Tortoise Attention Algorithm: A Novel Computational Tool for Measuring Children's Concentration

Lingqin Meng[1,2,*], Yang Gao[1,2], Wenjia Tan[4], Junhao Zhang[1,2], Joy Lee[5], Hui Lin[3], Lei Shen[3], Yitao Duan[3], Stella Christie[1,2]
[1]Department of Psychological and Cognitive Sciences, Tsinghua University, Beijing, China
[2]Tsinghua Laboratory of Brain and Intelligence, Tsinghua University, Beijing, China
[3]NetEase Youdao, Beijing, China
[4]Department of Computer and Information Science, University of Macau, Macao, China
[5]Graduate School of Education, Peking University, Beijing, China
[*] Correspondence to *menglq24@mails.tsinghua.edu.cn*

Abstract

This study introduces the tortoise attention algorithm (TAA), a novel computational tool for accurately measuring children's concentration during learning activities. The algorithm integrates weighted behavior duration and temporal stability metrics to calculate a comprehensive concentration score. We conducted three studies to evaluate the effectiveness of the algorithm. Study 1 demonstrated that TAA's concentration scores significantly predicted performance on math tasks but found no significant relationship with executive function performance, as measured by the ANT test. Study 2 revealed that TAA outperformed humans in predicting children's math performance, underscoring the algorithm's ability to mitigate biases inherent in human assessments of fidgeting behaviors. Study 3 further showed that while human evaluations of concentration were consistent across classroom and café settings, they failed to align with actual learning outcomes. These findings highlight that TAA provides an objective and reliable tool for evaluating concentration, enabling educators to refine teaching strategies and improve learning outcomes.

Keywords: fidgeting behavior; tortoise attention algorithm; concentration score; temporal stability; learning performance

Introduction

Slow but steady wins the race.
— Aesop's Fables, The Tortoise and the Hare

Concentration is a critical indicator of learning, commonly evaluated through observable behaviors such as fidgeting, mind wandering, and other off-task activities (Newmann, 1992; Donnelly et al., 2016; Yang, Zhao, Tian, & Xing, 2021). Fidgeting, in particular, is often perceived as a manifestation of inattention or spontaneous mind wandering (Carriere, Seli, & Smilek, 2013). Teachers and parents frequently encourage children to remain physically still and focused, assuming that such behavior reflects cognitive engagement and improves academic performance (Pas & Bradshaw, 2014; Novita, Schönmoser, & Lipowska, 2023; Mack, Scherrer, & Preckel, 2024). However, does empirical evidence substantiate this widely held assumption?

Prior research reveals a complex and often contradictory relationship between fidgeting and academic performance (Perrykkad & Hohwy, 2020). Some studies suggest that frequent fidgeting correlates with lower educational performance. For instance, Merrell and Tymms demonstrated that excessive fidgeting in children aged 4 to 6 significantly impaired their reading and mathematical abilities (Merrell & Tymms, 2001). Similarly, Hulac et al. observed that fidgeting behaviors in college students reduced cognitive efficiency during mathematical tasks (Hulac, Aspiranti, Kriescher, Briesch, & Athanasiou, 2021). These findings reinforce the common perception that fidgeting is a sign of distraction.

However, some studies showed the opposite results: that fidgeting may in fact, facilitate learning. For example, students with attention-deficit/hyperactivity disorder (ADHD) benefit from using fidget tools, which enhance sustained attention and classroom behavior (Aspiranti & Hulac, 2022). Additionally, researchers have found a positive association between fidgeting and performance in some cognitive tasks, such as working memory, particularly in "tip-of-the-tongue" situations (Pine, Bird, & Kirk, 2007; Sarver, Rapport, Kofler, Raiker, & Friedman, 2015). Rickman et al. suggested that fidgeting behaviors, such as doodling, enhance sustained attention and memory retention (Rickman, Johnson, & Miles, 2013; Andrade, 2010). However, not all evidence supports the beneficial effects of fidgeting; some studies have shown that fidgeting can impair memory retention or have no measurable impact on learning performance, depending on the task and population studied (Farley, Risko, & Kingstone, 2013; Graziano, Garcia, & Landis, 2020). Another study found that fidgeting during educational video-watching negatively affected memory retention in undergraduate students (Soares & Storm, 2020).

These mixed findings raise a fundamental question: why do different tasks yield varying relationships between fidgeting and learning outcomes? One potential explanation is that many previous works employed artificial fidgeting behavior rather than using naturally occurring ones (Brown, 1992). Most studies have commonly opted to induce fidgeting through tools such as fidget spinners, as these controlled environments provide a standardized and replicable framework for examining its effects (Burnet et al., 2021; Koiler, Schimmel, Bakhshipour, Shewokis, & Getchell, 2022; Narukawa et al., 2023). However, this emphasis on artificially induced fidgeting neglects the complexities of spontaneous fidgeting in natural contexts, which may more accurately represent real-world learning behaviors (Markowitz et al., 2023; Beaurenaut, Kovarski, Destais, Mennella, & Grèzes, 2024). For example, natural fidgeting may indicate various mental states, such as inattention, boredom, or arousal regulation, and serve different functions depending on the context (Lis et

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

al., 2010; Carriere et al., 2013; Spencer-Mueller & Fenske, 2024). Hence, when studies artificially manipulated fidgeting, the same imposed behavior may serve different cognitive functions—for instance, causing inattention in some cases while helping students reduce stress and enhance focus in others, thereby limiting the external validity of these findings (Perrykkad & Hohwy, 2020).

