BIROn - Birkbeck Institutional Research Online

ARTICLE

Error detection through mouse movement in an online adaptive learning environment

Susanne M. M. de Mooij[1] | Maartje E. J. Raijmakers[2] | Iroise Dumontheil[1,3] |
Natasha Z. Kirkham[1] | Han L. J. van der Maas[2]

[1]Centre for Brain and Cognitive Development, Department of Psychological Sciences, Birkbeck, University of London, London, UK

[2]Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

[3]Centre for Educational Neuroscience, University of London, London, UK

Correspondence
Susanne M.M. de Mooij, Centre for Brain and Cognitive Development, Department of Psychological Sciences, Birkbeck, University of London, Malet street, London WC1E 7HX, UK.
Email: susannedemooij94@gmail.com

Funding information
European Union's Horizon 2020 Marie Sklodowska-Curie Innovative Training Network, Grant/Award Number: 721895

Abstract

While response time and accuracy indicate overall performance, their value in uncovering cognitive processes, underlying learning, is limited. A promising online measure, designed to track decision-making, is computer mouse tracking, where mouse attraction towards different locations may reflect the consideration of alternative response options. Using a speedy arithmetic multiple-choice game in an online adaptive learning environment, we examined whether mouse movements could reflect arithmetic difficulties when error rates are low. Results showed that mouse movements towards alternative responses in correctly answered questions mapped onto the frequency of errors made in this online learning system. This mapping was stronger for the younger children, as well as for easy arithmetic problems. On an individual level, users showed more mouse movement towards their previously made response errors than towards other alternative options. This opens the possibility of adapting feedback and instruction on an individual basis through mouse tracking.

KEYWORDS
adaptive learning, arithmetic, educational games, error detection, mouse tracking

1 | INTRODUCTION

The learning process of essential skills, such as arithmetic, has a unique trajectory in every child. To create an effective teaching–learning environment, tailoring the instruction and feedback to the needs of individual learners is necessary (Federico, 2000). Emerging e-learning platforms and new technologies can accommodate these needs. But to individualize learning materials (Bray & Mcclaskey, 2010), we need to understand the underlying thinking processes that drive particular behavioural responses during learning, that is, we need to understand why a child may consistently have difficulties when faced with a particular type of problem. In the present study, mouse tracking was used to measure difficulties children might have during arithmetic problem-solving in an online learning environment.

1.1 | Individual differences in arithmetic difficulties

When solving problems, the students' error responses are thought to reflect their cognitive process and/or applied strategies (Ben-Zeev, 1998; Buwalda, Borst, van der Maas, & Taatgen, 2016; Savi, Deonovic, Bolsinova, Van Der Maas, & Maris, 2018). Rational errors are errors that are logically consistent and rule-based rather than being random (Ben-Zeev, 1998). Rational errors reflect the student consistently applying an incorrect procedure (Brown & Burton, 1978). For example, for

arithmetic problems, a common mistake is $3 \times 2 = 5$, where the numbers are added instead of multiplied. Being able to diagnose incorrect understanding, sometimes called misconceptions, by analysing systematic difficulties can ultimately help in individualizing education to students' state of knowledge. But the downside of this approach is that it requires students to make a sufficient number of mistakes.

Traditionally, tests in primary and secondary education are constructed such that students have a high probability of answering correctly. To keep students motivated, adaptive learning systems also choose a high success rate and, as a consequence, few error responses are available (Jansen et al., 2013). Having to rely on the limited number of incorrect responses registered per student makes it almost impossible to rapidly detect underlying systematic difficulties in adaptive learning systems. Studies have shown that during the retrieval of mental arithmetic, in addition to the correct association (e.g., $3 \times 2 = 6$), problems can produce false associations with incorrect answers (e.g., $3 \times 2 = 5$; Campbell, 1987; Domahs, Delazer, & Nuerk, 2006; Siegler, 1988). This means that even though a correct answer is eventually given, the student might still contemplate these incorrect answers, possibly associated with misconceptions. Correct and incorrect answers then compete during the decision-making process.

1.2 | Different ways to track arithmetic cognitive processes

Measures of neural activity (e.g., electroencephalography) or eye movements (eye tracking) have provided important insights into the dynamics of the learner's cognitive processes in mathematics (i.e., Artemenko et al., 2019; de Mooij, Kirkham, Raijmakers, van der Maas, & Dumontheil, 2020; Hinault & Lemaire, 2016; Huebner & LeFevre, 2018; Lai et al., 2013; Spüler et al., 2016). However, these measures are also laborious and expensive and, therefore, difficult to scale to large samples outside research laboratories, such as an online learning environment.

An emerging addition to these methods is mouse tracking, that is, the recording and tracking of computer mouse movements made by participants, with the aim of providing a continuous stream of information during the decision-making process (Dale, Kehoe, & Spivey, 2007; Freeman, 2018; Freeman, Dale, & Farmer, 2011; Hehman, Stolier, & Freeman, 2015; Song & Nakayama, 2009; Spivey & Dale, 2006; Stillman, Shen, & Ferguson, 2018). Mouse tracking, as a method, was first introduced by Spivey, Grosjean, and Knoblich (2005), who used it as a window into the internal cognitive process during language comprehension. Since the free-to-use software of Mousetracker was introduced by Freeman and Ambady (2010), it has become very popular in diverse domains of social science (for recent reviews, see Erb, 2018; Freeman, 2018; Stillman et al., 2018). This method has many practical advantages: it can be collected online (so reduces the need to bring participants into the lab), it is relatively inexpensive and has a much broader scope in terms of the number and variety of participants. These characteristics are especially useful

when studying young children for whom online data are often difficult to obtain.

Hand or mouse trajectories during mental arithmetic are also used to study numerical processing (Dotan & Dehaene, 2013; Faulkenberry, Witte, & Hartmann, 2018; Fischer & Hartmann, 2014; Marghetis, Núñez, & Bergen, 2014; Santens, Goossens, & Verguts, 2011). These studies showed that motor action is not only the end product of perceptual and cognitive processes but also that it can reveal how multiple representations are competing with each other during the problem-solving process (Schulte-Mecklenbeck, Kuehberger, & Johnson, 2019; Spivey, 2007; Spivey & Dale, 2006). For example, Faulkenberry (2014); Faulkenberry, Cruise, Lavro, and Shaki (2016) examined mouse movements during a numerical comparison task, where the numerical distance between two digits needed to be judged, while ignoring the physical size of the digits. These studies showed that in incongruent trials—when these two variables differed (e.g., in a 2–8 trial, the font size of the 2 was bigger than the font size of the 8)—the mouse path was curving more towards the incorrect response than in the congruent trials (e.g., when the 2 was smaller than the 8). This greater attraction towards the incorrect response, due to size congruity interference, is thought to reflect the response competition. Mouse tracking allows an examination of the strength of the attraction towards alternative options, without the need for an error to be made. What cannot be claimed is that mouse tracking is measuring the same as eye tracking: gaze usually precedes the hand/mouse movement in the decision-making process and therefore it is not expected to be synonymous with overt attention, measured with eye tracking (Franco-Watkins & Johnson, 2011; Guo & Agichtein, 2010; Huang & White, 2012; Spivey, 2007).

