Manipulating Interface Design Features Affects Children's Stop-and-Think Behaviours in a Counterintuitive-Problem Game

Running Head: HCI affects stopping-and-thinking behaviours

Andrea Gauthier*

UCL Knowledge Lab, Institute of Education, University College London, andrea.gauthier@ucl.ac.uk

Kaska Porayska-Pomsta

UCL Knowledge Lab, Institute of Education, University College London, k.porayska-pomsta@ucl.ac.uk

Iroise Dumontheil

Centre for Brain and Cognitive Development, Department of Psychological Sciences, Birkbeck College – University of London, i.dumontheil@bbk.ac.uk

Sveta Mayer

Institute of Education, University College London, s.mayer@ucl.ac.uk

Denis Mareschal

Centre for Brain and Cognitive Development, Department of Psychological Sciences, Birkbeck College – University of London, d.mareschal@bbk.ac.uk

The human-computer interaction (HCI) design of educational technologies influences cognitive behaviour, so it is imperative to assess how different HCI strategies support intended behaviour. We developed a neuroscience-inspired game that trains children's use of "stopping-and-thinking" (S&T)—an inhibitory control-related behaviour—in the context of counterintuitive science problems and tested the efficacy of four HCI features in supporting S&T: (1) a readiness mechanic, (2) motion cues, (3) colour cues, and (4) rewards/penalties. In a randomised eye-tracking trial with 45 7-to-8-year-olds, we found that the readiness mechanic increased S&T duration, that motion and colour cues proved equally effective at promoting S&T, that combining symbolic colour with the readiness mechanic may have a cumulative effect, and that rewards/penalties may have distracted children from S&T. Additionally, S&T duration was related to in-game performance. Our results underscore the importance of interdisciplinary approaches to educational technology research that actively investigates how HCI impacts intended learning behaviours.

• Human-centered computing ~ Interaction design ~ Empirical studies in interaction design • Human-centered computing ~ Human computer interaction (HCI) ~ Empirical studies in HCI • Human-centered computing ~ Human computer interaction (HCI) ~ HCI design and evaluation methods ~ User studies

* Corresponding author.

Additional Keywords and Phrases: human-computer interaction design, visual cues, inhibitory control, primary education, game-based learning

ACM Reference Format:

First Author's Name, Initials, and Last Name, Second Author's Name, Initials, and Last Name, and Third Author's Name, Initials, and Last Name. 2018. The Title of the Paper: ACM Conference Proceedings Manuscript Submission Template: This is the subtitle of the paper, this document both explains and embodies the submission format for authors using Word. In Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY. ACM, New York, NY, USA, 10 pages. NOTE: This block will be automatically generated when manuscripts are processed after acceptance.

1 INTRODUCTION

Computer-based learning activities offer varied approaches to support more inclusive self-regulated learning within and outside of the classroom [1]. They do this by adapting to the needs of individual learners, which can motivate children to engage—and stay engaged—within the learning environment [19,43,55], and by training skills relevant to self-regulated learning, e.g., [7]. When learning, self-regulated learners take control of their thoughts and actions, display metacognitive awareness of their learning strategies, have higher self-efficacy, and are more likely to achieve academically than their peers [7,60]. As such, promoting self-regulated learning has long been a focus in educational research. With more emphasis being placed on personalised, self-regulated, learning than ever before (e.g., with the abrupt rise in home-schooling provoked by the COVID-19 pandemic [75]), there is demand for more research into the application of (i) educational neuroscience insights to promote self-regulated learning, (ii) human-computer interaction (HCI) design strategies to effectively guide computer-based activities based on educational neuroscience, and (iii) artificial intelligence in education approaches to tailor these technologies to individual learners.

This paper presents work conducted as part of an educational neuroscience-focused project called UnLocke. UnLocke hypothesised that training primary school children to apply stopping-and-thinking (S&T) skills—a form of inhibitory control—when solving mathematics and science problems would promote counterintuitive reasoning and academic achievement, by enabling children to suppress (or inhibit) their intuitive but incorrect beliefs and misconceptions [57]. To this end, we developed a trivia game-based environment (called *Stop & Think*), for 7- to 10-year-olds, to train the application of S&T behaviours—a skill directly linked to self-regulation competencies [77]—when engaged in counterintuitive problem solving in science and maths [77]. As part of an Education Endowment Foundation randomised control trial efficacy evaluation [68], a teacher-led version of this non-adaptive software was assessed in a trial with 6,672 children from 89 schools across England. Children played *Stop & Think* as a whole class for a target of 12 minutes, 3 times per week, for 10 weeks, with the activity projected on an interactive whiteboard at the front of the room and with the usual class teacher acting as facilitator. Our research revealed that training children's use of within-domain S&T behaviours in this way leads to improved counterintuitive reasoning skills [77] and increased scores on standardised math and science tests [68,77].

Due to an increasing emphasis being placed on personalised, technology-mediated, learning strategies, the aim of the current study was to understand how to modify the existing software for use in an individualised context, where children play on the computer by themselves and the activity is not supported by a teacher. More specifically, we manipulated the HCI characteristics of the software and assessed which characteristics best fostered inhibitory control-related S&T behaviours in children, using behavioural and eye-tracking measures.

1.1 Inhibitory control and the *Stop & Think* game

Our brains use two distinct ways of reasoning that compete with each other [25,40,47]: (1) the *heuristic system*, that enables intuitive, quick decision-making in familiar situations; and (2) *the analytic system*, that operates slowly, sequentially, and logically, allowing for abstract reasoning. This second system acts to suppress the quick, intuitive decision-making of the heuristic system when we are engaged in logical tasks, via a subset of *inhibitory control* skills. Inhibitory control is a type of cognitive skill, known as an "executive function", that is foundational to self-regulated learning and life-long academic achievement [4,42,57,80]. Some forms of inhibitory control, specifically perceptual interference control, prior knowledge inhibition, and response inhibition, have also been found to be important for maths and science learning, e.g., [32], where inhibition of pre-existing beliefs or superficial perceptions is necessary for learning and applying new and counterintuitive knowledge [22,56,67,71]. This is because perceptual interference (e.g., the Earth is flat, because it looks flat [5]) and prior knowledge interference (e.g., 5 > 3, therefore -5 > -3 [13,37]) remain even after the child has learnt the new concepts, and are often recalled more quickly by the brain's heuristic system than the correct (counterintuitive) one. Furthermore, response inhibition also plays a role by allowing children to suppress motor responses and, thus, not give in immediately to perceptual and prior knowledge interference. As such, exercising "stopping and thinking" via the analytic system allows the child to engage perceptual and/or prior knowledge inhibition, giving them time to think about the correct concept.

The UnLocke project chose a game-based learning approach to capitalise on motivational aspects of games for young learners [30,63]. Games can be particularly useful in motivating prolonged engagement in repetitive tasks [27], such as those found in academic skills training, e.g., [26,49]. Meta-analyses have also consistently shown game-based learning interventions to have small to moderate advantages on cognitive outcomes over conventional learning approaches (e.g., lectures, drill-and-practice; [20,70,78]).

We developed a trivia-genre game-based learning environment, called *Stop & Think*, for 7- to 10-year-old children, in which they were given structured opportunities to repeatedly exercise S&T behaviours when solving counterintuitive maths and science problems [77]. The in-game concepts and activities were aligned with the National Curriculum for Years 3 and 5 in England. The game had two modes: (1) a TV trivia game-show mode and (2) a task mode that actively trained children to apply S&T behaviours. In the game-show mode, children were introduced vicariously to the in-game tasks and common maths and science misconceptions. The show-host character introduced quiz questions (**Figure 1A**), while three contestant characters indicated their readiness to respond by pressing buzzers. The reasoning behind contestants' answers was articulated through speech bubbles to scaffold the player's learning (**Figure 1C-F**). In the task mode (**Figure 1B**), the focus was on training children to apply within-domain S&T skills by having them pause for a few seconds (henceforth—the *S&T mechanic*) before allowing them to answer each question, to encourage them to suppress intuitive thinking and adopt analytic counterintuitive reasoning. An icon (the S&T icon, seen in **Figure 1B**) pulsated in the bottom left-hand corner of the screen to indicate to the child when they should be stopping-and-thinking before responding to a problem. *Stop & Think* offered 30 sessions in total, each comprising a maths component and a science component (presented in a random order), and wherein each component contained six problems centred around the same counterintuitive topic. Of these six problems, the first was "exploratory", allowing the user to submit up to four incorrect responses and providing increasingly more supportive feedback with each incorrect attempt. The last five problems were presented during a "repeated practice" phase, which allowed children to practice their S&T skills in more problems around the same topic (**Supplementary Material Figures S1**). Each of these practice problems allowed two attempts before providing the correct answer and moving onto the next problem. Each *Stop & Think* session was timed

for 12 minutes (six minutes in maths and six in science), so the number of problems that were attempted depended on the speed and accuracy of the player. In this experiment we focus upon studying the *Stop & Think* science component.