Thus, to understand accurately whether fidgeting helps or hinders concentration and learning, one must measure naturally occurring fidgeting behavior. However, studying natural fidgeting presents significant challenges. First, observing and coding such behaviors is labor-intensive, often requiring human coders to analyze fidgeting in classroom settings (Godwin et al., 2016). Second, quantifying natural fidgeting is difficult because the parameters that best indicate concentration are not well-defined. For example, researchers often use eye gaze as a proxy for attention, but it does not consistently reflect concentration—children may maintain eye contact with teachers while mind-wandering (Henderson & Ferreira, 2013). Moreover, subjective questionnaires often fail to predict fidgeting accurately, and fMRI results are frequently affected by motion artifacts from spontaneous head movements (Lis et al., 2010; Engelhardt et al., 2017). These challenges highlight the critical need for developing an innovative tool to measure fidgeting and concentration accurately.

Current Research

Here, we propose a novel computational tool for quantifying fidgeting and concentration during learning activities, named the tortoise attention algorithm (TAA). TAA uses the principles of temporal stability analysis, weighted behavior modeling, and Bayesian optimization. A crucial benchmark for TAA is to know whether its quantification of fidgeting-concentration actually correlates with learning outcomes. To do this, in Study 1 we asked children aged 7-13 years to do a math test and an Executive Function (EF) task, with the algorithm generated corresponding concentration (fidgeting) scores. The question is whether TAA's score correlates with children's math and EF performance. In Studies 2 and 3, we extended this benchmarking by comparing TAA with human perceptions of fidgeting. Specifically, we investigated whether adults could accurately predict learning outcomes based on children's fidgeting behaviors and how does the accuracy of human judgment compare to that of TAA.

Tortoise Attention Algorithm

TAA processes a dataset of thousands of labeled videos and recordings of children engaged in natural learning activities. It recognizes nine distinct behaviors: studying paper materials, interacting with a computer (including screen viewing and keyboard/mouse operations), using a phone, using a tablet, resting on the desk, leaving the seat, playing with toys, looking around, and eating or drinking. These behaviors were selected based on established literature linking specific actions to attention allocation in educational settings (Godwin et al., 2016; Trabelsi, Alnajjar, Parambil, Gochoo, & Ali, 2023; Zhang et al., 2023). TAA organizes the raw data into a time-stamped matrix, with each row representing an observed behavior at a specific moment.

The algorithm also introduces a smoothing module, aggregating the frequent behaviors within a 10 to 30-second sliding window to reduce noise and provide a reliable basis for analysis (Zhu, Huang, Xue, Mihaylova, & Chambers, 2022). The algorithm calculates the concentration score using two key metrics: the weighted average of behavior durations and the temporal stability of attention, combining them to create a comprehensive and adaptable measure of concentration across various learning scenarios (Killeen & Fetterman, 1988; Mayes, Gordon, Calhoun, & Bixler, 2014).

Calculation of Concentration Score

TAA integrates the weighted average of behavior durations (F_{avg}) and temporal stability of attention (S_{coef}) to derive a comprehensive concentration score. F_{avg} measures concentration by giving more weight to focus-related behaviors and less to distractions. The equation is given by:

$$F_{avg} = \frac{\sum(t_n \cdot V_n)}{y} \tag{1}$$

where t_n is the duration of behavior n, V_n represents the weight assigned to behavior n, and y is the total session duration.

The assignment of behavior weights (V_n) in TAA reflects insights from studies on attention and cognitive performance, which emphasize the differential impacts of various activities on sustained attention and task performance (Schweizer, Moosbrugger, & Goldhammer, 2005; Carver & Scheier, 2012; Lawson, Parrinello, & Ruff, 1992). The initial weights for each behavior were determined heuristically, guided by their observed relevance to positive learning outcomes. Behaviors associated with higher cognitive engagement, such as studying paper-based materials, are assigned higher weights; in contrast, distractive actions, including engaging with electronic devices, are assigned lower weights (Campbell, 1993; Permzadian & Credé, 2016).

S_{coef} measures the temporal stability of attention by evaluating the duration and distribution of uninterrupted focus blocks during a session. It accounts for the most extended continuous focus block ($Base_{score}$) and additional significant focus periods (Add_{score}), offering a comprehensive evaluation of focus quality. The formula for S_{coef} is:

$$S_{coef} = min\left[1, Base_{score} + Add_{score}\right] \tag{2}$$

In equation (2), $S_{coef} \in [0,1]$, where $Base_{score}$ is defined as:

$$Base_{score} = \begin{cases} c_1 \cdot K + c_2, & if \;\; y \geq T \\ \frac{K}{y} + c_3, & if \;\; y < T \end{cases} \tag{3}$$

where K represents the duration of the most extended uninterrupted focus block (in minutes), T is the session duration threshold, and c_1, c_2 and c_3 are constants.

This formulation integrates the absolute duration of the primary focus block (K) and its relative importance within the total session duration (y). For extended sessions ($y \geq T$), the score highlights the cumulative impact of K through a linear relationship ($c_1 \cdot K + c_2$), emphasizing the significance of sustained engagement. In shorter sessions ($y < T$), K is normalized by y, ensuring proportional evaluation relative to the session length. The c_3 stabilizes the score, preventing over-penalization in brief sessions.

