operationalizations of quitting are utilized, as they are hypothesized to be dependent on each other. Here, the MSSM is used to measure the probabilities of transitioning between persisting and quitting states. For the analysis of individual differences, no differentiation is made between soft-quitting or hard-quitting, rather, a quit is denoted when a user exits a game before the session has ended.

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 Multi-State and 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 9, which is the maximum amount of sequential errors possible before a game is complete.

4.4.1 Difficulty setting

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

Speed of responding The previous study and this replication included a covariate determining the speed of errors. However, given that an error is contained in such a response, it is possibly that is is related to the sequential error variable, giving rise to multicollinearity in the model. To avoid this in further analyses, we include a response time variable, denoting whether a response was fast or slow, regardless of its correctness. 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 [2].

Playing inside or outside school hours Additionally, we measured whether students play during school hours, between 07:00 and 15:00 on weekdays, or not. This is a binary variable and takes on the values 0 (inside school hours) or 1 (outside school hours).

User ability To estimate user ability, we extracted ability in a certain game, 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.

5 Data Availability

Due to the private nature of children's learning data, data are available from the corresponding author upon reasonable request.

[2] Given that fast errors showed a significant effect on quitting in the MSSM, we also fitted a model including the interaction between sequential errors and response time in the 2-state model. Results are described in section 3.2.2.

6 Code Availability

All code used for data analyses in this project is publicly available at `https://github.com/ann1ejohansson/three-strikes.git`. All resources concerning the current work can also be accessed at `https://ann1ejohansson.github.io/three-strikes/`.

References

[1] M. Janosz, I. Archambault, J. Morizot, and L. S. Pagani. School engagement trajectories and their differential predictive relations to dropout. *Journal of Social Issues*, 64(1):21–40, 2008. doi:10.1111/j.1540-4560.2008.00546.x.

[2] Y. Li and R. M. Lerner. Trajectories of school engagement during adolescence: Implications for grades, depression, delinquency and substance use. *Developmental Psychology*, 47(1):233–247, 2011. doi:10.1037/a0021307.

[3] K. Upadyaya and K. Salmela-Aro. Development of school engagement in association with academic success and well-being in varying social contexts. *European Psychologist*, 18(2):136–147, 2013. doi:10.1027/1016-9040/a000143.

[4] K. S. Blondal and S. Adalbjarnardottir. Student disengagement in relation to expected and unexpected educational pathways. *Scandinavian Journal of Educational Research*, 56(1):85–100, 2012. doi:10.1080/00313831.2011.568607.

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

[6] A. Istenič Starčič. Human learning and learning analytics in the age of artificial intelligence. *British Journal of Educational Technology*, 50(6):2974–2976, 2019. doi:10.1111/bjet.12879.

[7] J. M. Lodge and L. Corrin. What data and analytics can and do say about effective learning. *Npj Science of Learning*, 2(1):5, 2017. doi:10.1038/s41539-017-0006-5.

[8] F. R. Sullivan and P. K. Keith. Exploring the potential of natural language processing to support microgenetic analysis of collaborative learning discussions. *British Journal of Educational Technology*, 50(6):3047–3063, 2019. doi:10.1111/bjet.12875.

[9] I. Borrella, S. Caballero-Caballero, and E. Ponce-Cueto. Taking action to reduce dropout in MOOCs: Tested interventions. *Computers & Education*, 179:104412, April 2022. doi:10.1016/j.compedu.2021.104412.

[10] A. L. Duckworth, P. D. Quinn, and E. Tsukayama. Revisiting the factor structure of grit: A commentary on duckworth and quinn (2009). *Journal of Personality Assessment*, 103(5):573–575, 2021. doi:10.1080/00223891.2021.1942022. PMID: 34254861.

[11] M. Credé, C. Tynan, M, and P. D. Harms. Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology*, 113(3):492–511, 2017. doi:10.1037/pspp0000102.

[12] C. R. O'Neal, M. Y. Boyars, and L. W. Riley. Dual language learners' grit, engagement and literacy achievement in elementary school. *School Psychology International*, 40(6):598–623, 2019. doi:10.1177/0143034319875176.

[13] N. Oeri and C. M. Roebers. Task persistence in kindergarten children: Disentangling age from schooling effects. *British Journal of Developmental Psychology*, 39(1):217–230, 2021. doi:https://doi.org/10.1111/bjdp.12358.

[14] A. Kramer, H. M. Huizenga, A. C. K van Duijvenvoorde, and L. Krabbendam. Do I want to learn today? Day-to-day variations in adolescents' academic motivation and effort. *Learning and Motivation*, 85:101957, 2024. doi:https://doi.org/10.1016/j.lmot.2023.101957.

[15] H. Hou. 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, 2015. doi:https://doi.org/10.1016/j.chb.2015.02.010.

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

[17] A. O. Savi, van Klaveren C., and I. Cornelisz. Combating effort avoidance in computer adaptive practicing: Does a problem-skipping restriction promote learning? *Computers & Education*, 206: 104908, 2023. doi:https://doi.org/10.1016/j.compedu.2023.104908.

