OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

Your Brain on ChatGPT: Is Bilingualism the Solution?

Koko Kusano
koko.kusano@gmail.com

ABSTRACT

The widespread adoption of Artificial Intelligence (AI) Large Language Models (LLMs) such as ChatGPT has raised concerns about their cognitive impact. Research suggests that reliance on LLMs may impair not only users' perceived but also palpable ownership over their thinking, surpassing cognitive-offloading effects previously observed with other technologies. Recent studies found diminished cognitive activity in individuals using ChatGPT to write essays, compared to those using Google Search or internal cognition. These findings point to potential declines in executive functions inclusive of but not limited to working memory, attention, and decision-making.

In contrast, bilingualism has been linked to cognitive resilience. According to Abutalebi and Green (2007), bilinguals routinely manage dual linguistic systems, strengthening executive functions that extend beyond linguistic into non-linguistic domains.

Therefore, this study investigated how language background influences LLM engagement and cognitive performance. Focusing on individuals aged 18-25, it examined differences among trilinguals, bilinguals and monolinguals in their use of LLMs and their performance on either cognitive task testing executive functioning: the Simon or AX Continuous Performance Task (AX-CPT). Trilinguals were included to examine the scalability of the bilingual advantage to those managing three or more languages. The study found a hierarchy in LLM engagement, with trilinguals and bilinguals reporting higher levels of reliance, frequency of use, and a greater number of cognitive purposes (e.g. attention, cognitive flexibility) offloaded onto LLMs than monolinguals. LLM use was negatively correlated with performance on cognitive tasks. Bilinguals generally performed better than monolinguals on these tasks. However, the subsequent impact with language background remained statistically inconclusive. The findings did not support a definitive protective effect of bilingualism against LLM-induced cognitive decline.

February 2026
Vol 4. No 1.

INTRODUCTION

Artificial Intelligence (AI) Large Language Models (LLMs) and Their Cognitive Impacts

Artificial Intelligence (AI) Large Language Models (LLMs) — ChatGPT, Gemini, Claude and Copilot, to name a few — have become ubiquitous. Half of Americans use them now, more than a quarter use them daily, and 49% believe these tools are smarter than they are (*Close Encounters of the AI Kind*, 2025). While LLMs are not inherently more intelligent, by pure definition — in fact some studies highlight potential cognitive benefits of LLMs for users (Peláez-Sánchez et al., 2024), recent studies suggest that they could be making users less smart (Stadler et al., 2024; Gerlich, 2025). For instance, Stadler et al. (2024) found that students using ChatGPT produced fewer relevant recommendations and lower reasoning and argumentation than those using Google Search. Similarly, Gerlich (2025) identified a negative correlation between frequent AI usage and critical thinking skills. Although the metrics used in these studies are imperfect proxies for cognitive quality, they nonetheless signal a reduction in independent thought linked to LLM use.

The neurological impact of this reliance is evidenced by a comprehensive investigation conducted at MIT, which found that participants relying on LLMs when writing an essay, a cognitively demanding process involving both macro-level (e.g., organizing ideas, structuring arguments) and micro-level (e.g., word choice, grammar, syntax) thinking, demonstrated the weakest neural activity (Kosmyna et al., 2025). Electroencephalography (EEG) scans showed that participants writing independently had the strongest neural connectivity, especially in the alpha, theta, and delta frequency bands associated with concentration and attention (Abhang et al., 2016). This group also expressed the highest satisfaction and ownership of their work. In contrast, participants relying on LLMs demonstrated the weakest neural activity and the lowest engagement and satisfaction, indicating that reliance on these tools fosters a form of cognitive dependence that could be detrimental to human intellectual development and autonomy in the long run (Shanmugasundaram & Tamilarasu, 2023).

This shift in cognition due to technological advancements aligns with Sparrow et al.'s (2001) "Google Effect," used to describe the tendency of people to rely on computers for information. Sparrow found that individuals remember the source or where the information is stored when they expect future access to information rather than the information itself. This trend extends to GPS-enabled mobile applications which have reduced users' ability to form spatial memory, as they rely on external devices for route planning and orientation (Cheng et al., 2023). These developments follow a long tradition of humans offloading cognitive burdens, depending on external tools or bodily manipulation to minimize mental effort (Dunn & Risko, 2015). According to Sweller's (1988) Cognitive Load Theory, because our working memory is limited both in capacity and duration, we naturally seek to offload "secondary" knowledge, or knowledge acquired by conscious effort, to reduce mental strain (Sweller, 2024).

However, LLMs may represent a more expansive form of cognitive offloading because they are designed to minimize internal effort more aggressively than previous tools. Unlike search engines, which still entail cognitive effort such as discerning relevance, synthesizing information, and capitalizing on domain-specific prior knowledge (Stadler et al., 2024), LLMs provide explicit answers and parse data into digestible chunks by interpreting human-language queries (Kosmyna et al., 2025). Their personalization features, tailored to the individual likes and behavior of the user, can shape user decision-making by offering tailored suggestions or interpretations (Kearney et al., 2025). Such advancements pose grave questions about the future of human cognitive autonomy.

Importantly, individuals with strong foundational cognitive activity can effectively integrate LLMs as an aid without diminishing their cognitive engagement. Participants who first wrote an essay independently in the MIT study exhibited a vast expansion of neural connections across EEG frequency bands when they used LLMs in a final session (Kosmyna et al., 2025). This suggests that LLMs may increase productivity in already cognitively engaged individuals, a finding consistent with broader evidence that higher levels of educational attainment are positively correlated with stronger critical thinking skills, independent of AI tool usage (Gerlich, 2025).

Bilingualism and Its Cognitive Effects

In contrast to these risks, bilingualism has long been associated with enhanced cognitive resilience. Bilingualism is proficiency in two languages — though "competence" is itself a complex and variably defined entity, varying by context in acquisition, proficiency, and usage. This will be further clarified later in the manuscript.

The cognitive resilience of bilingualism is largely accounted for by the increased mental effort to manage two linguistic systems, as proposed in Abutalebi and Green's (2007) language control hypothesis (Abutalebi & Green, 2007). The cognitive processes underlying bilingual language use are multi-component. They include: selecting the appropriate language for communication, retrieving target words in that language, inhibiting interference from the other language, monitoring for potential intrusions (i.e., words or structures from the other language), and dynamically switching between languages as necessary. Abutalebi and Green argue that these processes rely on sustained executive control and attentional monitoring — processes also engaged in supporting non-linguistic cognitive functioning.

Indeed, a number of studies confirm what could be termed the bilingual cognitive advantage, especially in domains such as problem-solving, attentional control, metacognitive awareness, interference control, and divergent thinking (Yang, 2017).

For example, cognitive neuroscientist Ellen Bialystok found bilinguals to perform better on Witkin et al.'s (1971) Hidden Figures Test, a test of field independence, which measures the ability to detect separate parts in a complex visual field — a task that involves analytical processing (Bialystok, 1992). In another study, bilinguals outperformed monolinguals in a Simon task, demonstrating they are not just better in problem solving but also quicker and more precise responses, indicative of enhanced attentional control (Bialystok, 2006).

Beyond problem solving and attentional control, bilingualism has also been linked with greater metacognitive awareness. Bilinguals showed better correct knowledge of their own reading comprehension, as measured by a subset of the Nelson-Denny Reading Test, and greater awareness of working memory, as measured by an adapted version of Hannon and Daneman's (2001) Reading Span test (17). In terms of interference control, a meta-analysis by Donnelly et al. (2019) found that bilingual adults tend to exhibit a greater inhibitory control advantage — reflected in smaller differences in reaction times between trials with and without distracting information in tasks such as the Flanker, Simon, and Stroop (Donnelly et al., 2019). On the other hand, bilingual children in this same study demonstrated a more general executive control advantage, showing shorter average reaction times across all trial types. Additionally, bilinguals scored higher on the Abbreviated Torrance Test for Adults, a measure of divergent thinking, which includes the generation of original responses through the retrieval of unrelated information across different cognitive domains (Kharkhurin, 2008).

Still, the bilingual advantage is not uniformly supported. Some behavioral studies have failed to replicate bilingual and monolingual differences in verbal, visual, or visuospatial working memory (Gutiérrez-Clellen et al., 2004; Namazi & Thordardottir, 2010; Engel de Abreu, 2011). These inconsistencies suggest that the bilingual advantage may depend on task type, participant demographics, and operationalization of bilingualism.

In response to Abutalebi and Green's (2007) call for the application of functional neuroimaging to clarify these differences (Abutalebi & Green, 2007), a growing body of research has delineated a network of cortical and subcortical regions involved in bilingual language control and persisting in their wider cognitive advantage. Key regions include the dorsal anterior cingulate cortex (dACC) and pre-supplementary motor area (pre-SMA) complex, left prefrontal cortex, left caudate, bilateral inferior parietal lobules, as well as supporting regions such as the right prefrontal cortex, thalamus, putamen, and cerebellum (Abutalebi & Green, 2007).

It is an especially remarkable finding that bilinguals have delayed onset of dementia symptoms — by an average of 4.1 to 5.1 years compared to monolinguals (Calvo et al., 2023). This lag has been attributed to greater gray matter volume in bilinguals, for example in the ACC, a region central to conflict monitoring, learning through feedback, and goal-directed action and thus to regulating cognitive load (Boroujeni et al., 2022).

On the other hand, although bilinguals show late onset of symptoms, the progression of dementia tends to be faster after onset (Calvo et al., 2023). Studies have found greater brain atrophy in Alzheimer's-associated regions in the medial temporal lobe along with steeper declines in cerebral glucose metabolism among bilinguals as compared to monolinguals. Nonetheless, bilinguals outperform monolinguals with similar levels of brain degeneration on both short- and long-term verbal memory and visuospatial tasks. One explanation is that bilinguals establish numerous and stronger neural pathways, so their brains are less vulnerable to localized damage — a paradigm consistent with the cognitive reserve hypothesis.

MATERIALS AND METHODS

While prior research has explored the intersection of bilingualism and LLM usage — such as bilinguals' preference for models that code-mix (i.e., interchangeably use two languages) (Bawa et al., 2020) — the reciprocal effects of these two factors remain underexplored.

Therefore, this study aimed to illuminate whether bilingual cognitive strain influences behavioral patterns in LLM usage, and whether bilingualism can mitigate LLM-induced cognitive decline. By positioning bilingualism as a potential buffer against the cognitive downsides of AI interaction — just as it has been shown to mitigate age-related losses in some executive functions (Bialystok et al., 2004), this research offers insight into how linguistic diversity might shape more resilient and cognitively engaged LLM users.

HYPOTHESES

The hypotheses were set out as follows, in two phases:

<u>Part 1: Questionnaire</u>

Firstly, this study predicted that bilinguals use LLMs more than monolinguals do. While this seems contradictory to prior claims made on the bilingual advantage, the reality is that bilinguals experience more cognitive load than monolinguals due to their dual language background. Hence, it would not come as a surprise that bilinguals might actively seek cognitive offloading opportunities like the use of LLMs, just as they gesture at a higher rate than monolinguals (Nicoladis, 2007) and prefer LLMs that code-mix (Bawa et al., 2020). LLM self-reported reliance, usage frequency, and the number of offloaded executive functions were treated as behavioral proxies for the intensity of cognitive offloading, allowing for the second and third hypotheses testing whether the cognitive offloading induces cognitive decline, and whether the bilingual advantage offsets this said decline.

<u>Part 2: Cognitive Tasks</u>

Secondly, this study predicted a negative correlation between LLM use and performance on the Simon (Bialystok, 2006) and AX-CPT tasks (Beatty-Martínez et al., 2020), consistent with existing research on cognitive decline associated with offloading onto LLMs.

Thirdly, this study predicted that bilinguals would generally perform better than monolinguals on the executive function tasks, consistent with existing research of the bilingual cognitive advantage. The Simon task is known to measure inhibitory or interference control, much like bilinguals have to suppress one language while using another. The AX-CPT task is known to measure proactive (e.g. goal maintenance, conflict monitoring, and interference suppression) and reactive (e.g. response inhibition) control processes.

Assuming the previous prediction holds true, this study further hypothesized that bilinguals will show no significant differences in their performance on the Simon and AX-CPT tasks, regardless of LLM usage, because their enhanced executive control may buffer against cognitive impairment introduced by LLMs. In contrast, monolinguals are expected to exhibit significant performance differences on these tasks depending on their interaction with LLMs. However, if the results contradict this prediction, they would align with prior research findings (Stadler et al., 2024; Gerlich, 2025; Kosmyna et al., 2025) which posit that LLMs diminish cognitive engagement irrespective of language background. Such an outcome would imply either that bilingualism does not confer a cognitive advantage, or that its protective effects are insufficient to counteract the cognitive consequences of LLM reliance.

PARTICIPANTS

To test our hypotheses, we distributed an online experiment administered using Gorilla Experiment Builder to a target population consisting of 18 to 25-year-olds who had diverse uses of LLMs and diverse language learning backgrounds. We targeted this specific age group for a few reasons. According to Elon University's data 18 to 29-year-olds had the highest proportion of LLM users (*Close Encounters of the AI Kind*, 2025). Likewise, in prior research younger participants exhibited higher dependence on AI tools and lower critical thinking scores compared to older participants (Gerlich, 2025). This study capped the age group at 25, however, as this is the age at which brain maturation is said to halt (Arain et al., 2013).

The distribution period spanned from August 20, 2025 to September 15, 2025. The primary outlets for distributing the questionnaire were the experimenter's personal contacts and via SurveyCircle and SurveySwap, and participants voluntarily opted to take part. Participants were incentivized with a chance to win a USD20 Amazon gift card. This sampling method combined both opportunity and volunteer sampling.

The study included a total of 35 participants, all of whom were 18 to 25-year-olds. Among them, 17 were biologically male, whereas 20 were biologically female. Their educational levels ranged from "High School" (65.71%), "Some College" (17.14%), "Bachelor's Degree" (14.29%), and "Master's Degree" (0.29%). Most participants identified their occupation as "Student" (91.43%), though one identified as "Middle Management," and two as "Other" (i.e., "Actor" and "Artist").

In terms of LLM use, the majority of participants (94.29%) expressed they actively use LLMs. Participants were clarified that by "actively," it meant engaging with LLMs. For example, Google Gemini is often an integrated feature in Google Chrome, but simply having access to it did not count as "active" use. Among those who reported active use of LLMs, the percentage of participants using ChatGPT was 87.88%, Google Gemini 33.33%, DeepSeek 15.15%, Perplexity 15.15%, Microsoft Copilot 12.12%, Claude 0.30%, "Midjourney" 0.30%, "YouTube Summarizer AI Bot" 0.30%, "Notion AI" 0.30%, "Brand Website" 0.30%.

However, 2 participants expressed they did not actively use LLMs. One explained "[they do not] like using AI chatbot(s) unless absolutely necessary [and that they] prefer to make use of [their] own cognitive abilities to analyse, evaluate, and solve [their tasks]." The other explained "[they] don't because [they] find their information unreliable, inconsistent, and [that they are] worried about the chatbots' sources when they generate a response. Chat GPT for instance gets a lot of its information from Reddit and it constantly incorrectly solves math problems which leads [them] to question its reliability." Interestingly, they were both monolingual.

In terms of language background, 11 participants were monolingual (31.43%), 19 were bilingual (54.29%), and 5 others were trilingual (14.29%). Participants who were proficient in more than three languages were told to input as being trilingual. Trilinguals were included to examine whether the bilingual advantage extends to those managing three or more languages. This remains underexplored in existing literature (Schroeder & Marian, 2016) and serves to test the validity of the bilingual advantage.

Only participants who answered all questions on the questionnaire and completed all trials in the cognitive task they were allocated to were included.

STATEMENT OF INFORMED CONSENT

Participants gave informed consent. They were informed that the study aimed at "understanding how individuals aged 18-25 engage with AI chatbots." This description was sufficient to ensure informed consent while withholding the study's specific focus on linguistic background to minimize demand characteristics in participants' responses. The phrase "AI chatbots" was used to refer to LLMs to avoid jargon. Participants were fully informed of the study's duration, the confidentiality of their responses, the voluntary nature of their participation, and their right to withdraw on or before September 15, 2025. A point of contact was also given should they have any questions regarding the study.

MATERIALS

This study was designed to examine three key relationships: 1) the link between linguistic background and LLM usage, 2) the link between LLM usage and executive function, and 3) the interaction between language background and executive function in the context of LLM usage. To address these aims, the study was divided into two components: a questionnaire assessing participants' language profiles and LLM engagement, and a cognitive task randomly assigned as either the Simon or AX-CPT task.

<u>Part 1: Questionnaire</u>

Participants completed a structured questionnaire capturing demographic information, including age (between 18 and 25), gender, education level, and occupation. The survey also gathered data on participants' LLM usage patterns, such as frequency of use and typical conversation topics.

To assess linguistic background, participants self-identified as monolingual, bilingual, or trilingual. Recognizing that these categories can be reductive, the study incorporated a more nuanced approach to language profiling.

- **Age of Acquisition:** Participants who identified as bilingual or trilingual were asked to specify the age at which they acquired their second and/or third language(s). According to Muñoz's meta-analysis of nine studies examining the impact of age and exposure to foreign language learning in formal settings, students who began learning a foreign language later often demonstrated advantages in perception, oral fluency, lexical development, and morphological acquisition — at least in the short term (Vorstman et al., 2009). Later language acquisition has thus been linked to enhanced cognitive outcomes, particularly in areas of executive functioning. However, establishing a clear directional hypothesis remains challenging. While studies such as those by Muñoz and others suggest that later acquisition may confer certain cognitive benefits (Goodrich & Lonigan, 2017; Gutiérrez-Clellen et al., 2004), others such as D'Souza and D'Souza report early cognitive advantages in preverbal infants exposed to bilingual environments from birth. These early-exposed infants showed improved cognitive flexibility and adaptive learning, as evidenced by faster habituation, a stronger preference for novelty over familiarity, and enhanced anticipatory responses (D'Souza & D'Souza, 2016).

- **Context of Acquisition:** Participants also indicated the context(s) in which they learned their additional language(s), including formal (e.g., classroom instruction), informal (e.g., exposure at home or in the community), or a combination of both. Research by Vorstman et al. (2009) suggests that formal language learning environments may contribute more significantly to the cognitive advantages associated with bilingualism than informal ones, while combined exposure appears to yield the greatest benefit (Goodrich & Lonigan, 2017).
- **High and Balanced Language Proficiency:** Participants self-rated their proficiency in each language using a Likert-type scale ranging from 0 ("beginner") to 100 ("advanced"). According to Cummins' Common Underlying Hypothesis (1981), high and balanced proficiency across languages is crucial for cognitive benefits to emerge. As Cummins explains, "As proficiency in one language develops, so does language-independent knowledge (i.e., the common underlying proficiency) that supports the development of skills in both languages." (Goodrich & Lonigan, 2017)
- **Context of Language Use:** Participants also described the contexts in which they primarily used their second and/or third language(s). Based on the framework proposed by Beatty-Martínez et al. (2020), interactional contexts were categorized as integrated (i.e., frequent code-switching within a single setting or conversation), separate (i.e., distinct language use across different settings with minimal switching), or varied (i.e., a combination of integrated and separate contexts) (Beatty-Martínez et al., 2020). Beatty-Martínez et al. (2020) found that the bilingual cognitive advantage was most pronounced among individuals who used their languages in varied contexts, followed by those in integrated contexts, with the smallest advantage observed in separate contexts.

This layered approach allowed for a more comprehensive understanding of participants' linguistic profiles and their potential influence on cognitive engagement with LLMs.

<u>Part 2: Tasks</u>

Participants were randomly assigned to complete one of two cognitive tasks designed to measure executive functioning: the Simon task or the AX-CPT task.

Part 2.1: Task (Simon)

The Simon task, based on Simon and Rudell's (1967) work, assesses stimulus-response compatibility — specifically, how congruence between stimulus location and response direction affects reaction time (Simon & Rudell, 1967). In the original paradigm, participants responded to verbal cues ("right" or "left") presented to either ear by pressing corresponding keys.

This study employed a modified version of the task, as implemented via Gorilla Experiment Builder. Participants responded to 40 visual cues ("right" or "left") displayed to the right or left of a central fixation point by pressing the J key for "right" and the F key for "left." This variation maintains the core interference control mechanism while adapting it for visual presentation. In prior research, such as Bialystok's (Bialystok, 2006), the Simon task has served as a proxy for attentional control and problem-solving.

Part 2.2: Task (AX-CPT)

The AX-CPT task, based on the Dual Mechanisms of Control framework (Braver, 2012), assesses both proactive and reactive control (Beatty-Martínez et al., 2020). Proactive control involves advanced preparation and sustained maintenance of contextual cues, linked to lateral prefrontal cortex (PFC) activity and high working memory. In contrast, reactive control relies on transient retrieval of context across a broader frontoparietal network (Gonthier et al., 2016).

This study employed the version developed by Ophir, Nass, Wagner, and Posner (2009) (Ophir et al., 2009). Participants were presented with 60 red cue-probe letter pairs. They were instructed to respond "yes" only to AX sequences (X following A); all other combinations (AY, BX, BY) "no." To increase cognitive load, three white distractor letters appeared between cue and probe. Proactive control predicts better BX but poorer AY performance, while reactive control predicts the opposite. BY trials served as a baseline, with fast and accurate responses expected (Polizzotto et al., 2018).

AX trials comprised 70% of all trials to strengthen target expectancy and bias toward A-cues and X-probes. The remaining trial types (AY, BX, BY) each occurred 10%.