TAA incorporates Add_{score} to reward focus blocks exceeding a predefined threshold duration. Based on the speed-accuracy trade-off model, the weight of Add_{score} is dynamically adjusted as a hyperparameter to capture the contributions of additional focus blocks in varying session contexts (Lo & Wang, 2006; Schultz, 2015). Add_{score} is defined as:

$$Add_{score} = \sum_{n=1}^{N} (c_4 \cdot K_n + c_5) \cdot \mathbb{I}(K_n > t_{thr}) \tag{4}$$

In equation (4), K_n represents the duration of the n-th focus block, and $\mathbb{I}(K_n > t_{thr})$ is an indicator function that equals 1 if $K_n > t_{thr}$, and 0 otherwise. The term c_4 quantifies the proportional contribution of block duration, calibrated based on its empirical relationship with concentration stability, while c_5 provides a baseline reward to include shorter yet meaningful blocks. Finally, t_{thr} denotes the minimum focus block duration required for consideration.

The final concentration score integrates F_{avg} and S_{coef} to provide a robust and comprehensive evaluation of concentration, defined as:

$$F_{score} = F_{avg} \cdot S_{coef} \tag{5}$$

This score ranges from 0 to 1, capturing the intensity of concentration and the temporal stability across the session. By integrating weighted averages with stability metrics, TAA provides a nuanced and adaptable representation of children's concentration, suitable for diverse learning scenarios.

Bayesian Optimization

To optimize TAA parameters, we applied Bayesian optimization to maximize the correlation between the calculated F_{score} and children's learning outcomes (ξ), such as test scores (Frazier, 2018; X. Wang, Jin, Schmitt, & Olhofer, 2023).

$$\mathcal{L}(\theta) = Corr\left(F_{\text{score}}(\theta), \xi\right) \tag{6}$$

where $\theta = \{c_1, c_2, c_3, c_4, c_5, t_{thr}\}$ represents the parameter set.

The surrogate model $g(\theta)$ is modeled as a Gaussian Process (GP), expressed as:

$$g(\theta) \sim GP\left(\mu(\theta), k\left(\theta, \theta'\right)\right) \tag{7}$$

In equation (7), $g(\theta)$ is defined by a mean function $\mu(\theta)$ and a kernel (covariance) function $k\left(\theta, \theta'\right)$. Specifically, $\mu(\theta)$ represents the prior expectation of $g(\theta)$ over the parameter space θ, serving as the baseline prediction before incorporating observed data. $k\left(\theta, \theta'\right)$ quantifies the similarity between two parameter sets θ and θ', determining the covariance structure of the GP. We adopt the Radial Basis Function (RBF) kernel:

$$k\left(\theta, \theta'\right) = \sigma^2 \exp\left(-\frac{\left\|\theta - \theta'\right\|^2}{2l^2}\right) \tag{8}$$

where σ^2 is the signal variance, l is the length scale.

The acquisition function, based on Expected Improvement (EI), selects the next parameter set for evaluation:

$$a(\theta) = \mathbb{E}\left[\max\left(0, g(\theta) - g\left(\theta^+\right)\right)\right] \tag{9}$$

Finally, the surrogate model predicts the current best value as $g\left(\theta^+\right)$. The next parameter set θ_{t+1} is selected by maximizing the acquisition function:

$$\theta_{t+1} = \arg\max_{\theta} a(\theta) \tag{10}$$

Before implementing TAA in our studies, we conducted a pilot evaluation to validate its accuracy in detecting fidgeting behaviors. Three laboratory research assistants performed specific fidgeting behaviors while studying during a 5-minute assessment to test the algorithm's detection capabilities. Results showed that the TAA accurately identified most behaviors, such as playing with pens, using mobile phones, touching hair, and looking around, but struggled with subtle actions like doodling or nail-biting. This preliminary validation confirmed TAA's capability to detect the most relevant fidgeting behaviors while identifying specific limitations in detecting subtle movements—insights that guided our methodology and interpretation.

Study 1

Study 1 investigates the effectiveness of TAA in measuring children's concentration during learning tasks. Specifically, we asked whether the concentration scores generated by the algorithm are associated with performance on tasks requiring sustained attention and executive functioning, such as the International Mathematics Assessment System (IMAS) and the Attention Network Test (ANT). Building on the established link between attention and learning performance, we hypothesize that higher concentration scores will correlate positively with better outcomes in these learning tasks (Bull & Scerif, 2001; Blair & Raver, 2015; Bull & Lee, 2014).

Method

We experimented at the Tsinghua Child Cognition Center, collaborating with Youdao. We recruited 35 children aged between 7 and 13 years (M = 9.05, SD = 2.13) by disseminating experiment-related information through our recruitment platform. Each participant sat at a standardized desk equipped with a camera-integrated TAA lamp (Figure 1) while a second camera in the corner recorded the entire session. We tested the accuracy of TAA in measuring fidgeting behaviors by using it to calculate concentration scores during two tasks: the IMAS and the ANT. We chose the IMAS because researchers frequently use it to assess overall learning performance, and its design aligns with children's grade levels and abilities (Dubuc, Aubertin-Leheudre, & Karelis, 2020; Clements & Sarama, 2020). We employed the ANT to assess three distinct attentional networks in children: alerting, orienting, and executive control (Fan, McCandliss, Sommer, Raz, & Posner, 2002; Blankenship et al., 2019). The ANT utilized in this study is a shortened adaptation of the original version from Jin Fan's website in 2005 (Fan, McCandliss, Fossella, Flombaum, & Posner, 2005).

