A Correlational Study of the Relationship Between ChatGPT Usage & Cognitive Flexibility

Victoria Szymkiewicz, Edgar Carvalho, Ezra Ford

Duke Vertices
Volume 2, Issue 2

Article Synopsis

As artificial intelligence tools like ChatGPT become increasingly prevalent in educational settings, it is crucial to understand how their usage influences students' cognitive processes. Szymkiewicz et al. investigate the relationship between the frequency of ChatGPT usage and cognitive flexibility among university students. They find that frequent users (those who used ChatGPT for more than 50% of their academic work) exhibit greater cognitive flexibility compared to infrequent users, as evidenced by less of a decrease in response accuracy when switching between two tasks. While further research is needed to fully understand the implications of these results, they may help to destigmatize the use of AI tools in educational contexts.

Graphic by Amanda Li

A Correlational Study of the Relationship Between ChatGPT Usage & Cognitive Flexibility

Victoria Szymkiewicz[1,4], Edgar Carvalho[2,4], Ezra Ford[3,4]

[1]Duke University

[2]Northwestern University

[3]Pomona College

[4]Danish Institute of Study Abroad, Copenhagen

https://doi.org/10.55894/dv3.13

Acknowledgements

The authors extend their sincere gratitude to Angela Mastropasqua and Nicola Menale for their invaluable mentorship and astute guidance throughout the research process. We also express our appreciation to all the participants whose engagement and contribution were essential to the advancement of this research.

Abstract

Amidst the growing integration of AI into educational practices, concerns arise regarding how unregulated interaction with AI outside the structured classroom environment may influence university students' cognitive functions. This research examines the differential effects of ChatGPT usage intensity on university students' cognitive flexibility. The researchers hypothesize a negative correlation between ChatGPT usage and cognitive flexibility, measurable by increased response times and decreased accuracy between the two tasks of the Arrow Switch Test. We employed this test to contrast the performance of frequent (participants who used ChatGPT ≥ 50% of the time on academic assignments) and infrequent (participants who used ChatGPT ≤ 35% of the time on academic assignments) groups. For the purpose of this study, infrequent users were used as a control group. Both groups displayed a statistically significant decrease in task accuracy—highlighting the Arrow Switch Task's efficacy in assessing cognitive flexibility—and response time between the first and second task. Notably, infrequent users of ChatGPT demonstrated a larger decline in accuracy and response time following changed task conditions compared to frequent users. This finding calls for further investigation into the longitudinal effects of AI tools on learning processes, necessitating a larger sample size and a more granular analysis of usage patterns to understand the subtleties of AI's impact on cognitive flexibility.

Keywords: artificial intelligence, education, cognitive flexibility, correlation, ChatGPT

INTRODUCTION

With swift advancements in technology, Artificial Intelligence (AI) tools like ChatGPT have become commonplace in educational settings. The prospective usage of AI inside classrooms has been promising with preliminary research showing that educational AI could improve the quality of education college students receive (Alam, 2022). AI has also proven to be useful to personalize education through resources like media recommendations, as it increases students' engagement and abilities (Huang et al., 2023).

Furthermore, AI has been shown to increase academic performance, self-efficacy, and motivation in students when used as a tool to provide real-time feedback during the completion of supplemental course review sheets (Lee et al., 2022). Research on AI as a classroom tool has only begun, and there is a gap exploring how AI not regulated by educators affects student performance. Although AI is anticipated to be a beneficial tool for educators and students within controlled classroom settings, the focus of this paper is on the potential negative consequences of students' unrestricted use of AI for academic purposes outside classroom settings (Timms, 2016).

Specifically, unrestricted ChatGPT usage could lead students to complete academic tasks with less cognitive involvement. ChatGPT has been characterized by scholars as high-tech plagiarism leading to learning avoidance (Chomsky et al. 2023). As such, the researchers set out to explore how unrestricted usage of ChatGPT on academic tasks influences students' cognitive flexibility, a critical aspect of learning and academic performance. Cognitive flexibility—the ability to adapt thinking and approach to varying tasks—is a significant predictor of academic success (Kercood et al., 2017). This adaptability encompasses not only the ability to switch between tasks or topics with ease but also involves the aptitude for understanding and applying concepts across different contexts. For instance, it enables students to leverage mathematical formulas learned in one scenario to solve science problems in another, or to draw on historical events to deepen their analysis of literature. This research will examine whether and to what extent ChatGPT usage correlates with this crucial intellectual capability.