PROCEDURE

All participants completed Part 1: Questionnaire (*Figure 1*). However, Part 2: Tasks utilized an independent measures design, where half of participants only completed the Simon and the other half only the AX-CPT task. Each participant completed the assigned task in a single sitting as they were restricted to a set duration of response time. They were not able to continue editing their responses if they exceeded 2 hours from the start of the task.

Task	Participants	
Simon	Eligible* Participants (18)	Non-Eligible Participants (1)

	Biological Sex: • Male 8 • Female 10 Education: • High School 14 • Some College 3 • Bachelor's Degree 1 Occupation: • Student 18 LLM Use: • Yes 17 • No 1 Language Background: • Monolingual 4 • Bilingual 13 • Trilingual 1	Biological Sex: • Male 1 • Female 0 Education: • Some College 1 Occupation: • Student 1 LLM Use: • Yes 1 • No 0 Language Background: • Monolingual 0 • Bilingual 1 • Trilingual 0
AX-CPT	<u>Eligible* Participants (13)</u> Biological Sex: • Male 7 • Female 6 Education: • High School 6 • Some College 2 • Bachelor's Degree 4 • Master's Degree 1 Occupation: • Student 10 • Middle Management 1 • Other (i.e., "Actor" and "Artist") 2 LLM Use: • Yes 12 • No 1 Language Background:	<u>Non-Eligible Participants (3)</u> Biological Sex: • Male 2 • Female 1 Education: • High School 3 Occupation: • Student 3 LLM Use: • Yes 3 • No 0 Language Background: • Monolingual 0 • Bilingual 0 • Trilingual 3

	● Monolingual 7 ● Bilingual 5 ● Trilingual 1	

Figure 1: Table illustrating Participants' Quota on Simon and AX-CPT tasks.
**More than or equal to 50% overall task accuracy.*

The study was administered online using a questionnaire designed with Gorilla Experiment Builder. Gorilla Experiment Builder was an ideal software for our experimental design, as it was an effective tool for administering tasks with some existing templates that could be used.

DATA ANALYSIS

All statistical analyses were performed using R Version 4.5.1 (2025-06-13) to evaluate participant responses to the questionnaire and participant performance on either the Simon or AX-CPT task.

Coding and Scoring

Participants' demographic, LLM-related, and linguistic variables were coded using structured ordinal scales.

Demographic Variables:

- Educational Level was ranked as:
  "High School" = 1, "Some College" = 2, "Bachelor's Degree" = 3, "Master's Degree" = 4, "Doctorate" = 5
- Occupational Level was ranked as:
  "Student" = 1, "Worker" = 2, "Specialist" = 3, "Middle Management" = 4, "Top Management" = 5

These ranking schemes were adapted from Gerlich's coding system (Gerlich, 2025). Any education or occupational levels not represented within these predefined categories were left excluded.

LLM-related Variables:

- LLM Reliance was measured on a continuous scale from 0 to 100, reflecting participants' self-reported dependence on LLMs.
- LLM Usage Frequency, defined as participants' self-reported frequency of interaction with LLMs, was coded on a scale from 0 to 12:
  "Never" = 0, "Every month" = 1, "Every 2-3 weeks" = 2, "Every week" = 3, "Every 2-3 days" = 4, "Every day" = 5, "Every 6-10 hours" = 6, "Every 2-5 hours" = 7, "Every hour" = 8, "Every 45 minutes" = 9, "Every 30 minutes" = 10, "Every 15 minutes" = 11, "Every 1 minute" = 12.
- Participants were also asked about the cognitive purposes for which they use LLMs. These purposes were grouped into five domains: Attention, Cognitive Flexibility Inhibition, Monitoring, and Working Memory. Each domain was represented by two descriptive statements (*Figure 2*).

Cognitive Purpose	**Description**
Attention	● To maintain focus on a specific topic ● To break down complex information into manageable parts
Cognitive Flexibility	● To explore different perspectives or ideas ● To switch between tasks or topics efficiently
Inhibition	● To avoid impulsive decisions ● To filter out irrelevant information
Monitoring	● To check my understanding or progress ● To get feedback on my work or ideas
Working Memory	● To hold and manipulate information while problem-solving ● To help me remember steps or instructions
Other - please specify:	

Figure 2: Table illustrating Operationalization of Cognitive Domains.

Linguistic Variables:

- Age of L2/L3 Acquisition, referring to the age at which participants acquired their second and/or third language (if identified as bilingual or trilingual), was coded as:
  "Before Age 1" = 1, "Age 1-5" = 2, "Age 6-10" = 3, "Age 11-15" = 4, "Age 16-20" = 5, "Age 21-25" = 6.
- Context of L2/L3 Acquisition was coded based on the environment in which the language(s) were learned:
  "Non-formal setting (i.e., at home, with relatives and friends)" = 1, "Formal setting (i.e., at school, at language center)" = 2, "Both Non-formal and Formal settings" = 3.
- L1/L2/L3 Proficiency was measured on a continuous scale from 0 to 100, reflecting participants' self-reported proficiency in each reported language.
- Language Use Context, defined as the predominant setting in which participants use their reported languages, was coded as:
  "Separate Language Context (i.e., distinct language use per setting with minimal switching)" = 1, "Integrated Language Context (i.e., frequent switching between languages within the same setting)" = 2, "Varied Language Context (i.e., a combination of Separate and Integrated Language Contexts)" = 3.

RESULTS

1) Large Language Models (LLMs) Use and Language Background

Trilinguals reported the highest LLM reliance (*Figure 3*) and usage frequency (*Figure 4*), followed by bilinguals and monolinguals. A Kruskal-Wallis test confirmed statistically significant differences in reliance (**$\chi^2(2) = 6.6714$, $p = 0.0356$**) and usage frequency (**$\chi^2(2) = 6.2705$, $p = 0.0435$**). Trilinguals also utilized LLMs for a broader range of cognitive purposes (*Figure 5*).

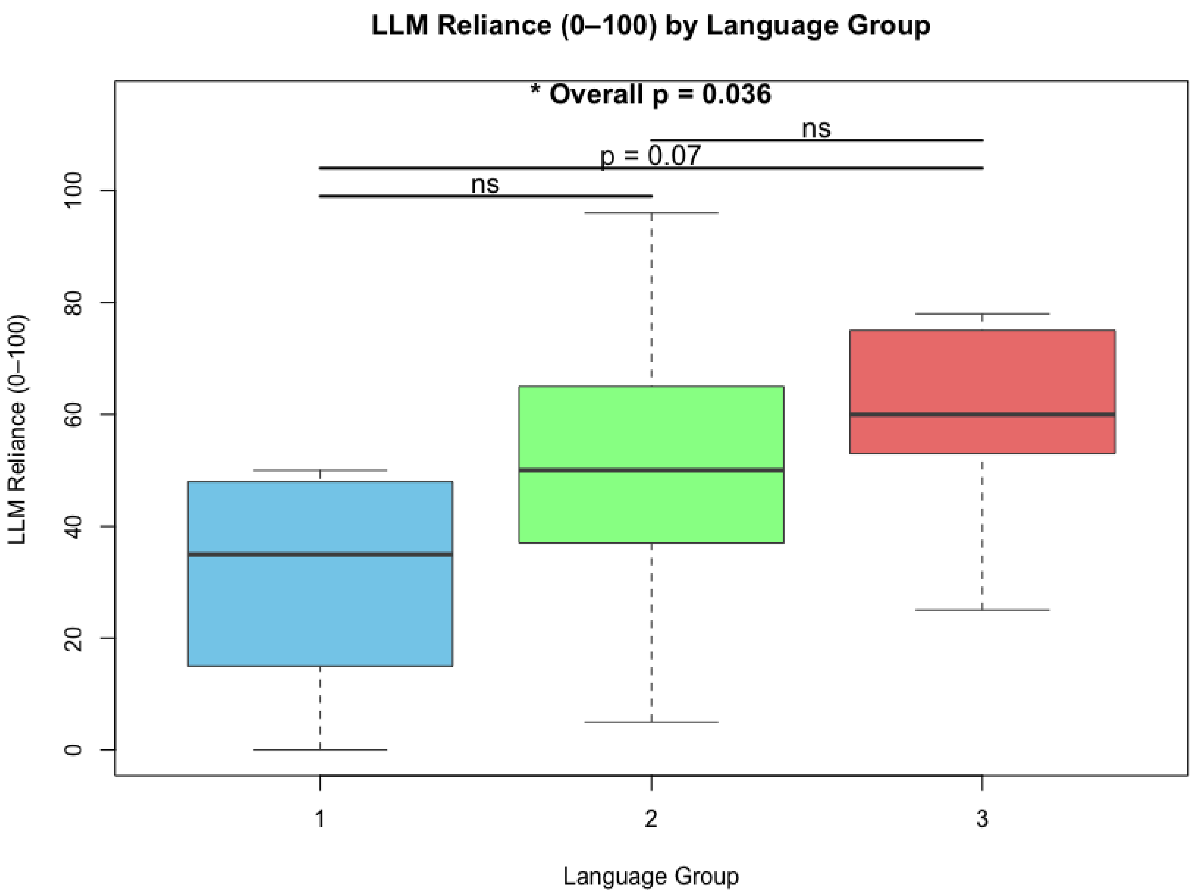

*Figure 3: Boxplot illustrating LLM Reliance (0-100) by Language Group (monolingual = 1, bilingual = 2, trilingual = 3). *Critical value set at 0.05.*

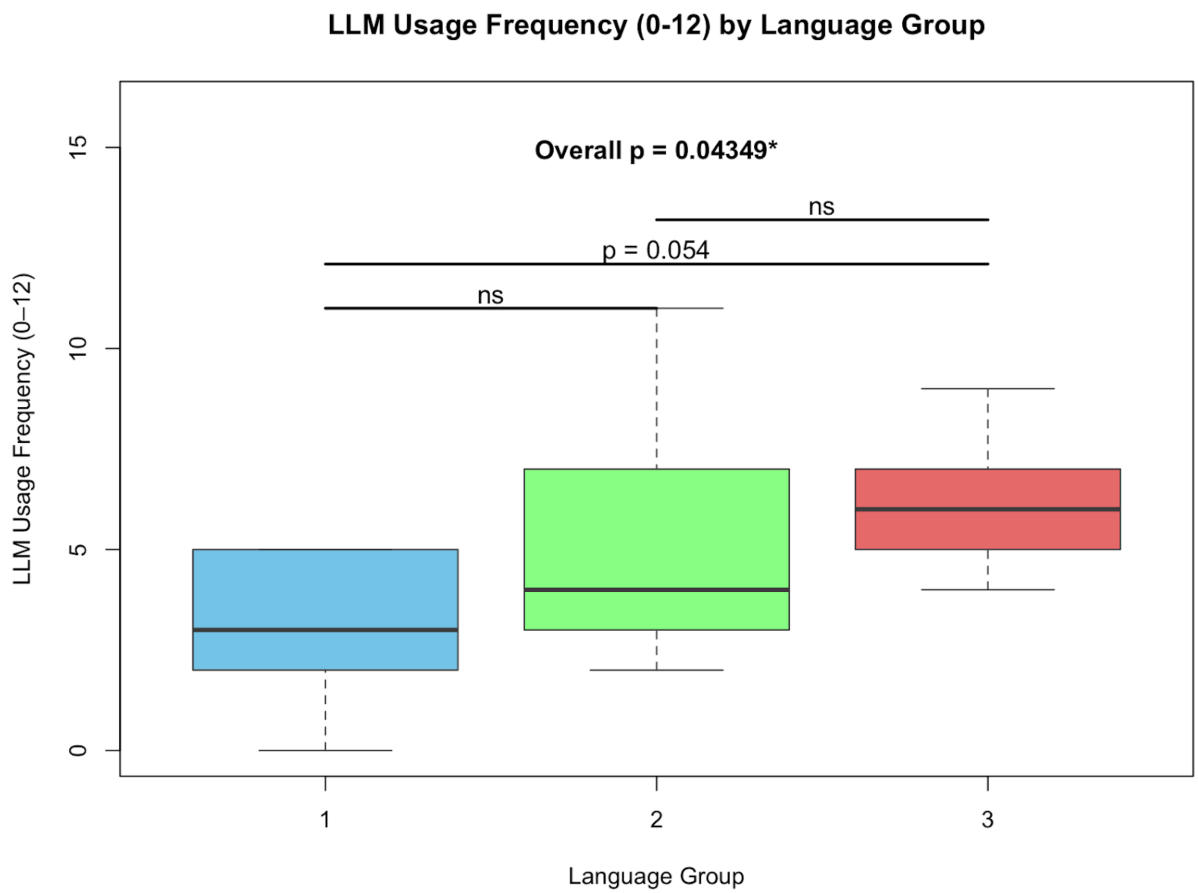

*Figure 4: Boxplot illustrating LLM Usage Frequency (never = 0 - every 1 minute = 12) by Language Group (monolingual = 1, bilingual = 2, trilingual = 3). *Critical value set at 0.05.*

Cognitive Purpose	**Percentage Distribution of Participants (%) using LLMs for each Cognitive Purpose by Language Group**		
	1	**2**	**3**
Attention	63.64	73.68	100.00
Cognitive Flexibility	36.36	63.16	40.00
Inhibition	18.18	36.84	20.00
Monitoring	54.55	68.42	80.00
Working Memory	9.09	42.11	20.00

*Figure 5: Table illustrating Percentage Distribution of Participants using LLMs for each Cognitive Purpose by Language Group (monolingual = 1, bilingual = 2, trilingual = 3). *Red font indicates more than half of the participants in each language group used LLMs for each cognitive purpose.*

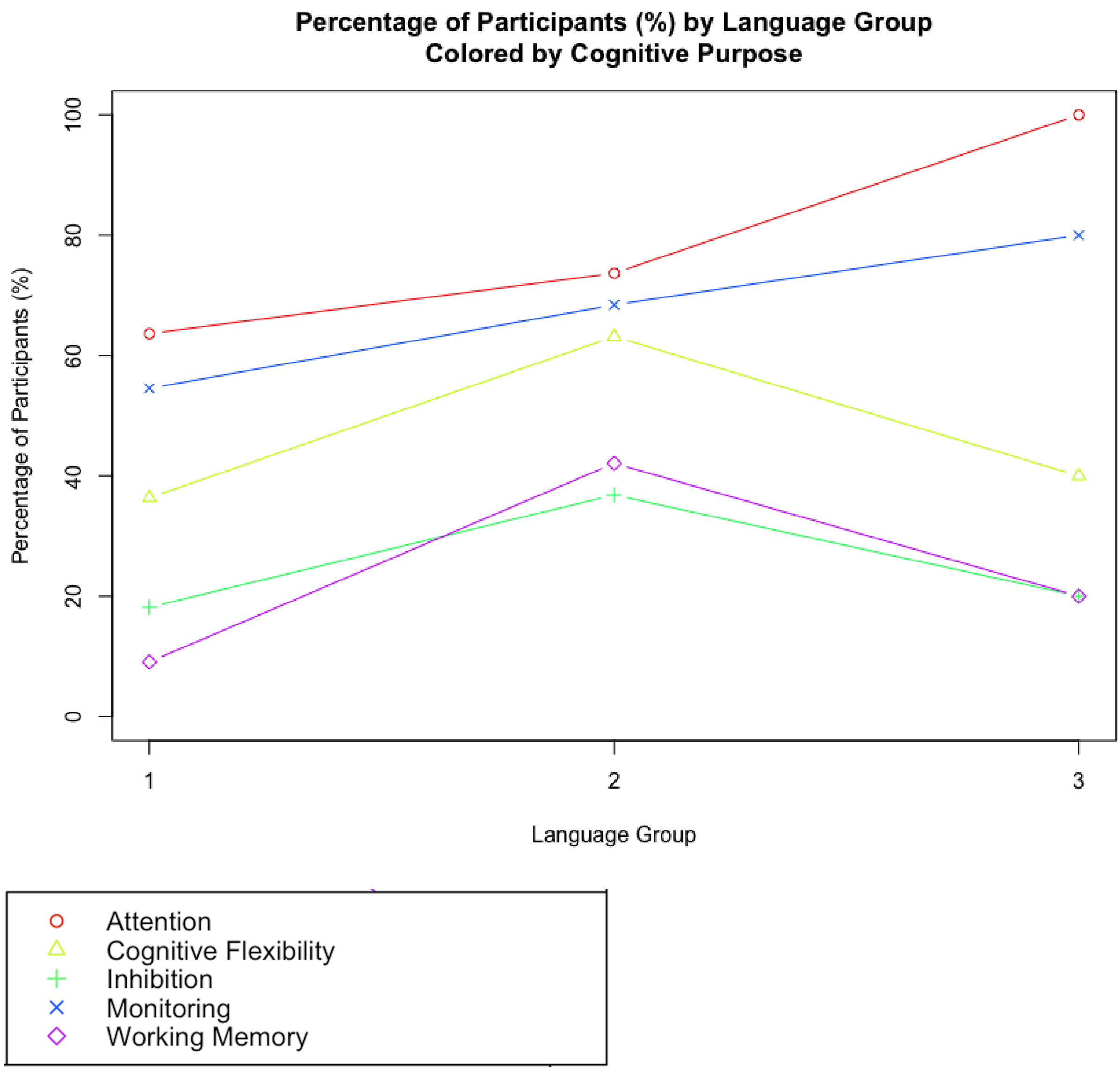

Figure 6: Graph illustrating Percentage Distribution of Participants (%) using LLMs for each Cognitive Purpose by Language Group (monolingual = 1, bilingual = 2, trilingual = 3).

These analyses were repeated using alternative measures of language background, including: 1) age of L2/L3 acquisition (*Figures 7-9*), 2) context of L2/L3 acquisition (*Figures 10-12*), 3) self-reported proficiency in L1, L2, and L3 (*Figures 13-15*) context of language use (*Figures 16-18*). As the proficiency gap between a participant's first and second language increased, LLM reliance decreased, as did the number of cognitive purposes offloaded onto the LLMs (*Figures 13-15*). Of the 35 participants surveyed, two reported no use of LLMs; both were monolingual.

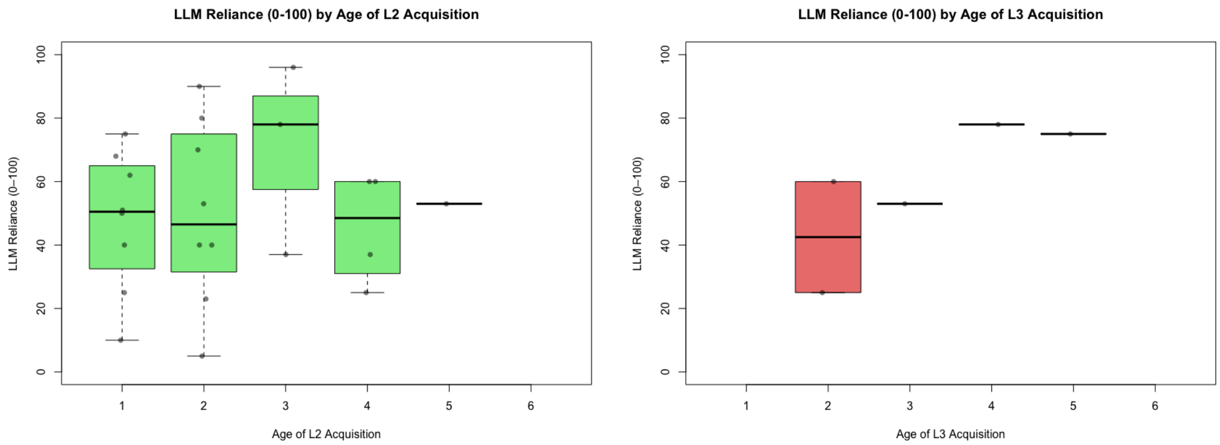

Figure 7: Boxplot illustrating LLM Reliance (0-100) by Age of L2/L3 Acquisition (before age 1 = 1, age 1-5 = 2, age 6-10 = 3, age 11-15 = 4, age 16-20 = 5, age 21-25 = 6).

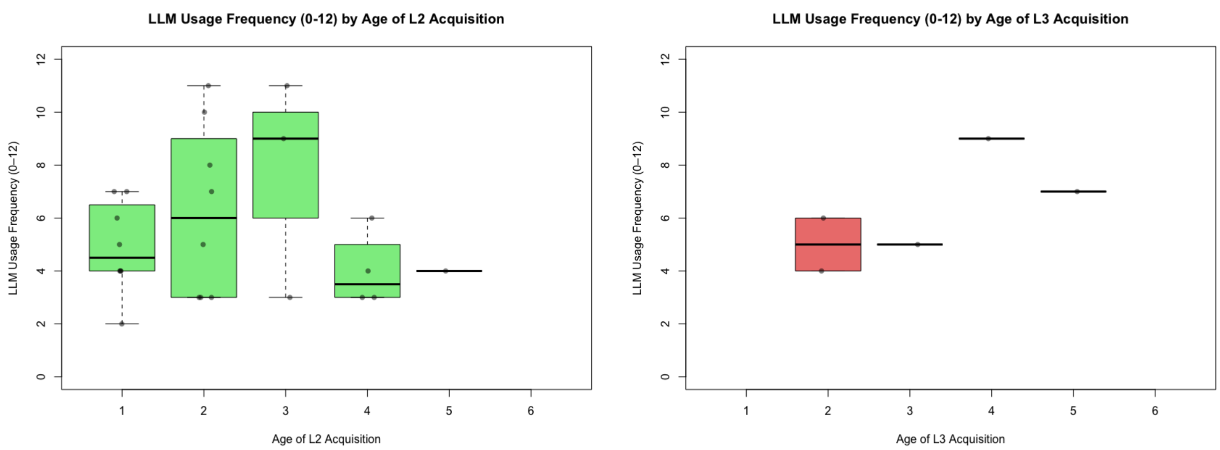

Figure 8: Boxplot illustrating LLM Usage Frequency (never = 0 - every 1 minute = 12) by Age of L2/L3 Acquisition (before age 1 = 1, age 1-5 = 2, age 6-10 = 3, age 11-15 = 4, age 16-20 = 5, age 21-25 = 6).