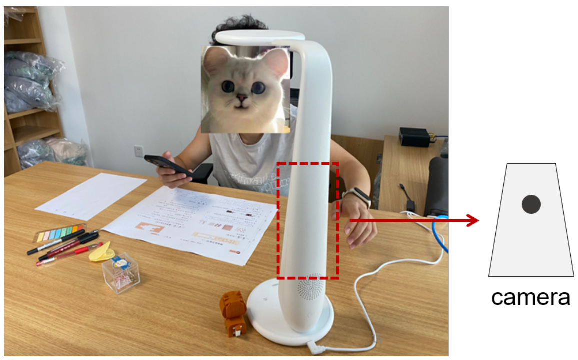

Figure 1: Experimental setup for Study 1. The desk lamp, positioned in front of the child, incorporates a camera that records in real-time and utilizes TAA to calculate the child's concentration score.

Each child completed two tasks: a 30-minute paper-based IMAS and a 15-minute tablet-based ANT, designed to assess math performance and executive function, respectively. Participants entered a comfortable room with a neatly arranged desk and a cozy sofa. The desk contained a two-page mathematics assessment and an A4 sheet for calculations or notes. Positioned on the left side, a TAA-integrated lamp recorded movement patterns, subsequently quantifying these into fidgeting metrics with optimal visual calibration. Five manipulable items (ruler, writing instruments, Lego figurine, hair accessory, and paper clip container) occupied the desk to enhance the naturalistic quality of the experimental environment and provide interactive stimuli during the assessment. The experimenter introduced the experimental protocol, seated the participant, set a 30-minute timer, and exited the room. Upon returning after 30 minutes, the experimenter provided a short break before introducing the ANT, explaining the instructions, and guiding the participant through a practice session. After ensuring participant comprehension, the experimenter again left the room, allowing the official ANT task to commence.

Results and Discussion

We first evaluated the accuracy of children on the IMAS (children's math scores) and the ANT (children's EF). Next, we analyzed the relationship between concentration scores (TAA-Concentration) and their performance on these tests. We found a significant positive correlation between TAA-Concentration and children's math scores ($r = 0.451$, $p = 0.009$), which indicates that higher TAA-concentration scores are associated with better math performance.

Our analysis revealed that S_{coef} is the primary driver of the significant correlation between TAA-Concentration and math performance ($r = 0.509$, $p = 0.005$). This parameter integrates absolute and relative concentration metrics, allowing it to dynamically adapt to varying session lengths while filtering out noise from brief attentional fluctuations (Altmann & Kamide, 2007). This adaptability aligns well with the cognitive demands of mathematics, which require sustained focus for complex reasoning and multi-step problem-solving. Prior research supports this relationship: Blair and Raver demonstrated that extended attention is critical for success in analytical tasks (Blair & Raver, 2015), while Sarver et al. showed that sustained attention enhances task monitoring and persistence (Sarver et al., 2015).

In contrast, we found no significant correlation between TAA metrics and children's executive functioning as measured by ANT. This divergence likely stems from fundamental differences in task demands: IMAS requires a sustained focus that aligns with TAA's measurement approach, while ANT emphasizes rapid attentional switching and inhibitory control across brief trials—cognitive mechanisms that operate on significantly shorter timescales. Future research might investigate this hypothesis by incorporating additional executive function measures that require sustained cognitive control, thereby clarifying the specific dimensions of attention that TAA most effectively captures.

Given the established link between EF and math performance, we examined this relationship in our sample (Bull & Lee, 2014). As expected, we found a positive correlation between children's EF accuracy and math scores ($r = 0.386$, $p = 0.048$), and a negative correlation between EF reaction time and math scores ($r = -0.556$, $p = 0.003$). These findings confirm that both components of EF performance relate to mathematical ability, which may help explain why TAA's focus on sustained attention captures different cognitive aspects than those measured by the ANT task.

Study 2

Study 1 demonstrates that TAA assessment of concentration predicts math performance with relatively high accuracy. To further our benchmarking of TAA, in Study 2, we examined whether humans can accurately predict children's test performance based on their natural behaviors. Our objective is to determine whether human judgments of concentration can reliably predict learning outcomes, an issue not thoroughly addressed in previous research. We also seek to compare the accuracy of the algorithm's predictions with humans.

Method

To achieve this, we extracted 30-second video clips from the 30-minute recordings of each child completing the paper-based IMAS, as evaluated by TAA in Study 1 ($N = 35$). Since it is impractical for human raters to watch the full 30 minutes of footage for each of the 35 children, we selected a representative slice by extracting 30-second clips from the midpoint of each recording. This choice ensures that the selected clips represent typical behavior during the session while reducing cognitive load on the evaluators. The rationale for using short clips was to simulate real-world scenarios where educators often form judgments based on limited observations (Tversky & Kahneman, 1974; Kraus & Keltner, 2009).

We then recruited 593 adults (parents or teachers, $N = 593$; aged 24–35 years, $M = 33.97$, $SD = 3.94$) online from the same recruitment platform to rate each child on a 1–5 scale based on the following questions: (1) "How focused do you think this child is?" (1 = very unfocused, 5 = very focused); (2) "How would you rate this child's posture?" (1 = very poor, 5 = excellent); (3) "How well do you think this child performed on the test?" (1 = very poor, 5 = excellent); and (4) "How much do you think this child fidgeted?" (1 = not at all, 5 = very much). Adults rated each child randomly, and the four evaluation questions appeared randomly.