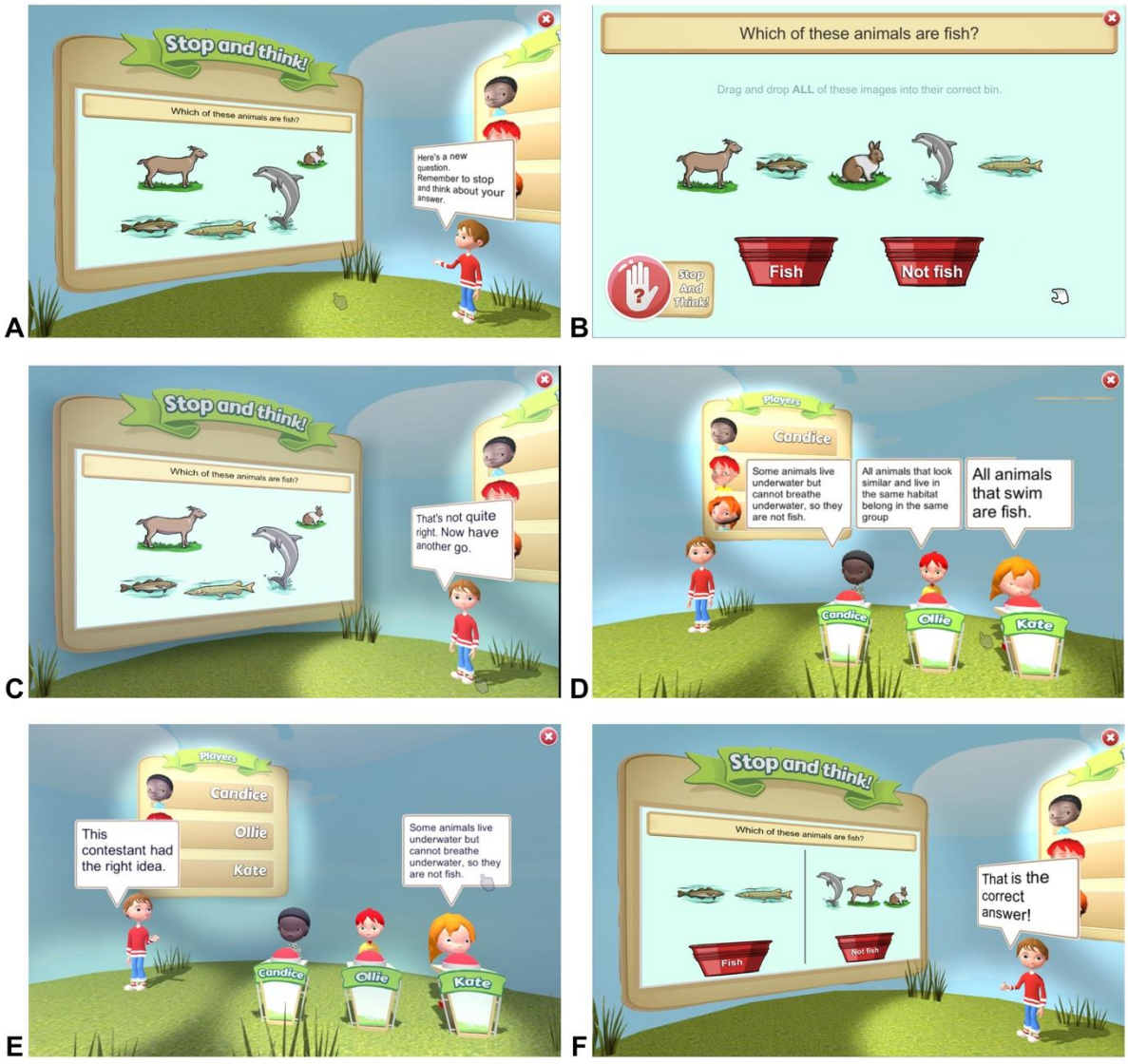

Figure 1. Overview of current non-adaptive support provided to *Stop & Think* players. (A) Game-show mode – show-host character introduces a new problem; (B) Task mode – player completes the counterintuitive problem; (C) Game-show host encourages the player to try again upon the first incorrect attempt; (D) Upon a second incorrect attempt, the player is shown the reasoning of other

contestants (one of whom has the correct idea) before trying again; (E) if the player is again incorrect, they are shown who had the correct idea so that this idea could then be applied; (F) Feedback to a correct attempt.

1.2 HCI features that support individualised training

The impact of specific HCI and design decisions on game-based learning outcomes is rarely investigated [14,20,28,29]. There is a myriad of ways in which HCI design influences interactions and training in individualised computer- and game-based learning environments, from the aesthetic of the graphics, usability of the interface, to the organisation and flow of information on the screen [62]. However, because we were designing for (a) small children, (b) S&T training, and (c) use in schools within a limited timeframe, we chose to focus on the following four specific features: (i) a readiness mechanic, (ii) pre-attentive visual cues, including motion and colour, and (iii) a reward and penalty system, as a game-specific feature. These features were identified as potentially impactful through a co-design workshop—involving developmental neuroscientists, artificial intelligence and HCI researchers, game designers, and educators—to investigate ways in which the HCI design of our computer-based intervention could be enhanced to promote the application of S&T behaviours. Below, we describe how each feature might impact S&T training in this context.

1.2.1 Readiness indication as a mandatory interaction

In a typical classroom, children raise their hands to indicate their readiness to provide answers. This type of behaviour is not commonplace in computer game-based activities, where there is an emphasis on play, exploration, and experimentation [30]. However, when exercising S&T behaviour, a similar mechanic to hand-raising could encourage metacognitive awareness and monitoring, allowing children to state their intentionality, or readiness to answer, after giving them time to stop and think before answering. This hypothesis is in line with the theory of planned behaviour, where goals are accomplished by a series of more-or-less well-planned actions that are, in turn, controlled by goal-seekers' intentions [2,3]. Hand-raising may play a role in such planning, and a mandatory interaction mechanic that amplifies such intentional behaviour may provide an important means for S&T training.

1.2.2 Pre-attentive cues: motion and colour

Pre-attentive—or perceptually salient—features are processed by our visual systems before conscious attention, thereby effectively guiding our gaze to areas of interest [76]. Our visual sensory systems are particularly attuned to motion in our periphery because of the evolutionary advantage that motion detection conveyed to our early ancestors [61,76]. Persistent motion, like blinking or pulsating, may consistently grab a user's attention, so that important notifications cannot be ignored [8,76]. In some cases, this may not be desirable, as motion may distract from the task at-hand. In contrast, there is some suggestion that only the appearance of *new* moving objects in the field of view grabs attention, and that more persistent motion can be ignored easily [24,39,76]. Colour can also deliver pre-attentive cues, e.g., by acting as a symbol to represent culturally recognised actions, emotions, or states [31,76]. Children develop a symbolic understanding of colour at a young age [18]. For instance, the traffic light metaphor is commonly used in early childhood education to teach "stop and go" types of activities, such as in nutritional education programs, e.g., [44] where a green light can indicate healthy foods ("go" behaviour) and a red light—unhealthy foods ("stop" behaviour). Some open-learner models [17,33] also use this metaphor to indicate to young learners how well they are performing within intelligent tutoring systems. This suggests that symbolic colour may be well interpreted and applied by children within different contexts. Thus, the pre-attentive features of motion and symbolic colour are potential tools to guide S&T behaviour in a computerised self-regulated learning environment, e.g., by flagging learners to "stop and think" at appropriate intervals.

1.2.3 Reward systems and penalties

Reward systems are implemented in games to increase players' motivation by relaying performance information and offering tangible rewards [9,30,63]. Reward systems may also promote S&T and broader metacognitive competencies by encouraging self-evaluation and self-monitoring through presenting scores of recently completed problems and through delivering new problems wherein the player can improve previous scores [30]. Finally, penalties play a critical role in game-based learning by creating a sense of risk and consequences for actions, thereby contextualising the value of rewards [30,46,52]. As such, reward systems and penalties in a learning environment might assist in S&T training by motivating children to stop-and-think for longer to improve performance and, in turn, obtain a greater reward.

1.3 The current study

The current study sought to understand how HCI features (as described above) either support or hinder children's S&T behaviour in an individual-use context, through interaction (e.g., click) and eye-tracking data. Since the 1980s, eye-tracking has been used to assess the usability of computer interfaces [69]. Conventional methods of understanding on-screen attentional allocation (e.g., through users' self-reports and button clicks) are biased toward conscious processes [69], where users purposefully attend to on-screen elements in an effort to achieve a goal or complete a task. However, visual attention does not entirely depend on conscious control; in other words, users often do not realise where they are looking [69]. In contrast to conventional methods, eye-tracking technology can more reliably determine allocation of attention through gaze data by capturing information on fixation frequency and duration on areas of interest, amongst other metrics. The eye-mind hypothesis [45] postulates that people "attend to and process the visual information that they are currently looking at", with the caveat being that the visual environment must be relevant to the cognitive task undertaken [41,45,64–66]. As such, this fixation information can help determine what visual information is being processed by children during cognitive tasks in digital environments.

Throughout this article we use the term "S&T behaviour" to refer to the children's use of the "stop-and-think" skill during the game. S&T behaviour is exhibited when children (i) "stop" to consider the problem, as evidenced by ceasing interactivity (clicks) in the software and focusing their gaze on the question presented; and (ii) "think" about their answer before responding, as evidenced by sustained non-interaction and gaze focused on the presented answer-objects (e.g., animals and buckets as depicted in **Figure 1B**). "Better" S&T behaviours in this context refer to: increased time spent stopping-and-thinking, increased fixations on on-screen answer-objects and question-textbox elements, decreased clicks made when supposed to be engaged in stopping-and-thinking (i.e., "invalid clicks"), and decreased fixations on the S&T icon.

The objectives (O) of the research were to:

1. Determine how various HCI features (as described above) promoted—or interfered with—children's use of S&T in the game;
2. Demonstrate how children's S&T behaviours are associated with their accuracy on the learning activities; and
3. Use the findings from O1-O2 to conceptualise the design of an adaptive system to underly *Stop & Think*, to support individual children's use of S&T based on their unique behaviours.

In relation to O1, we hypothesised (i) that the addition of the readiness mechanic would improve children's S&T-related behaviours, (ii) that motion would be more distracting than symbolic colour, pulling fixations away from answer-objects and increasing fixations on the pulsating S&T icon, and (iii) that the reward/penalty mechanic would improve S&T behaviours due to its motivational influence, thereby increasing answer accuracy. For O2, we hypothesised that

children's average recorded S&T time, as well as fixations on answer/question elements, would be positively related to their answer accuracy, and that invalid clicks would be negatively related to answer accuracy. Confirmation or rejection of the above hypotheses will help to inform O3.

To keep the length of this manuscript manageable, we offer more details of this study in our Supplementary Materials (SM) document.

2 METHOD AND MATERIALS

2.1 Stimuli

We developed four versions of the *Stop & Think* software that implemented the S&T mechanic through different visual and interactive designs (**Figure 2**), to evaluate how various HCI features impacted S&T behaviours. We label these conditions here as **Motion**, **Motion+**, **Colour+**, and **Colour+Reward**, where "+" indicates the readiness interaction mechanic. The **Motion** condition is the baseline condition and, like the version of *Stop & Think* used in the teacher-led trial [68], uses motion to promote S&T behaviour. While the game-show host reads the question, the S&T icon pulsates in the bottom left corner of the screen and continues for five seconds after narration finishes. During these five seconds, the screen is "locked", after which the icon disappears, and the screen becomes interactive. **Motion+** builds on the first condition by adding mandatory intentional interaction before the player can submit their answer. After the five-second enforced S&T time, the pulsating icon is replaced by a button that reads "I'm ready to answer!". Once pressed, the screen becomes interactive. **Colour+** also includes the readiness mechanic design feature but uses colour instead of motion to prompt the child to engage in S&T behaviour. The common analogy of traffic lights is used: (1) a red "Stop" icon appears in the bottom left corner of the screen as the question is being narrated, and the background is shaded with a red hue; (2) the red icon changes to amber "Think" icon, and the edges of the background change to yellow. After five seconds of enforced thinking, the "I'm ready" button pops out the side of the icon; (3) once the button is pressed, the screen is unlocked and the yellow icon and background change briefly to green for two seconds before both fading away. **Colour+Reward** builds on Colour+ by integrating simple rewards and penalties in the form of game tokens. Each correctly answered question rewards a base value of one token; multiple consecutive correct answers during repeated-practice problems earn a bonus multiplier, which gains one multiplier point for each correct answer. The bonus multiplier value is reset (the penalty) when an incorrect answer is given. Tokens are awarded in game-show mode when the host gives feedback regarding answer accuracy (depicted in **Section SM1**).