1.3 | The current study

In the current study, we implemented mouse tracking outside the laboratory. Specifically, mouse movements of primary school children were tracked while doing arithmetic exercises in an online adaptive practice environment (Straatemeier, 2014). The aim of this study was twofold. First, it serves as a validation that mouse movements can reflect the competition, at the cognitive level, between multiple answers during the resolution of arithmetic problems, in an ecologically valid setting. Critically, we hypothesized that the extent of mouse movements towards non-selected incorrect options would relate positively to how frequently these errors are made in a speedy arithmetic multiple-choice game. Second, we examined whether we could detect individual systematic patterns of mouse attraction towards certain errors made in the past. The ultimate aim of this study is to develop a measure of attraction towards systematic difficulties that can be used to adapt feedback and instruction to every unique learning trajectory.

For this measure of attraction, we developed a mouse-tracking method to analyse mouse movement in more complex daily-life tasks. Most of the previous mouse-tracking studies have used a strict two-response options design; existing measures of mouse-trajectory dynamics have been developed for this design, such as the maximum deviation (MD) away from the correct response (Hehman et al., 2015).

Some studies have used designs with four-response choices (see for example Cloutier, Freeman, & Ambady, 2014; Koop & Johnson, 2013), but in analyses have only selected the trajectories where there was deviation towards one of the alternative options and not multiple. In the current study, we have designed a method with five alternative response options, where attraction towards multiple answer options is allowed. During the process of solving a mathematical problem, individual children can have different error-related associations, but importantly, a child can also have multiple error-related associations. By presenting more than one alternative option, we can track the whole problem-solving process, where a variation of error-related associations is also possible.

2 | METHODS

2.1 | Participants

For this study, 90,000 children, aged between 5 and 13 years old (*M* = 10.2 y), were randomly selected from a pool of users playing actively in an online learning environment (*N* = 180,000 users). The participants of the online learning environment are mainly children from primary schools that have bought accounts for their students. Participants had to have logged in to the environment in the last 3 months before data collection to increase the chance that these children would play the game during the 6 weeks of data collection. Not all students were selected to be tracked through mouse movements to limit the load on the database. In this online learning environment, students can decide for themselves when they want to play, which games they want to play, and how many problems of a particular game they want to play. After recording the game for a total of 6 weeks, 1,590 different users (*M* = 10.3 y; *SD* = 1.46; 46% female) had played the selected arithmetic problems in our task and their trajectories could be used for analyses. As can be seen in Table 1, users were predominantly 8 years of age and older, since the selected problems required basic knowledge of, and practice with, all mathematical operations.

2.2 | Materials and equipment

The responses and the mouse movement data were collected within an online adaptive learning environment for practicing mathematics called "Math Garden" (www.oefenweb.com), used by over 180,000 primary school children in the Netherlands. This rich source of information has served as the ideal basis for numerous substantive and methodological papers. These topics range from replicating effects predicted by theoretical models about mathematics (van der Ven, Straatemeier, Jansen, Klinkenberg, & van der Maas, 2015), and cognitive strategies and models (Buwalda et al., 2016; Hofman, Visser, Jansen, Marsman, & van der Maas, 2018; Jansen & van der Maas, 2002), to longitudinal studies showing mutualism in mathematical abilities (Hofman, Kievit, Stevenson, Molenaar, & Van Der Maas, 2018; Ou et al., 2019) and the ideographic approach (Hofman, Jansen, de Mooij,

Stevenson, & van der Maas, 2018). In Math Garden, every student has his/her virtual garden, where each plant represents a game from a different domain, such as addition, multiplication or percentages. Student abilities and item difficulties are estimated—based on speed and accuracy—using the item response theory, where an Elo rating system adaptively matches students to items on-the-fly (for more detail, see Klinkenberg, Straatemeier, & Van Der Maas, 2011; Maris & van der Maas, 2012). Based on a student's current ability to estimate, the difficulty level is determined so that the student has a fixed probability of answering correctly to ensure students remain motivated (Jansen et al., 2013; Straatemeier, 2014). Children can choose between three difficulty levels (with expected probabilities correct of .6, .75 and .9).
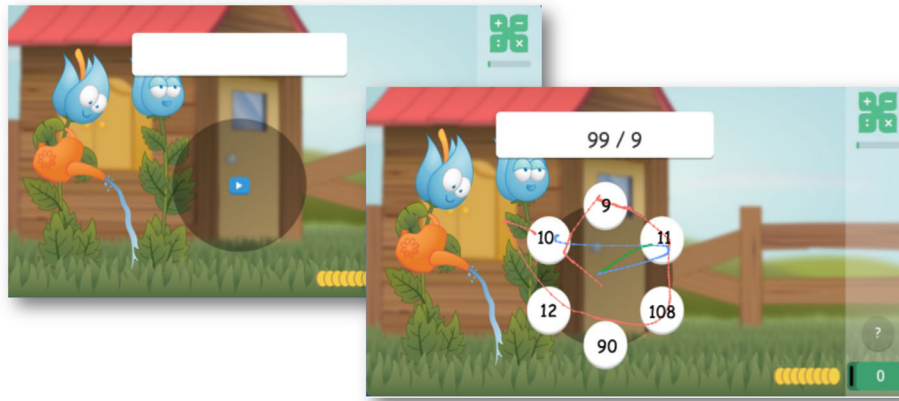
2.2.1 | Arithmetic speed mix game

Speed mix, one of the 24 games available in Math Garden, was used for the current study. In each game session, 10 problems with a mix of four different operations (i.e., addition, subtraction, multiplication and division) are presented. In this game, students are asked to click one of the six answer options within 8 s. The remaining time is visualised as virtual coins counting down on the bottom right of the screen (see Figure 1). The student is rewarded with the remaining coins on the screen after giving a correct response; an incorrect response leads to the remaining coins being subtracted from their total points. No coins are given or lost when failing to answer within the 8-second time limit. This way of scoring is known as the "High Speed High Stakes" rule, which has excellent psychometric qualities (Maris & van der Maas, 2012). This task was chosen for this study because (a) the students practice core arithmetic skills, basic tools essential for solving more complex maths problems; (b) it has a multiple-choice design instead of giving a response through a keypad so that the mouse trajectory towards the different response options can be investigated; (c) the students are under time pressure to answer (i.e., 8 s), which promotes movement of the mouse before reaching a decision (Kieslich & Henninger, 2017; Scherbaum & Kieslich, 2018).