Results and Discussion

Even from brief 30-second video clips, adults consistently judged whether a child was concentrating, with high internal reliability (Cronbach's $\alpha = 0.957$-0.963). We analyzed the relationship between concentration and fidgeting ratings to verify that fidgeting highly influenced these judgments. As expected, concentration negatively correlated with perceived fidgeting ($r = -0.905$, $p < 0.001$), confirming that fidgeting was a primary indicator in adults' concentration evaluations. Similarly, adults believed that children who concentrated more (i.e., fidgeted less) would perform better on math tests, shown by the strong correlation between adult concentration scores and predicted math performance ($r = 0.974$, $p < 0.001$).

Interestingly, this correlation does not hold when compared to children's actual performance in mathematics tasks. Specifically, human ratings of concentration and fidgeting show no significant correlation with children's actual math scores, as shown in Table 1. In contrast, concentration scores generated by TAA accurately predict children's math performance (Study 1), whereas adult ratings fail to do so (Study 2). These findings suggest that humans and TAA perceive the relationship between fidgeting and concentration differently—TAA demonstrates better accuracy in predicting whether a child's fidgeting behavior impacts their learning outcomes, using math performance as a benchmark.

The reliance on thin-slicing—a tendency to form quick impressions based on limited information—compromises the accuracy of human predictions (Ambady & Rosenthal, 1992; Kang & Bodenhausen, 2015; Galesic et al., 2021). This discrepancy may stem from a limitation in our design: while TAA analyzes the full video, human raters are restricted to a 30-second clip. This difference likely accounts for the performance gap, suggesting that human predictions could improve with longer observation periods. Future studies might explore whether extending observation time or using multiple segments could help reduce the impact of thin-slicing (Kirillov et al., 2023).

Study 3

We were somewhat surprised to find that TAA outperformed humans in accuracy, prompting us to investigate whether human judgments of fidgeting vary across different contexts (Goethe, Sørum, & Johansen, 2022). Previous research indicates that individuals rely on familiar environmental cues to form judgments when evaluating others' behavior (Gilovich, Griffin, & Kahneman, 2002; Kraus & Keltner, 2009; Mellers, McCoy, Lu, & Tetlock, 2024). Classroom settings offer a structured and familiar environment with well-defined behavioral norms, such as sitting still and focusing on tasks (Ames, 1992). We speculate that these norms may help evaluators more readily associate observed behaviors with concentration. Informal settings like cafés introduce more variability in behavioral cues, such as movement or environmental interactions, which might challenge humans' ability to interpret these behaviors accurately. To test this hypothesis, we conducted Study 3, recruiting a new group of adults to rate the same 35 videos used in Studies 1 and 2, but with the background altered to depict either a classroom or a café.

Method

A new group of parents ($N = 657$, 90% female, $M = 34.49$, $SD = 4.19$) and teachers ($N = 109$, 92.7% female, $M = 34.57$, $SD = 4.24$), recruited from the same platform, rated the same videos as in Study 2, but with modified backgrounds depicting either classrooms or cafés. Parents' firstborn children averaged 5.42 years of age ($SD = 3.01$), with 55.1% boys and 44.9% girls. Most teachers (98.2%) had children of their own. Researchers increasingly recognize cafés as popular venues for after-school studying, making them a suitable comparison group (Purwadi & Manurung, 2020). Each participant rated six 30-second videos randomly, including three with a café setting and three with a classroom setting, to partially control for rating variations among adult evaluators.

Results and Discussion

Teachers and parents show no differences in their judgments of concentration and fidgeting. These judgments remain consistent across both café and classroom contexts. Similar to the findings in Study 2, these judgments do not accurately predict actual math scores (café: parents, $t = 1.029$, $p = 0.312$; teachers, $t = 0.461$, $p = 0.648$; classroom: parents, $t = -1.424$, $p = 0.165$; teachers, $t = -1.490$, $p = 0.146$).

The findings from Study 3 reveal that human evaluations of concentration and fidgeting remain consistent across classroom and café settings. This suggests that teachers and parents may rely on fixed heuristics—such as equating stillness with concentration—when evaluating children's attention, irrespective of the contextual setting (Gilovich et al., 2002; Oeberst & Imhoff, 2023). This reliance could obscure subtle

Table 1: Correlation between TAA-Concentration and other variables. ($*: p < 0.05$, $**: p < 0.01$)

	TAAC	TAAS	HRC	HRF	HRP	CMS	CEFA	CEFRT
TAA-Concentration (TAAC)	1							
TAA-S_{coef} (TAAS)	0.970**	1						
Human-Rated Concentration (HRC)	-0.08	-0.15	1					
Human-Rated Fidgeting (HRF)	0	0.1	-0.905**	1				
Human-Rated Performance (HRP)	-0.081	-0.143	0.974**	-0.890**	1			
Children's Math Scores (CMS)	0.451*	0.509**	-0.029	0.1	-0.068	1		
Children's EF Accuracy (CEFA)	0.296	0.305	0.099	0.019	0.127	0.386*	1	
Children's EF Reaction Time (CEFRT)	-0.151	-0.16	0.038	-0.106	0.046	-0.556**	-0.101	1

behavioral variations that arise in different learning environments (Sisk, Remington, & Jiang, 2019).