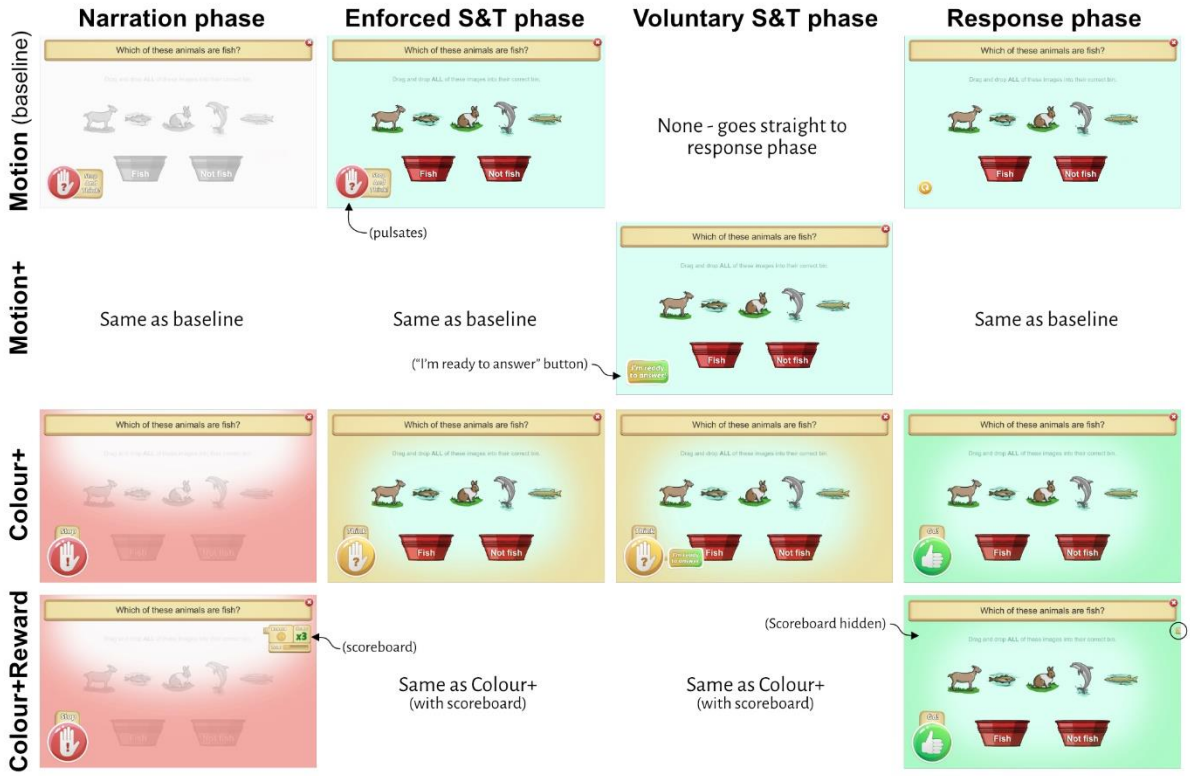

Figure 2. Differences in HCI design between conditions during *Stop & Think* science problems. **Motion**: motion visual cue; **Motion+**: motion visual cue + readiness mechanic; **Colour+**: colour visual cue + readiness mechanic; **Colour+Reward**; colour visual cue + readiness mechanic + rewards/penalties.

2.2 Participants

A total of 45 (19 girls, 26 boys) 7- to 8-year-olds with no known special educational needs drawn from two English primary state-schools were included in the analysis, with 11 randomly allocated to each of conditions Motion, Motion+, and Colour+Reward, and 12 to Colour+. Participants' mean age was 7.76 years (SD = 0.32). Groups were homogenous in terms of age and gender composition (**Section SM2 Tables S1**, **S2**). Participants were recruited through a 20-minute interactive presentation about simple neuroscience two weeks before the study. Opt-in consent was obtained from the children's parents/guardians. Children were offered stickers and a certificate of participation as incentives.

2.3 Procedure

Participants were excused from their lessons in pairs to participate in the study, which took about 30 minutes to complete. This included a brief introduction, calibration of the eye-tracker, a two-minute video tutorial on how to play, and 12 minutes of playtime (6 minutes in maths and 6 in science). Two eye-tracking stations were set up on identical laptops running Tobii Studio 3.3.2, Tobii X2-30 [73] compact eye-trackers, with child-sized computer mice and headphones. Laptop stations were set up facing each other so that children could not see each other's screen, in a quiet room. Tobii Studio was used to launch the game and record the screen using the "Screen Recording" Tobii element.

While eye-tracking took place, click-stream data was also digitally tracked, including interactions with answer-objects, click attempts made during the S&T mechanic, and correct/incorrect answers. The researcher gave simple instruction on how to use the software when requested by the child (e.g., "drag and drop objects", or "click on in the textbox"), but did not give any instruction pertaining to the S&T mechanic, nor did they provide feedback or help regarding correct/incorrect responses. The same science topics—the categorisation of fish, birds, and mammals—was chosen for all participants (see **Section SM3** for visuals). Whilst children also played a maths topic (addition of fractions with a common denominator), we have excluded those data from analysis, since this topic proved too difficult for our target group, obscuring any relationships between our measures and actual stopping-and-thinking.

Using Tobii Studio, eye-tracking recordings were parsed into four gameplay phases (or "scenes") for each problem attempt, based on segments of interaction related to S&T behaviour (**Figure 2**): (1) Question narration phase, (2) Enforced S&T phase (i.e. first five seconds after narration), (3) Voluntary S&T phase (i.e. after first five seconds, before "I'm ready to answer!" button is pressed, in the Motion+, Colour+, and Colour+Reward conditions only), and (4) Response phase (whilst interactions are enabled but the child has yet to make their first click). Gaze and interaction data that occurred *after* the child first interaction with an answer object during the Response phase were not considered.

2.4 Measures

This study focuses on interaction and eye-tracking measures from the science component of the *Stop & Think* game.

2.4.1 Interaction measures

In-game performance. In-game performance was calculated based on correct and incorrect answers given. To account for the game allowing multiple attempts for each problem, the single science exploratory problem was worth a total of four possible points, and the five repeated-practice problems was worth a total two possible points, each reducing in value by one point for every incorrect attempt. Therefore, the total possible science score was 14 points.

Time spent in gameplay phases. Average time spent in gameplay phases was calculated by summing time spent in the narration phase, enforced S&T phase, voluntary S&T phase, and response phase, up until the first click on an answer object is made in the response phase.

S&T time. Average total S&T time was calculated by (i) identifying the "time-to-first-click" in each problem attempt during narration, enforced S&T, voluntary S&T, and response phases of the task mode, (ii) summing these times across all phases, then (iii) averaging over all science problem attempts. We considered first-clicks in all phases of gameplay, rather than only during voluntary S&T and response phases, because the HCI features may also affect S&T behaviours before a response is allowed; so, the first click during these earlier phases would indicate that the child had stopped stopping-and-thinking early on. Steps (i) and (ii) in this calculation are visualised in **Figure 3** to help clarify what was performed and why. The figure depicts four different scenarios of children interacting in the same problem, where the same total amount of time has elapsed from the beginning of the narration phase (0 s) to when the child begins valid interactions with answer objects (13 s in this example). While the total elapsed time is the same across all four scenarios (13 s), the total amount of time spent stopping-and-thinking is arguably very different, which is why we looked at all phases across the task mode. Additionally, we compared S&T time as a proportion of time spent in the gameplay phases, calculated by dividing S&T time by total time spent in gameplay phases.

Invalid clicks. Invalid clicks were tallied as clicks occurring during narration, enforced S&T, and voluntary S&T phases of gameplay. The screen was "locked" from interactions during these phases, so attempting to interact with

answer-objects would result in an "invalid click", indicating that the child was "doing", rather than stopping-and-thinking.