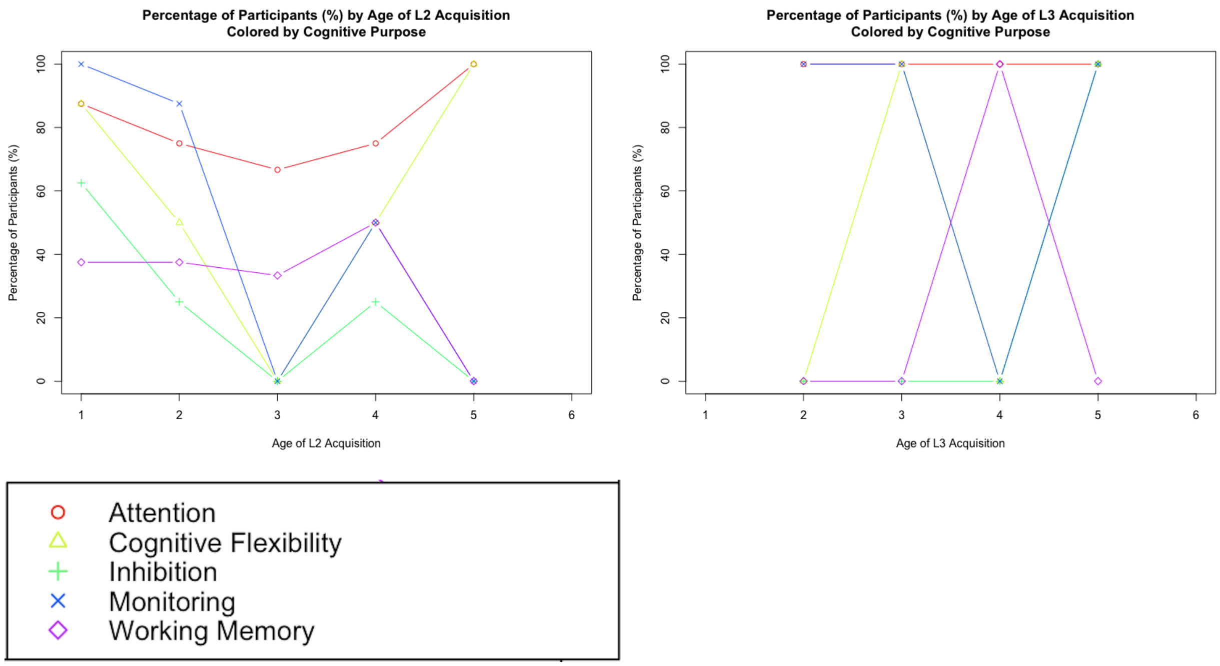

Figure 9: Graph illustrating Percentage Distribution of Participants (%) using LLMs for each Cognitive Purpose by Age of L2/L3 Acquisition (before age 1 = 1, age 1-5 = 2, age 6-10 = 3, age 11-15 = 4, age 16-20 = 5, age 21-25 = 6).

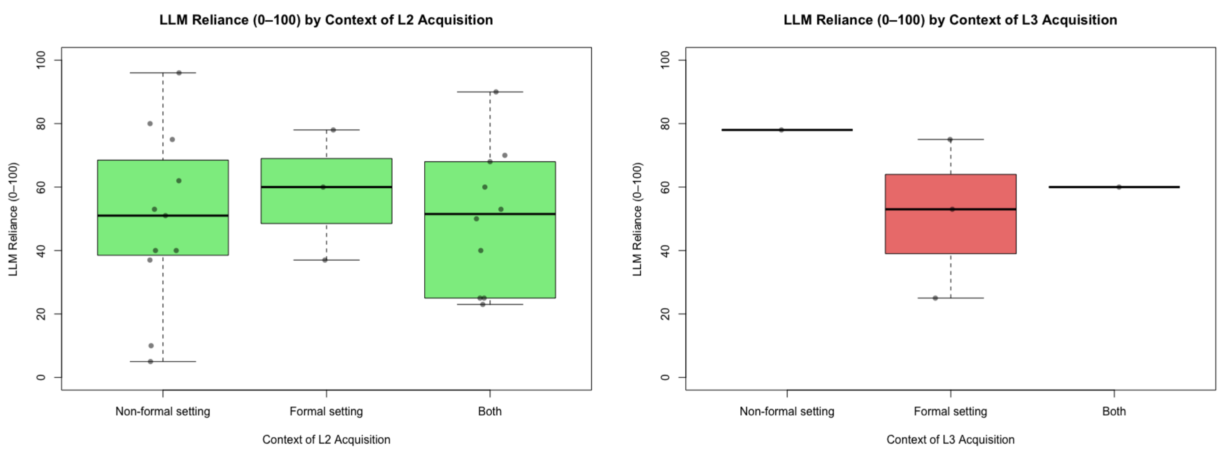

Figure 10: Boxplot illustrating LLM Reliance (0-100) by Context of L2/L3 Acquisition.

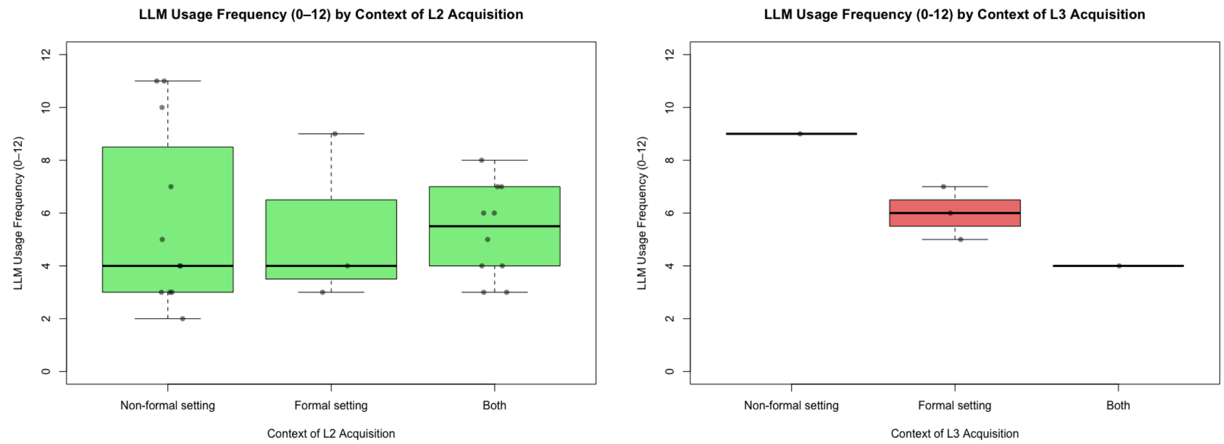

Figure 11: Boxplot illustrating LLM Usage Frequency (never = 0 - every 1 minute = 12) by Context of L2/L3 Acquisition.

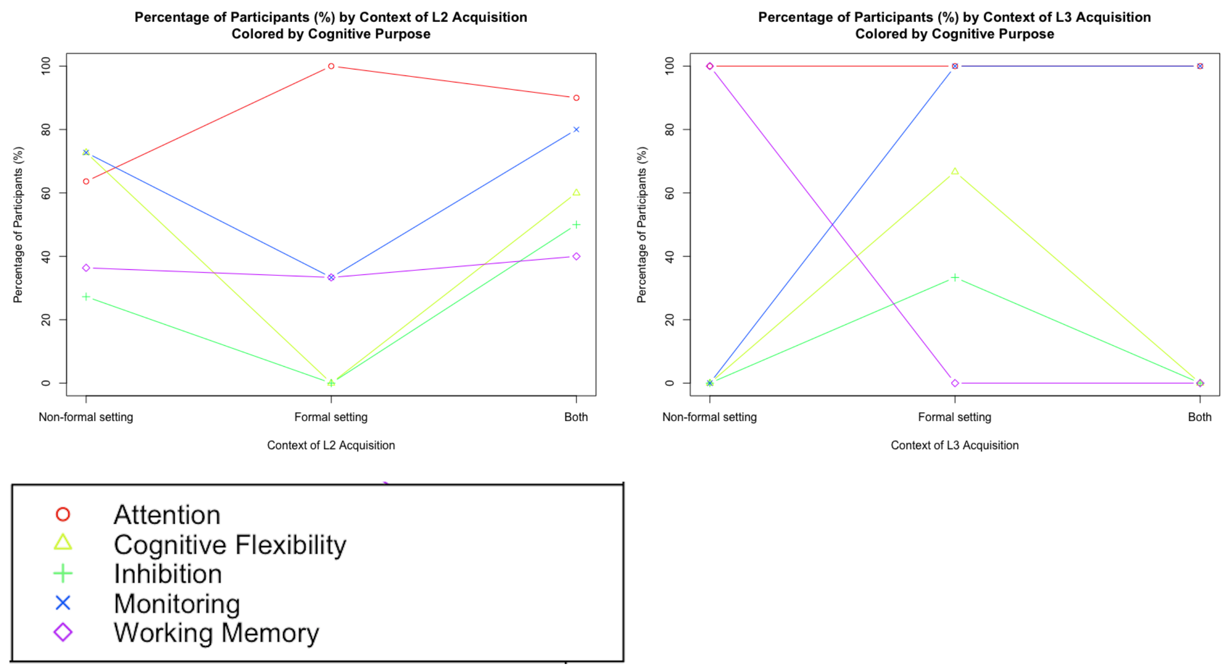

Figure 12: Graph illustrating Percentage Distribution of Participants (%) using LLMs for each Cognitive Purpose by Context of L2/L3 Acquisition.

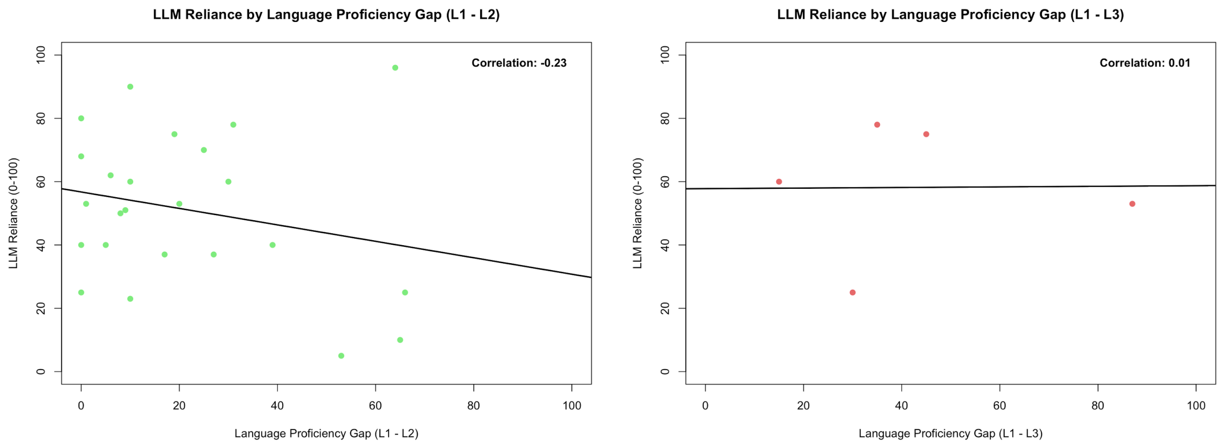

Figure 13: Graph illustrating LLM Reliance (0-100) by Language Proficiency Gap (L1 - L2 / L1 - L3).

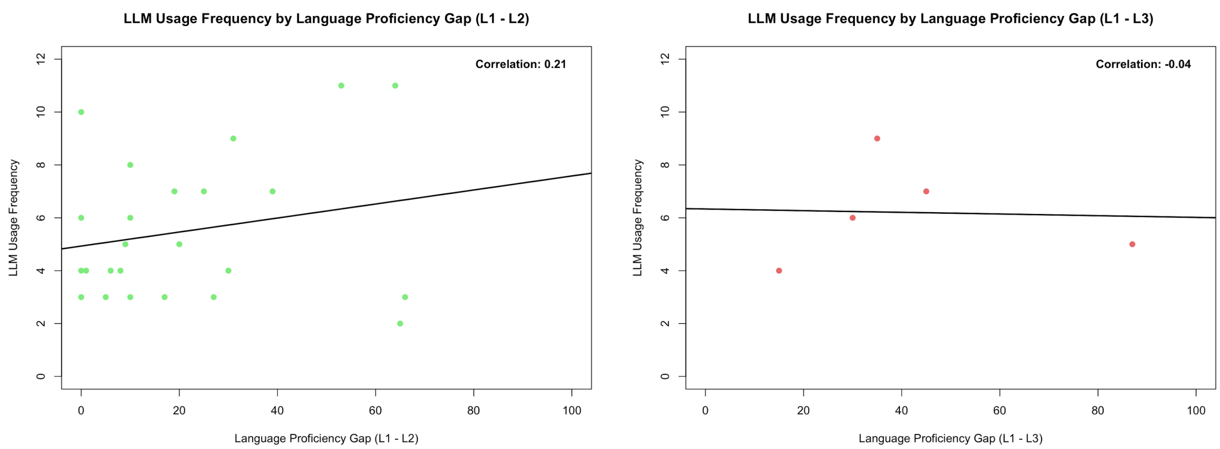

Figure 14: Graph illustrating LLM Usage Frequency (0-12) by Language Proficiency Gap (L1 - L2 / L1 - L3).

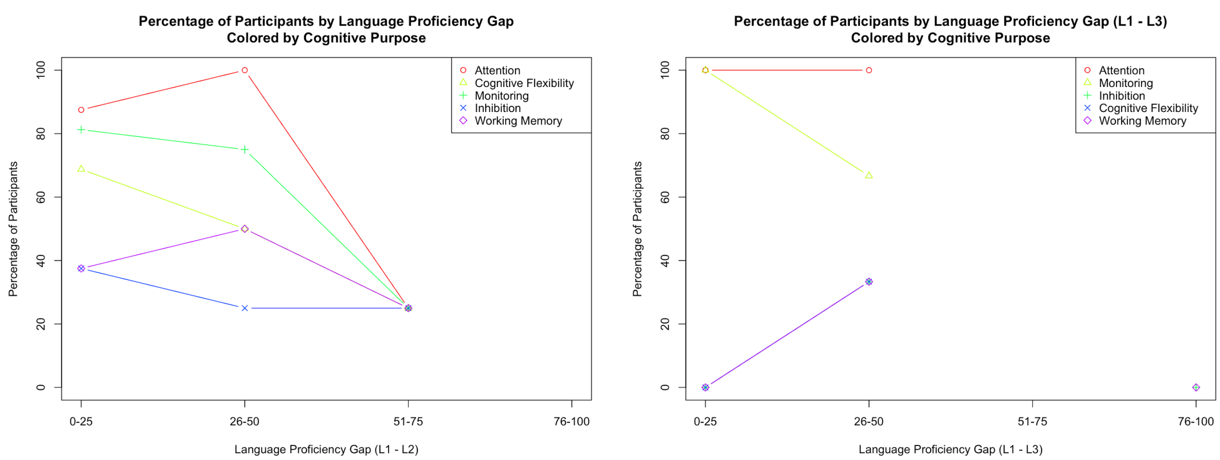

Figure 15: Graph illustrating Percentage Distribution of Participants (%) using LLMs for each Cognitive Purpose by Language Proficiency Gap (L1 - L2 / L1 - L3).

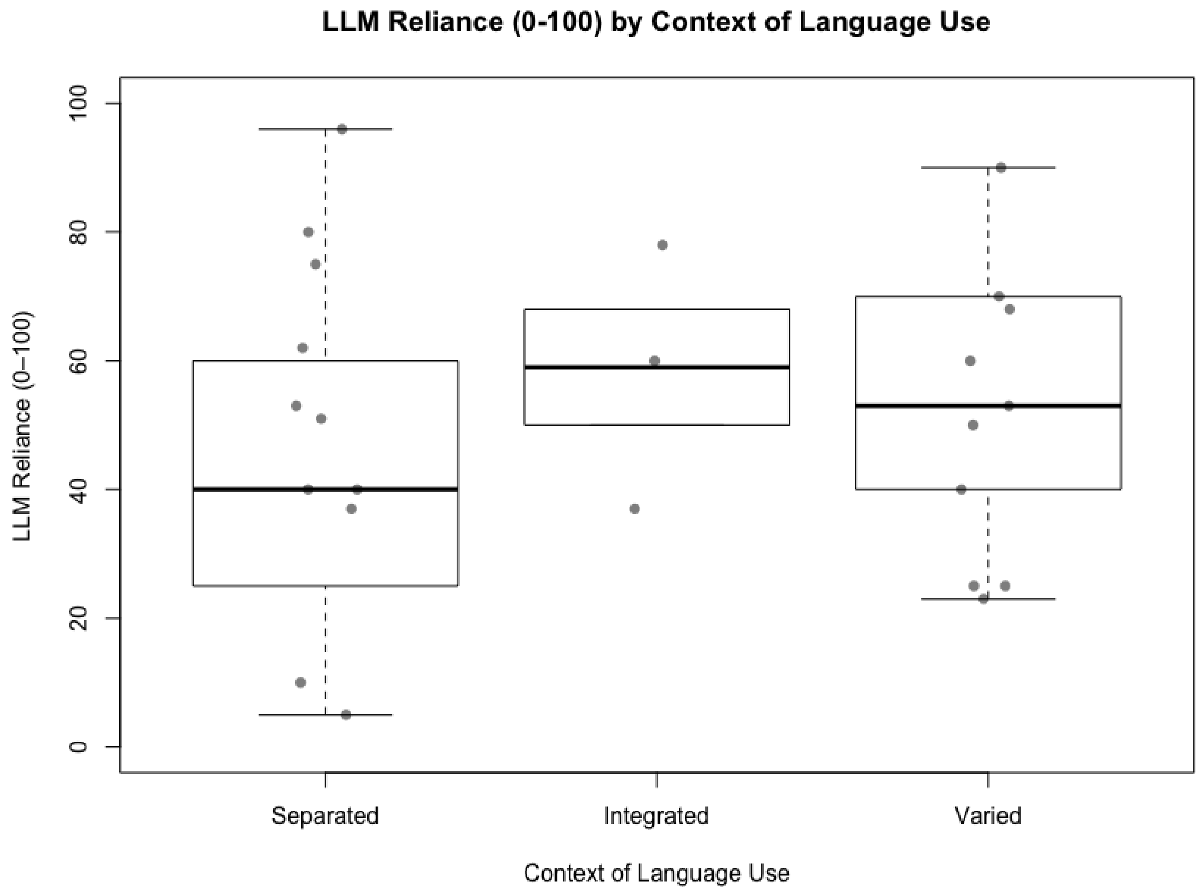

Figure 16: Boxplot illustrating LLM Reliance (0-100) by Context of Language Use.

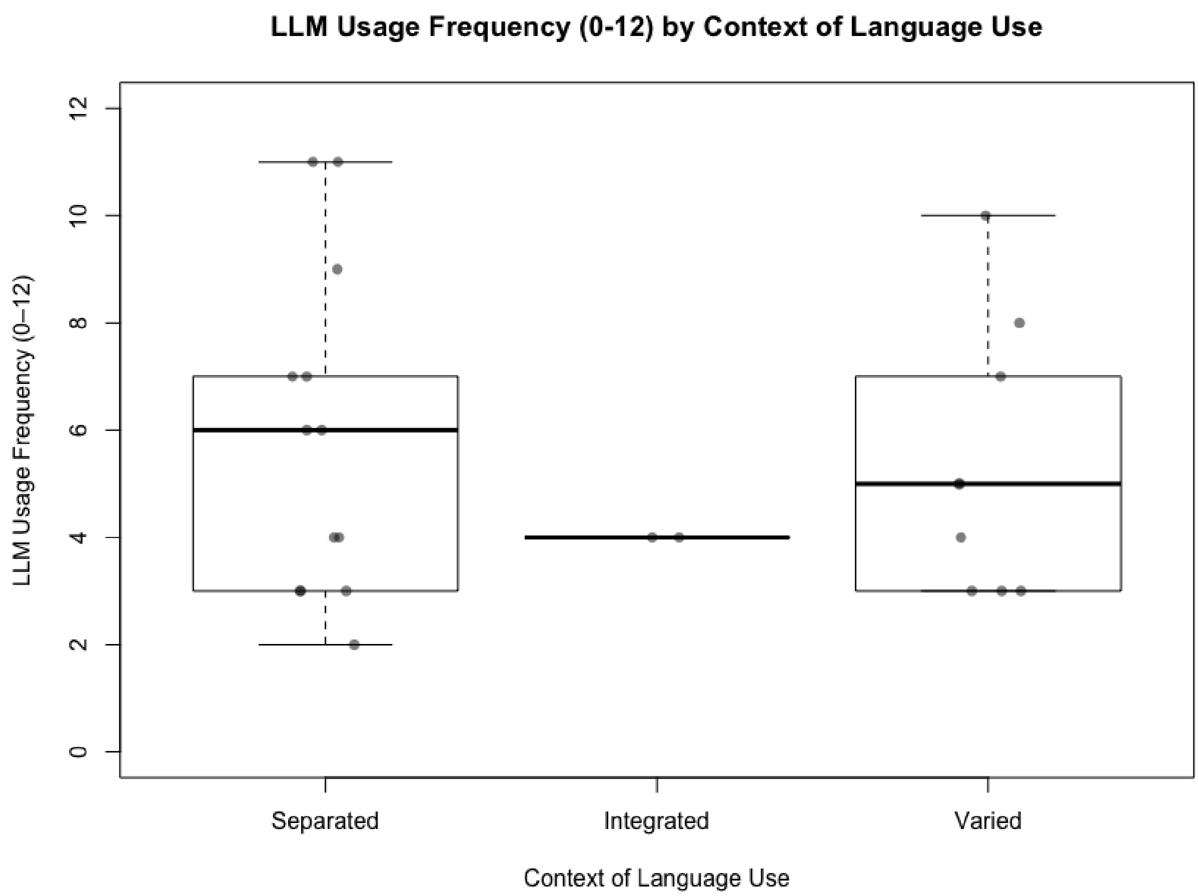

Figure 17: Boxplot illustrating LLM Usage Frequency (never = 0 - every 1 minute = 12) by Context of Language Use.