The original speed mix game design was adapted for this mouse-tracking study. At the start of each problem, a start screen was added, with a blue button in the middle that needed to be clicked before the arithmetic problem and answer options are shown. This ensured that every mouse trajectory started in the same position and with equal distance from all the answer options (Figure 1). The answer options

TABLE 1 Distribution of age and gender in the sample

Age	Proportion of the sample (%)	Female %
5–7 years	3.3	43%
8 years	10.8	43.1
9 years	15.0	44.4
10 years	23.1	44.8
11 years	23.4	46.7
12 years	20.7	48.4
13 years	3.7	50.0

FIGURE 1 Design arithmetic speed mix task| Left: Start screen shown before every problem administration. Students needed to click the blue button in the middle to reveal the problem. Right: Example arithmetic problem: "99/9". The students clicked one of six answer options (40 pixels) including one correct and five alternative options. The answer options were placed randomly in one of six locations every time the problem was administered. The question mark on the bottom right can be clicked when the student does not know the answer to the question. In addition, three examples of mouse trajectories are shown [Colour figure can be viewed at wileyonlinelibrary.com]

for a given arithmetic problem were the same for every participant but were randomly placed across the six location boxes every time the problem was presented to a participant. A trial ended with the participant either clicking on one of the answer options, with a click on the question mark (Figure 1) to skip the trial (i.e., when the student does not know the answer) or when the time limit was exceeded.

The speed mix game contains over a thousand different problems. To reduce the server storage load and simplify the analyses, we selected 36 questions to be tracked. Nine problems were selected per operand use (i.e., addition, subtraction, multiplication and division). The problems were chosen based on the frequency of errors made for each problem and the problem difficulty, see Table 2. The problem difficulty was based on the average Elo rating (see Klinkenberg et al. 2011) averaged over a year of data collection within Math Garden before the start of this study. The frequency of errors was calculated ass the proportion of times a particular incorrect response was given to a problem over the course of the year before starting the study.

2.2.2 | Mouse tracking

A mouse-tracking tool was built on the website of Math Garden for this study in JavaScript. On every administered problem, the x- and y-coordinates of the mouse, the timestamp and the answer option locations were recorded at a sample rate of approximately 35 Hz (35 measurements per second).

2.3 | Procedure

For 6 weeks, the participants were tracked when playing the speed mix game. Students played in their natural environment independently, either in- or outside school, but mostly within school hours. No

teachers or parents were involved in this study. Schools and families with accounts are informed that Math Garden collects diverse categories of data, such as mouse-tracking data, using some for research analysis. Children (their parents or schools) can opt out of being part of the research done in the practice system and are, therefore, not included in this study. All data were anonymized before analysis. This procedure was approved by the university department's Ethics Review Board. Math Garden can be played on different kinds of devices, including touchscreens and tablets. Sessions tracked in this study that were played using touchscreen devices were excluded from analyses. Since Math Garden is a child-directed setting, children could decide for themselves whether they would play the speed mix game and for how long they would play. This means that during our data collection, not every child has seen (or answered) all the possible problems; some children may have only seen one question, and some children may have performed the same problem multiple times.

2.4 | Measures

2.4.1 | Response errors

To determine, in a stable and robust way, which errors are made most often, the responses administered for the last 2 years by all users in Math Garden were examined for the chosen 36 questions (see Table 2). From the $N = 607,125$ collected responses, 70% were answered correctly and less than 1% were not answered on time. In the remaining $N = 128,000$ the student chose one of the five alternative incorrect responses. The overall proportion of error occurrence per incorrect response was calculated first across all users in Math Garden, to compute associations to mouse attraction. For example, the incorrect answer, 512, was given 13.5% of the time for the problem, 5 + 508, that is, the most frequent error, see Table 2). For the

TABLE 2 Percentage correct and incorrect responses for some of the tracked arithmetic questions of all the users of Math Garden who have played in the last 2 years, as well as the corresponding difficulty level and overall error percentage

Arithmetic problems	Difficulty level [−15 15]	Frequency errors (percentage errors)	Correct answer	Error 1	Error 2	Error 3	Error 4	Error 5
5 + 508	0.8	93,406 (28%)	**513**	**512**	**503**	**515**	**523**	**558**
			78.4%	13.5%	3.3%	2.7%	1.4%	0.6%
7 + 80	−5.4	74,775 (26%)	**87**	**78**	**88**	**807**	**150**	**15**
			86.8%	10.3%	1.1%	0.9%	0.5%	0.4%
55 + 66	4.6	33,276 (27%)	**121**	**111**	**131**	**120**	**116**	**115**
			64.9%	28.2%	3.1%	1.5%	1.3%	1%
9−7	−5.9	89,467 (25%)	**2**	**3**	**1**	**4**	**16**	**0**
			84.5%	8.8%	2.6%	2.3%	1.2%	0.6%
14−5	−5.2	90,022 (27%)	**9**	**11**	**10**	**19**	**8**	**7**
			83.4%	9.4%	2.3%	2%	1.8%	1.1%
55−8	0.9	91,703 (26%)	**47**	**46**	**57**	**17**	**53**	**50**
			80.2%	10.9%	5.1%	1.3%	1.9%	0.6%
5 × 5	−6.4	69,766 (25%)	**25**	**10**	**35**	**20**	**30**	**24**
			82.9%	6.8%	4.1%	3.6%	2%	0.6%
7 × 9	−1.0	76,903 (29%)	**63**	**64**	**72**	**54**	**70**	**16**
			76.5%	10.2%	6.7%	5.1%	0.9%	0.6%
50 × 12	2.4	60,520 (26%)	**600**	**550**	**580**	**60**	**500**	**58**
			77.3%	9.6%	5%	4.1%	3.2%	0.8%
18/2	−1.8	78,647 (28%)	**9**	**6**	**8**	**7**	**10**	**20**
			71.7%	17.4%	8.2%	1.6%	0.6%	0.6%
36/4	−0.6	92,387 (27%)	**9**	**8**	**7**	**32**	**40**	**10**
			74.2%	16.3%	6.6%	1.2%	0.8%	0.8%
99/9	−2.7	83,306 (29%)	**11**	**10**	**9**	**90**	**12**	**108**
			67.9%	14%	11.1%	4.8%	1.3%	0.9%

Note: Highlighted in bold are the arithmetic questions and the corresponding answer option choices.

individual mapping analyses, the proportion of error occurrence per incorrect response was calculated for every individual student.