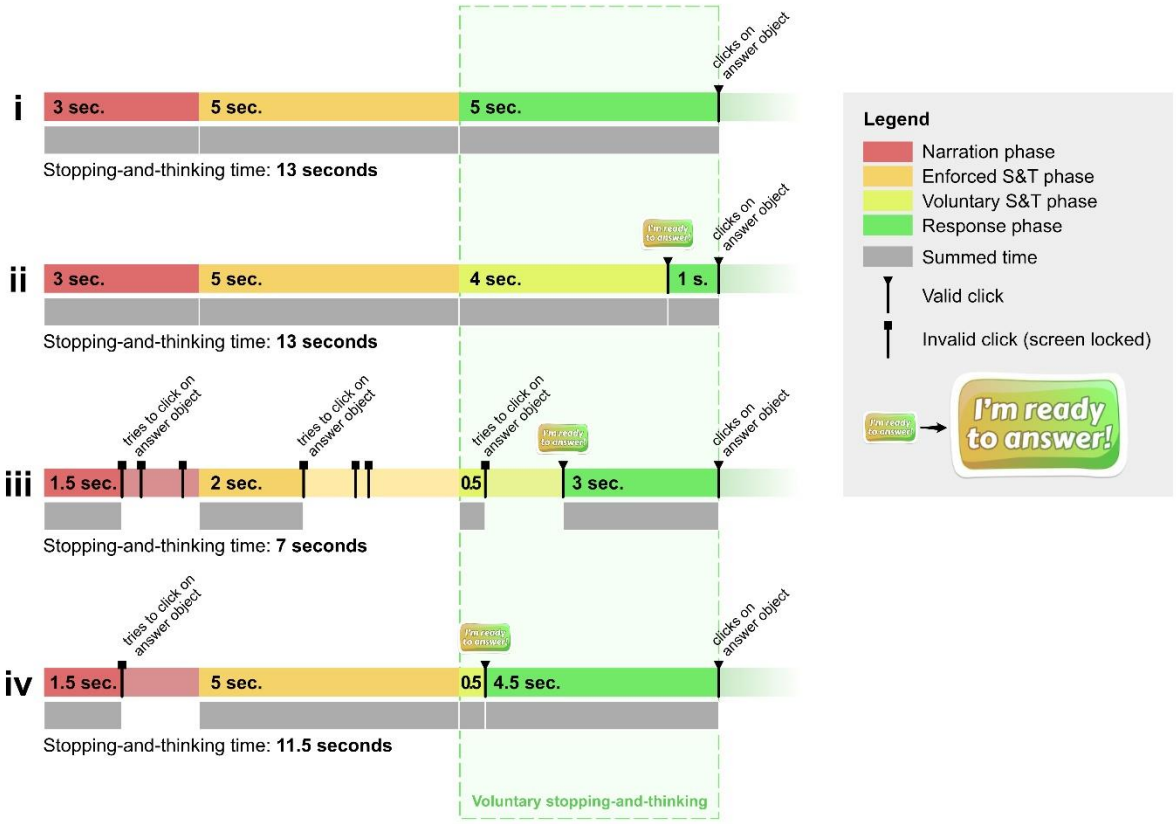

Figure 3. S&T time visualisation. (i) A child in the baseline Motion condition appropriately exercises their S&T skills by refraining to click throughout the narration and enforced S&T phase, then continues to stop-and-think well into the response phase. **(ii)** A child in one of the '+' conditions (Motion+/Colour+/Colour+Reward) appropriately exercises their S&T skills by refraining to click throughout the narration and enforced S&T phase, continues to stop-and-think into the voluntary S&T phase, before clicking the "I'm ready to answer" button and interacting with answer objects shortly thereafter. **(iii)** A child in one of the '+' conditions does not exhibit S&T behaviour by attempting to interact while the screen is locked and initially ignoring the "I'm ready to answer". **(iv)** A child in one of the '+' conditions tries to interact with answer objects during narration, but then realises they should be stopping-and-thinking and remedies their behaviour. This individual also ignores the purpose of the "I'm ready" button, clicking it immediately, but then continues to stop-and-think into the response phase, which we recognise as equally valid.

2.4.2 Eye-tracking measures.

Areas of interest. Areas of interest (AOI) were set for (i) **answer-objects** (i.e., animal images and sorting buckets), (ii) **S&T/I'm ready icon**, (iii) **question textbox**, and (iv) **scoreboard** (Colour+Reward condition only). Fixation data were analysed in terms of mean total fixation duration per problem, where a fixation is defined as lasting for 60 ms or longer. The fixation data were summed over all phases of gameplay, up until the first valid click made by the child during response phase, and then averaged over all science problem attempts. By including response phase fixations, this calculation controls for the lack of voluntary S&T phase in the Motion condition to make a direct comparison to other

conditions. Fixation data were analysed in terms of their actual values, as well as percentages of total fixation duration (e.g., time fixated on answer-objects AOI x 100 / total fixations made on the screen). Finally, we provide data on fixation frequency on AOI in the SM, since these do not add substantially to inferences made solely on duration data.

Time not fixating. Finally, the duration of time children spent *not* fixating on the screen was calculated by (i) summing fixation duration on all AOI and other non-AOI areas of the screen during all four phases of gameplay (up until the first valid click), then (ii) subtracting this value from the total raw duration of the four gameplay phases (e.g., this total duration would be 13 s as visualised in **Figure 3**). This value represents time spent looking off-screen (or not fixating at all) during S&T training. We examined this value as a raw number, as well as a percentage of total time over the four gameplay phases.

3 RESULTS

All analyses were performed in SPSS v.26 ($\alpha = .05$), using non-parametric analyses to compensate for our small sub-group sample sizes ($n < 15$) [34]. We used Kruskal-Wallis (*KW*) tests to compare between multiple conditions, Bonferroni-adjusted Mann-Whitney U (*U*) tests for post-hoc pairwise comparisons, and Spearman correlations (*rho*) to assess the relationships between in-game performance score and other metrics across conditions.

3.1 Effect of design characteristics on S&T behaviour

3.1.1 Interaction data

In-game performance. Participants completed (whether correctly or incorrectly because they ran out of attempts or time) an average of 3.82 out of 6 (SD = 1.03) unique science problems in 7.09 (SD = 1.69) attempts, within the six minutes of playtime allotted to science. The number of unique science problems completed and attempted was similar across conditions (**Section SM4 Tables S3**, **S4**).

The overall mean science performance was 4.82 (*SD* = 2.49) out of 14. There was no statistically significant difference between conditions ($KW = 6.64$, $p = .084$), though children in the Colour+Reward condition performed marginally better than those in other conditions (**Table 1**).

Table 1. **Interaction data**: Descriptive statistics of in-game performance scores on science problems (maximum score = 14), raw time spent stopping-and-thinking, and proportion of time spent stopping-and-thinking in gameplay phases.

	In-game performance		S&T time		% time stopping-and-thinking	
Condition	*M* (SD)	Range	*M* (SD)	Range	*M* (SD)	Range
Motion	4.45 (2.38)	2-9	8.45 (1.20)	6.64-10.11	57.88 (5.23)	48.26-63.64
Motion+	4.55 (3.24)	0-9	13.11 (2.31)	10.14-17.24	62.61 (12.90)	29.52-78.66
Colour+	4.00 (2.44)	0-7	13.35 (4.33)	8.57-21.26	61.49 (9.48)	47.96-76.48
Colour+Reward	6.36 (0.92)	5-8	12.76 (3.31)	9.85-21.60	67.39 (6.37)	55.67-74.42

Time spent in gameplay phases. Children averaged 19.49 (SD = 6.28) seconds across the four gameplay phases in each problem. The total average time spent in gameplay phases was significantly different between conditions ($KW = 28.09$, $p < .001$). The baseline Motion condition averaged 13.60 s (*SD* = .04) in the gameplay phases, which is less than those in Motion+ who spent on average 22.94 s (*SD* = 8.75, $U = 24.91$, $p < .001$), those in Colour+ who spent 21.76 s (*SD*

= 3.49, U = 26.00, p < .001), and those in Colour+Reward who spent 19.46 s (SD = 5.02, U = 16.27, p = .022). Time in '+' conditions were similar (U's < 9.73).

S&T time. Total average S&T time is reported in **Table 1**. Children averaged 11.95 s (SD = 3.58) of stopping-and-thinking per science problem, which represents on average 62.3% of the time spent in the gameplay phases. This is on average 5.49 s (SD = 3.20) beyond the enforced S&T phase (the green shaded area in **Figure 3**). There was a significant effect of condition on S&T time (KW = 20.87, p < .001). Bonferroni-adjusted pairwise comparisons showed that conditions Motion+ (U = 22.73, p < .001), Colour+ (U = 19.05, p = .003), and Colour+Reward (U = 20.09, p = .002) had significantly longer S&T times than the baseline Motion condition (no other pairwise comparison was significant, U's < 3.68). This pattern was also observed when looking at time beyond the enforced S&T phase only (**Section SM5 Tables S5**, **S6**), indicating a direct effect of the readiness mechanic. Condition also had a marginal effect on the percentage of time spent stopping-and-thinking during the gameplay phases (KW = 7.67, p = .053), where there is some indication that those in the Colour+Reward condition may have spent a greater proportion of their training time without trying to respond than other groups (**Table 1**).

Invalid Clicks. Children made an average of 1.28 (SD = 1.04) invalid clicks per problem. There were no differences between conditions (KW = 6.64, p = .084) (**Section SM6 Table S7**).

3.1.2 Eye-fixation data

Answer objects AOI. A significant difference between conditions was observed in duration of fixation on the answer objects AOI (KW = 15.96, p = .001, **Table 2**). Bonferroni-adjusted pairwise comparisons revealed significantly longer fixation on answer objects for the Colour+ condition than the Motion condition (U = 20.26, p = .001) and the Colour+Reward condition (U = 16.27, p = .018). All other comparisons were non-significant (U's < 11.46). Fixation frequency revealed similar results (**Section SM7 Table S8**).

A different perspective was gained by examining the percentage of time spent on answer objects out of the total fixation duration time, instead of the raw durations. Again, a significant difference was observed (KW = 25.08, p < .001, **Table 2**), but here the baseline Motion condition was associated with proportionally longer fixations on answer objects than the Motion+ (U = 24.27, p < .001), Colour+ (U = 14.56, p = .047), and Colour+Reward (U = 24.18, p < .001) conditions. Other comparisons were non-significant (U's < 9.71). This is perhaps expected given that children are likely to look at the S&T icon to press the "I'm ready" button in '+' conditions, thereby diverting a proportion of their fixation durations away from answer objects.

Table 2. **Fixation data**: Raw total fixation duration and percent fixation duration on **answer object** and **S&T/I'm Ready icon**, as well as raw time and percentage time **not fixating on the screen**.

Raw fixation duration (s)						
	Answer objects AOI		S&T/I'm Ready icon AOI		Not fixating on screen	
Condition	*M* (SD)	Range	*M* (SD)	Range	*M* (SD)	Range
Motion	5.20 (1.59)	2.52-7.98	0.30 (0.23)	0.02-0.67	3.00 (0.40)	1.14-5.92
Motion+	7.19 (2.31)	2.41-10.04	2.07 (0.65)	0.79-3.16	8.52 (8.98)	2.06-33.04
Colour+	9.97 (3.33)	5.70-16.13	1.89 (1.35)	0.00-4.10	5.31 (2.04)	2.84-10.14
Colour+Reward	6.06 (4.04)	2.38-16.37	1.79 (0.92)	0.77-3.32	6.53 (3.17)	2.67-11.95
Percent fixation duration (%)						
Motion	91.03 (3.81)	85.14-96.86	5.71 (4.58)	0.36-13.92	32.49 (7.58)	8.11-67.41
Motion+	71.84 (5.88)	60.79-77.96	25.18 (5.27)	16.62-33.86	39.30 (22.85)	14.53-81.89

Colour+	80.35 (8.31)	65.74-91.33	17.75 (10.10)	0.00-28.81	29.06 (8.63)	18.74-43.38
Colour+Reward	71.40 (10.78)	54.73-88.98	25.61 (9.48)	7.09-44.30	44.48 (19.63)	18.86-71.47

S&T/I'm ready icon AOI. Analysis of fixations on the S&T icon confirm this interpretation. A significant difference was found in S&T/I'm ready icon fixation duration (KW = 21.24, p < .001, **Table 2**). Children in the baseline Motion condition, which lacked the "I'm ready" button, looked at the icon less than those in the Motion+ (U = 23.18, p < .001), Colour+ (U = 19.14, p = .003), and Colour+Reward (U = 19.91, p = .002) conditions. Comparisons between '+' conditions were non-significant (U's < 4.05). Fixation frequency on the icon revealed similar results (**Section SM8 Table S9**).