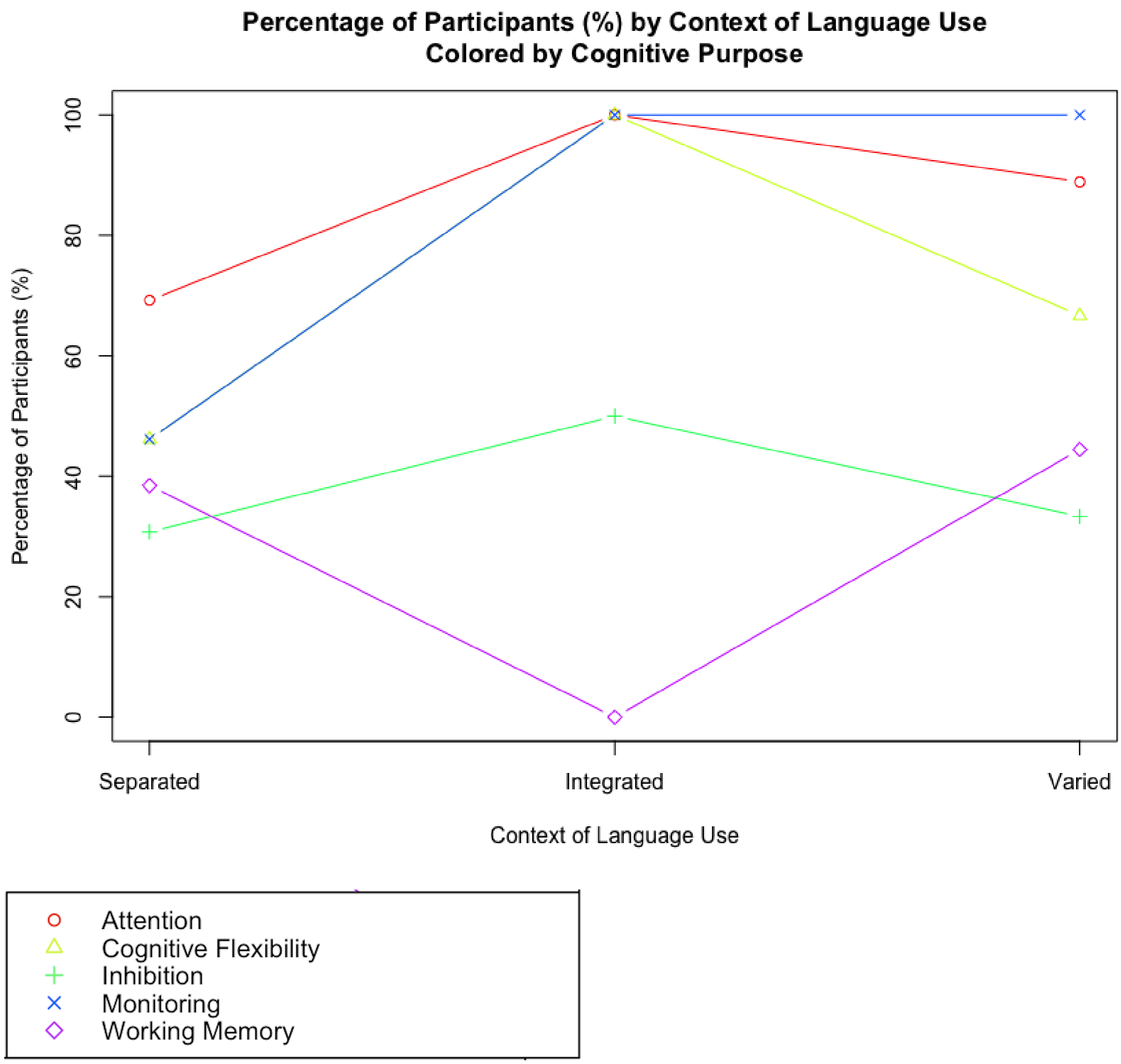

Figure 18: Graph illustrating Percentage Distribution of Participants (%) using LLMs for each Cognitive Purpose by Context of Language Use.

2) Large Language Models (LLMs) Use and Performance in Cognitive Tasks (Simon and AX-CPT)

Both Simon and AX-CPT tasks appear to provide valid measures. A repeated-measures ANOVA on Simon task results revealed a trend toward differences in accuracy across trial types, F(3,51) = 2.512, *p = 0.069*, though this did not reach statistical significance at the critical value 0.05 (*Figure 19*). Accuracy on the AX-CPT task differed significantly across trial types, F(3,33) = 3.12, *p* = 0.039 (*Figure 21*). There was a significant overall effect of trial type on reaction time, F(3,33) = 7.70, *p* = 0.00049 (*Figure 22*). Pairwise comparisons revealed that participants responded significantly faster on BY trials compared to

AY trials (p = 0.005). There were also trend-level differences suggesting faster responses on BY trials compared to AX (p = 0.054) and BX (p = 0.064) trials.

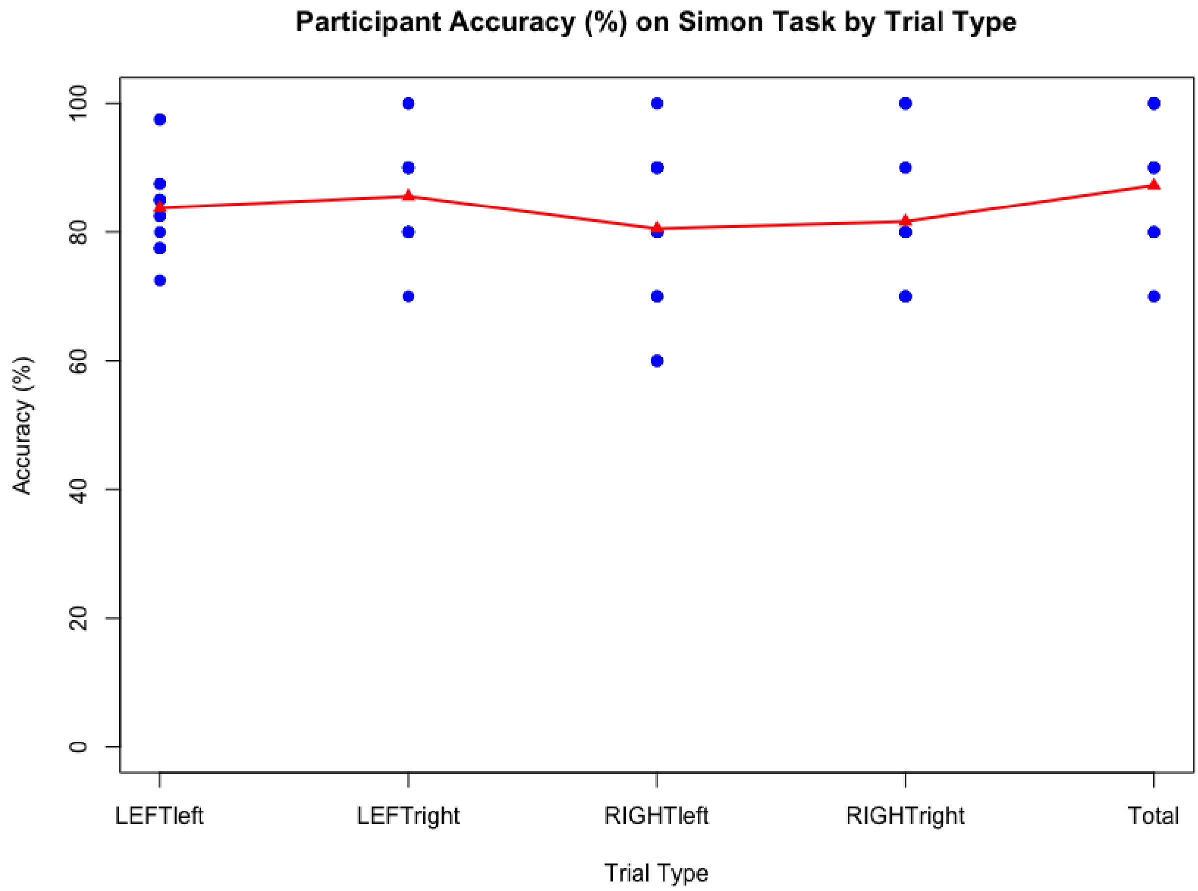

*Figure 19: Graph illustrating Participant Accuracy (%) on Simon Task by Trial Type. *Red line represents the mean value.*

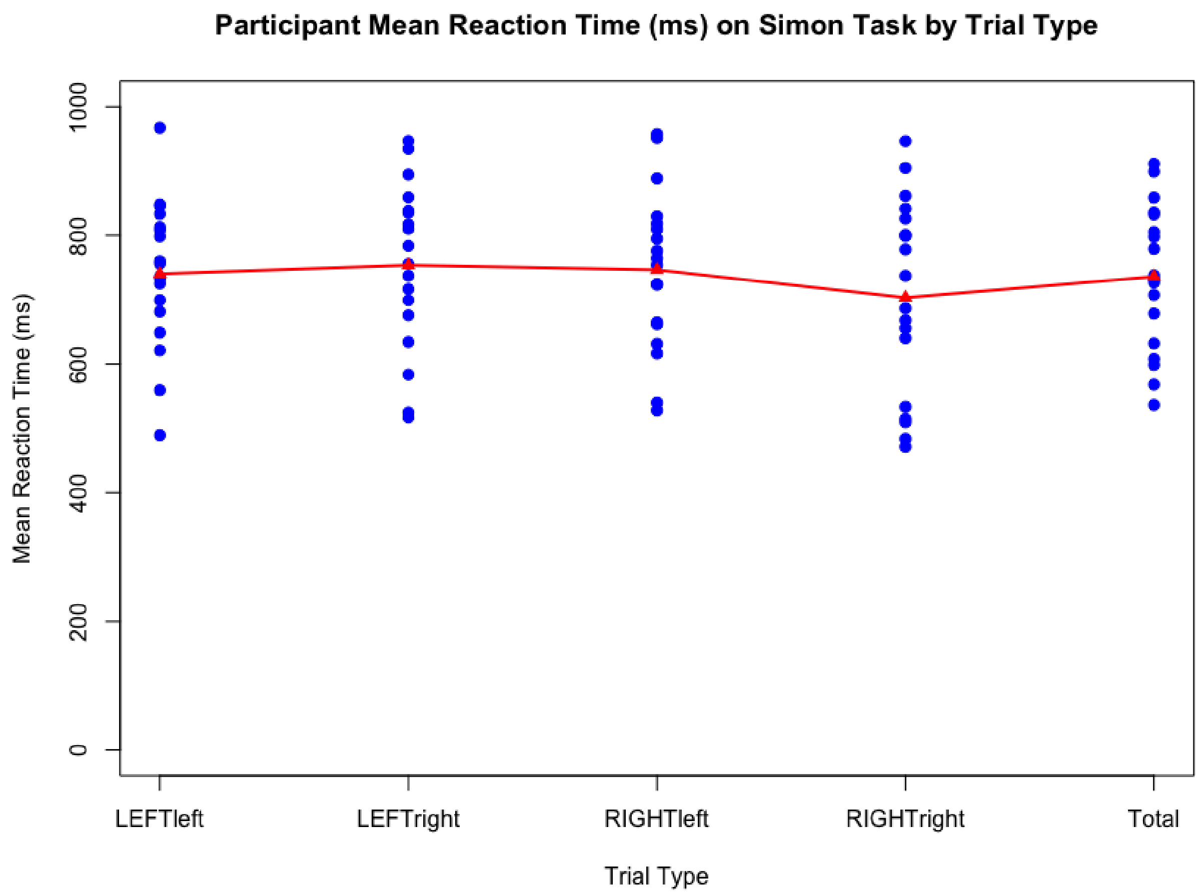

*Figure 20: Graph illustrating Participant Mean Reaction Time (ms) on Simon Task by Trial Type. *Red line represents the mean value.*

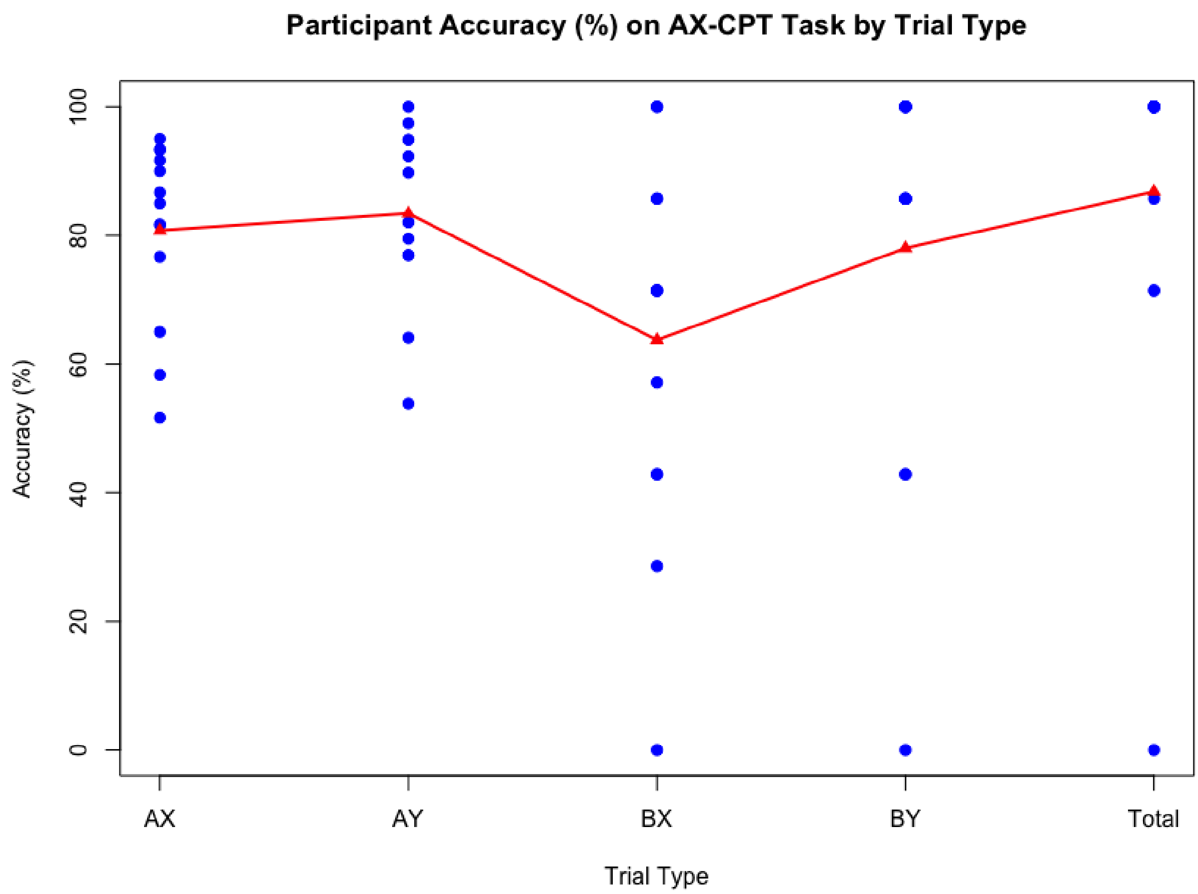

*Figure 21: Graph illustrating Participant Accuracy (%) on AX-CPT Task by Trial Type. *Red line represents the mean value.*

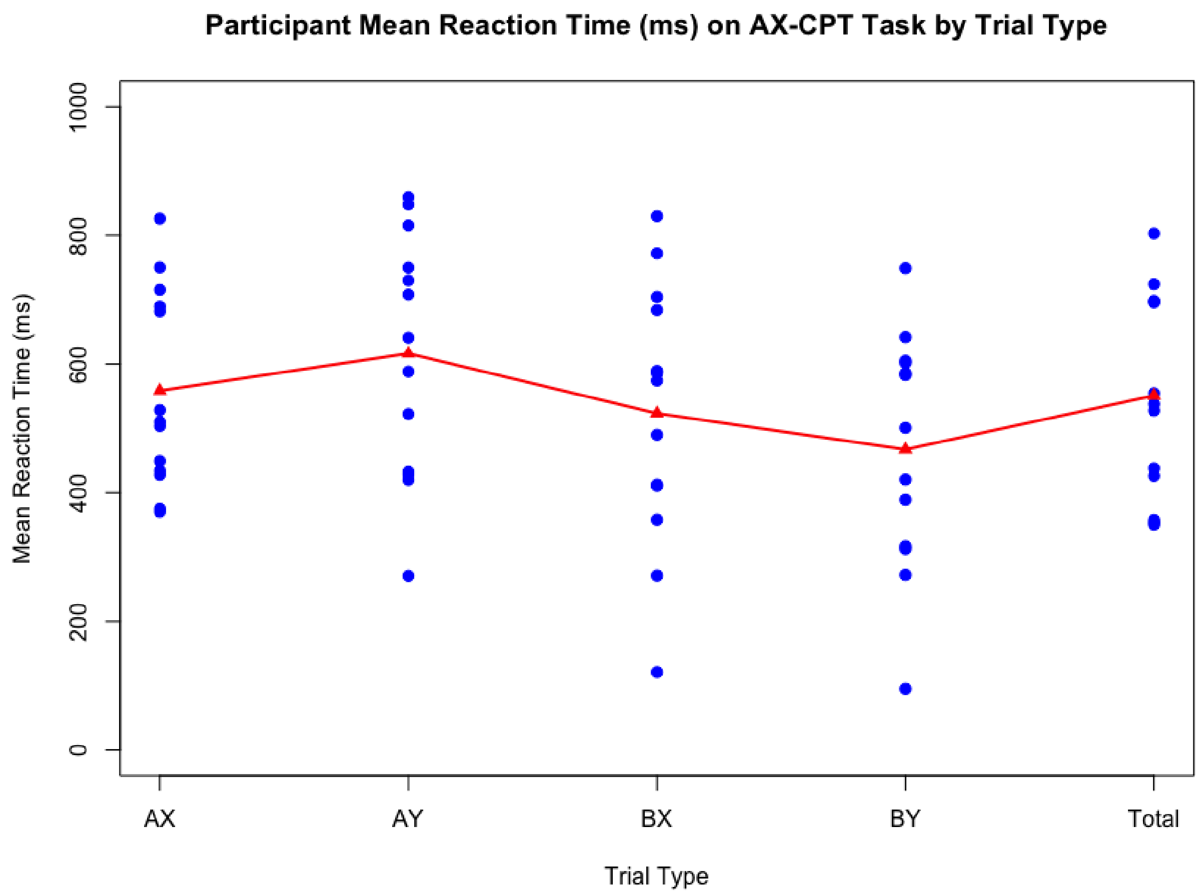

*Figure 22: Graph illustrating Participant Mean Reaction Time (ms) on AX-CPT Task by Trial Type. *Red line represents the mean value.*

Participants who were more active LLM users performed worse on these tasks (*Figures 23, 25, 27, 29, 31, 33*), as indicated by the predominantly negative correlations except in graphs framed in red. Moreover, more active LLM users spent less time completing the tasks (*Figures 24, 26, 28, 30, 32, 34*). Moderate negative correlation coefficients were observed across all graphs depicting participant accuracy in BY trials relative to LLM engagement on the AX-CPT tasks (*Figures 29, 31, 35*).