We hypothesize that frequent use of ChatGPT for academic purposes decreases cognitive flexibility, observable through increased response time and decreased accuracy between the two tasks of the Arrow Switch Test. The Arrow Switch Test was implemented in order to quantify cognitive flexibility. This test required participants to look at an arrangement of arrows, and, based on the color of the arrows, respond accordingly. One color would require the participants to respond with the direction of the rightmost arrow in the arrangement, while the other color would require the participants to respond with the direction of the leftmost arrow. Halfway through the test, participants were prompted to switch how they responded so that the color cues were flipped. The portions of the test before and after the switch were labeled as Task 1 and Task 2, respectively. Both accuracy and response times were recorded. By investigating this relationship, the study seeks to contribute to the broader discourse on the role AI plays in influencing cognitive processes.

METHODS

Participants

The study sample comprised 22 American undergraduate students studying at the Danish Institute for Study Abroad, selected primarily due to accessibility and their willingness to participate voluntarily. There was no intentional imbalance in gender participation: 10 females participated, and 12 males participated.

The participants were divided into two groups based on their self-reported usage of ChatGPT for academic purposes: the control group (14 students) reported using ChatGPT less than 35% of the time, while the experimental group (8 students) used ChatGPT more than 50% of the time. Students were asked to measure their usage of ChatGPT in relation to all of their classes dating back to ChatGPT's launch on November 30, 2022. The division was intended to contrast the cognitive flexibility between frequent and infrequent users of ChatGPT. Both females and males were present in both the control and experimental groups.

Inclusion criteria required that participants had access to AI tools in their academic environment and that they spoke English fluently. Students with diagnosed cognitive impairments affecting task performance were excluded to maintain result integrity.

Materials

An initial survey was conducted through Google Forms to quantify each student's usage frequency of ChatGPT for academic tasks. Participants were asked what percentage of their academic work they complete using ChatGPT, and when they do use it, what percentage of the time they integrate ChatGPT generated text into their work. "Use" was defined as prompting ChatGPT for any type of response without regard for whether this response would be directly integrated (i.e. copy and pasted) into their academic work or not. Integration of content was defined as copying and pasting unaltered content and incorporating altered content (paraphrasing, restructuring, rewording, reworking, adapting, or rewriting) into personal work as opposed to using ChatGPT for solidifying one's understanding of concepts.

A computerized Arrow Switch Test was developed to quantify cognitive flexibility. This task was designed to record response accuracy and time, providing a measure of the participant's ability to shift cognitive strategies. The task was run in a controlled laboratory setting on desktop computers equipped with E-Prime software to ensure precise timing and data collection. Participants interacted with the task using standard keyboards using the hand of their preference.

Arrow Switch Test

The Arrow Switch Test was modeled after the principles of the Wisconsin Card Sort Test (Grant & Berg, 1948). The Arrow Switch Test required participants to focus on a series of five arrows of the same color, either purple or orange, displayed on a computer screen. Participants were tasked with responding to the arrows' orientation by pressing designated keys on a keyboard. The responses were tied to the colors of the arrows, requiring participants to employ attention to detail and color-orientation association (i.e., if the arrows were orange, the participants pressed the arrow key corresponding to the direction of the rightmost arrow, and if the arrows were purple, the leftmost arrow of the five).

The first task, Task 1, of the test consisted of 42 trials, after which the investigation entered its critical phase: the switch. At this juncture, the previously learned color-response associations were reversed without prior notice to the participants. They were explicitly informed of the new associations and thereafter completed 42 more trials under these new conditions (Task 2). This sudden reversal in task requirements was designed to measure cognitive flexibility, challenging the participants' ability to adapt to new rules and to modify their cognitive strategies accordingly. Therefore, Task 1 and Task 2 together comprise the singular Arrow Switch Task that participants undertake.