When looking at proportion of fixation duration on the S&T icon, we again observe a significant difference (KW = 20.34, p < .001, **Table 2**), wherein the baseline Motion condition children fixated proportionally less on the S&T icon in comparison to Motion+ (U = -22.09, p < .001) and Colour+Reward (U = -21.64, p = .001). There was also a trend for children in the Motion condition to fixate for proportionally less time on the S&T icon than those in the Colour+ condition (U = 14.12, p = .060), but the difference did not reach significance.

Question textbox AOI. Fixation duration and the percentage fixation duration on the question-box AOI was overall low, and equally low across groups (*Raw duration*: KW = 2.99, p = .394; *Percent*: KW = 0.66, p = .882), indicating that the children tended not to read the questions while it was being narrated. The average fixation duration on the question AOI across conditions was 0.20 s (SD = 0.21) per problem, which is just 2.13% (SD = 2.09) of all fixations; refer to **Section SM9 Table SM10** for details within each condition.

Scoreboard AOI. Participants in the Colour+Reward condition fixated on the scoreboard an average of 0.04 s (SD = 0.05) per problem, which accounts for only 0.05% (SD = 0.07%) of fixations made during the four gameplay phases. This indicates that the scoreboard as an interface element did not distract visually from the training task. We did not assess scoreboard fixations outside of the main gameplay phases (e.g., when the game show host awarded tokens, **Figure S1**).

Time not fixating. The actual time children spent not fixating on the screen was statistically different between conditions (KW = 11.75, p = .008, **Table 2**). Children in the baseline Motion condition spent less time not fixating on the screen than those in the Motion+ (U = 16.09, p = .024) and Colour+Reward (U = 17.09, p = .014) conditions, but were similar to the Colour+ condition (U = 10.49, p = .334). Time spent not fixating on the screen was similar between '+' conditions (U's < 6.60). However, the difference in percentage of time spent not fixating on the screen between conditions did not reach significance (KW = 6.44, p = .092). This analysis controls for the fact that children in the Motion condition generally spent less time across the four phases of gameplay than others. The trend in this result points to the Colour+ condition possibly spending less time not fixating the Colour+Reward in this regard.

3.2 Relationship between S&T behaviours and in-game performance

There was a significant positive relationship between in-game performance and average S&T time (rho = 0.299, p = .046; **Figure 4A**), as well as between in-game performance and stopping-and-thinking as a percentage of total gameplay time (rho = .483, p = .001; **Figure 4B**). There was no relationship between invalid clicks in science problems and in-game performance (rho = -0.085, p = .579). Furthermore, there were no relationships between in-game performance and any AOI fixation measure, nor were there relationships between time not fixating on the screen (raw and percent) and in-game performance (refer to **Section SM10 Table S11**).

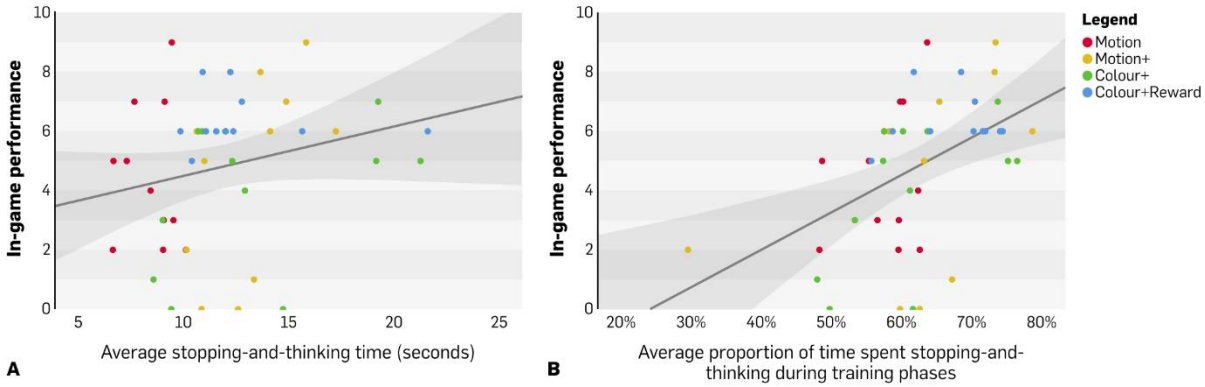

Figure 4. Association between in-game performance and (A) average S&T time and (B) proportion of time spent stopping-and-thinking, with 95% confidence intervals indicated in the shaded area.

4 DISCUSSION

This study explored how HCI design decisions impacted children's implementation of the "stop-and-think" instructions as measured through gaze and in-game behaviour, as well as their performance on science problems. Below, we discuss evidence of (i) the impact of HCI features on S&T behaviours, (ii) the relationship between S&T behaviours and in-game performance, (iii) the application of these findings to the design of an adaptive system to support personalised training in *Stop & Think*, and (iv) the limitations and future directions of this work.

4.1 Effect of design characteristics on S&T behaviour

First and foremost, the type of interface condition to which children were exposed did not significantly impact in-game performance on science problems. We surmise that this is because a single session of S&T training may not be sufficient to forge a meaningful relationship between S&T behaviours and actual performance; in previous trials, this training occurred three times per week over 10 weeks [68,77]. It is possible that, if children continued to train with these conditions over a longer period of time, greater differences in performance might be observed, spurred by the differential S&T behaviours that they generate (elaborated below).

Secondly, a major finding was that the addition of the mandatory, readiness mechanic through the "I'm ready to answer!" button (in the three '+' conditions) led to significant increases in (a) overall time spent in gameplay training, (b) average S&T time, and (c) proportion of time spent stopping-and-thinking during training, compared to the baseline condition without this feature. This supports our hypothesis that integrating a mechanic that allows children to indicate when they are ready to answer (like hand-raising) may encourage planning of, and commitment to, future interactions and intentions [2,3], thereby enhancing S&T training. It may be argued that the additional time taken to "think" in the '+'conditions may have resulted from the extra step needed to click the "I'm ready" button. The '+' conditions generated significantly more fixations on the S&T icon than the baseline condition, so a greater proportion of fixations was diverted away from answer objects, which may be considered a deficit of the readiness mechanic. However, participants in the Motion condition differed from others, on average, by 4-5 seconds of S&T time each problem, which is more than what would be reasonably expected from having to click a button—and is longer than the additional 1-2 seconds diverted by the S&T icon and "I'm ready" button in the '+' conditions. As such, we feel that this additional stopping-and-time was representative of true "thinking" time.

Thirdly, the comparison of Motion+ and Colour+ revealed that the use of symbolic colour and motion were equally effective at promoting S&T behaviours. Both conditions encouraged similar average S&T times and fixation duration on answer-objects, suggesting that they were equally effective at focusing children on the S&T task. Interestingly, the two conditions also prompted similar fixations on the S&T icon, meaning that, contrary to some research [8], the motion of the icon in Motion+ did not distract in an obvious way from the thinking task. This might be explained by children acclimatising to the motion with practice or that, because the icon is a persistent visual element in the user's peripheral vision during the S&T mechanic, children's attention was not constantly drawn to it [24,39,76]. Yet, there may have been an interaction effect between the use of symbolic colour and readiness mechanic that encouraged better S&T-related fixations. For instance, the Colour+ condition generated significantly more fixations on answer objects than the baseline Motion condition, whereas no difference was observed between Motion and Motion+. Additionally, children Motion+ spent proportionally more time looking at the S&T icon than those in the Motion condition, whilst the difference between Colour+ and Motion did not reach significance. There is a long-standing notion that the "whole is greater than the sum of its parts" when it comes to interface design, so we may be observing this here [10,12,38,59].

Finally, contrary to our hypothesis, the rewards and penalty system (tokens and bonus multipliers in Colour+Reward) did not positively impact S&T behaviours in obvious ways. The Colour+Reward condition generated fewer fixations on answer items in comparison to Colour+, in which children were exposed to the same stimulus without the reward system. Additionally, whilst Colour+Reward and Colour+ conditions generated similar fixations on the S&T icon and similar time not fixated, the Colour+Reward condition performed less well in these metrics in comparison to the baseline Motion condition, indicating a dampening effect of the reward feature on the effectiveness of Colour+. These findings would suggest that the reward and penalty system interfered with this S&T behaviour. This is dissimilar to [58] who found that the visibility of time pressure/scoring in a maths game increased children's eye-fixations on problem-relevant features (e.g., question and answer elements). We attribute the increased fixations on the S&T icon to an increased level of excitement generated by the scoring mechanic, that moved children toward more frequently checking for the presence of the "I'm ready" button (attached to the icon) and subsequent ability to play. However, it is possible that this behaviour may have remedied itself over several play-sessions; perhaps participants did not have enough exposure to the scoring mechanic over the 12-minute playtime to develop an appreciation of its value [48,49], or apply it in competitive scenarios [36]. Contrastingly, perhaps the penalty of losing the bonus multiplier upon a wrong answer was not dramatic enough to instil a sense of action-consequence in participants [30,46], thus motivating S&T behaviour. Alternatively, perhaps the scoring mechanic was ineffective because it did not provide feedback on the S&T behaviour itself, only whether the answer was correct. Feedback is crucial for engaging children in meta-learning [6], so game-based learning and AI-in-education interventions often use scoring mechanisms and dynamic visual displays (e.g., in open-learner models) to support learners' recovery from errors, heighten engagement, and promote reflective thinking [48,79]. A breadth of research indicates that feedback is pivotal for successful game-based and self-regulated learning, but how, when, and what content should be provided as feedback to foster deep learning remains inconclusive [48]. While positive effects of the reward and penalty system on S&T behaviour were not observed, children in the Colour+Reward condition nonetheless performed marginally better than those in other conditions on problem accuracy and on proportion of time spent stopping-and-thinking. This again begs an additional metric (e.g., body language and facial expressions) to help explain how other factors, such as motivation and engagement, which may have been enhanced with the reward/penalty feature, may be related to performance.