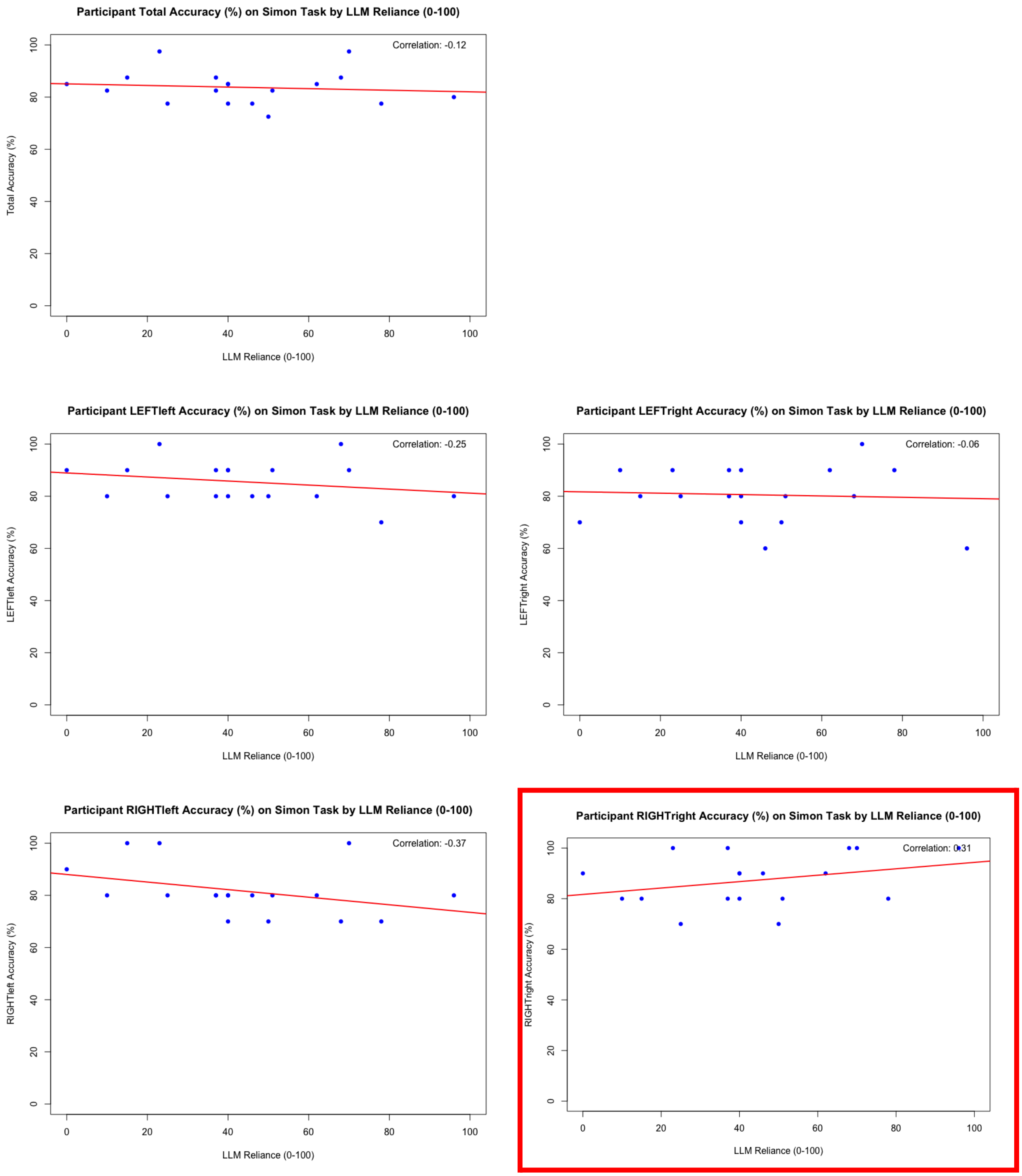

Figure 23: Graph illustrating Participant Accuracy (%) on Simon Task by LLM Reliance (0-100).

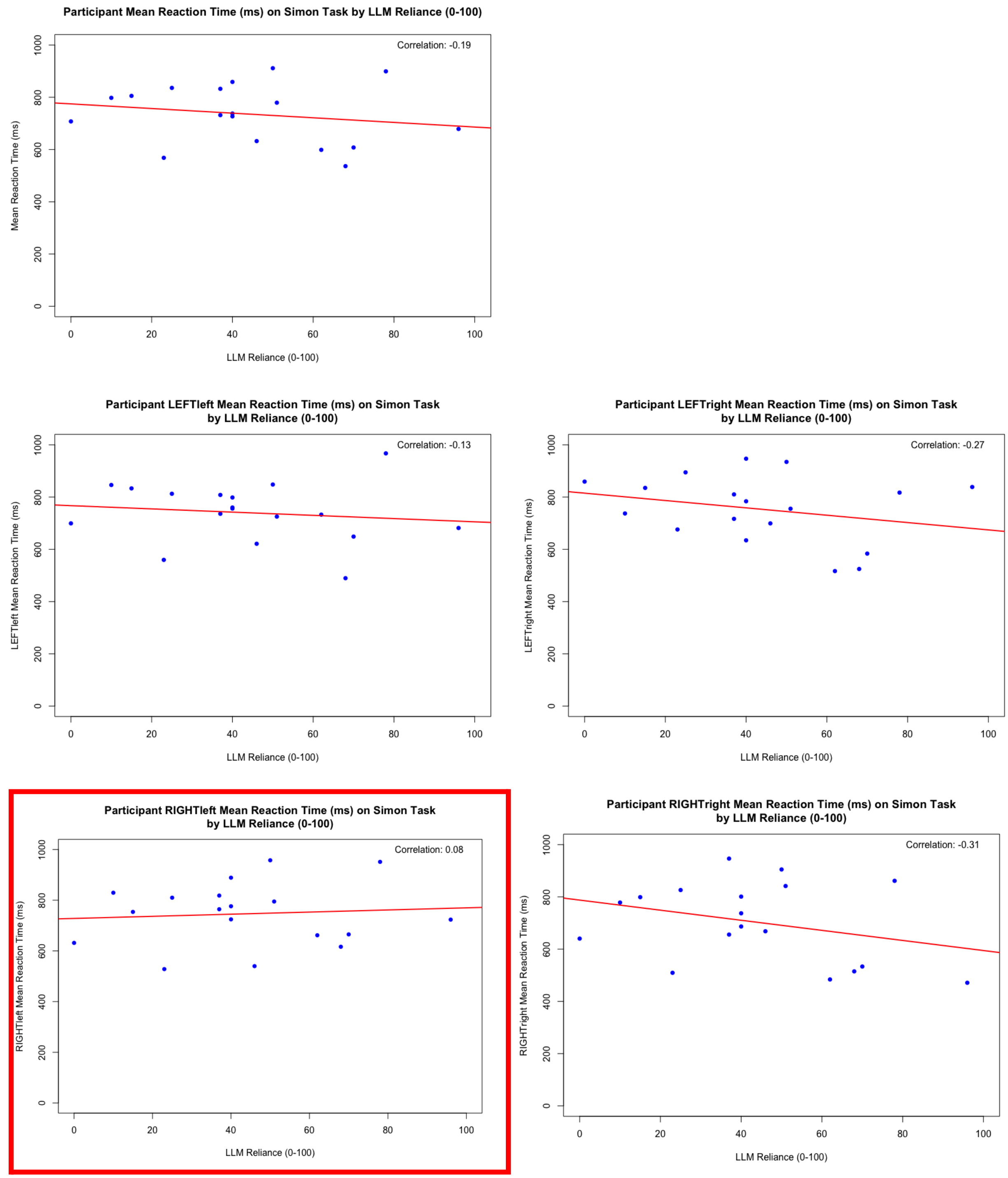

Figure 24: Graph illustrating Participant Mean Reaction Time (ms) on Simon Task by LLM Reliance (0-100).

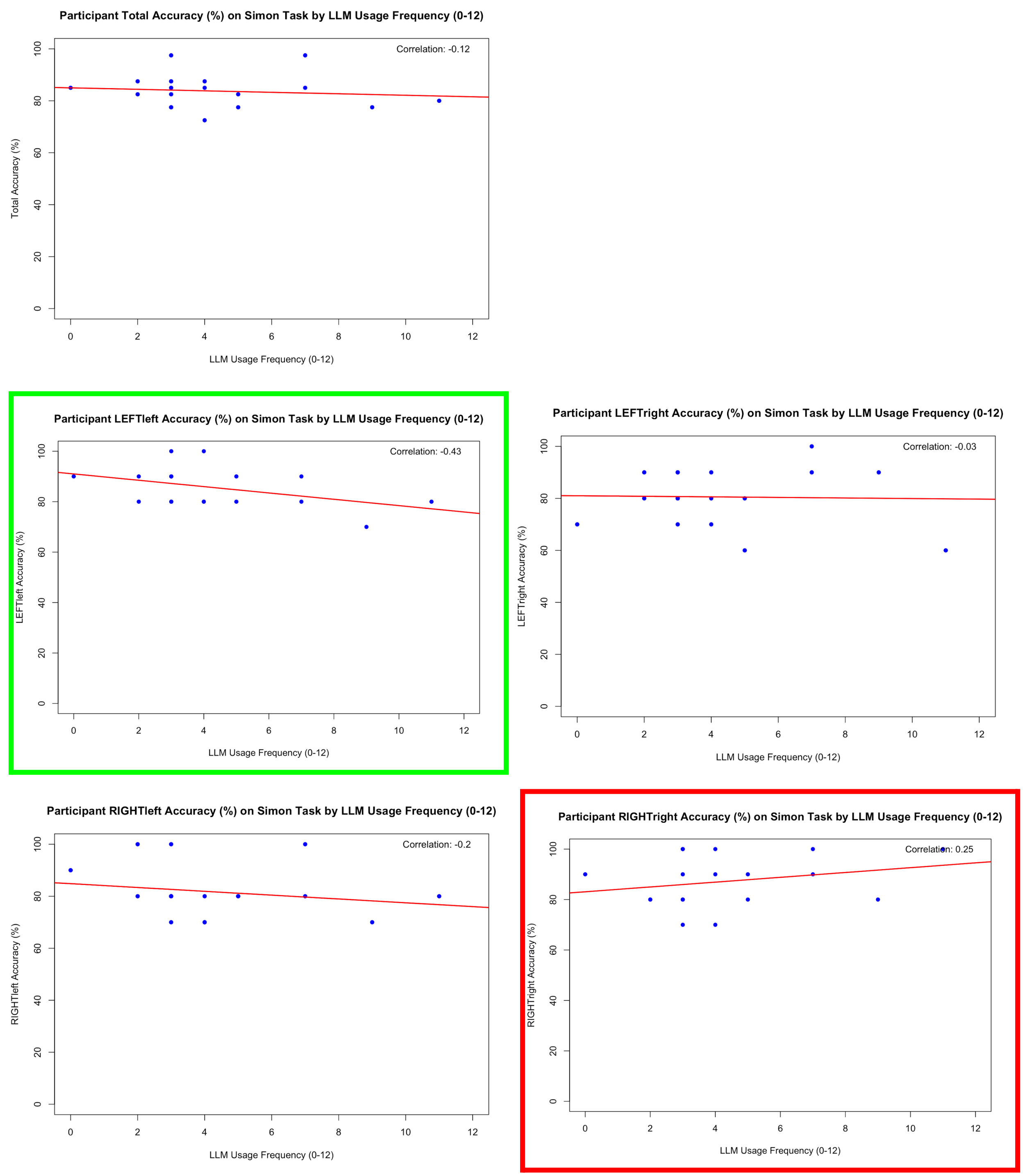

Figure 25: Graph illustrating Participant Accuracy (%) on Simon Task by LLM Usage Frequency (0-12).

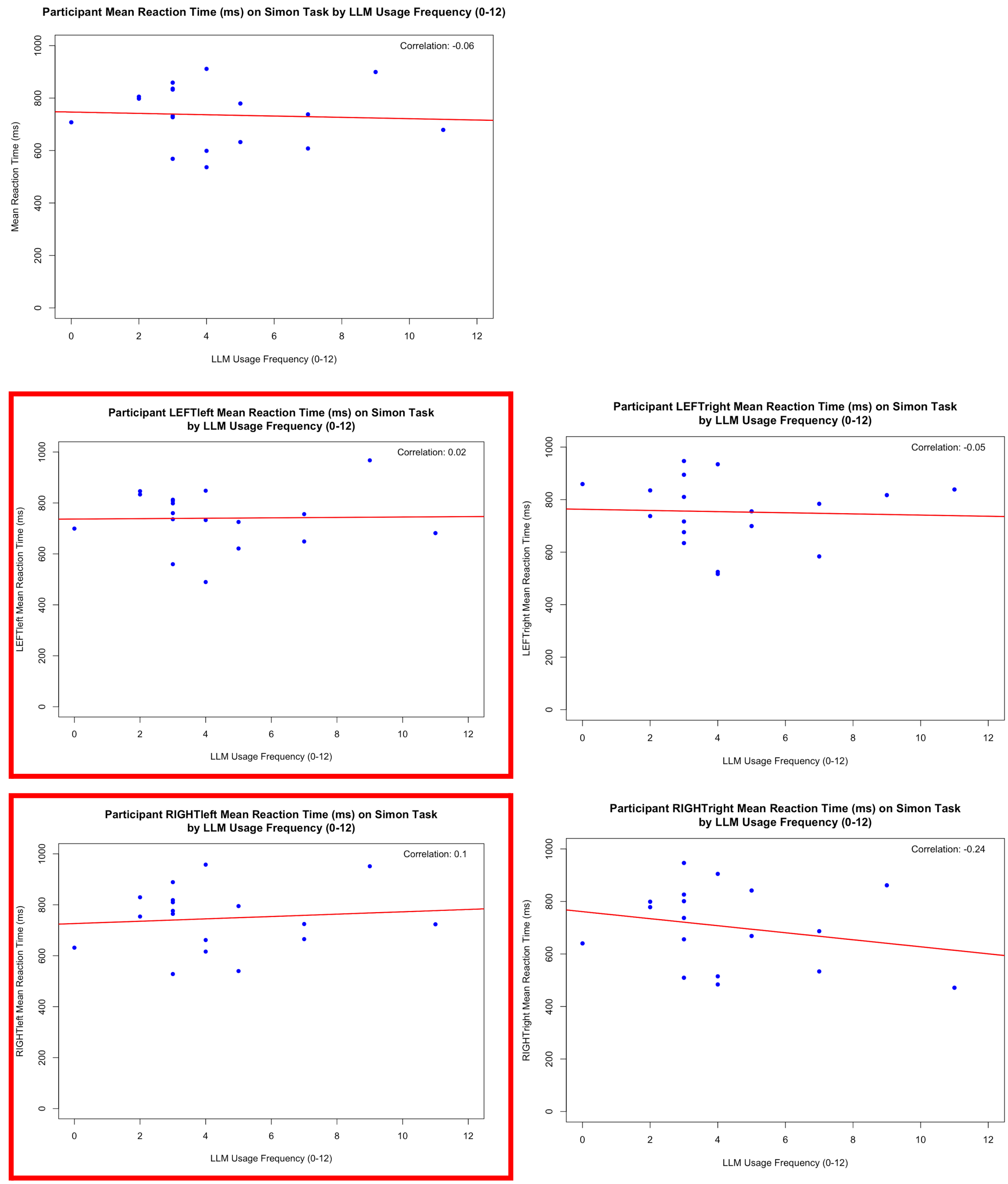

Figure 26: Graph illustrating Participant Mean Reaction Time (ms) on Simon Task by LLM Usage Frequency (0-12).

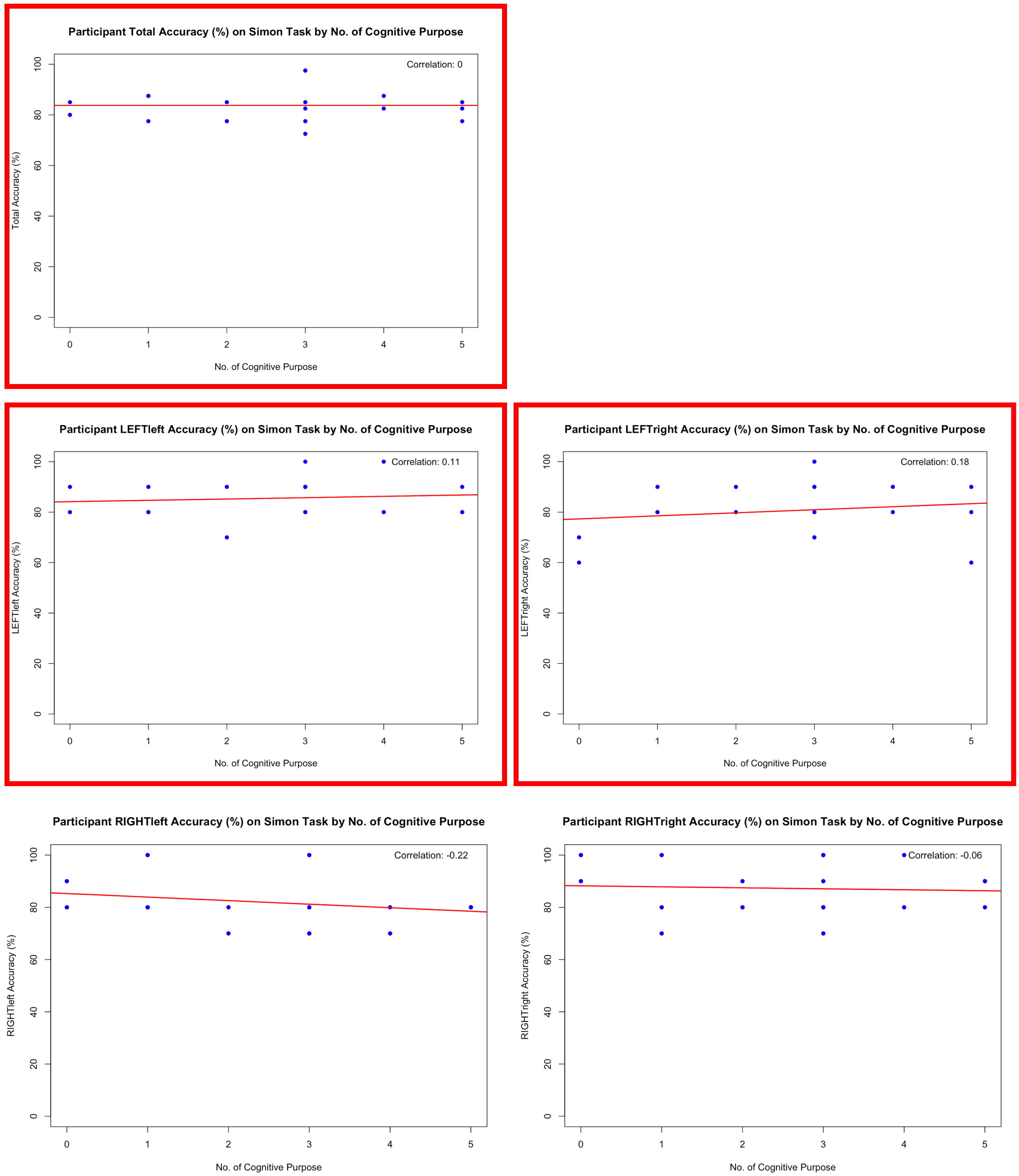

Figure 27: Graph illustrating Participant Accuracy (%) on Simon Task by No. of Cognitive Purpose.

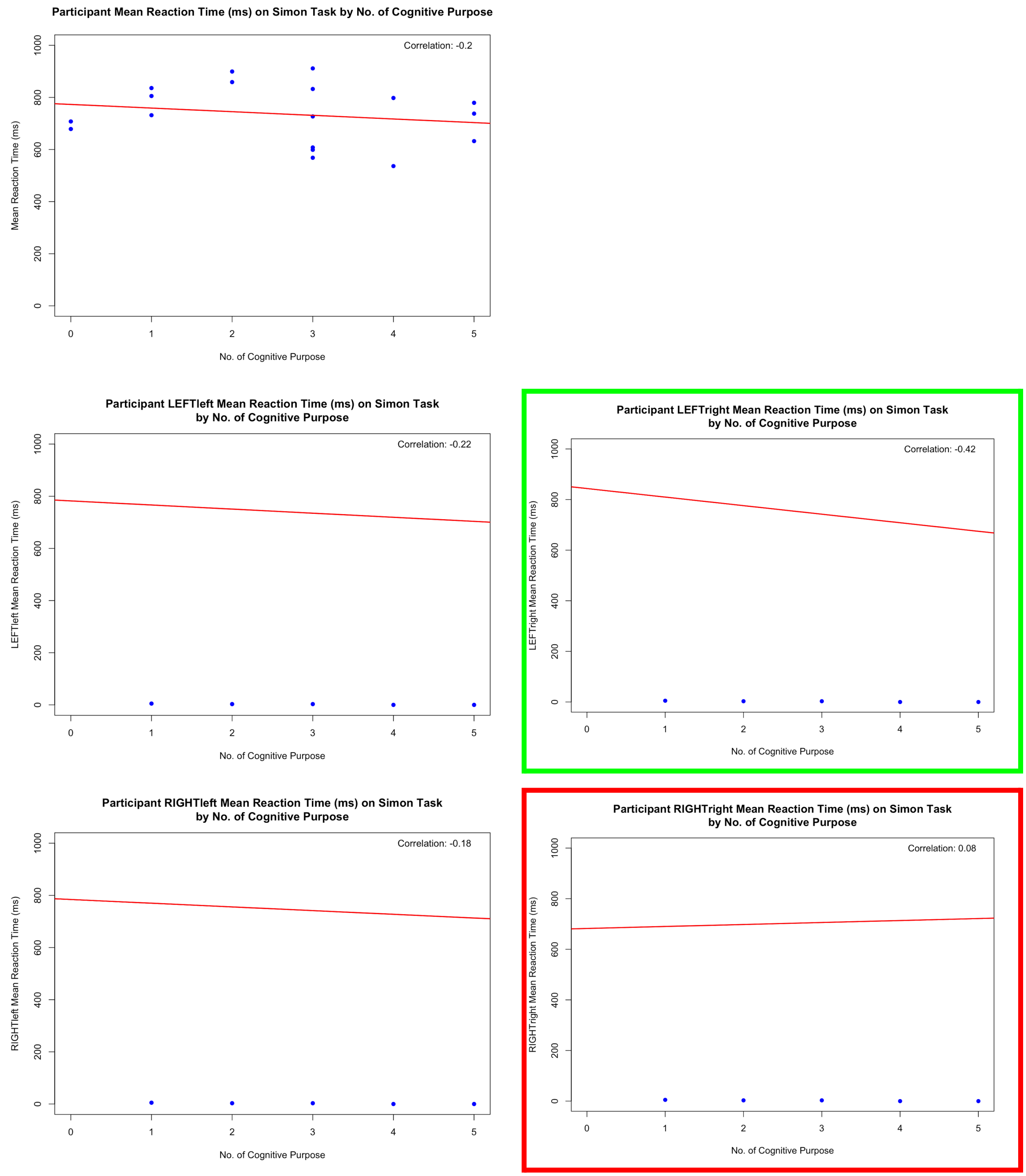

Figure 28: Graph illustrating Participant Mean Reaction Time (ms) on Simon Task by No. of Cognitive Purpose.