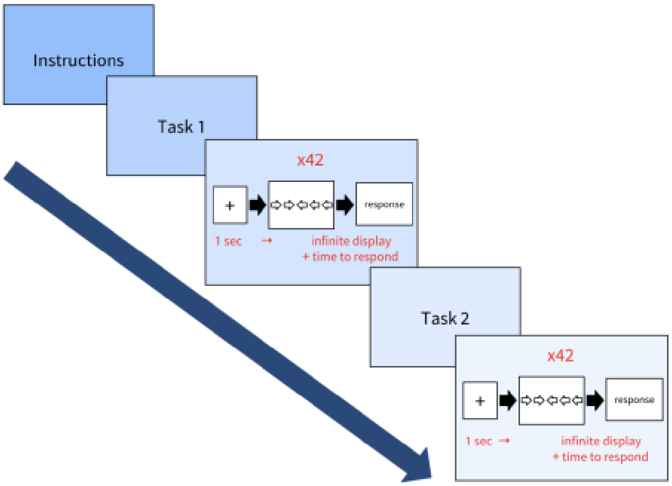
Procedure

Before participation, students completed an informed consent process outlining the study's objectives, anonymous and confidential treatment of participant data, and the right to withdraw at any time without penalty. Participants then completed the online survey quantifying their frequency of ChatGPT usage on academic tasks. Following the survey, participants were brought into the laboratory within the same week and introduced to the Arrow Switch Test. The entire task sequence was conducted in a quiet and controlled laboratory environment. Participants completed the test individually in a temperature-controlled environment with a singular desk, chair, and computer. This environment was maintained to ensure the accuracy of response data and to minimize external variables affecting the participants' performance.

Throughout the task, response times and accuracy were recorded for each participant through E Prime software. These data were central to the study's analysis, providing objective measures of the impact of ChatGPT usage on cognitive flexibility.

Figure 1

Arrow Switch Test



Analyses

The initial analysis involved conducting both Wilcoxon Tests and t-Tests to determine the significance of the data. However, it became apparent that the small sample size precluded the assumption of normality, leading to non-normal distribution of the data. Consequently, the skewness of the data varied, being either left or right, depending on the specific variable and group under examination. A Wilcoxon Test tests for differences between group means of independent samples when the data is not normally distributed. Therefore, due to its accurate reflection of the data, only the Wilcoxon Test was used for analysis. Accuracies of individuals were averaged across the 42 trials in the two tasks separately. Response times of correct answers only were analyzed.

RESULTS

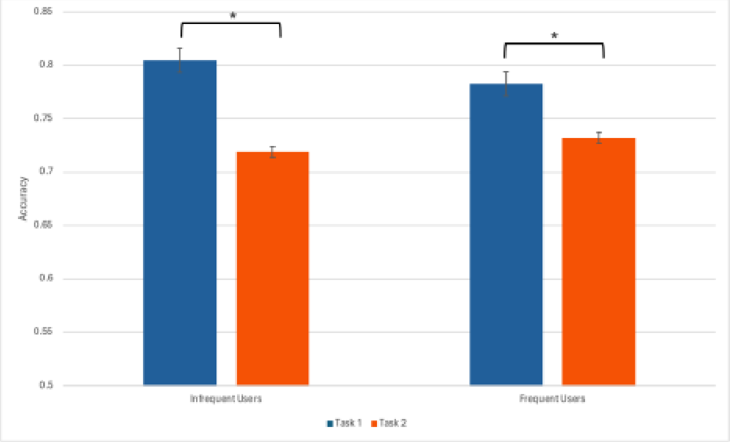
The evaluation of cognitive flexibility through the Arrow Switch Test revealed significant decreases in performance in terms of accuracy and response times across tasks and user groups.