4.2 S&T behaviours and in-game performance

We considered that (i) more time spent stopping-and-thinking, (ii) a greater proportion of training time spent stopping-and-thinking, (iii) fewer clicks before a response was allowed, (iv) more/longer fixations on answer-objects, and (v) fewer fixations on the S&T icon, reflected children's use of their S&T skills in the task and attempts to follow instructions to "stop and think" before responding. Two of these "S&T behaviours" were found to positively associate with children's performance on counterintuitive science problems.

Specifically, the average stopping-and-thinking, as well as the proportion of training time spent stopping-and-thinking, was significantly correlated with in-game performance on science problems, suggesting that engaging in S&T heightened children's chances of success when encountering counterintuitive problems [15,77]. However, we failed to observe any relationships between gaze data and performance, which contradicts other research relating gaze to attentional focus [41,45,64–66]. It is unclear why this was the case. Given that performance was strongly correlated with S&T time, we expected that fixations on answer objects would also be significantly related to performance, since increased fixations should naturally follow with increased time-on-task. It could be that children were looking elsewhere than on-screen, but still have been on-task and "thinking": we experienced some children looking skyward, closing/squinting their eyes while thinking. This is reflected by the high percentage of time "not fixated" on the screen. Gestures and body language are a mechanism for cognitive change and an indication of thinking taking place [16,21,35]. Hence, body language and gestures (not measured in this study) that diverted children's eyes away from the screen may account for why fixations on answer objects and other metrics were not related to performance, while S&T time was.

4.3 Towards an adaptive system for *Stop & Think*

Our results indicate three aspects of the game that might be adapted to support personalised S&T training and aid children in transitioning to un-cued self-regulated learning scenarios: (i) the enforced S&T mechanic; (ii) the difficulty of the content or level of support given to the child; and (iii) the visual cues (e.g., S&T icon). We found that time spent stopping-and-thinking was positively correlated with performance in science problems. An adaptive system might use the player's average S&T time together with answer correctness as a measure to calibrate the optimal S&T time and the level of scaffolding needed (e.g., enforced vs. voluntary S&T mechanic). However, we cannot say for certain that longer S&T times necessarily mean that children are exercising S&T; for some more new or more advanced topics, it could mean that the child is confused and needs support, or that their mind is wandering. The adaptive system should seek to differentiate between players meaningfully engaging in the desired behaviour and being 'lost'. Calibrating the enforced S&T mechanic based on children's average voluntary S&T time and answer correctness to infer S&T engagement may be sufficient to simplistically adapt hints or adjust content difficulty. However, gaining an understanding of relationships between gameplay, S&T behaviour, and performance *over time* (i.e. repeated exposure to S&T training tasks) is likely to require the application of machine learning, e.g., hidden Markov models [74]—an approach that we are presently exploring.

Finally, the fixation data for the baseline Motion condition suggests that participants glanced at the S&T icon but did not focus their attention on it (average total fixation duration was 0.3 seconds per problem). Thus, it seems participants used the icon as intended, as a reference for thinking and interaction phases of gameplay. By gauging when learners become more adept at S&T behaviour and less reliant on cues, an adaptive system should, ideally, scaffold away the icon and "I'm ready" button to transition children to performing S&T in un-cued environments, exemplifying self-regulated learning.

4.4 Limitations and future directions

While this study successfully highlighted how various design characteristics can influence S&T behaviour in an individualised, game-based environment, our results come with some limitations. First and foremost is our relatively small sample size. While 45 participants is larger than many published eye-tracking studies, e.g., [23,50,51,54,72], the subgroups based on stimulus condition are much smaller, yielding non-normally distributed data requiring non-parametric analyses. This limits our ability to investigate potential confounding factors, such as gender or preferences for gaming activities.

Secondly, we did not perform a pre-assessment of children's knowledge and misconceptions prior to their interaction in *Stop & Think* because of the priming effect it would have had on knowledge-recall within the game. However, performing a pre-assessment would have helped to support assumptions made about the relationships between S&T-related interactions and in-game problem accuracy.

Thirdly, this study measured S&T behaviour based on interactions with on-screen elements and fixations on interface elements. While time spent stopping-and-thinking was found to be significantly related to performance on science problems, suggesting that it might be a good measure for reflective thinking, the fixation data did not support previously identified relationships between fixations and attentional focus [64–66]. As such, additional metrics for measuring thinking relevant to the current project should be considered in future work, such as physical gestures which may have diverted eye-gaze away from the screen (e.g., looking skyward) or facial expressions which may reveal the child's frustration or confusion [16,21].

Finally, we looked at S&T behaviours during a single training session in a programme that would normally take place over 30 sessions (see [68]). We did not measure long-term changes related to children learning to apply S&T behaviours to science problem-solving in the real world. While our results are promising in terms of the usefulness of certain HCI features (e.g., readiness mechanic) in encouraging S&T-use, longitudinal research is needed to fully appreciate their potential effects on children's use of S&T skills in out-of-lab—and out-of-game—contexts.

Future longitudinal research might also investigate the possible integration of open-learner modelling and data visualisations with scoring mechanics, to promote metacognitive competencies and enhance self-regulation in children, as the current research demonstrated a lack of impact from a simple accuracy-based reward system. These findings highlight the pressing need for iterative HCI design research into the major pedagogical features of any computer-based learning environment, e.g., the impact of different scoring mechanics and feedback displays.

5 IMPLICATIONS AND CONCLUSION

Computer-based, adaptive, learning environments enable learning to take place either at school or at home by supporting the unique needs of individual learners. However, the design of these environments requires careful consideration. This paper provides a multidisciplinary perspective on this issue—the first is from an educational neuroscience perspective, suggesting that training domain-specific use of S&T improves children's counterintuitive reasoning, which is critical to successful self-regulated learning [56,77]; the second is from a human-computer interactions perspective, highlighting the acute need to evaluate how the visual and interactive designs of digital environments support the intended learning or training goals [62,76]; and the third is from an artificial intelligence in education perspective, identifying ways in which technology might adapt to support educational neuroscience training, based on HCI findings.

Unfortunately, these themes are often not explored together in educational research, where the adaptive, visual, and HCI design of cognitive interventions often receive little attention. Reproducing educational neuroscience laboratory procedures in live classrooms is notoriously difficult, and many practitioners claim even undesirable, since the partly

unpredictable ecology of classroom settings stands in direct contradiction with the settings of randomised controlled trials and prescriptive procedures needed to achieve the same outcomes [11]. Computerising such interventions is often seen in psychology and educational neuroscience as a feasible way in which the consistency of their delivery, and by extension, the desirable learning outcomes may be guaranteed more readily, e.g., [53]. However, although computerising interventions is standard practice in cognitive neuroscience and psychology research, there is a notable and substantial disconnect between the intervention programmes developed and research in technology-mediated learning, HCI practices, and adaptive learning environments. This means that many computerised cognitive interventions lack appropriate interactive and adaptive elements, and often ignore key design and knowledge engineering principles/methods at the intersection of HCI and artificial intelligence in education that might be critical to delivering the cognitive neuroscience interventions as intended. For example, both visualisations and the use of such visualisation as part of adaptive feedback in an intervention environment are critical in ensuring that any perceptual or priming interference with the task is reduced to a minimum [62,76]. This is critical in the contexts such as S&T training on counterintuitive science and maths concepts, where the perceptual and prior beliefs interference has been identified as the main obstacle for the success of such training [5,13,37]. We need to be reminded that the design, rather than the medium of delivery, ultimately predicts learning outcomes [20,28]; when cognitive interventions are delivered via interactive technology, such as games, the design of the HCI and adaptive system are paramount.

This experiment investigated the effect of four interface design characteristics on children's application of S&T skills in a game-based learning environment. We found that that a readiness mechanic increased children's time spent stopping-and-thinking, that persistent motion and symbolic colour were equally effective at promoting S&T-related behaviour but that the combination of symbolic colour and the readiness mechanic may have a cumulative effect, and that the reward/penalty mechanic may have distracted from the cognitive task or may not have provided enough detail to effectively promote S&T behaviours. Additionally, children's time spent stopping-and-thinking (both in raw duration and percent) was related to their performance on science problems, supporting previous research touting S&T-training to support academic performance [68,77]. Our results provide insights about how the *Stop & Think* game should adapt to support the learning activity, by e.g., adapting enforced S&T time and visual thinking prompts in response to answer accuracy and voluntary S&T time through machine learning techniques. In conclusion, this work explicitly bridged educational neuroscience, HCI and artificial intelligence in education research by acting as an intermediary step between (a) implementing established educational neuroscience principles in digital education, (b) determining which HCI design features best support children's use of S&T to improve their counterintuitive problem-solving, and (c) conceptualising how the *Stop & Think* game should adapt to bolster personalised training. To that effect, it supports the shift in pedagogy toward personalised, technology-mediated, self-regulated learning, that we are currently experiencing.

6 ACKNOWLEDGEMENTS

We would like to thank the primary school administrators, teachers, and children who participated in this project for their support of the research. Beyond the authors cited above, the UnLocke project team consists of Michael Thomas, Derek Bell, Emily Farran, Andrew Tolmie, and Annabel Page, who we thanks for their contributions to this research. This work was funded by the Wellcome Trust and the Education Endowment Foundation.

References

[1] M Ahmad, L A Rahim, and N I Arshad. 2014. A review of educational games design frameworks: An analysis from software engineering. *2014 Int. Conf. Comput. Inf. Sci. ICCOINS 2014 - A Conf. World Eng. Sci. Technol. Congr. ESTCON 2014 - Proc.* (2014).

DOI:https://doi.org/10.1109/ICCOINS.2014.6868452

[2] Icek Ajzen. 1985. From Intentions to Actions: A Theory of Planned Behaviour. In *Action-control*, Kuhl J. and Beckmann J. (eds.). Springer, Berlin, Heidelberg, 11–39. DOI:https://doi.org/https://doi.org/10.1007/978-3-642-69746-3_2

[3] Icek Ajzen. 1991. The theory of planned behavior. *Organ. Behav. Hum. Decis. Process.* 50, 2 (December 1991), 179–211. DOI:https://doi.org/https://doi.org/10.1016/0749-5978(91)90020-T

[4] Nicholas P. Allan, Laura E. Hume, Darcey M. Allan, Amber L. Farrington, and Christopher J. Lonigan. 2014. Relations Between Inhibitory Control and the Development of Academic Skills in Preschool and Kindergarten: A Meta-Analysis. *Dev. Psychol.* 50, 10 (2014), 2368–2379. DOI:https://doi.org/10.1037/a0037493

[5] Michael Allen. 2014. *Misconceptions in primary science*.