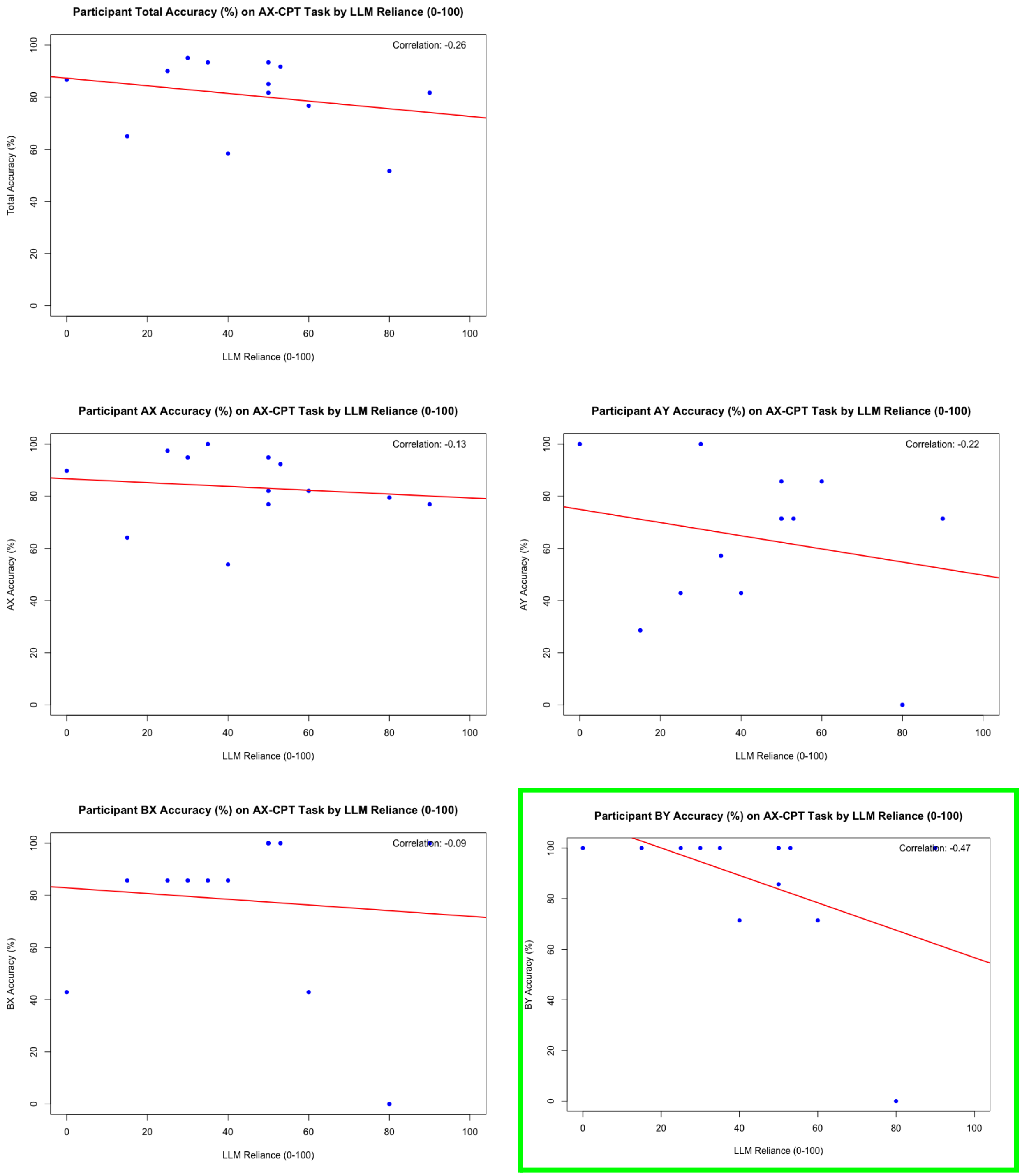

Figure 29: Graph illustrating Participant Accuracy (%) on AX-CPT Task by LLM Reliance (0-100).

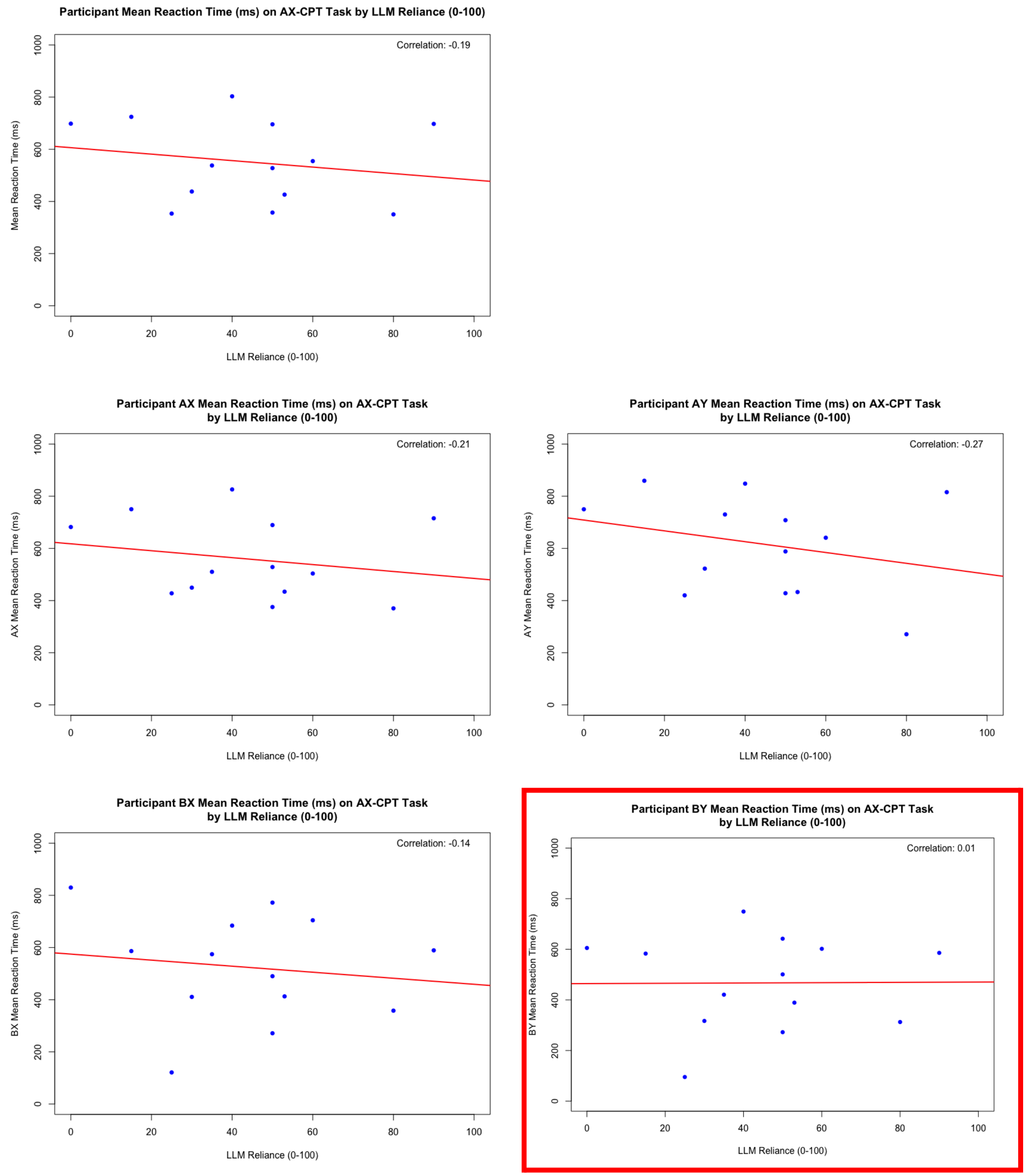

Figure 30: Graph illustrating Participant Mean Reaction Time (ms) on AX-CPT Task by LLM Reliance (0-100).

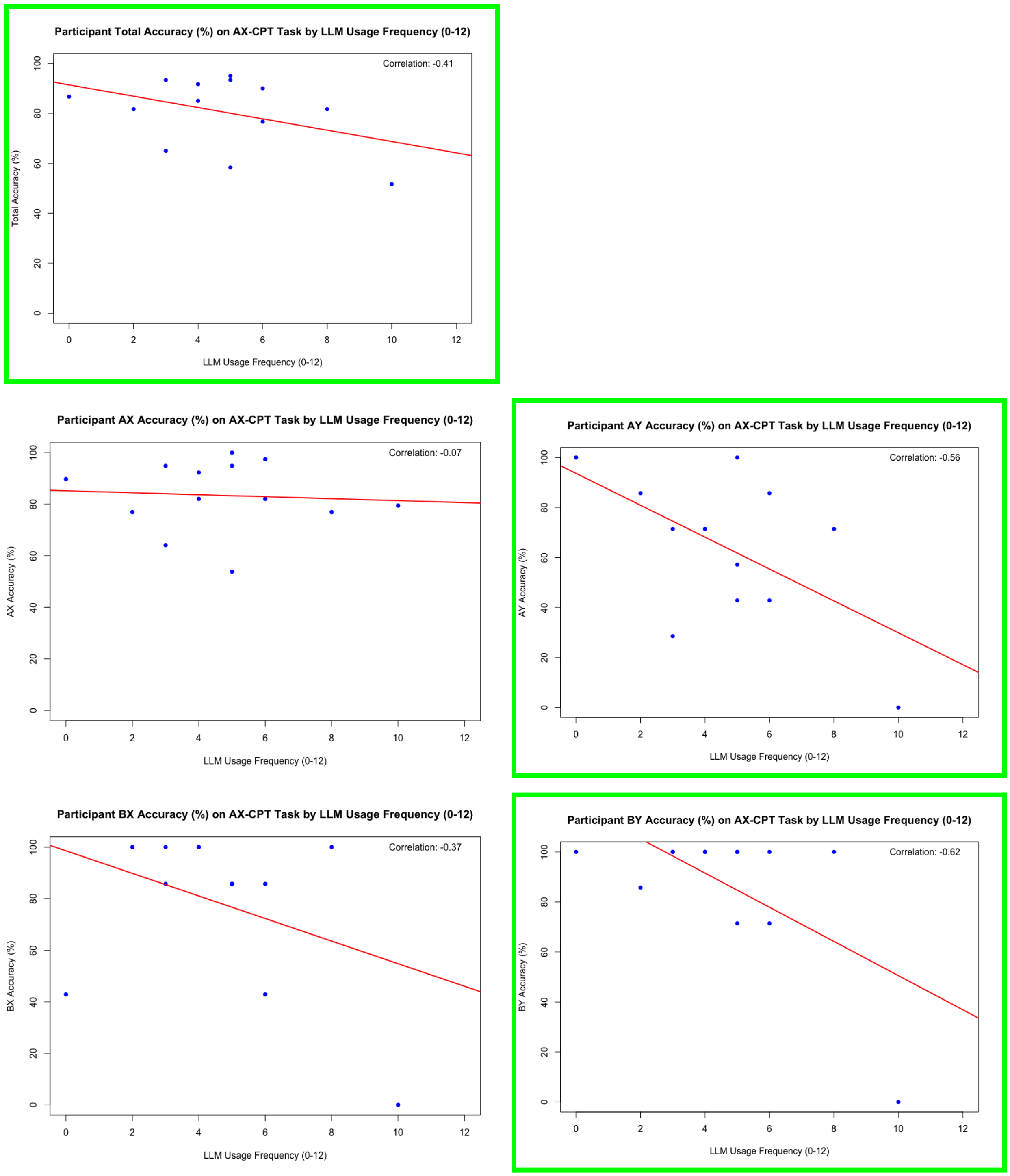

Figure 31: Graph illustrating Participant Accuracy (%) on AX-CPT Task by LLM Usage Frequency (0-12).

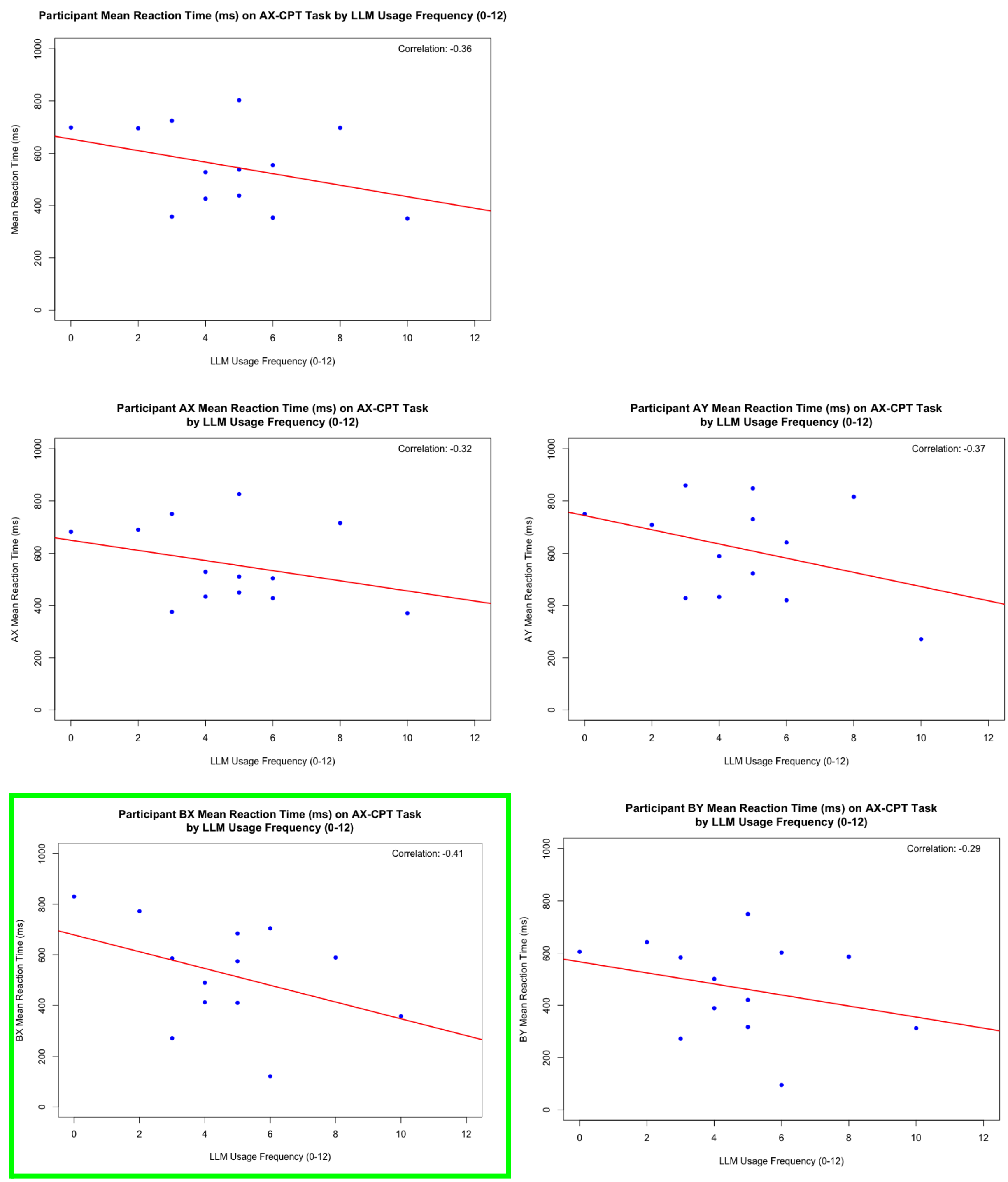

Figure 32: Graph illustrating Participant Mean Reaction Time (ms) on AX-CPT Task by LLM Usage Frequency (0-12).

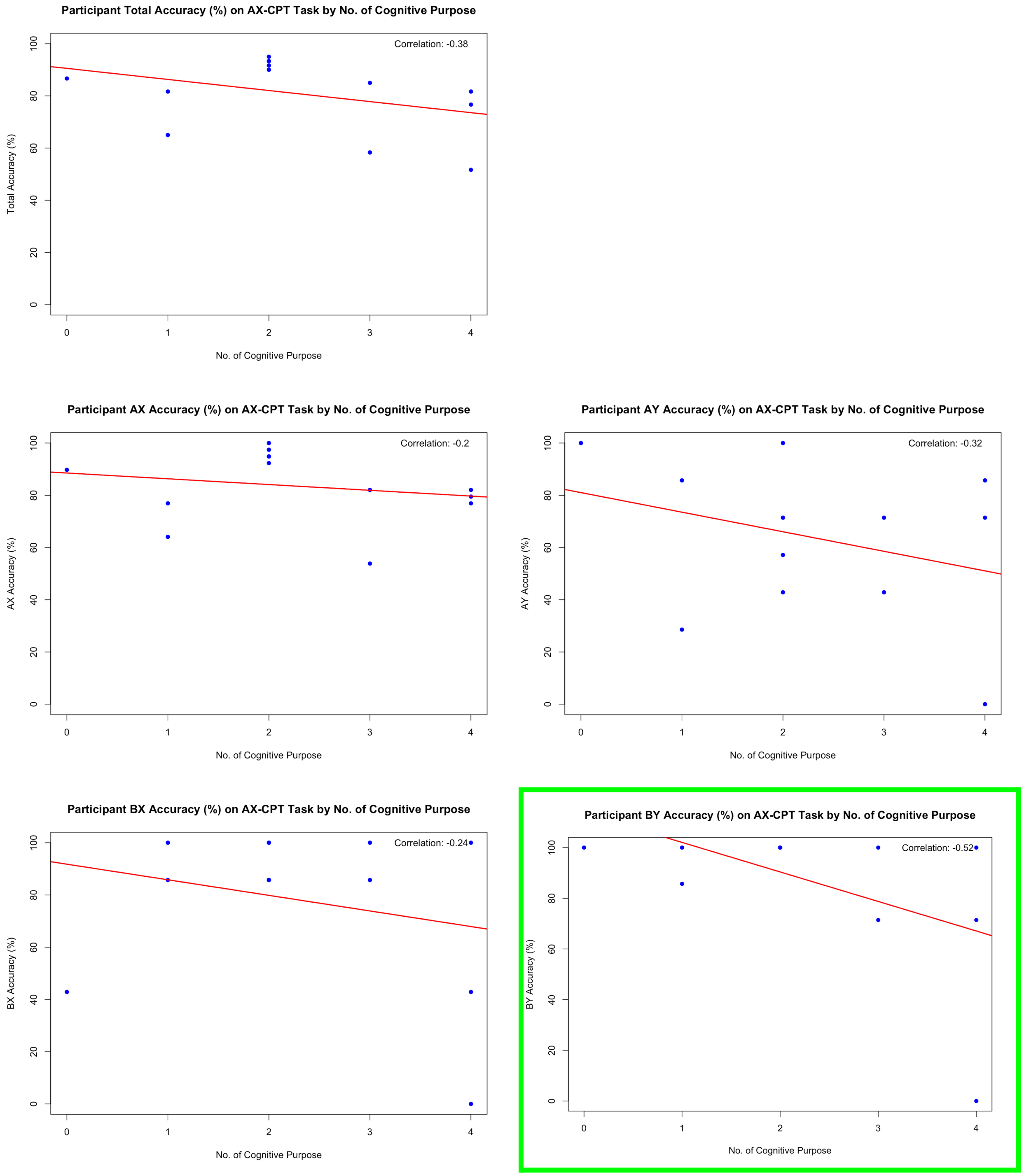

Figure 33: Graph illustrating Participant Accuracy (%) on AX-CPT Task by No. of Cognitive Purpose.