Accuracy Analysis

Comparing the accuracy across all participants between Task 1 and Task 2 revealed a significant decrease from an average accuracy of 79.7% in Task 1 to 72.4% in Task 2. In this case the t-score = 22.0 and p = 0.0001 which meant that the mean accuracy from Task 1 was statistically significantly different from the mean accuracy of Task 2, illustrating that the Arrow Switch Test effectively measured cognitive flexibility by initially acclimating participants to Task 1 before requiring them to change their responses in Task 2. This expected drop in accuracy underscores the test's ability to challenge cognitive adaptability. Upon segmenting participants into infrequent and frequent users, the decline in accuracy between Task 1 and Task 2 remained significant for both groups, as illustrated in Figure 2. To compare across groups, eight infrequent users were randomly selected and compared with eight frequent users, revealing that the average accuracy drop from Task 1 to Task 2 was -0.086 for infrequent users and -0.051 for frequent users, indicating a greater decline in accuracy among infrequent users (see Figure 3). This difference in accuracy is calculated by subtracting Task 1 scores from Task 2 scores. The Wilcoxon Test results for the comparison between the difference of accuracy between Tasks in frequent users versus infrequent users yielded p = 0.0 and t-score = 4.0. Yet, considering each task separately, no statistical significance was found between the accuracies of infrequent and frequent users, suggesting that the overall usage was not correlated with different levels of performance in individual tasks.

Figure 2

Accuracies in the Arrow Switch Test Separated by Usage

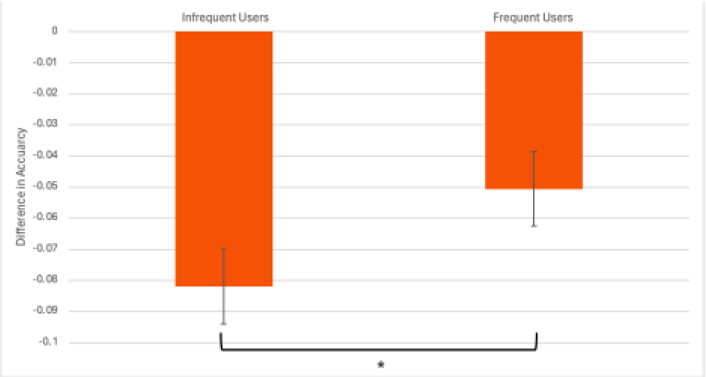


Note. The average accuracy for infrequent users on Task 1 was 80.5% and on Task 2 was 71.9%. This difference was significant (p < 0.01 and t-score = 1.0) according to the Wilcoxon Test results. The average accuracy for frequent users on Task 1 was 78.3% and on Task 2 was 73.2%. This difference was significant (p < 0.01 and t-score = 8.0) according to the Wilcoxon Test results. The same analyses for accuracy were done with ChatGPT integration, however no statistically significant results were found.

* indicates a p < 0.01

Figure 3

Differences in Accuracies on the Arrow Switch Test Separated by Usage



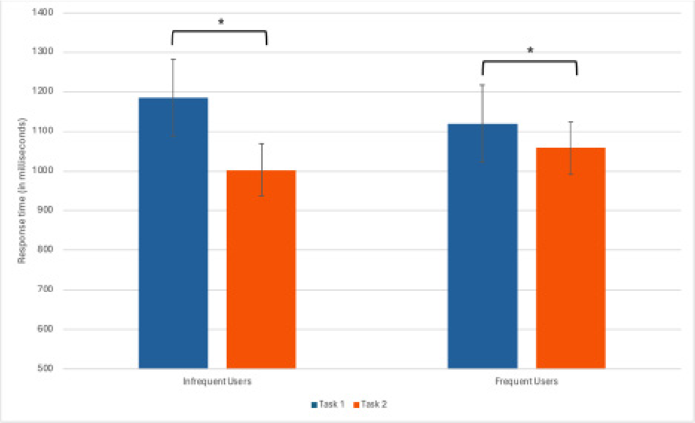
Note. * indicates p < 0.01

Response Time Analysis

When comparing the response time of all users in Task 1 and Task 2, there was a significant decrease in response time between the two tasks (p < 0.01 and t-score = 58.0) with the average response time in Task 1 being 1.163 seconds and the average response time in Task 2 being 1.022 seconds. The mean value of the response time for Task 1 across all users was statistically significantly different from the mean response time for Task 2. When participants were divided into infrequent and frequent users, the significant difference in response times between the two tasks persisted within both groups (refer to Figure 4). To compare across groups, eight infrequent users were randomly selected to compare the differences in response times in the two tasks between infrequent users and frequent users. The eight randomly selected infrequent users had an average difference in response time of -132.8 milliseconds between the two tasks while frequent users had an average difference in response time of -61.29 milliseconds (see Figure 5). Regarding the differences in response times between the two tasks, the Wilcoxon test results found that there was a significant difference between infrequent users and frequent users (p < 0.01 and t-score = 17.0). However, when comparing the response times of infrequent and frequent users within each task, there was no statistically significant difference between these two groups: their response times in Task 1 were not found to be significantly different from each other, and similarly, their response times in Task 2 were not significantly different from each other.