[6] Susan Askew and Caroline Lodge. 2004. Gifts, ping-pong and loops – linking feedback and learning. In *Feedback For Learning*, Susan Askew (ed.). Routledge, London, 1–17. DOI:https://doi.org/https://doi.org/10.4324/9780203017678

[7] Roger Azevedo, Nicholas V. Mudrick, Michelle Taub, and Amanda E. Bradbury. 2019. Self-regulation in computer-assisted learning systems. In *The Cambridge handbook of cognition and education*, John Dunlosky and Katherine A. Rawson (eds.). Cambridge University Press, Cambridge, 587–618. DOI:https://doi.org/https://doi.org/10.1017/9781108235631.024

[8] L. Bartram, Colin Ware, and T. Calvert. 2003. Moticons: Detection, distraction and task. *Int. J. Human–Computer Stud.* 58, 5 (2003), 515–545.

[9] Wendy L. Bedwell, Davin Pavlas, Kyle Heyne, Elizabeth H. Lazzara, and Eduardo Salas. 2012. Toward a taxonomy linking game attributes to learning: An empirical study. *Simul. Gaming* 43, 6 (2012), 729–760. DOI:https://doi.org/10.1177/1046878112439444

[10] Randolph Bias, Paul Marty, and Ian Douglas. 2012. Usability/User-Centered Design in the iSchools: Justifying a Teaching Philosophy. *J. Educ. Libr. Inf. Sci.* 53, 4 (2012), 274–289.

[11] Gert Biesta. 2013. *The beautiful risk of education / Gert J. J. Biesta*. Paradigm, Boulder, Colo. ; London.

[12] Patricia A Billingsky. 1993. Reflections on ISO 9241: Software usability may be more than the sum of its parts. *StandardView* 1, 1 (1993), 22–25.

[13] Bofferding. 2019. Understanding negative numbers. In *Constructing numbers*. Springer, Cham, 251–277.

[14] Elizabeth A. Boyle, Thomas M. Connolly, Thomas Hainey, and James M. Boyle. 2012. Engagement in digital entertainment games: A systematic review. *Comput. Human Behav.* 28, 3 (2012), 771–780. DOI:https://doi.org/10.1016/j.chb.2011.11.020

[15] Annie Brookman-Byrne, Denis Mareschal, Andrew K. Tolmie, and Iroise Dumontheil. 2018. Inhibitory control and counterintuitive science and maths reasoning in adolescence. *PLoS One* 13, 6 (2018), 1–19. DOI:https://doi.org/10.1371/journal.pone.0198973

[16] Neon Brooks and Susan Goldin-Meadow. 2016. Moving to Learn: How Guiding the Hands Can Set the Stage for Learning. *Cogn. Sci.* 40, 7 (2016), 1831–1849. DOI:https://doi.org/10.1111/cogs.12292

[17] Susan Bull, Manveer Mangat, Andrew Mabbott, Abdallatif Abu Issa, and Josie Marsh. 2005. Reactions to inspectable learner models: seven year olds to university students. In *Proceedings of Workshop on Learner Modelling for Reflection, International Conference on Artificial Intelligence in Education*, Amsterdam, The Netherlands, 1–10.

[18] Esther Burkitt, Martyn Barrett, and Alyson Davis. 2003. Children's colour choices for completing drawings of affectively characterised topics. *J. Child Psychol. Psychiatry Allied Discip.* 44, 3 (2003), 445–455. DOI:https://doi.org/10.1111/1469-7610.00134

[19] Thomas K. F. Chiu and Cher Ping Lim. 2020. Strategic Use of Technology for Inclusive Education in Hong Kong: A Content-Level Perspective. *ECNU Rev. Educ.* (2020), 209653112093086. DOI:https://doi.org/10.1177/2096531120930861

[20] Douglas. B. Clark, E. E. Tanner-Smith, and S. S. Killingsworth. 2016. Digital Games, Design, and Learning: A Systematic Review and Meta-Analysis. *Rev. Educ. Res.* 86, 1 (2016), 79–122. DOI:https://doi.org/10.3102/0034654315582065

[21] Susan Wagner Cook, Howard S. Friedman, Katherine A. Duggan, Jian Cui, and Voicu Popescu. 2017. Hand Gesture and Mathematics Learning: Lessons From an Avatar. *Cogn. Sci.* 41, 2 (2017), 518–535. DOI:https://doi.org/10.1111/cogs.12344

[22] A Diamond and K Lee. 2011. Interventions Shown to Aid Executive Function Development in Children 4 to 12 Years Old. *Science (80-.).* 333, 6045 (2011), 959–964. DOI:https://doi.org/10.1126/science.1204529

[23] Aboozar Eghdam, Johanna Forsman, Magnus Falkenhav, Mats Lind, and Sabine Koch. 2011. Combining usability testing with eye-tracking technology: Evaluation of a visualization support for antibiotic use in intensive care. *Stud. Health Technol. Inform.* 169, June 2016 (2011), 945–949. DOI:https://doi.org/10.3233/978-1-60750-806-9-945

[24] James T. Enns, Erin L. Austen, Vincent Di Lollo, Robert Rauschenberger, and Steven Yantis. 2001. New objects dominate luminance transients in setting attentional priority. *J. Exp. Psychol. Hum. Percept. Perform.* 27, 6 (2001), 1287–1302. DOI:https://doi.org/10.1037/0096-1523.27.6.1287

[25] Jonathan St.B.T Evans. 2003. In two minds: dual-process accounts of reasoning. *Trends Cogn. Sci.* 7, 10 (2003), 454–459. DOI:https://doi.org/10.1016/j.tics.2003.08.012

[26] Varvara Garneli, Michail Giannakos, and Konstantinos Chorianopoulos. 2017. Serious games as a malleable learning medium: The effects of narrative, gameplay, and making on students' performance and attitudes. *Br. J. Educ. Technol.* 48, 3 (2017), 842–859. DOI:https://doi.org/10.1111/bjet.12455

[27] Andrea Gauthier, Michael Corrin, and Jodie Jenkinson. 2015. Exploring the influence of game design on learning and voluntary use in an online vascular anatomy study aid. *Comput. Educ.* 87, September (September 2015), 24–34. DOI:https://doi.org/10.1016/j.compedu.2015.03.017

[28] Andrea Gauthier and Jodie Jenkinson. 2018. Designing productively negative experiences with serious game mechanics: Qualitative analysis of game-play and game design in a randomized trial. *Comput. Educ.* 127, 2018 (2018), 66–89. DOI:https://doi.org/10.1016/j.compedu.2018.08.017

[29] Matthew Gaydos. 2015. Seriously Considering Design in Educational Games. *Educ. Res.* 44, 9 (2015), 1–6. DOI:https://doi.org/10.3102/0013189X15621307

[30] James Paul Gee. 2007. *What Video Games Have To Teach Us About Learning And Literacy* (2nd ed.). Palgrave MacMillan, New York, New York, USA.

[31] Nils Gehlenborg and Bang Wong. 2012. Points of view: Mapping quantitative data to color. *Nat. Methods* 9, 8 (2012), 769. DOI:https://doi.org/10.1038/nmeth.2134

[32] Camilla Gilmore, Sarah Keeble, Sophie Richardson, and Lucy Cragg. 2015. The role of cognitive inhibition in different components of arithmetic. *ZDM - Math. Educ.* 47, 5 (2015), 771–782. DOI:https://doi.org/10.1007/s11858-014-0659-y

[33] Sylvie Anne-Sophie Girard. 2011. Traffic Lights and Smiley Faces: Do children learn mathematics better with affective Open- Learner Modelling tutors? University of Bath.

[34] Stephanie Glen. 2014. Non Parametric Data and Tests (Distribution Free Tests). *StatisticsHowTo.com: Elementary Statistics for the rest of us.* Retrieved August 10, 2020 from https://www.statisticshowto.com/parametric-and-non-parametric-data/

[35] Susan Goldin-Meadow and Melissa A. Singer. 2003. From Children's Hands to Adults' Ears: Gesture's Role in the Learning Process. *Dev. Psychol.* 39, 3 (2003), 509–520. DOI:https://doi.org/10.1037/0012-1649.39.3.509

[36] Thomas Hainey, Thomas M. Connolly, Elizabeth A. Boyle, Amanda Wilson, and Aisya Razak. 2016. A systematic literature review of games-based learning empirical evidence in primary education. *Comput. Educ.* 102, February 2009 (2016), 202–223. DOI:https://doi.org/10.1016/j.compedu.2016.09.001

[37] Alice Hansen, D. Drews, J. Dudgeon, F. Lawton, and L. Surtees. 2017. *Children's errors in mathematics* (4th editio ed.).

[38] Jorg Henseler. 2015. *Is the whole more than the sum of its parts?: On the interplay of merketing and design research.*

[39] Anne P. Hillstrom and Steven Yantis. 1994. Visual motion and attentional capture. *Percept. Psychophys.* 55, 4 (1994), 399–411. DOI:https://doi.org/10.3758/BF03205298

[40] Olivier Houdé, Laure Zago, Emmanuel Mellet, Sylvain Moutier, Arlette Pineau, Bernard Mazoyer, and Nathalie Tzourio-Mazoyer. 2000. Shifting from the Perceptual Brain to the Logical Brain: The Neural Impact of Cognitive Inhibition Training. *J. Cogn. Neurosci.* 12, 5 (2000), 721–728. DOI:https://doi.org/10.1162/089892900562525

[41] Jukka Hyönä. 2010. The use of eye movements in the study of multimedia learning. *Learn. Instr.* 20, 2 (April 2010), 172–176. DOI:https://doi.org/10.1016/j.learninstruc.2009.02.013

[42] Robin Jacob and Julia Parkinson. 2015. The Potential for School-Based Interventions That Target Executive Function to Improve Academic Achievement: A Review. *Rev. Educ. Res.* 85, 4 (2015), 512–552. DOI:https://doi.org/10.3102/0034654314561338

[43] Aidan Jones and Ginevra Castellano. 2018. Adaptive Robotic Tutors that Support Self-Regulated Learning: A Longer-Term Investigation with Primary School Children. *Int. J. Soc. Robot.* 10, 3 (2018), 357–370. DOI:https://doi.org/10.1007/s12369-017-0458-z

[44] Taejin Jung, Jiancheng Huang, Linda Eagan, and Diane Oldenburg. 2019. Influence of school-based nutrition education program on healthy eating literacy and healthy food choice among primary school children. *Int. J. Heal. Promot. Educ.* 57, 2 (2019), 67–81. DOI:https://doi.org/10.1080/14635240.2018.1552177

[45] Marcel A. Just and Patricia A. Carpenter. 1980. A theory of reading: From eye fixations to comprehension. *Psychol. Rev.* 87, 4 (1980), 329–354. DOI:https://doi.org/10.1037//0033-295x.87.4.329

[46] Jesper Juul. 2009. Fear of failing? the many meanings of difficulty in video games. In *The video game theory reader 2*, Mark J. P. Wolf and Bernard Perron (eds.). Routledge, New York, 237–252. DOI:https://doi.org/10.1017/CBO9781107415324.004

[47] Daniel Kahneman. 2011. *Thinking, Fast and Slow*. Penguin Press, New York.