General Discussion

We developed TAA as a tool to aid learning. Given the wideheld concern of educators—whether parents at homes or teachers at schools—to make sure that students are "concentrating" during their studies, we developed an algorithm that can measure naturally occurring fidgeting behaviors. The algorithm identifies nine distinct activity patterns from behavioral data and calculates a concentration score based on duration and stability. We fine-tuned its parameters using Bayesian optimization, ensuring a strong correlation with children's learning outcomes. Our findings indicate that TAA is relatively effective, with its concentration scores significantly correlating with children's actual math scores, using IMAS as a benchmark. Interestingly, when comparing algorithmic predictions with human assessments, we found that humans were notably less accurate in predicting children's performance based on observations of concentration and fidgeting. Specifically, adults often equate fidgeting with a lack of concentration, yet our results suggest that this assumption does not necessarily translate to poorer math performance.

A key finding of this research is the significant role of S_{coef} in driving the predictive power of TAA's concentration scores. This result underscores the importance of maintaining extended periods of focused engagement, particularly in tasks requiring analytical reasoning and problem-solving. Unlike human evaluations that often rely on momentary behavioral cues, S_{coef} provides a more objective and reliable measure of sustained concentration by balancing absolute concentration duration with its proportional significance relative to the session length, particularly in structured tasks like mathematics.

While S_{coef} demonstrates predictive validity, the concentration score formula warrants refinement. Although F_{avg} effectively differentiates focus-related from distraction-related behaviors, its multiplicative interaction with S_{coef} may constrain measurement precision. This equal weighting assumption overlooks potential task-dependent variations in relative importance. For instance, in time-constrained tasks, the frequency and duration of focus-related behaviors captured by F_{avg} might carry greater significance, and S_{coef} could play a more dominant role in tasks demanding sustained effort (Kurzban, Duckworth, Kable, & Myers, 2013; Iberl & Ulrich, 2023). Therefore, we could introduce an adjustable weighting scheme calibrated using empirical data, allowing the concentration score to more accurately reflect the distinct cognitive demands of various tasks and align the formula more closely with observed performance patterns (F. Wang et al., 2021).

TAA shows promise as an educational tool, particularly given our finding that humans often misjudge children's concentration levels. However, several limitations merit attention before implementation. While TAA predicts math performance better than human observers, it requires validation across more diverse samples spanning different ages, cultural backgrounds, and cognitive profiles. Children use varied fidgeting patterns for attention regulation, demanding consideration of individual differences.

The current implementation of TAA can identify most behaviors with reasonable concordance with human coding, but struggles with complex multitasking scenarios and certain nuanced actions. For instance, some children's data showed 'eating/drinking' codes when children placed fingers near their mouths, while children touching their chins while concentrating were misclassified with lower-weighted 'face-touching' behaviors. We also observed that 'looking at the camera' constituted a high proportion of looking-around behaviors, though the relationship between camera-glancing and concentration remains under investigation. These technical issues highlight the optimization needed before translating TAA into practical products.

TAA embodies principles analogous to Aesop's fable—slow but steady wins the race, and effective attention measurement requires persistence and consistency. Rather than privileging rapid, flashy performances reminiscent of the hare's initial speed, TAA prioritizes sustained engagement and temporal stability as high-weight predictors of learning outcomes. The parallel between Aesop's allegory and educational assessment illuminates a crucial insight—academic success often stems not from conspicuous behavioral indicators but from sustained attentional persistence that genuinely facilitates learning.

Acknowledgments

We thank anonymous reviewers for their valuable comments. We also thank Jinyun Lyu and Xiaojie Tian for insightful discussions. This work was funded by NetEase Youdao, Tsinghua University Grant 20222000227. Junhao Zhang is currently employed at Iflytek Co., Ltd., but this work was conducted at Tsinghua University.

References

Altmann, G. T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of memory and language*, *57*(4), 502–518.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological bulletin*, *111*(2), 256.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of educational psychology*, *84*(3), 261.

Andrade, J. (2010). What does doodling do? *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *24*(1), 100–106.

Aspiranti, K. B., & Hulac, D. M. (2022). Using fidget spinners to improve on-task classroom behavior for students with adhd. *Behavior Analysis in Practice*, *15*(2), 454–465.

Beaurenaut, M., Kovarski, K., Destais, C., Mennella, R., & Grèzes, J. (2024). Spontaneous instrumental approach-avoidance learning in social contexts in autism. *Molecular Autism*, *15*(1), 33.

Blair, C., & Raver, C. C. (2015). School readiness and self-regulation: A developmental psychobiological approach. *Annual review of psychology*, *66*(1), 711–731.

Blankenship, T. L., Slough, M. A., Calkins, S. D., Deater-Deckard, K., Kim-Spoon, J., & Bell, M. A. (2019). Attention and executive functioning in infancy: Links to childhood executive function and reading achievement. *Developmental science*, *22*(6), e12824.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The journal of the learning sciences*, *2*(2), 141–178.

Bull, R., & Lee, K. (2014). Executive functioning and mathematics achievement. *Child Development Perspectives*, *8*(1), 36–41.

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental neuropsychology*, *19*(3), 273–293.

Burnet, K., Blackwell, J., Kelsch, E., Hanson, E. D., Stone, K., Fryer, S., … Stoner, L. (2021). Cerebrovascular function response to prolonged sitting combined with a high-glycemic index meal: A double-blind, randomized cross-over trial. *Psychophysiology*, *58*(8), e13830.

Campbell, J. (1993). A theory of performance. *Personnel selection in organizations/Jossey-Bass*.

Carriere, J. S., Seli, P., & Smilek, D. (2013). Wandering in both mind and body: individual differences in mind wandering and inattention predict fidgeting. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *67*(1), 19.