Figure 4

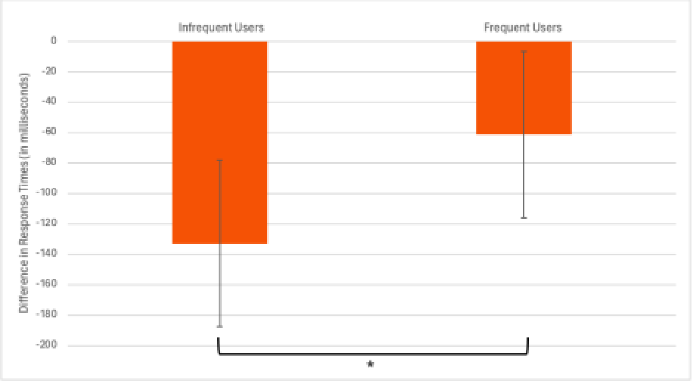
Response Times in the Arrow Switch Test Separated by Usage



Note. The average response time for infrequent users on Task 1 was 1.186 seconds and on Task 2 was 1.002 seconds. This difference was significant (p < 0.01 and t-score = 17.0) according to the Wilcoxon Test results. The average response time for frequent users on task 1 was 1.120 seconds and on task 2 was 1.059 seconds. This difference was also significant (p < 0.01 and t-score = 13.0) according to the Wilcoxon Test results. The same analyses for response time were done with ChatGPT integration, but no statistically significant results were found. * indicates p < 0.01

Figure 5

Differences in Response Times on the Arrow Switch Test Separated by Usage



Note. * indicates p < 0.01

DISCUSSION

The results do not support the hypothesis that there is a negative correlation between ChatGPT usage and cognitive flexibility as measured by increased response times and decreased accuracy on a cognitive flexibility task. The infrequent users demonstrated a larger average difference of both accuracy and response time between the two tasks compared to the frequent users. In addition, response times for all groups decreased between Task 1 and Task 2. This suggests participants may have felt more comfortable responding to Task 2 given its similarity to Task 1. There was no significant difference observed between frequent and infrequent users in each task, suggesting that both groups have comparable abilities in handling standard and switched tasks independently. However, the significant difference between the two groups' change in accuracy in the two tasks indicates that frequent users may possess greater cognitive flexibility compared to infrequent users, leading to less of a performance drop when the task rules are changed (Figure 3). This finding directly contradicts the original hypothesis, which anticipated that infrequent ChatGPT users would exhibit superior cognitive flexibility. Furthermore, no statistically significant data was found that correlated the amount of integration of ChatGPT generated responses with accuracy or response times between the two tasks.

One possible explanation for the observed enhanced cognitive flexibility among frequent ChatGPT users could be related to the concept of 'transfer of learning.' Transfer of learning refers to the application of knowledge and skills acquired in one context to a different context (Perkins & Salomon, 1992). Frequent interaction with the dynamic and variable environment of ChatGPT may have facilitated the development of transferable cognitive skills, such as adaptability and cognitive flexibility, which can be applied to novel tasks like the Arrow Switch Test (Barnett & Ceci, 2002). However, the correlation reported in this research cannot be assumed to be a direct connection and other possible variables could mediate this relationship.

Another explanation could be linked to the notion of 'desirable difficulties' in learning. Desirable difficulties refer to learning conditions that may initially impede performance but lead to long-term retention and transfer (Bjork & Bjork, 2011). The challenges encountered by frequent ChatGPT users in navigating between AI-generated and human-generated content could act as desirable difficulties, promoting deeper processing and more flexible cognitive strategies (McDaniel & Butler, 2011).