[48] Fengfeng Ke, Valerie Shute, Kathleen M. Clark, and Gordon Erlebacher. 2019. *Interdisciplinary Design of Game-based Learning Platforms*. DOI:https://doi.org/10.1007/978-3-030-04339-1

[49] Harri Ketamo and Kristian Kiili. 2010. Conceptual Change Takes Time: Game Based Learning Cannot be Only Supplementary Amusement. *J. Educ. Multimed. Hypermedia* 19, 4 (2010), 399–419. Retrieved from http://ezproxy.lib.swin.edu.au/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=iih&AN=60012400&site=ehost-live&scope=site

[50] Asma Ben Khedher, Imène Jraidi, and Claude Frasson. 2018. Static and dynamic eye movement metrics for students' performance assessment. *Smart Learn. Environ.* 5, 1 (2018). DOI:https://doi.org/10.1186/s40561-018-0065-y

[51] Kristian Kiili, Harri Ketamo, and Michael D Kickmeier-rust. 2014. Eye Tracking in Game-based Learning Research and Game Design. *Int. J. Serious Games* 1, 2 (2014), 51–65.

[52] Daniel King, Paul Delfabbro, and Mark Griffiths. 2009. Video Game Structural Characteristics: A New Psychological Taxonomy. *Int. J. Ment. Health Addict.* 8, 1 (April 2009), 90–106. DOI:https://doi.org/10.1007/s11469-009-9206-4

[53] Torkel Klingberg. 2010. Training and plasticity of working memory. *Trends Cogn. Sci.* 14, 7 (2010), 317–324. DOI:https://doi.org/10.1016/j.tics.2010.05.002

[54] Joy Yeonjoo Lee, Jeroen Donkers, Halszka Jarodzka, and Jeroen J.G. van Merriënboer. 2019. How prior knowledge affects problem-solving performance in a medical simulation game: Using game-logs and eye-tracking. *Comput. Human Behav.* 99, May (2019), 268–277. DOI:https://doi.org/10.1016/j.chb.2019.05.035

[55] Angeliki Leonardou, Maria Rigou, and John Garofalakis. 2019. Opening user model data for motivation and learning: The case of an adaptive multiplication game. *CSEDU 2019 - Proc. 11th Int. Conf. Comput. Support. Educ.* 1, Csedu (2019), 383–390. DOI:https://doi.org/10.5220/0007735603830390

[56] Denis Mareschal. 2016. The neuroscience of conceptual learning in science and mathematics. *Curr. Opin. Behav. Sci.* 10, (2016), 114–118. DOI:https://doi.org/10.1016/j.cobeha.2016.06.001

[57] megan M. McClelland, Claire E. Cameron, Shannon B. Wanless, and Amy Murray. 2007. Executive function, behavioural self-regulation, and social-emotional competence. In *Comtemporary Perspectives on Social learning in Early Childhood Education*, Olivia Saracho and Bernard Spodek (eds.). IAP, 83–107.

[58] Susanne M.M. de Mooij, Natasha Z. Kirkham, Maartje E.J. Raijmakers, Han L.J. van der Maas, and Iroise Dumontheil. 2020. Should online math learning environments be tailored to individuals' cognitive profiles? *J. Exp. Child Psychol.* 191, (2020), 104730. DOI:https://doi.org/10.1016/j.jecp.2019.104730

[59] Harold G. Nelson and Erik Stolterman. *The Design Way: Intentional Change in an Unpredictable World*. Educational Technology Publications, Englewood Cliffs, New Jersey.

[60] John L. Nietfeld. 2017. The Role of Self-Regulated Learning in Digital Games. In *Handbook of Self-Regulation of Learning and Performance*, D.H. Schunk and J. A. Greene (eds.). Routledge, Abingdon, Oxon, 271–284. DOI:https://doi.org/10.4324/9781315697048-18

[61] Harold E. Petersen and Doris J. Dugas. 1972. The Relative Importance of Contrast and Motion in Visual Detection. *Human Factors: The Journal of Human Factors and Ergonomics Society 14*, 207–216. DOI:https://doi.org/10.1177/001872087201400302

[62] Jennifer Preece, Yvonne Rogers, and Helen Sharp. 2002. *Interaction Design: beyond human-computer interaction*. John Wiley & Sons, Ltd. DOI:https://doi.org/10.1007/3-540-34874-3_10

[63] J.-N. Jean-Nicolas Proulx, Margarida Romero, and Sylvester Arnab. 2017. Learning Mechanics and Game Mechanics Under the Perspective of Self-Determination Theory to Foster Motivation in Digital Game Based Learning. *Simul. Gaming* 48, 1 (2017), 81–97. DOI:https://doi.org/10.1177/1046878116674399

[64] Thomas Roderer, Saskia Krebs, Corinne Schmid, and Claudia M. Roebers. 2012. The Role of Executive Control of Attention and Selective Encoding for Preschoolers' Learning. *Infant Child Dev.* 21, (2012), 146–159. DOI:https://doi.org/10.1002/icd

[65] Thomas Roderer and Claudia M. Roebers. 2014. Can you see me thinking (about my answers)? Using eye-tracking to illuminate developmental differences in monitoring and control skills and their relation to performance. *Metacognition Learn.* 9, 1 (2014), 1–23. DOI:https://doi.org/10.1007/s11409-013-9109-4

[66] Claudia M. Roebers, Corinne Schmid, and Thomas Roderer. 2010. Encoding strategies in primary school children: Insights from an eye-tracking approach and the role of individual differences in attentional control. *J. Genet. Psychol.* 171, 1 (2010), 1–21. DOI:https://doi.org/10.1080/00221320903300361

[67] Laurence Rousselle, Emmanuelle Palmers, and Marie-Pascale Noël. 2004. Magnitude comparison in preschoolers: what counts? Influence of perceptual variables. *J. Exp. Child Psychol.* 87, 1 (2004), 57–84. DOI:https://doi.org/10.1016/j.jecp.2003.10.005

[68] Palak Roy, Simon Rutt, Claire Easton, David Sims, Sally Bradshaw, and Stephen McNamara. 2019. *Stop and Think: Learning Counterintuitive Concepts*. Retrieved from www.educationendowmentfoundation.org.uk

[69] M Schiessl, S Duda, a Thölke, and R Fischer. 2003. Eye tracking and its application in usability and media research. *MMI-interaktiv J.* July (2003), 1–10. Retrieved from http://www.researchgate.net/publication/26411583_Eye_tracking_and_its_application_in_usability_and_media_research/file/60b7d51b71597a5db5.pdf

[70] Traci Sitzmann. 2011. A Meta-Analytic Examination of the Instructional Effectiveness of Computer-Based Simulation Games. *Pers. Psychol.* 64, 2 (June 2011), 489–528. DOI:https://doi.org/10.1111/j.1744-6570.2011.01190.x

[71] Ruth Stavy and Reuven Babai. 2010. Overcoming intuitive interference in mathematics: insights from behavioral, brain imaging and intervention studies. *ZDM* 42, 6 (2010), 621–633. DOI:https://doi.org/10.1007/s11858-010-0251-z

[72] Ana Susac, Andreja Bubic, Jurica Kaponja, Maja Planinic, and Marijan Palmovic. 2014. Eye Movements Reveal Students' Strategies in Simple Equation Solving. *Int. J. Sci. Math. Educ.* 12, 3 (2014), 555–577. DOI:https://doi.org/10.1007/s10763-014-9514-4

[73] Tobii Technology. 2015. Tobii Pro Studio 3.3.2. Retrieved from www.tobii.com

[74] Berna Haktanirlar Ulutas, N. Fırat Özkan, and Rafał Michalski. 2019. Application of hidden Markov models to eye tracking data analysis of visual quality inspection operations. *Cent. Eur. J. Oper. Res.* (2019). DOI:https://doi.org/10.1007/s10100-019-00628-x

[75] Rojin Vishkaie. 2020. Design in the Pandemic: Building Resilience for the Digital Divide in Education. *Interactions*, 36–37. Retrieved from http://boardgamegeek.com/boardgame/137136/pandemic-in-the-lab

[76] Colin Ware. 2013. *Information Visualization: Perception for Design* (3rd Editio ed.). Elsevier, Inc., Waltham, MA.

[77] Hannah R Wilkinson, Claire Smid, Su Morris, Emily K Farran, Iroise Dumontheil, Sveta Mayer, Andrew Tolmie, Derek Bell, Ka Porayska-pomsta, Wayne Holmes, Denis Mareschal, Michael S C Thomas, and The Unlocke Team. 2019. Domain-Specific Inhibitory Control Training to Improve Children ' s Learning of Counterintuitive Concepts in Mathematics and Science. *J. Cogn. Enhanc.* in press (2019). DOI:https://doi.org/https://doi.org/10.1007/s41465-019-00161-4

[78] Pieter Wouters, Christof van Nimwegen, Herre van Oostendorp, and Erik D. van der Spek. 2013. A meta-analysis of the cognitive and motivational effects of serious games. *J. Educ. Psychol.* 105, 2 (2013), 249–265. DOI:https://doi.org/10.1037/a0031311

[79] Yu chu Yeh, Han Lin Chang, and Szu Yu Chen. 2019. Mindful learning: A mediator of mastery experience during digital creativity game-based

learning among elementary school students. *Comput. Educ.* 132, 64 (2019), 63–75. DOI:https://doi.org/10.1016/j.compedu.2019.01.001

[80] Philip David Zelazo, Clancy B. Blair, and Michael T. Willoughby. 2016. Executive Function: Implications for Education. NCER 2017-2000. *Natl. Cent. Educ. Res.* (2016). Retrieved from https://eric.ed.gov/?id=ED570880%0Ahttps://eric.ed.gov/?q=gender+differences+distraction&pg=2&id=ED570880