2.4.2 | Mouse movements

The number of trajectories collected per user in this study varied from one to 48, including repetitions of the same administered problems ($M = 2.89$, $SD = 3.82$). In six weeks, a total of 6,443 trajectories were collected. Firstly, trajectories with more than three mouse locations ($N = 5,653$) were selected, to remove trajectories with minimal movement. Secondly, trajectories with uncontrolled movements, (i.e., mouse locations 150 pixels away from the answer options or question box) were excluded (remaining $N = 5,269$). Thirdly, only the mouse trajectories of the correctly answered problems were analysed, $N = 3,906$ trajectories ($N = 1,590$ users). The reason for this was twofold: (a) The correctly answered problems are in principle not informative of errors when only analysing reaction time and/or accuracy on a behavioural level; (b) incorrectly answered

problems contain actual mouse clicks within alternative answer options.

The main analysis was focused on whether there were mouse movements towards alternative answer options before the right answer was chosen. The trajectories were first smoothed with a 10-point smoothing window using the Savitzky–Golay filter from the "trajr" package (McLean & Skowron Volponi, 2018) in R (Team, 2013), which is particularly suitable to preserve the shape of the trajectory while removing high-frequency squiggles.

To analyse a complex design with five alternative options, we combined two measures to detect mouse attraction per trajectory. Firstly, a static method, where simply the number of points inside each answer box (radius of 40px) is counted, in line with how eye movements are typically classified (Schulte-Mecklenbeck et al., 2019), see Figure 2a (or Appendix 1.1 in Data S1 for more details). Secondly, a dynamic method, where the mouse directions are classified as movement towards an answer box, from one mouse location to the other, inspired from animal movement analyses (see for example Michelot, Langrock, & Patterson, 2017), and visualised in Figure 2b (or Appendix 1.1 in Data

FIGURE 2 Two measures of mouse attraction. In the static method (a) the mouse locations in the answer option boxes (40 px) are counted, that is, three mouse locations in answer box. 15. Dynamic mouse method (b) counts the mouse movements towards the answer options, that is, 4 mouse moves towards answer 15 [Colour figure can be viewed at wileyonlinelibrary.com]

(a)
Static mouse method

3 mouse points in answer 15
5 mouse points in answer 16

(b)
Dynamic mouse method

4 mouse moves towards answer 15
4 mouse moves towards answer 16
2 mouse moves towards answer 14
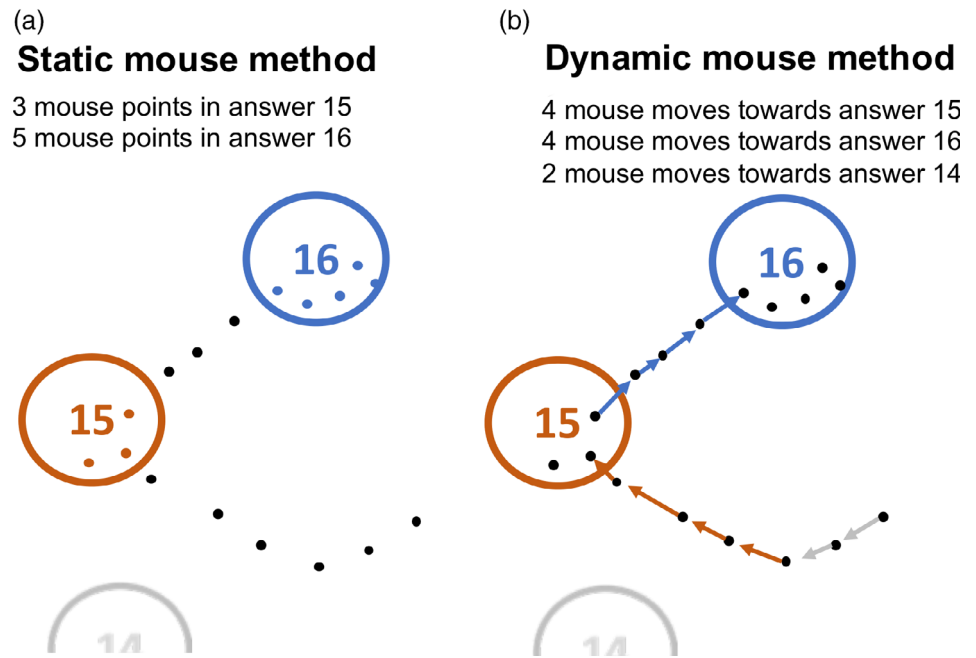
S1 for more details). The advantage of adding the dynamic method is that there is still detection of movement towards an attractor, regardless of where the mouse is located at that point. For every trajectory, the findings in these methods were combined to ensure that the competition between the choices can be analysed from the start of the trajectory to the end. This was done by adding the number of movements from the dynamic method that was not inside an answer box, with the mouse locations obtained from the static method. The mouse locations and movements that could not be associated with any of the incorrect answer options were excluded from the analyses.

3 | RESULTS

3.1 | Response accuracy

Of all responses, 74% were answered correctly. This meant that 26% of the trials and their mouse trajectories could be examined. The students answered correctly on average in 4.1 s (SD = 1.6 s), where the maximum time to respond is 8 s.

3.2 | Mouse trajectories

Per trajectory, M = 115.9 mouse locations (SD = 107.6) were collected, that is, around 35 measurements per second. Of all mouse locations, 22% were associated with the correct answer option—14% was inside the box and 7% was directed towards the option; 47% could not be directly associated with any of the answer options and 31% were associated with one of the five incorrect answer options (16% inside and 15% towards the answer option box). Of this 31%, the proportion of mouse locations associated with each incorrect answer option was

averaged across the participants for each arithmetic problem. An example of three different types of trajectories from different children for the same question can be found in Figure 1: Movement straight towards the error response (green line); movement towards alternative option (blue line); a clockwise movement (red line).