Through the Arrow Switch Test, a decline in accuracy between tasks was observed across all participants, indicating a measurable impact on cognitive performance when transitioning between tasks. The observed decrease in accuracy is consistent with the established literature indicating that cognitive performance can be influenced by the introduction of new task conditions, particularly when these involve a switch in cognitive strategy (Cañas et al., 2006). The decrease in response time was unexpected but can be explained by the fact that both tasks were similar, and thus participants may have had a false sense of comfort and familiarity on the second task despite the rule switch.

It is important to note that the Arrow Switch Test does not require higher-level thinking or integration of information. It is possible that frequent ChatGPT users have advanced adaptability skills for routine tasks but would be at a disadvantage with tasks requiring higher level cognition. In academic contexts, frequent users may use ChatGPT to perform the higher level analysis that infrequent users would perform themselves, and thus infrequent users would be more accustomed to tasks requiring a higher level of cognition.

Interestingly, the Wilcoxon Test findings suggest that infrequent users of ChatGPT may approach cognitive tasks with a strategy that favors speed over accuracy. This group demonstrated a significant decrease in accuracy yet faster response times on the second task, indicating a potential over-reliance on initial task learning, which did not transfer well when the task conditions changed. One plausible explanation for this observation could be rooted in the cognitive processing styles of the infrequent ChatGPT users. These individuals might favor a heuristic approach to problem-solving, which relies on intuitive, rule-of-thumb strategies that are faster but less precise. This heuristic processing style is generally more efficient in terms of response time but can lead to errors when the task complexity increases or when there is a need to adapt to new rules or conditions (Hjeij & Vilks, 2023).

In contrast, frequent users exhibited a significantly smaller change in response time, implying a more measured approach to the task switch. The results suggest that these individuals may prioritize maintaining accuracy over increasing speed, which could reflect a different aspect of cognitive flexibility—namely, the ability to maintain performance stability in the face of changing task demands (Braem et al., 2018). It is also possible an interest in maximizing performance underlies both more frequent use of ChatGPT and more effort made on the task.

It is also important to note that while the convenience sampling employed in this study offers valuable insights, it also introduces potential limitations, including a selection bias and the inherent challenges of self-report measures. These limitations may affect the generalizability of the findings and the accuracy of reported behaviors and attitudes. Future research should aim to address these limitations by employing more diverse and representative sampling methods and by incorporating objective measures of ChatGPT usage.

The study also did not categorize participants based on their ChatGPT subscription levels, such as GPT Pro, which may provide more human-like interactions. This limitation suggests that future research could explore how different levels of AI sophistication, afforded by various subscription models, influence cognitive flexibility and user dependency.

In general, further research with a larger sample size will be necessary to assume normality and to solidify results. Potentially confounding variables should be tested for such as concrete intelligence measures (i.e. IQ). Future research should aim to explore the boundaries and specificities of this observed phenomenon. It would be useful to examine whether the increased adaptability skills of frequent ChatGPT users extend to more complex cognitive tasks, which require deeper analytical thinking and comprehension. To explore if the benefits observed in simpler tasks extend to more complex scenarios, future research could incorporate tasks demanding higher-order cognitive processes, like problem-solving and critical thinking. Examples include tasks that integrate cognitive flexibility with complex puzzles, testing participants' abilities in logical deduction and strategic planning, or applying cognitive flexibility in critical evaluation scenarios, requiring the analysis of debate arguments to differentiate between strong and weak evidence. Additionally, given the relatively recent introduction of ChatGPT, longitudinal studies could be valuable to assess the impact of prolonged and sustained use on cognitive flexibility. The duration of ChatGPT usage should be investigated as a potential variable influencing the adaptability of cognitive strategies.

Furthermore, it is crucial to acknowledge the stigma associated with the use of ChatGPT in academic contexts. This stigma, often stemming from concerns about academic integrity and the appropriate use of AI in educational environments, could influence participants to underreport their engagement with ChatGPT. Implementing more precise and objective methods to quantify ChatGPT usage would enhance the reliability of research findings, but this approach presents potential challenges regarding privacy concerns. Future studies must therefore balance the need for accurate data collection with ethical considerations surrounding the privacy and autonomy of participants.