Carver, C. S., & Scheier, M. F. (2012). *Attention and self-regulation: A control-theory approach to human behavior*. Springer Science & Business Media.

Clements, D. H., & Sarama, J. (2020). *Learning and teaching early math: The learning trajectories approach*. Routledge.

Donnelly, J. E., Hillman, C. H., Castelli, D., Etnier, J. L., Lee, S., Tomporowski, P., … Szabo-Reed, A. N. (2016). Physical activity, fitness, cognitive function, and academic achievement in children: a systematic review. *Medicine and science in sports and exercise*, *48*(6), 1197.

Dubuc, M.-M., Aubertin-Leheudre, M., & Karelis, A. D. (2020). Lifestyle habits predict academic performance in high school students: The adolescent student academic performance longitudinal study (asap). *International journal of environmental research and public health*, *17*(1), 243.

Engelhardt, L. E., Roe, M. A., Juranek, J., DeMaster, D., Harden, K. P., Tucker-Drob, E. M., & Church, J. A. (2017). Children's head motion during fmri tasks is heritable and stable over time. *Developmental cognitive neuroscience*, *25*, 58–68.

Fan, J., McCandliss, B. D., Fossella, J., Flombaum, J. I., & Posner, M. I. (2005). The activation of attentional networks. *Neuroimage*, *26*(2), 471–479.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of cognitive neuroscience*, *14*(3), 340–347.

Farley, J., Risko, E. F., & Kingstone, A. (2013). Everyday attention and lecture retention: the effects of time, fidgeting, and mind wandering. *Frontiers in psychology*, *4*, 619.

Frazier, P. I. (2018). A tutorial on bayesian optimization. *arXiv preprint arXiv:1807.02811*.

Galesic, M., Bruine de Bruin, W., Dalege, J., Feld, S. L., Kreuter, F., Olsson, H., … van Der Does, T. (2021). Human social sensing is an untapped resource for computational social science. *Nature*, *595*(7866), 214–222.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge university press.

Godwin, K. E., Almeda, M. V., Seltman, H., Kai, S., Skerbetz, M. D., Baker, R. S., & Fisher, A. V. (2016). Off-task behavior in elementary school children. *Learning and Instruction*, *44*, 128–143.

Goethe, O., Sørum, H., & Johansen, J. (2022). The effect or non-effect of virtual versus non-virtual backgrounds in digital learning. In *Human interaction, emerging technologies and future systems v: Proceedings of the 5th interna-*

tional virtual conference on human interaction and emerging technologies, ihiet 2021, august 27-29, 2021 and the 6th ihiet: Future systems (ihiet-fs 2021), october 28-30, 2021, france (pp. 274–281).

Graziano, P. A., Garcia, A. M., & Landis, T. D. (2020). To fidget or not to fidget, that is the question: A systematic classroom evaluation of fidget spinners among young children with adhd. *Journal of attention disorders*, *24*(1), 163–171.

Henderson, J., & Ferreira, F. (2013). *The interface of language, vision, and action: Eye movements and the visual world*. Psychology Press.

Hulac, D. M., Aspiranti, K., Kriescher, S., Briesch, A. M., & Athanasiou, M. (2021). A multisite study of the effect of fidget spinners on academic performance. *Contemporary School Psychology*, *25*, 582–588.

Iberl, B., & Ulrich, R. (2023). On estimating the frequency of a target behavior from time-constrained yes/no survey questions: A parametric approach based on the poisson process. *Psychological Methods*.

Kang, S. K., & Bodenhausen, G. V. (2015). Multiple identities in social perception and interaction: Challenges and opportunities. *Annual review of psychology*, *66*(1), 547–574.

Killeen, P. R., & Fetterman, J. G. (1988). A behavioral theory of timing. *Psychological review*, *95*(2), 274.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... others (2023). Segment anything. In *Proceedings of the ieee/cvf international conference on computer vision* (pp. 4015–4026).

Koiler, R., Schimmel, A., Bakhshipour, E., Shewokis, P. A., & Getchell, N. (2022). The impact of fidget spinners on fine motor skills in individuals with and without adhd: An exploratory analysis. *Journal of behavioral and brain science*, *12*(3), 82–101.

Kraus, M. W., & Keltner, D. (2009). Signs of socioeconomic status: A thin-slicing approach. *Psychological science*, *20*(1), 99–106.

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and brain sciences*, *36*(6), 661–679.

Lawson, K. R., Parrinello, R., & Ruff, H. A. (1992). Maternal behavior and infant attention. *Infant Behavior and Development*, *15*(2), 209–229.

Lis, S., Baer, N., Stein-en Nosse, C., Gallhofer, B., Sammer, G., & Kirsch, P. (2010). Objective measurement of motor activity during cognitive performance in adults with attention-deficit/hyperactivity disorder. *Acta Psychiatrica Scandinavica*, *122*(4), 285–294.

Lo, C.-C., & Wang, X.-J. (2006). Cortico–basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nature neuroscience*, *9*(7), 956–963.

Mack, E., Scherrer, V., & Preckel, F. (2024). How smart is my child? the judgment accuracy of parents regarding their children's cognitive ability. *Child Development*.

Markowitz, J. E., Gillis, W. F., Jay, M., Wood, J., Harris, R. W., Cieszkowski, R., ... others (2023). Spontaneous behaviour is structured by reinforcement without explicit reward. *Nature*, *614*(7946), 108–117.