3.3 | Mouse attraction to incorrect responses on group level

Firstly, we investigated whether frequent response errors made in the past (i.e., in the past 2 years by all users of Math Garden) would also attract the greatest number of mouse movements in this study. Therefore, a Pearson correlation between the response error rate and the number of mouse movements was calculated for each arithmetic problem (Figure 3a). These Pearson r correlations were transformed using a Fisher Z transformation ($z = 0.5 \times \ln ((1 + r)/(1 - r))$). Findings revealed that the correlations ranged across the arithmetic problems from .45 to .70 with an average of $M = .66$ ($\bar{r} = .57$), which was significantly different from zero, $t(35) = 42.33$, $p < .001$. These high correlations mean that there were more mouse movements towards the errors frequently made in the Math Garden sample, than towards less frequent errors, in line with what was hypothesized. A linear regression showed that these correlations were a function of the problem difficulty (based on the estimated Elo rating, see Klinkenberg et al., 2011) where a higher correlation between the error rate and mouse movements was found when the problems were relatively easier, $B = -0.56$, $t(34) = 3.54$, $p = .001$ (Figure 3c). The average age of the users also associated positively with the size of the correlation, $B = -0.61$, $t(34) = 4.47$, $p < .001$ (Figure 3b). This suggests that mouse movements mapped better on to the underlying difficulties of younger children, who, on average, play somewhat easier problems.
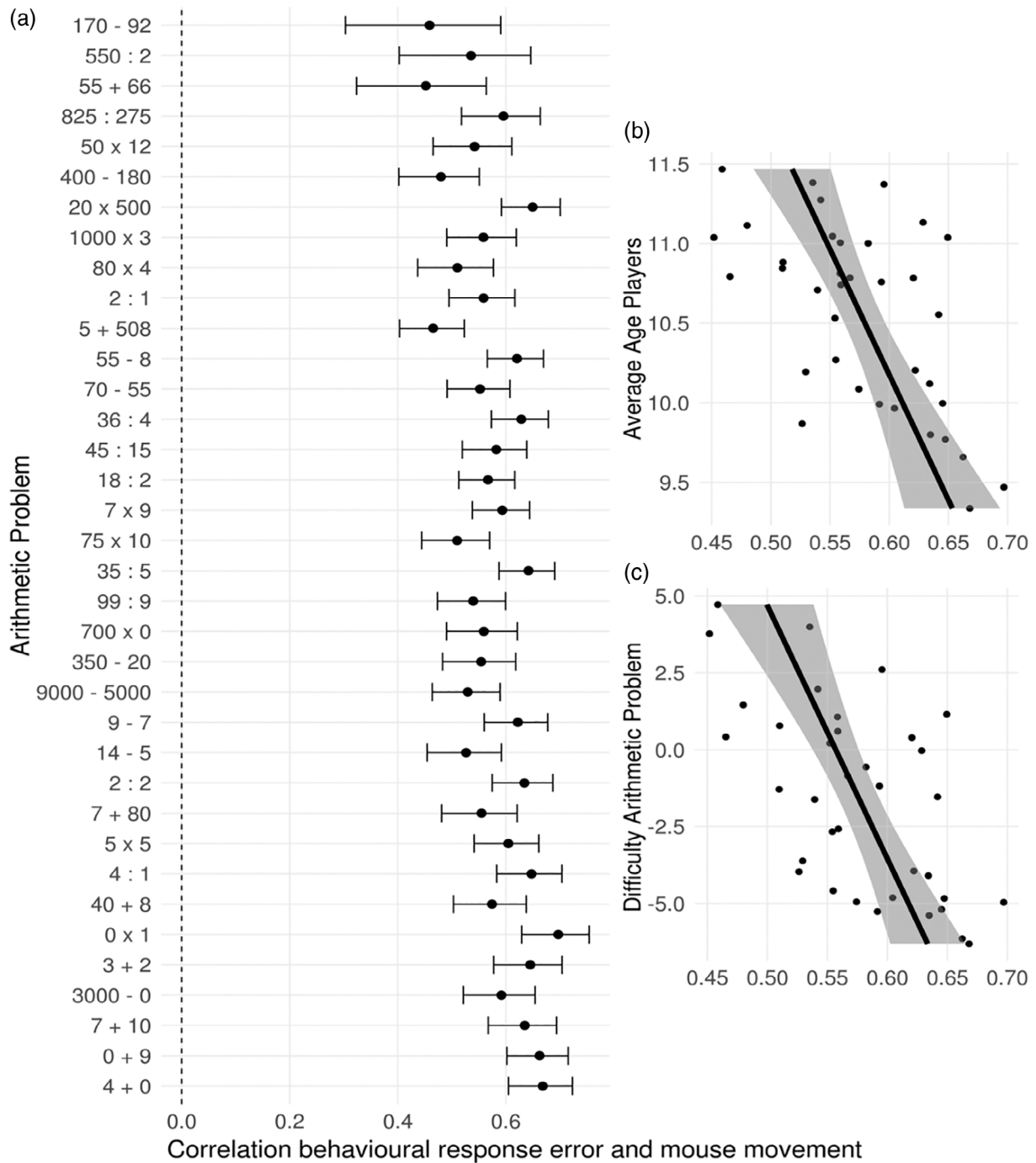
FIGURE 3 Correlations between the proportion mouse movements towards the incorrect answer options in correctly answered trials and the corresponding error rate calculated based on all Math Garden users: (a) per arithmetic question (ordered from high to low problem difficulty). The error bars show the 95% confidence interval of the correlation test. (b) Linear regression line showing that the average age of the players associated negatively with the size of the correlation; (c) Linear regression line showing that problem difficulty also associated negatively with the size of the correlation. The 95% confidence intervals are displayed as the shaded area around the mean for both lines

3.4 | Individual mapping of mouse movement on errors made in the past

In the analyses above, error response rates were calculated based on all Math Garden users, to ensure reliable error rate rankings. Next, we wanted to test whether individuals who made certain response errors in the past would also show mouse trajectories towards these specific incorrect responses, even when the final response was correct.