CONCLUSION

This study's exploration into the relationship between ChatGPT usage and cognitive flexibility has opened new avenues for understanding how interaction with AI may influence or may correlate with cognitive functions. Contrary to the initial hypothesis, the findings suggest that frequent users of ChatGPT may exhibit enhanced cognitive flexibility compared to infrequent users. In an era where AI is becoming increasingly embedded in our daily lives, the findings suggesting enhanced cognitive flexibility among frequent users of ChatGPT offer a counter-narrative to concerns about the potential cognitive drawbacks of frequent AI interaction. This research provides a nuanced perspective on how technology, often perceived as a crutch, might instead be fostering certain cognitive skills in users. The implication that engagement with AI could enhance adaptability and flexibility in cognitive tasks has profound implications for how we perceive, integrate, and utilize AI in various sectors, particularly in education and workforce development. This intriguing result not only challenges preconceived notions about the impact of AI interaction on cognitive abilities but also invites further investigation into the nuances of this relationship.

REFERENCES

1. Alam, A. (2022). Employing adaptive learning and intelligent tutoring robots for virtual classrooms and smart campuses: reforming education in the age of artificial intelligence. In *Advanced Computing and Intelligent Technologies: Proceedings of ICACIT 2022* (pp. 395-406). Singapore: Springer Nature Singapore. https://doi/org/10.1007/978-981-19-29 80-9_32

2. Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. Psychological Bulletin, 128(4), 612-637. https://doi.org/10.1037/0033-2909.128.4.612

3. Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), Psychology and the real world: Essays illustrating fundamental contributions to society (pp. 56-64). Worth Publishers.

4. Braem, S., & Egner, T. (2018). Getting a Grip on Cognitive Flexibility. *Current Directions in Psychological Science*, 27(6), 470-476. https://doi.org/10.1177/0963721418787475 Canas, J. J., Fajardo, I., & Salmeron, L. (2006). Cognitive flexibility. *International encyclopedia of ergonomics and human factors*, 1(3), 297-301. https://doi.org/10.1201/9780849375477 Chomsky, N., Roberts, I., & Watumull, J. (2023). Noam Chomsky: The False Promise of ChatGPT. The New York Times, 8.

5. Grant, D. A., & Berg, E. A. (1948). Wisconsin Card Sorting Test [Database record]. APA PsycTests. https://doi.org/10.1037/t31298-000

6. Hjeij, M., Vilks, A. A brief history of heuristics: how did research on heuristics evolve?. *Humanit Soc Sci Commun* 10, 64 (2023). https://doi.org/10.1057/s41599-023-01542-z

7. Huang, A. Y., Lu, O. H., & Yang, S. J. (2023). Effects of artificial Intelligence–Enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. *Computers & Education, 194*, 104684. https://doi.or g/10.1016/j.

compedu.2022.104684

8. Kercood, S., Lineweaver, T. T., Frank, C. C., & Fromm, E. D. (2017). Cognitive flexibility and its relationship to academic achievement and career choice of college students with and without attention deficit hyperactivity disorder. *Journal of Postsecondary Education and Disability, 30*(4), 329-344.

9. Lee, Y. F., Hwang, G. J., & Chen, P. Y. (2022). Impacts of an AI-based chabot on college students' after-class review, academic performance, self-efficacy, learning attitude, and motivation. *Educational technology research and development, 70*(5), 1843-1865. https://doi.org/10.1007/s11423-022-10142-8

10. McDaniel, M. A., & Butler, A. C. (2011). A contextual framework for understanding when difficulties are desirable. In A. S. Benjamin (Ed.), Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork (pp. 175-198). Psychology Press.

11. Perkins, D. N., & Salomon, G. (1992). Transfer of learning. In T. N. Postlethwaite & T. Husen (Eds.), International encyclopedia of education (2nd ed., pp. 6452-6457). Pergamon Press.

12. Sinnott, J., Hilton, S., Wood, M., & Douglas, D. (2020). Relating flow, mindfulness, cognitive flexibility, and postformal thought: Two studies. *Journal of Adult Development, 27*, 1-11. https://doi.org/10.1007/s10804-018-9320-2

13. Timms, M. J. (2016). Letting artificial intelligence in education out of the box: educational cobots and smart classrooms. *International Journal of Artificial Intelligence in Education, 26*, 701-712. https://doi.org/10.1007/s40593-016-0095