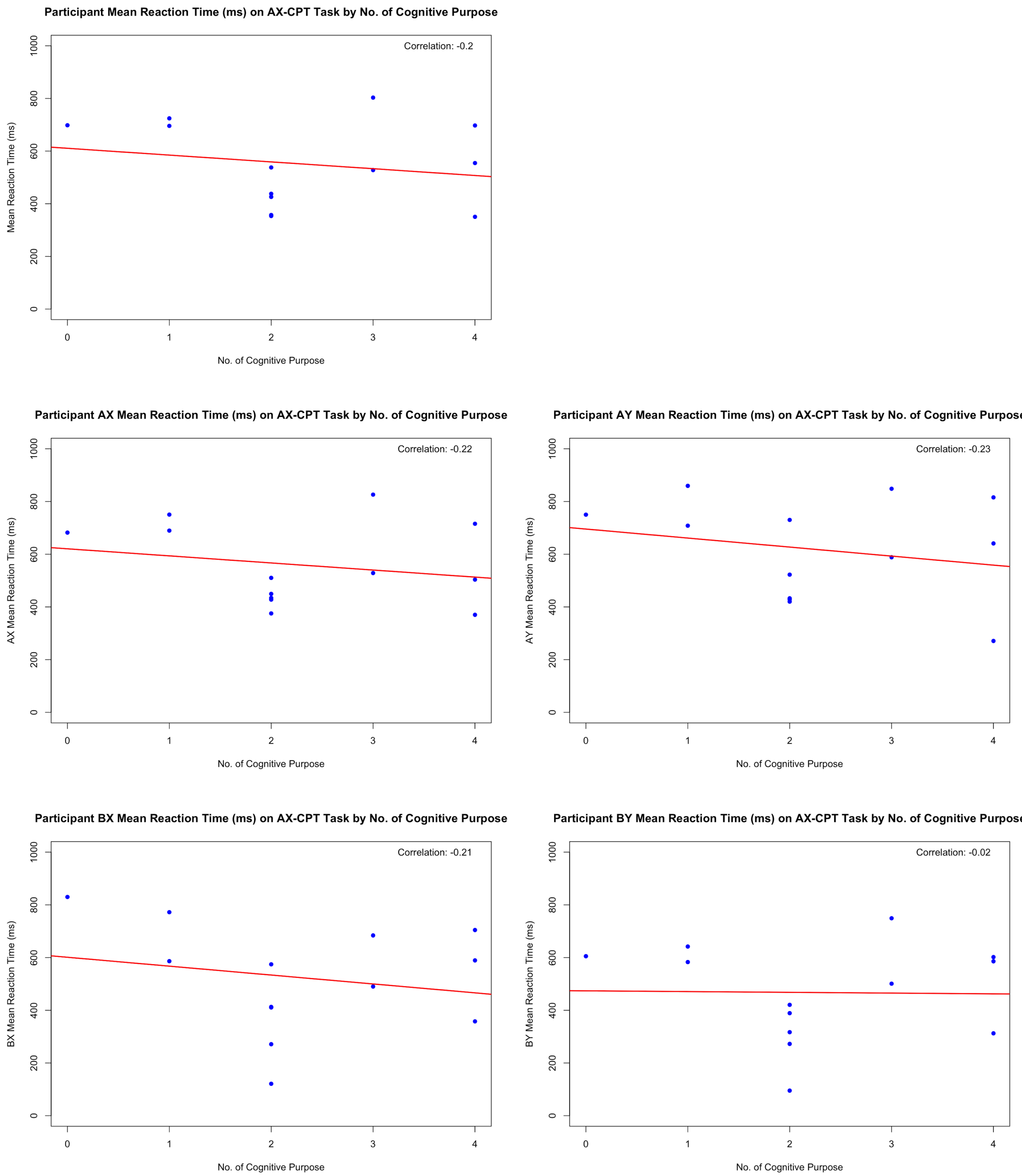

Figure 34: Graph illustrating Participant Mean Reaction Time (ms) on AX-CPT Task by No. of Cognitive Purpose.

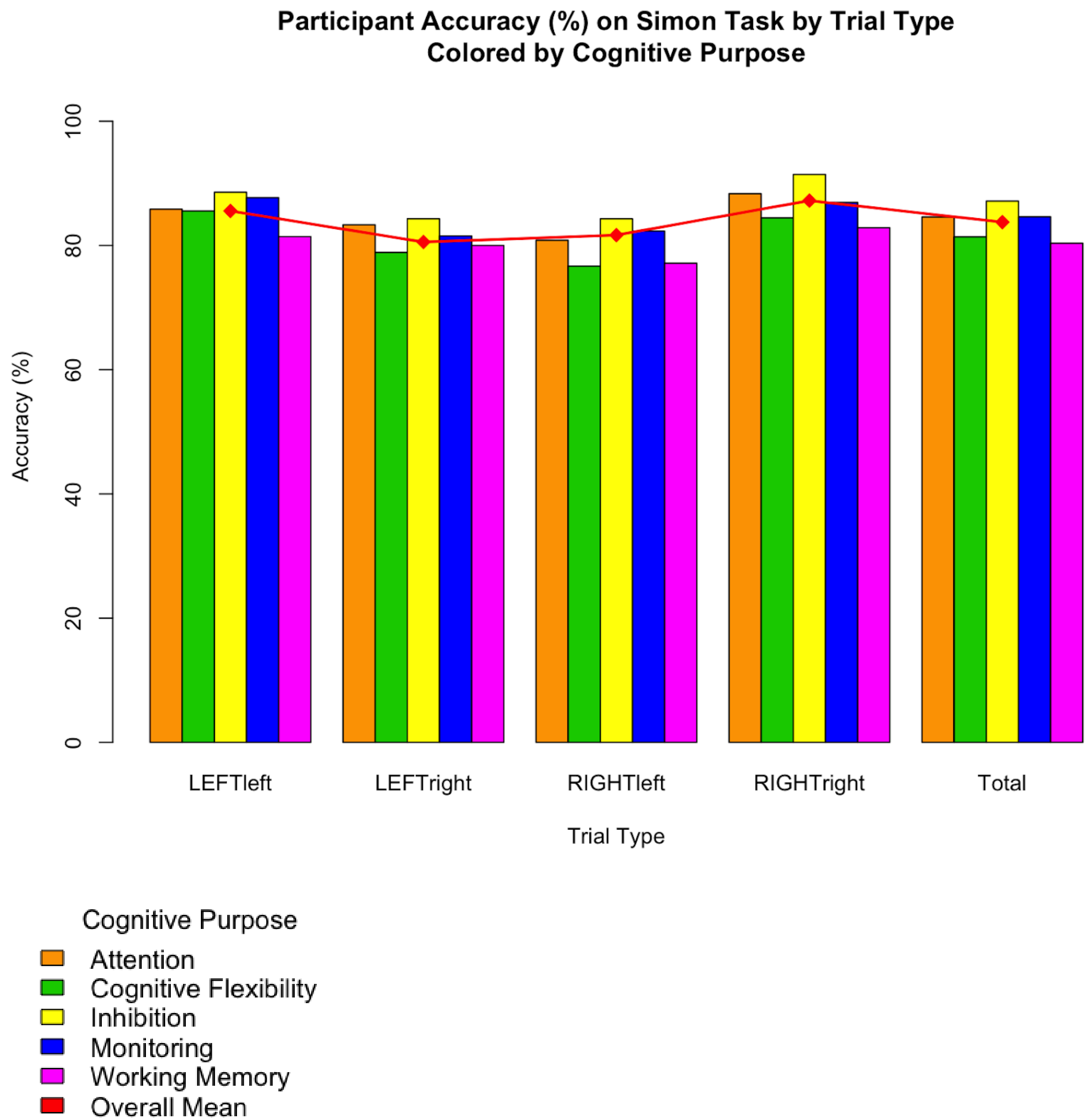

*Figure 35: Graph illustrating Participant Accuracy (%) on Simon Task by Trial Type Colored by Cognitive Purpose. *Red line represents the mean value.*

Participants who relied on LLMs for working memory showed the poorest overall performance on the AX-CPT Task (*Figure 36*). They performed worst on both AY and BX trials. Across groups, LLM use for any cognitive purpose was linked to poorer performance on AY trials compared to BX trials. Similarly, as participants offloaded more cognitive functions to LLMs, their accuracy on AY trials declined, while BX accuracy remained largely stable (*Figure 38*). Greater frequency of LLM use was also associated with reduced AY accuracy, with BX accuracy showing little change (*Figure 40*).

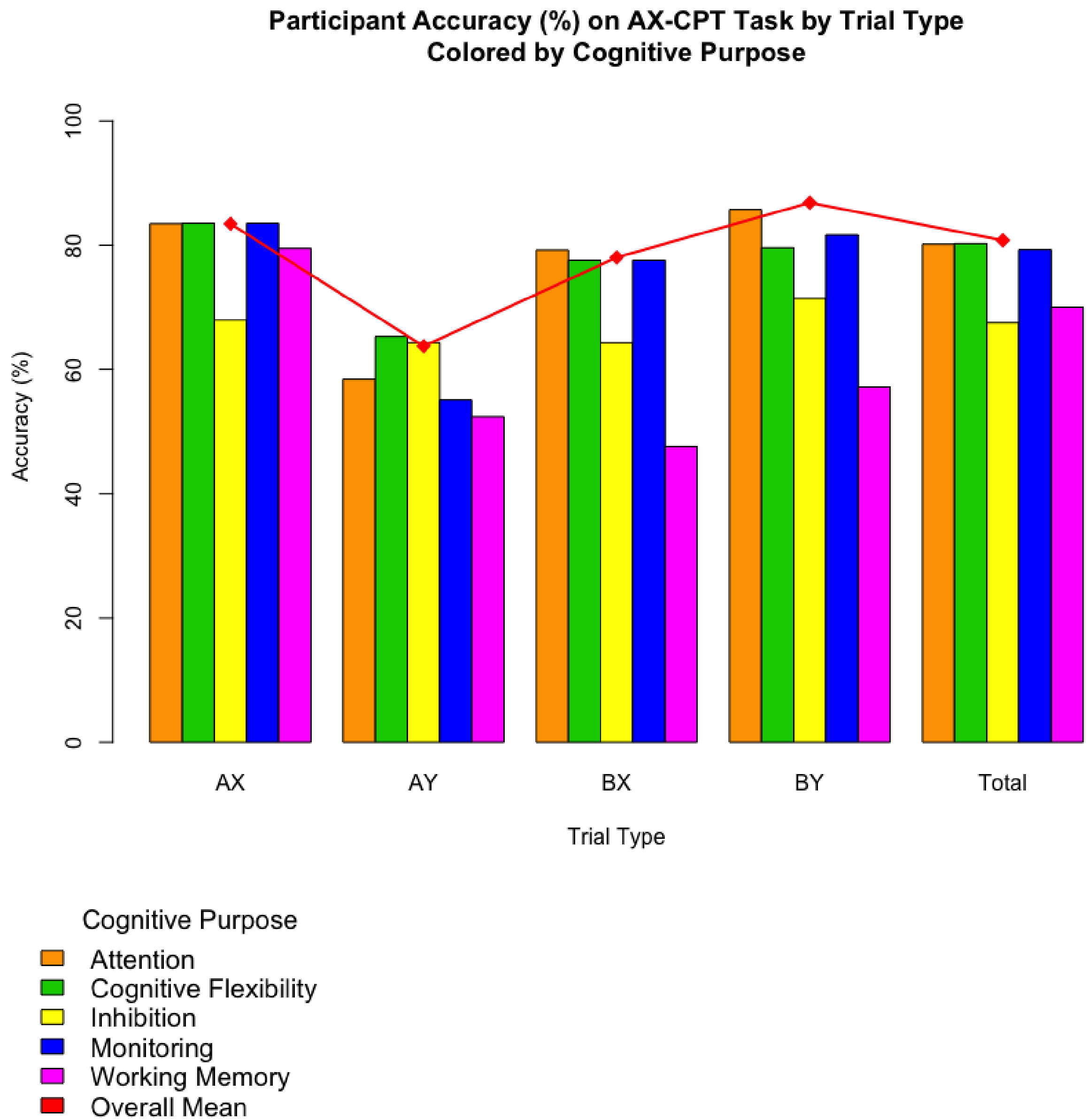

*Figure 36: Graph illustrating Participant Accuracy (%) on AX-CPT Task by Trial Type Colored by Cognitive Purpose. *Red line represents the mean value.*

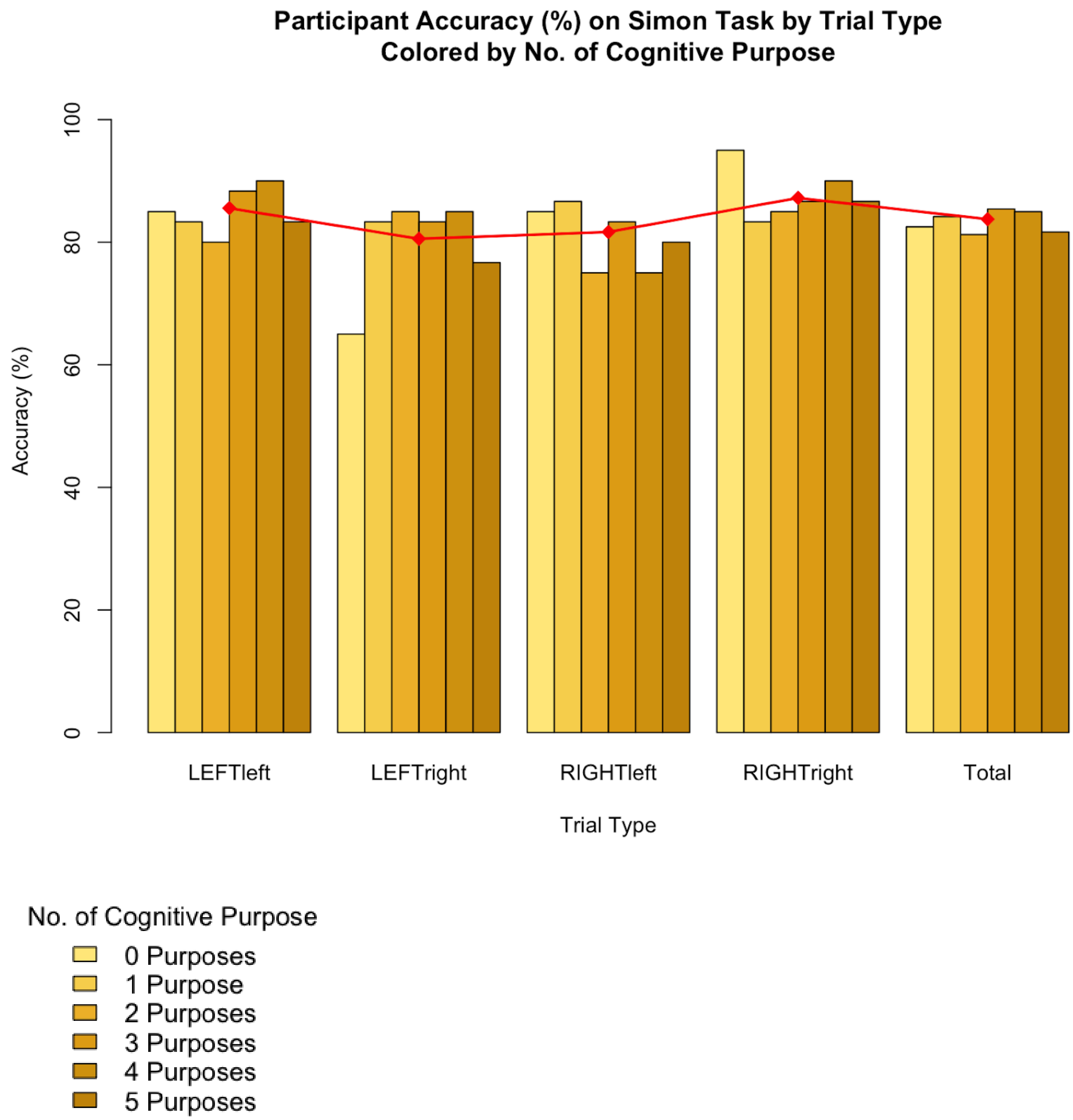

*Figure 37: Graph illustrating Participant Accuracy (%) on Simon Task by Trial Type Colored by No. of Cognitive Purpose. *Red line represents the mean value.*

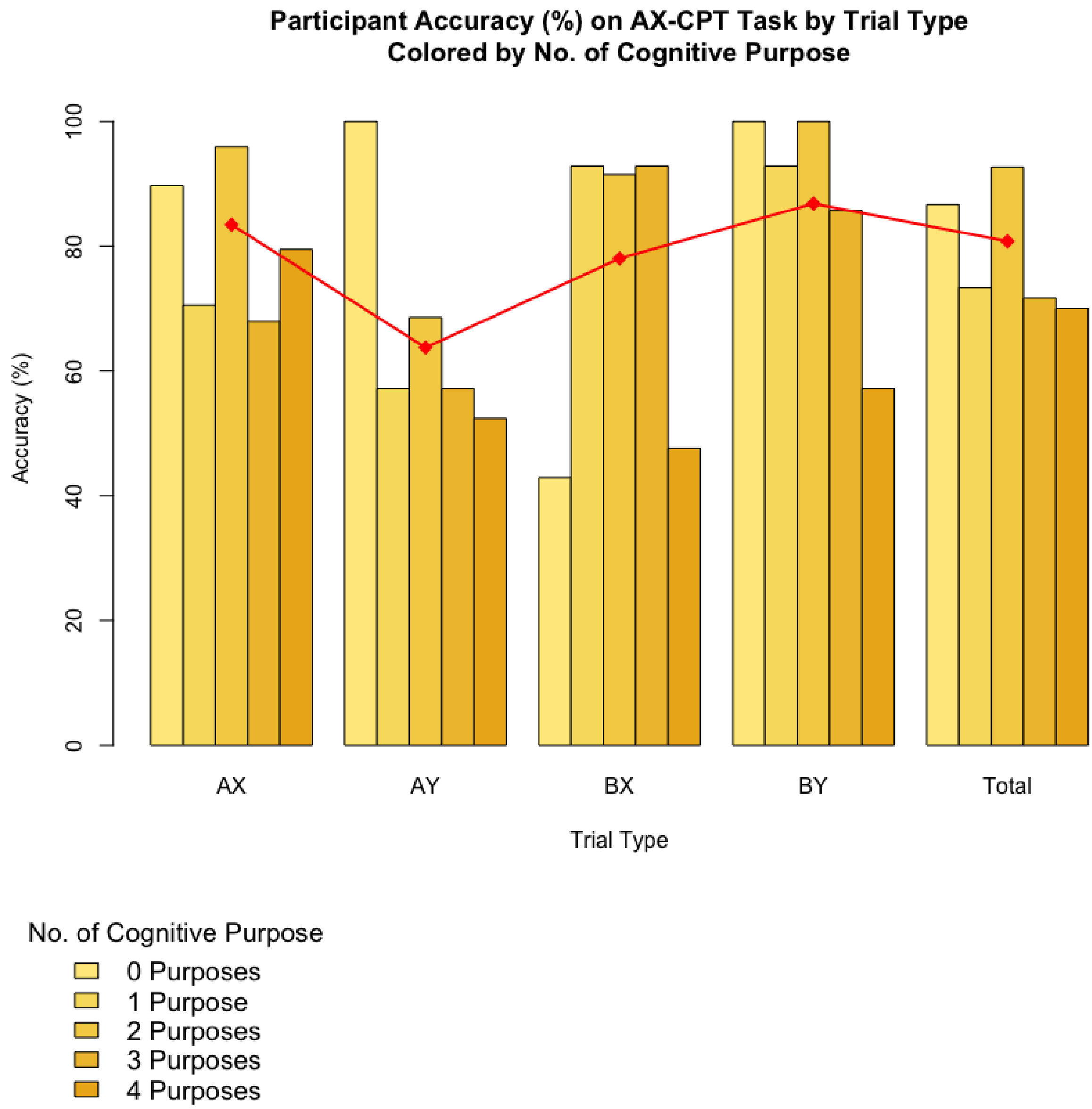

*Figure 38: Graph illustrating Participant Accuracy (%) on AX-CPT Task by Trial Type Colored by No. of Cognitive Purpose. *Red line represents the mean value.*

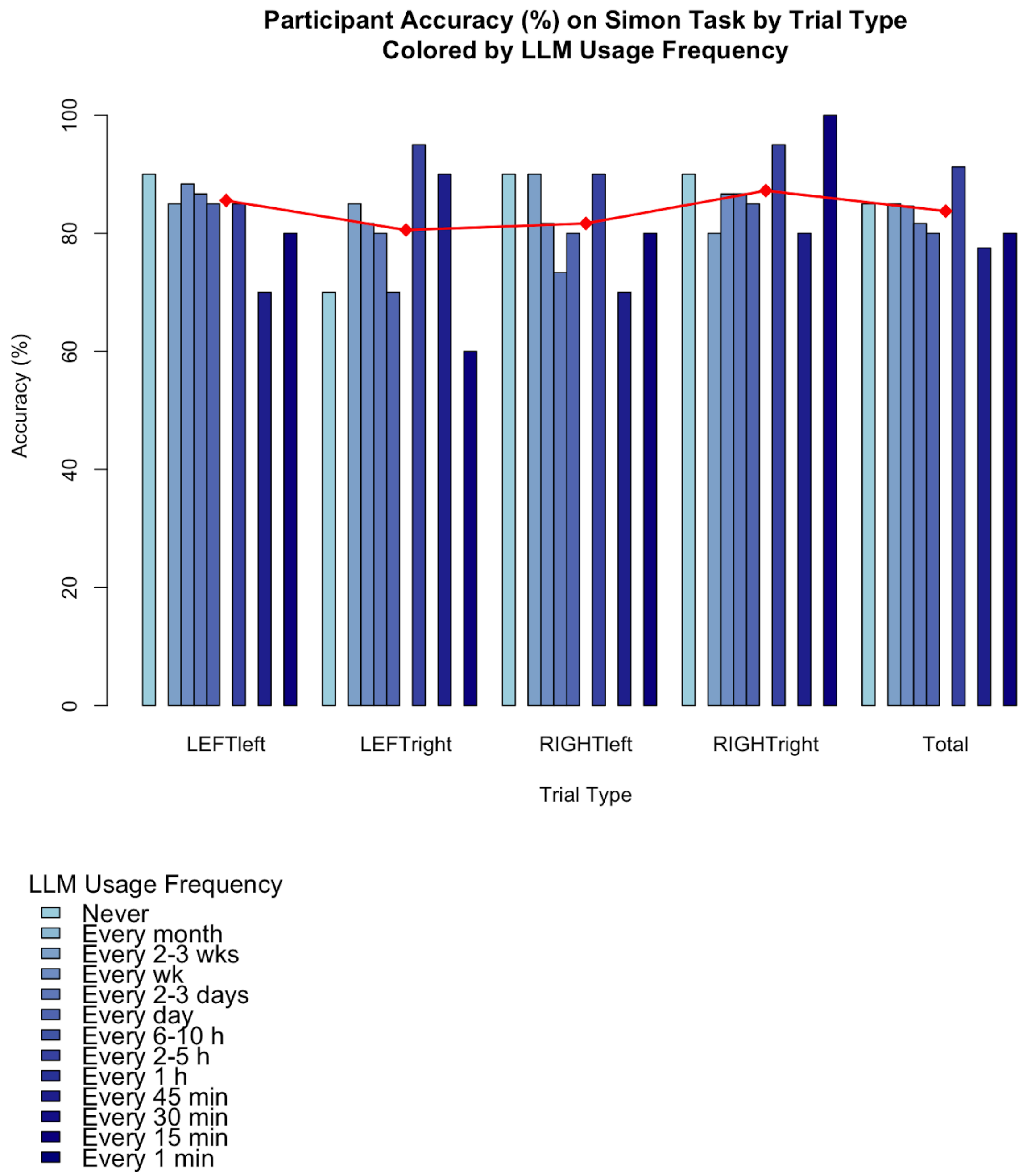

*Figure 39: Graph illustrating Participant Accuracy (%) on Simon Task by Trial Type Colored by LLM Usage Frequency. *Red line represents the mean value.*