Mayes, S. D., Gordon, M., Calhoun, S. L., & Bixler, E. O. (2014). Long-term temporal stability of measured inattention and impulsivity in typical and referred children. *Journal of Attention Disorders*, *18*(1), 23–30.

Mellers, B. A., McCoy, J. P., Lu, L., & Tetlock, P. E. (2024). Human and algorithmic predictions in geopolitical forecasting: Quantifying uncertainty in hard-to-quantify domains. *Perspectives on Psychological Science*, *19*(5), 711–721.

Merrell, C., & Tymms, P. B. (2001). Inattention, hyperactivity and impulsiveness: Their impact on academic achievement and progress. *British journal of educational psychology*, *71*(1), 43–56.

Narukawa, S., Nishimura, M., Kuze, I., Ohno, I., Fukunaga, M., Kobayasi, K. I., & Murai, S. A. (2023). Cortico-striatal activity associated with fidget spinner use: an fmri study. *Scientific Reports*, *13*(1), 15860.

Newmann, F. M. (1992). *Student engagement and achievement in american secondary schools.* ERIC.

Novita, S., Schönmoser, C., & Lipowska, M. (2023). Parent and teacher judgments about children's mathematics and reading competencies in primary school: Do parent judgments associate with children's educational outcomes? *Learning and Individual Differences*, *105*, 102302.

Oeberst, A., & Imhoff, R. (2023). Toward parsimony in bias research: A proposed common framework of belief-consistent information processing for a set of biases. *Perspectives on Psychological Science*, *18*(6), 1464–1487.

Pas, E. T., & Bradshaw, C. P. (2014). What affects teacher ratings of student behaviors? the potential influence of teachers' perceptions of the school environment and experiences. *Prevention science*, *15*, 940–950.

Permzadian, V., & Credé, M. (2016). Do first-year seminars improve college grades and retention? a quantitative review of their overall effectiveness and an examination of moderators of effectiveness. *Review of Educational Research*, *86*(1), 277–316.

Perrykkad, K., & Hohwy, J. (2020). Fidgeting as self-evidencing: A predictive processing account of non-goal-directed action. *New Ideas in Psychology*, *56*, 100750.

Pine, K. J., Bird, H., & Kirk, E. (2007). The effects of prohibiting gestures on children's lexical retrieval ability. *Developmental Science*, *10*(6), 747–754.

Purwadi, Y. S., & Manurung, E. M. (2020). Cafes: new learning and knowledge production space for millennial students. *Journal of Economics and Business*, *3*(1).

Rickman, S., Johnson, A., & Miles, C. (2013). The impact of chewing gum resistance on immediate free recall. *British Journal of Psychology*, *104*(3), 339–346.

Sarver, D. E., Rapport, M. D., Kofler, M. J., Raiker, J. S.,

& Friedman, L. M. (2015). Hyperactivity in attention-deficit/hyperactivity disorder (adhd): Impairing deficit or compensatory behavior? *Journal of abnormal child psychology*, *43*(7), 1219–1232.

Schultz, W. (2015). Neuronal reward and decision signals: from theories to data. *Physiological reviews*, *95*(3), 853–951.

Schweizer, K., Moosbrugger, H., & Goldhammer, F. (2005). The structure of the relationship between attention and intelligence. *Intelligence*, *33*(6), 589–611.

Sisk, C. A., Remington, R. W., & Jiang, Y. V. (2019). Mechanisms of contextual cueing: A tutorial review. *Attention, Perception, & Psychophysics*, *81*, 2571–2589.

Soares, J. S., & Storm, B. C. (2020). Putting a negative spin on it: Using a fidget spinner can impair memory for a video lecture. *Applied Cognitive Psychology*, *34*(1), 277–284.

Spencer-Mueller, E. K., & Fenske, M. J. (2024). Note-taking for the win: Doodling does not reduce boredom or mind-wandering, nor enhance attention or retention of lecture material. *Quarterly Journal of Experimental Psychology*, *77*(8), 1780–1796.

Trabelsi, Z., Alnajjar, F., Parambil, M. M. A., Gochoo, M., & Ali, L. (2023). Real-time attention monitoring system for classroom: A deep learning approach for student's behavior recognition. *Big Data and Cognitive Computing*, *7*(1), 48.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, *185*(4157), 1124–1131.

Wang, F., Li, J., Li, Z., Ke, D., Du, J., Garcia, C., & Rodriguez, J. (2021). Design of model predictive control weighting factors for pmsm using gaussian distribution-based particle swarm optimization. *IEEE Transactions on Industrial Electronics*, *69*(11), 10935–10946.

Wang, X., Jin, Y., Schmitt, S., & Olhofer, M. (2023). Recent advances in bayesian optimization. *ACM Computing Surveys*, *55*(13s), 1–36.

Yang, X., Zhao, X., Tian, X., & Xing, B. (2021). Effects of environment and posture on the concentration and achievement of students in mobile learning. *Interactive Learning Environments*, *29*(3), 400–413.

Zhang, D., Chen, C., Tan, F., Qian, B., Li, W., He, X., & Lei, S. (2023). Multi-view and multi-scale behavior recognition algorithm based on attention mechanism. *Frontiers in Neurorobotics*, *17*, 1276208.

Zhu, F., Huang, Y., Xue, C., Mihaylova, L., & Chambers, J. (2022). A sliding window variational outlier-robust kalman filter based on student's t-noise modeling. *IEEE Transactions on Aerospace and Electronic Systems*, *58*(5), 4835–4849.