To investigate this, a Pearson correlation per user, averaged over all problems, between their mouse movement and past error response rate was calculated. Correlations with missing values (either no history of previously made errors or mouse trajectories) were removed, as well as users that contained a correlation for only one arithmetic

problem. The Fisher Z transformed Pearson r correlations for all individuals (N = 101) were analysed with a t test (Figure 4). This test showed that the average correlation was significantly higher than zero, Fisher Z transformed M = .42 ($\bar{r}$ = .23), t(100) = 4.23, p < .001. This means that users showed more mouse movement towards their previously made response errors than to other incorrect options.

Furthermore, some checks were made to ensure that this finding was robust. Firstly, the average correlation per user over all problems was weighted by how many errors the person had made in their history of playing in Math Garden before data collection, assuming that problems with more error responses are more likely to reflect a consistent conceptual or procedural difficulty (e.g., a tendency to confuse multiplication and addition; see Appendix 2.1 in Data S1). The second check was to weight the correlation by how many mouse trajectories were collected from the user for a particular problem (Appendix 2.1 in Data S1). When a user had not given an incorrect response to a particular problem within a year from when the mouse trajectory was registered, the data were excluded since the player had presumably mastered the problem and it presented no difficulty. The checks show stable correlations irrespective of how much data were collected and when the errors were made.

4 | DISCUSSION

The aim of this study was to investigate whether it is possible to detect difficulties children encounter when solving arithmetic problems, without relying on the errors themselves. This was done by tracking mouse movements, a method intended to measure the competition between responses during the process of problem-solving. This study is, to our knowledge, the first to implement mouse tracking outside the lab in an online adaptive learning system. We analysed the mouse movements towards multiple attractors, which were the incorrect response options of multiple-choice arithmetic problems. Our findings showed that, even when the final response was correct, the mouse trajectories revealed attraction towards the errors most frequently made by students playing in Math Garden. All ages and arithmetic problems showed this high attraction, but the mapping of the mouse was stronger for younger children and easy arithmetic problems. Furthermore, we found that individuals who had made certain errors in the past (past 2 years before data collection) still showed more mouse movement towards these errors than to other attractors.

Thus, implementing mouse tracking in an online learning environment allows us to study whether students have systematic reasoning problems in solving arithmetic problems. This will enable us to give targeted feedback and instruction on the learning process, after both an incorrect and correct response. To give an example: if the method detects that a student is consistently attracted to an answer corresponding to mistaking addition for multiplication, it would greatly benefit the learning process to present a reminder to carefully check the operand, and a series of problems where the operations are constantly mixed, to practice switching between operations.

The strength of investigating the use of mouse tracking in an online learning environment is that the sample size is both large and heterogeneous in terms of math ability, age and background.
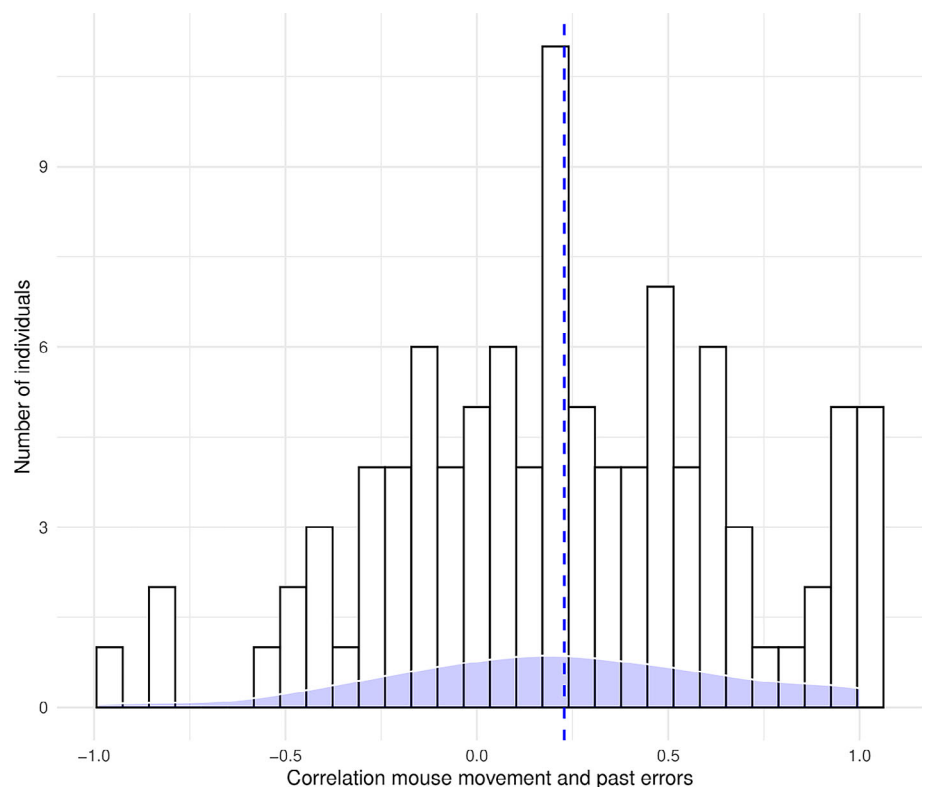
FIGURE 4 Histogram showing the individual correlations between mouse movement and response errors made in the past. The blue dashed line shows the average correlation (r = .23) [Colour figure can be viewed at wileyonlinelibrary.com]

Secondly, the mouse movements can be tracked for all its users daily for thousands of different maths problems at the same time throughout their whole primary school trajectory. Thirdly, the students learn these skills in their natural environment, either at school or at home. This limitless access to Math Garden for students in diverse circumstances ensures that our results are robust and would not only be reproducible in a lab-controlled experiment.

Other studies have used the large-scale Math Garden database to categorise and detect systematic errors (Savi et al., 2018; Straatemeier, 2014), but they had to rely on the incorrect responses a child picked when they made an error. Since online adaptive environments require a high success rate for motivational reasons, errors are rare; in this study (N = 6,443) an incorrect response was given for 20% of the problems administered. Our findings show that mouse movement can also reveal information about these errors in the other 80% of the responses, which would help with the diagnosis of systematic difficulties at an individual level.