[18] A. O Savi, N. M. Ruijs, G. K. J Maris, and H. L. van der Maas. Delaying access to a problem-skipping option increases effortful practice: Application of an a/b test in large-scale online learning. *Computers & Education*, 119:84–94, 2018. doi:https://doi.org/10.1016/j.compedu.2017.12.008.

[19] N. ten Broeke, A. D. Hofman, J. Kruis, S. M. M. de Mooij, and H. L. J. van der Maas. Predicting and reducing quitting in online learning [preprint]. *Open Science Framework*, 2022. doi:10.31219/osf.io/htzvm.

[20] E. A. Skinner, J. P. Graham, H. Brule, N. Rickert, and T. A. Kindermann. 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, 2020. doi:10.1177/0165025420924122.

[21] O. Viberg, M. Khalil, and M. Baars. Self-regulated learning and learning analytics in online learning environments: a review of empirical research. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, LAK '20, pages 524–533, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450377126. doi:10.1145/3375462.3375483.

[22] T. J. Cleary, J. Slemp, and E. R. Pawlo. Linking student self-regulated learning profiles to achievement and engagement in mathematics. *Psychology in the Schools*, 58(3):443–457, 2021. doi:10.1002/pits.22456.

[23] B. J. Zimmerman. Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2):64–70, May 2002. doi:10.1207/s15430421tip4102_2.

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

[25] P. R. Pintrich. The role of goal orientation in self-regulated learning. In *Handbook of Self-Regulation*, pages 451–502. Academic Press, 2000. doi:10.1016/B978-012109890-2/50043-3.

[26] A. E. Miller, J. M. Watson, and D. L. Strayer. Individual differences in working memory capacity predict action monitoring and the error-related negativity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(3):757–763, 2012. doi:10.1037/a0026595.

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

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

[29] A. B. Arnett, C. Rhoads, and T. M. Rutter. Reduced error recognition explains post-error slowing differences among children with attention deficit hyperactivity disorder. *Journal of the International Neuropsychological Society*, 28(8):810–820, 2022. doi:10.1017/S1355617721001065.

[30] M. Rueppel, K. A. Mannella, K. D. Fitzgerald, and H. S. Schroder. Post-error slowing in anxiety and obsessive-compulsive disorders. *Cognitive, Affective, & Behavioral Neuroscience*, 22(3):610–624, June 2022. doi:10.3758/s13415-021-00976-9.

[31] E. A. Skinner, M. J. Zimmer-Gembeck, J. P. Connell, J. S. Eccles, and J. G. Wellborn. Individual differences and the development of perceived control. *Monographs of the Society for Research in Child Development*, 63(2):i–231, 1998. URL `http://www.jstor.org/stable/1166220`.

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

[33] C. S. Dweck. *Self-Theories: Their Role in Motivation, Personality, and Development*. Psychology Press, 1st edition, 1999. doi:10.4324/9781315783048.

[34] C. H. Utman. Performance effects of motivational state: A meta-analysis. *Personality and Social Psychology Review*, 1(2):170–182, 1997. doi:10.1207/s15327957pspr0102_4.

[35] K. Murayama, T. Goetz, L. Malmberg, R. Pekrun, A. Tanaka, and A. J. Martin. Within-person analysis in educational psychology: Importance and illustrations. In David W. Putwain and Katy Smart, editors, *BJEP Monograph Series II: Part 12 The Role of Competence and Beliefs in Teaching and Learning*. British Psychological Society, 2017. doi:10.53841/bpsmono.2017.cat2023.6.

[36] J. E. Knowles, C. Frederick, and A. Whitworth. *merTools: Tools for Analyzing Mixed Effect Regression Models*, 2024. URL `https://CRAN.R-project.org/package=merTools`. R package version 0.6.2.

[37] E. C. Tolman. *Purposive behavior in animals and men*. University of California Press, 1932.

[38] K. Lewin. Field theory, 1948.

[39] A. Achtziger and P. M. Gollwitzer. *Motivation and volition in the course of action*. Springer, 2018.

[40] J. A. Leonard, D. M. Lydon-Staley, S. D. S. Sharp, H. Z. Liu, A. T. Park, D. S. Bassett, A. L. Duckworth, and A. P. Mackey. Daily fluctuations in young children's persistence. *Child Development*, 93(2):e222–e236, 2022. doi:https://doi.org/10.1111/cdev.13717.

[41] N. Judd, M. Aristodemou, and R. Kievit. Cognitive variability is ubiquitous and distinct from mean performance across eleven tasks with over 7 million trials. *Open Science Framework*, 2023.

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

[43] S. Klinkenberg, M. Straatemeier, and H. L. J. van der Maas. 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, 2011. doi:10.1016/j.compedu.2011.02.003.

[44] G. Maris and H. L. J. van der Maas. Speed-accuracy response models: Scoring rules based on response time and accuracy. *Psychometrika*, 77(4):615–633, 2012.

[45] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2024. URL `https://www.R-project.org/`.

[46] C. H. Jackson. Multi-state models for panel data: The msm package for r. *Journal of Statistical Software*, 38(8), 2011. doi:10.18637/jss.v038.i08.

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