There are some limitations to this study. First, collecting data in a naturalistic setting, such as Math Garden, can cause the data to be noisy. There is no way to control the circumstances within which the child is performing the task. For example, they can be distracted while playing the game or have a bad internet connection. Equipment is not standardized since the mouse used is different for every school and home. It is, therefore, necessary to collect a large dataset, such as in this study, to be very strict in terms of removing noisy mouse trajectories and unfinished game sessions. Second, many children practice their skills in online learning environments on tablets and touchscreens. Mouse tracking cannot be used for such devices. A third limitation is that mouse-tracking studies are bound to a specific design, and not applicable to all games and experimental tasks. Every trial needs to have the same starting point and multiple-choice options located at an equal distance from the starting point. Finally, mouse tracking is difficult to combine and/or compare with other process-tracing methods such as eye tracking in large online studies. Some laboratory studies have compared these two modalities and found sufficient correlation between gaze and cursor positions, but there is also a substantial variation (Franco-Watkins & Johnson, 2011; Huang & White, 2012; Lohse & Johnson, 1996; Quétard et al., 2016). For example, with prolonged cursor fixation, the attention and gaze might be somewhere else on or off the screen, which would not be picked up by the mouse metrics. Unfortunately, eye tracking is currently very complicated to implement in an online learning environment (partly due to privacy issues).

For further research, it would be interesting to investigate individual cases with both eye and mouse tracking in a laboratory setting to see how these methods interact and complement each other with regard to signalling the attractiveness of error options in math questions.

CONFLICT OF INTEREST

Han L. J. van der Maas is a full professor of Psychological Methods at the University of Amsterdam and founder of Oefenweb, the company that operates Math Garden. However, this study does not aim to demonstrate the effectiveness of training in Math Garden.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/jcal.12483.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Susanne M. M. de Mooij 🔟 https://orcid.org/0000-0001-8928-0851

REFERENCES

Artemenko, C., Soltanlou, M., Bieck, S. M., Ehlis, A. C., Dresler, T., & Nuerk, H. C. (2019). Individual differences in math ability determine neurocognitive processing of arithmetic complexity: A combined fNIRS-EEG study. *Frontiers in Human Neuroscience*, *13*(July), 1–13. https://doi.org/10.3389/fnhum.2019.00227

Ben-Zeev, T. (1998). Rational errors and the mathematical mind. *Review of General Psychology*, *2*(4), 366–383. https://doi.org/10.1037/1089-2680.2.4.366

Bray, B., & Mcclaskey, K. (2010). Personalization vs differentiation vs individualization. Retrieved from http://www.my-ecoach.com/online/resources/925/PersonalizationvsDifferentiationvsIndividualization.pdf

Brown, J. S., & Burton, R. R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science*, *2*(2), 155–192. https://doi.org/10.1016/S0364-0213(78)80004-4

Buwalda, T. A., Borst, J. P., van der Maas, H., & Taatgen, N. A. (2016). *Explaining mistakes in single digit multiplication*: *A cognitive model*. Proceedings of the 14th international conference on cognitive modeling, Volume 1. pp. 131–136.

Campbell, J. I. D. (1987). Network interference and mental multiplication. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(1), 109–123. https://doi.org/10.1037/0278-7393.13.1.109

Cloutier, J., Freeman, J. B., & Ambady, N. (2014). Investigating the early stages of person perception: The asymmetry of social categorization by sex vs. age. *PLoS One*, *9*(1), e84677. https://doi.org/10.1371/journal.pone.0084677

Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, *35*(1), 15–28.

de Mooij, S. M. M., Kirkham, N. Z., Raijmakers, M. E. J., van der Maas, H. L. J., & Dumontheil, I. (2020). Should online math learning environments be tailored to individuals' cognitive profiles? *Journal of Experimental Child Psychology*, *191*, 104730. https://doi.org/10.1016/j.jecp.2019.104730

Domahs, F., Delazer, M., & Nuerk, H. C. (2006). What makes multiplication facts difficult: Problem size or neighborhood consistency? *Experimental Psychology*, *53*(4), 275–282. https://doi.org/10.1027/1618-3169.53.4.275

Dotan, D., & Dehaene, S. (2013). How do we convert a number into a finger trajectory? *Cognition*, *129*(3), 512–529. https://doi.org/10.1016/j.cognition.2013.07.007

Erb, C. D. (2018). The developing mind in action: Measuring manual dynamics in childhood. *Journal of Cognition and Development*, *19*(3), 233–247. https://doi.org/10.1080/15248372.2018.1454449

Faulkenberry, T. J. (2014). Hand movements reflect competitive processing in numerical cognition. *Canadian Journal of Experimental Psychology*, 68(3), 147–151. https://doi.org/10.1037/cep0000021

Faulkenberry, T. J., Cruise, A., Lavro, D., & Shaki, S. (2016). Response trajectories capture the continuous dynamics of the size congruity effect ★. *Acta Psychologica*, 163, 114–123. https://doi.org/10.1016/j.actpsy.2015.11.010

Faulkenberry, T. J., Witte, M., & Hartmann, M. (2018). Tracking the continuous dynamics of numerical processing: A brief review and editorial.

Federico, P. A. (2000). Learning styles and student attitudes toward various aspects of network-based instruction. *Computers in Human Behavior*, 16(4), 359–379. https://doi.org/10.1016/S0747-5632(00)00021-2

Fischer, M. H., & Hartmann, M. (2014). Pushing forward in embodied cognition: May we mouse the mathematical mind? *Frontiers in Psychology*, 5, 1–4. https://doi.org/10.3389/fpsyg.2014.01315

Franco-Watkins, A. M., & Johnson, J. G. (2011). Applying the decision moving window to risky choice: Comparison of eye-tracking and mouse-tracing methods. *Judgment and Decision Making*, 6(8), 740–749. https://doi.org/10.1016/B978-012369442-3/50198-7

Freeman, J. B. (2018). Doing psychological science by hand. *Current Directions in Psychological Science*, 27(5), 315–323. https://doi.org/10.1177/0963721417746793

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods*, 42(1), 226–241. https://doi.org/10.3758/BRM.42.1.226

Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in motion reveals mind in motion, 2, 1–6. https://doi.org/10.3389/fpsyg.2011.00059

Guo, Q., & Agichtein, E. (2010). *Towards predicting web searcher gaze position from mouse movements*. Conference on Human Factors in Computing Systems – Proceedings. pp. 3601–3606. https://doi.org/10.1145/1753846.1754025

Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes and Intergroup Relations*, 18(3), 384–401. https://doi.org/10.1177/1368430214538325

Hinault, T., & Lemaire, P. (2016). What does EEG tell us about arithmetic strategies? A review. *International Journal of Psychophysiology*, 106, 115–126. https://doi.org/10.1016/j.ijpsycho.2016.05.006

Hofman, A. D., Jansen, B., de Mooij, S., Stevenson, C., & van der Maas, H. (2018). A solution to the measurement problem in the idiographic approach using computer adaptive practicing. *Journal of Intelligence*, 6(1), 14. https://doi.org/10.3390/jintelligence6010014

Hofman, A. D., Kievit, R. A., Stevenson, C. E., Molenaar, D., & Maas, H. L. J. Van Der. (2018). The dynamics of the development of mathematics skills: A comparison of theories of developing intelligence.

Hofman, A. D., Visser, I., Jansen, B. R. J., Marsman, M., & van der Maas, H. L. J. (2018). Fast and slow strategies in multiplication. *Learning and Individual Differences*, 68(406), 30–40. https://doi.org/10.1016/j.lindif.2018.09.007

Huang, J., & White, R. (2012). *User see, user point*: Gaze and cursor alignment in web search. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1341–1350. https://doi.org/10.1145/2207676.2208591

Huebner, M. G., & LeFevre, J. A. (2018). Selection of procedures in mental subtraction: Use of eye movements as a window on arithmetic processing. *Canadian Journal of Experimental Psychology*, 72(3), 171–182. https://doi.org/10.1037/cep0000127

Jansen, B. R. J., Louwerse, J., Straatemeier, M., Van der Ven, S. H. G., Klinkenberg, S., & Van der Maas, H. L. J. (2013). The influence of experiencing success in math on math anxiety, perceived math competence, and math performance. *Learning and Individual Differences*, 24, 190–197. https://doi.org/10.1016/j.lindif.2012.12.014

Jansen, B. R. J., & van der Maas, H. L. J. (2002). The development of children's rule use on the balance scale task. *Journal of Experimental Child Psychology*, 81(4), 383–416. https://doi.org/10.1006/jecp.2002.2664

Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods*, 49(5), 1652–1667. https://doi.org/10.3758/s13428-017-0900-z

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

Koop, G. J., & Johnson, J. G. (2013). The response dynamics of preferential choice. *Cognitive Psychology*, 67(4), 151–185. https://doi.org/10.1016/j.cogpsych.2013.09.001

Lai, M. L., Tsai, M. J., Yang, F. Y., Hsu, C. Y., Liu, T. C., Lee, S. W. Y., ... Tsai, C. C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review*, 10(88), 90–115. https://doi.org/10.1016/j.edurev.2013.10.001

Lohse, G. L., & Johnson, E. J. (1996). A comparison of two process tracing methods for choice tasks. *Organizational Behavior and Human Decision Processes*, 68(1), 28–43. https://doi.org/10.1006/obhd.1996.0087

Marghetis, T., Núñez, R., & Bergen, B. K. (2014). Doing arithmetic by hand: Hand movements during exact arithmetic reveal systematic, dynamic spatial processing. *Quarterly Journal of Experimental Psychology*, 67(8), 1579–1596. https://doi.org/10.1080/17470218.2014.897359

Maris, G., & van der Maas, H. L. J. (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

McLean, D. J., & Skowron Volponi, M. A. (2018). Trajr: An R package for characterisation of animal trajectories. *Ethology*, 124(6), 440–448. https://doi.org/10.1111/eth.12739

Michelot, T., Langrock, R., & Patterson, T. (2017). moveHMM an R package for the analysis of animal movement data. *Methods in Ecology and Evolution*, 7, 1308–1325. https://doi.org/10.1111/2041-210x.12578

Ou, L., Hofman, A. D., Simmering, V. R., Bechger, T., Maris, G., & van der Maas, H. L. J. (2019). *Modeling person-specific development of math skills in continuous time*: New evidence for mutualism. The 12th international conference on educational data mining. pp. 372–377.

Quétard, B., Quinton, J. C., Mermillod, M., Barca, L., Pezzulo, G., Colomb, M., & Izaute, M. (2016). Differential effects of visual uncertainty and contextual guidance on perceptual decisions: Evidence from eye and mouse tracking in visual search. *Journal of Vision*, 16(11), 28. https://doi.org/10.1167/16.11.28

Santens, S., Goossens, S., & Verguts, T. (2011). Distance in motion: Response trajectories reveal the dynamics of number comparison. *PLoS One*, 6(9), e25429. https://doi.org/10.1371/journal.pone.0025429

Savi, A. O., Deonovic, B. E., Bolsinova, M., Maas, H. L. J. Van Der, & Maris, G. K. J. (2018). Automated diagnosis of misconceptions in single digit multiplication. https://doi.org/10.31234/osf.io/53muj

Scherbaum, S., & Kieslich, P. J. (2018). Stuck at the starting line: How the starting procedure influences mouse-tracking data. *Behavior Research Methods*, 50(5), 2097–2110. https://doi.org/10.3758/s13428-017-0977-4

Schulte-Mecklenbeck, M., Kuehberger, A., & Johnson, J. G. (2019). *A handbook of process tracing methods* (2nd ed.). New York, NY: Routledge. https://doi.org/10.4324/9781315160559

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General*, 117(3), 258–275. https://doi.org/10.1037/0096-3445.117.3.258

Song, J. H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends in Cognitive Sciences.*, 13, 360–366. https://doi.org/10.1016/j.tics.2009.04.009

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, *102*(29), 10393–10398. https://doi.org/10.1073/pnas.0503903102

Spivey, M. J. (2007). *The continuity of the mind*. New York, NY: Oxford University Press. https://doi.org/10.1093/acprof

Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science*, *15*(5), 207–211. https://doi.org/10.1111/j.1467-8721.2006.00437.x

Spüler, M., Walter, C., Rosenstiel, W., Gerjets, P., Moeller, K., & Klein, E. (2016). EEG-based prediction of cognitive workload induced by arithmetic: A step towards online adaptation in numerical learning. *ZDM-Mathematics Education*, *48*(3), 267–278. https://doi.org/10.1007/s11858-015-0754-8

Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How mouse-tracking can advance social cognitive theory. *Trends in Cognitive Sciences*, *22*(6), 531–543. https://doi.org/10.1016/j.tics.2018.03.012

Straatemeier, M. (2014). Math Garden: A new educational and scientific instrument. Retrieved from http://hdl.handle.net/11245/1.417091

Team, R. C. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

van der Ven, S. H. G., Straatemeier, M., Jansen, B. R. J., Klinkenberg, S., & van der Maas, H. L. J. (2015). Learning multiplication: An integrated analysis of the multiplication ability of primary school children and the difficulty of single digit and multidigit multiplication problems. *Learning and Individual Differences*, *43*, 48–62. https://doi.org/10.1016/j.lindif.2015.08.013

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.