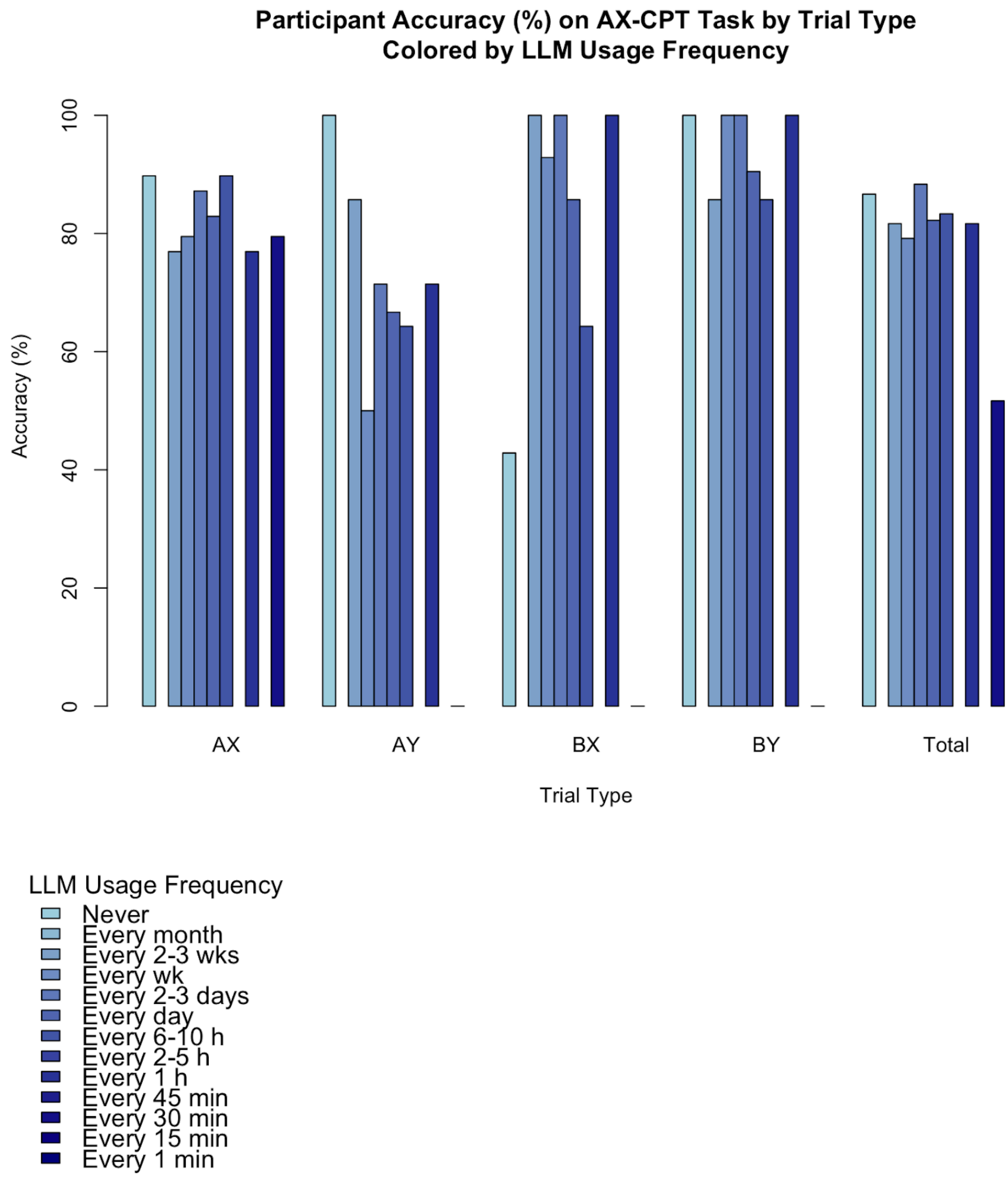

*Figure 40: Graph illustrating Participant Accuracy (%) on AX-CPT Task by Trial Type Colored by LLM Usage Frequency. *Red line represents the mean value.*

3) Large Language Models (LLMs) Use, Language Background, and Performance in Cognitive Tasks (Simon and AX-CPT)

In both the Simon and AX-CPT tasks, bilinguals showed higher accuracy compared to monolinguals, though trilinguals displayed fluctuating accuracy depending on trial type (*Figures 41-42*).

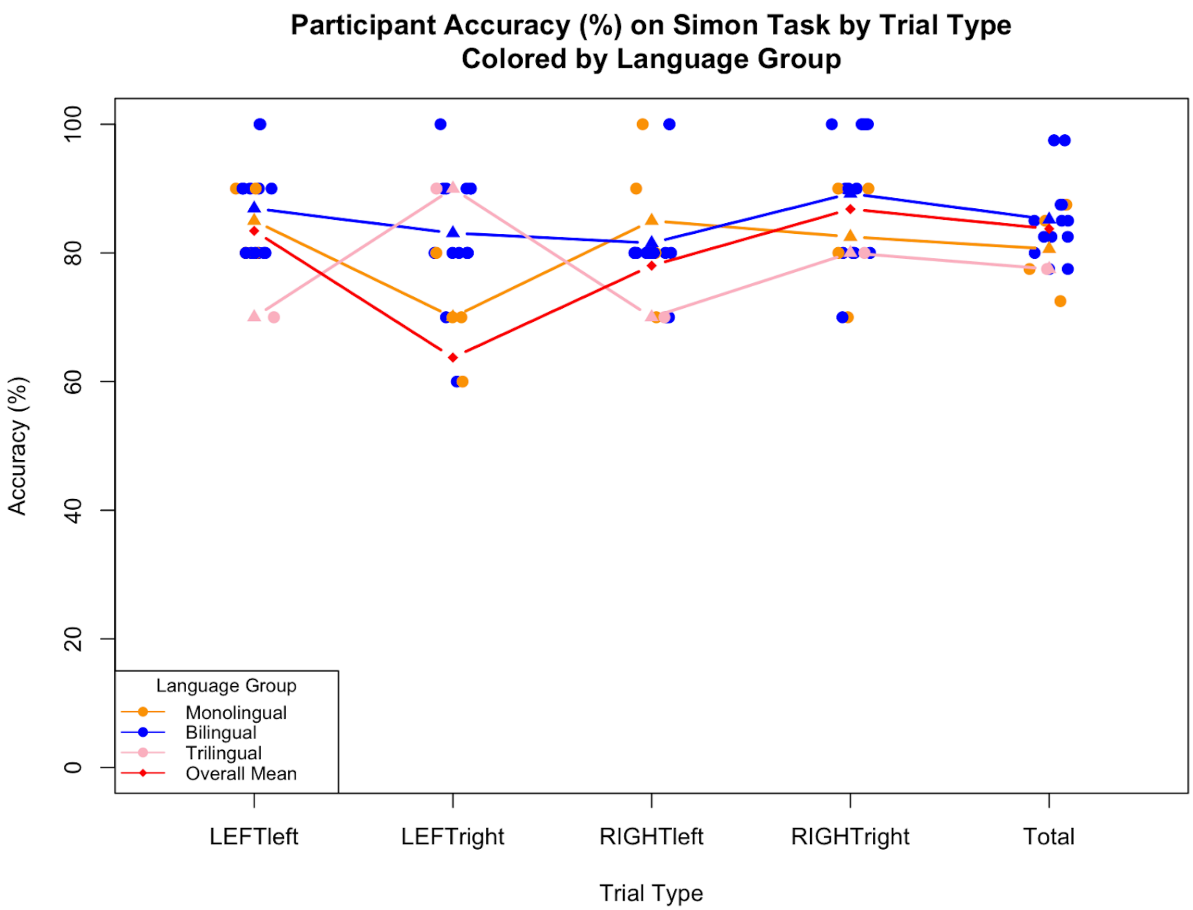

Figure 41: Graph illustrating Participant Accuracy (%) on Simon Task by Trial Type Colored by Language Group.

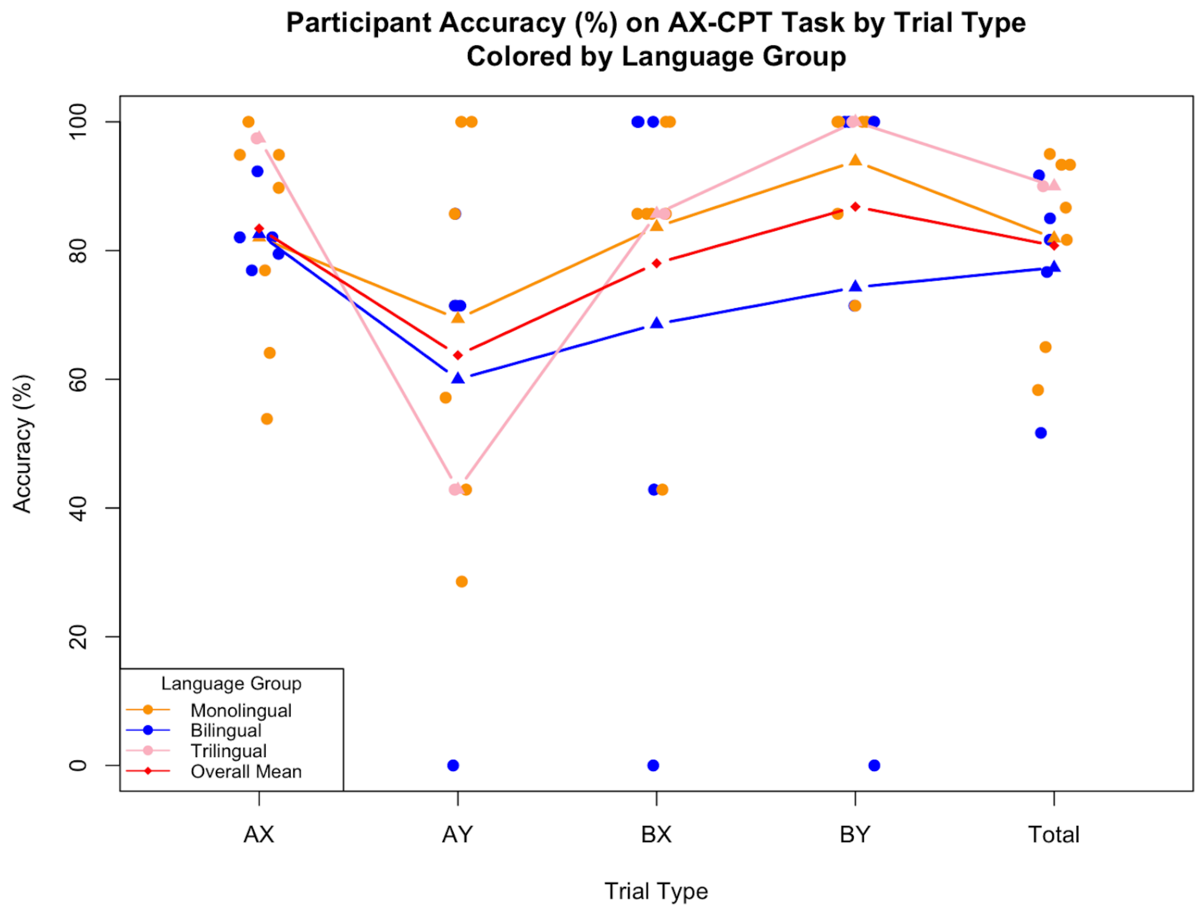

Figure 42: Graph illustrating Participant Accuracy (%) on AX-CPT Task by Trial Type Colored by Language Group.

In the Simon task, bilingual accuracy remained stable regardless of LLM reliance, while monolingual performance varied with LLM reliance (*Figures 43-44*). Bilinguals displayed smaller differences across reliance groups (typically $d < 1$), with a few moderate effects (e.g., 0-25 vs. 26-50 on RIGHTleft, $d = 1.05$). Monolinguals showed consistently large Cohen's d effect sizes across trial types (typically $d > 1$), except RIGHTright ($d = 0.45$).

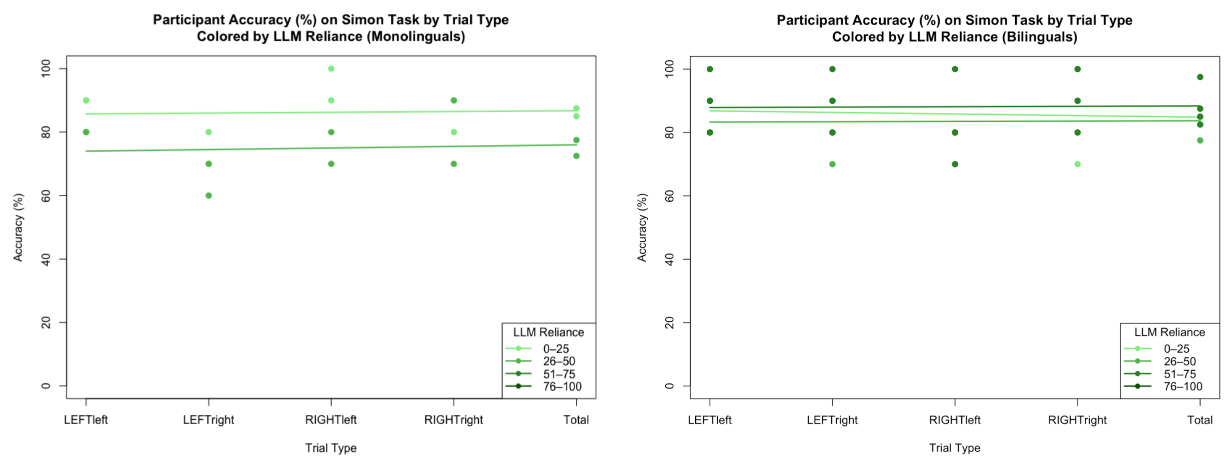

Figure 43: Graph illustrating Participant Accuracy (%) on Simon Task by Trial Type Colored by LLM Reliance for monolinguals and bilinguals.

Trial Type	Cohen's *d* (Monolinguals)	Cohen's *d* (Bilinguals)		
	0-25 vs 26-50	0-25 vs 26-50	0-25 vs 51-75	26-50 vs 51-75
LEFTleft	Inf	0.077	-0.333	-0.592
LEFTright	1.414	0.633	-0.105	-0.618
RIGHTleft	2.828	1.047	0.345	-0.505
RIGHTright	0.447	-0.394	-0.719	-0.505
Total	4.025	0.316	-0.263	-0.894

Figure 44: Table illustrating Cohen's d effect sizes for differences in Participant Accuracy (%) on Simon Task by Trial Type across LLM Reliance for monolinguals and bilinguals.

However, AX-CPT results contradicted Simon results (*Figures 45-46*). Bilinguals showed larger variability across reliance levels, while monolinguals showed smaller variability. For bilinguals, Cohen's *d* were generally large ($d > 1$), except BX and BY trials. For monolinguals, effects were mostly small ($d < 1$), except BX, which showed a stronger effect.

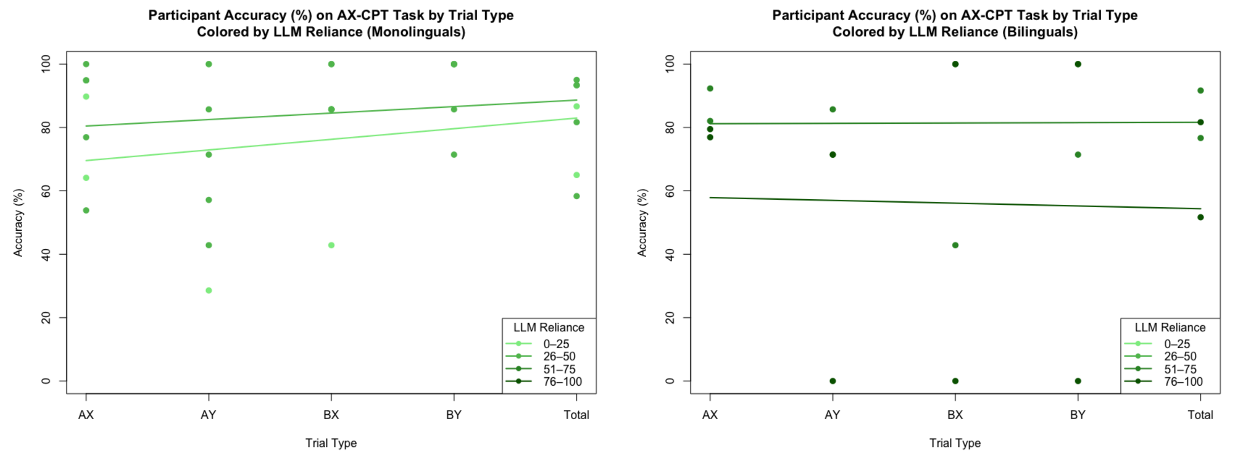

Figure 45: Graph illustrating Participant Accuracy (%) on AX-CPT Task by Trial Type Colored by LLM Reliance for monolinguals and bilinguals.

Trial Type	Cohen's d (Monolinguals)	Cohen's d (Bilinguals)
	0-25 vs 26-50	51-75 vs 76-100
AX	-0.386	1.698
AY	-0.183	1.177
BX	-1.226	0.372
BY	0.949	0.687
Total	-0.552	1.043

Figure 46: Table illustrating Cohen's d effect sizes for differences in Participant Accuracy (%) on AX-CPT Task by Trial Type across LLM Reliance for monolinguals and bilinguals.

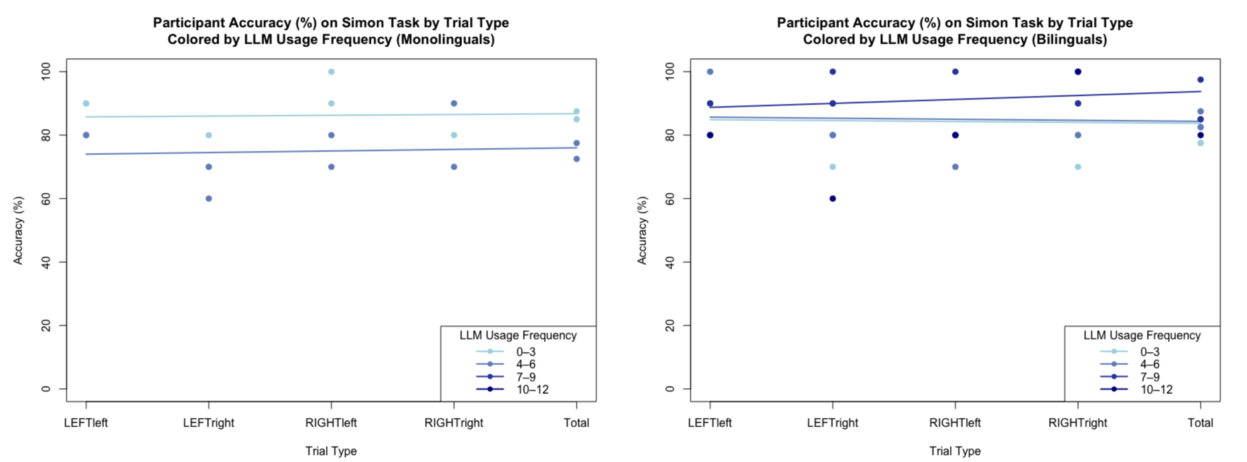

Figure 47: Graph illustrating Participant Accuracy (%) on Simon Task by Trial Type Colored by LLM Usage Frequency for monolinguals and bilinguals.

Trial Type	Cohen's d (Monolinguals)	Cohen's d (Bilinguals)		
	0-3 vs 4-6	0-3 vs 4-6	0-3 vs 7-9	4-6 vs 7-9
LEFTleft	Inf	-0.347	0.286	0.548
LEFTright	1.414	-0.067	-1.621	-1.871
RIGHTleft	2.828	0.573	-0.866	-1.414
RIGHTright	0.447	-0.389	-0.857	-0.548
Total	4.025	-0.117	-0.968	-1.137

Figure 48: Table illustrating Cohen's d effect sizes for differences in Participant Accuracy (%) on Simon Task by Trial Type across LLM Usage Frequency for monolinguals and bilinguals.

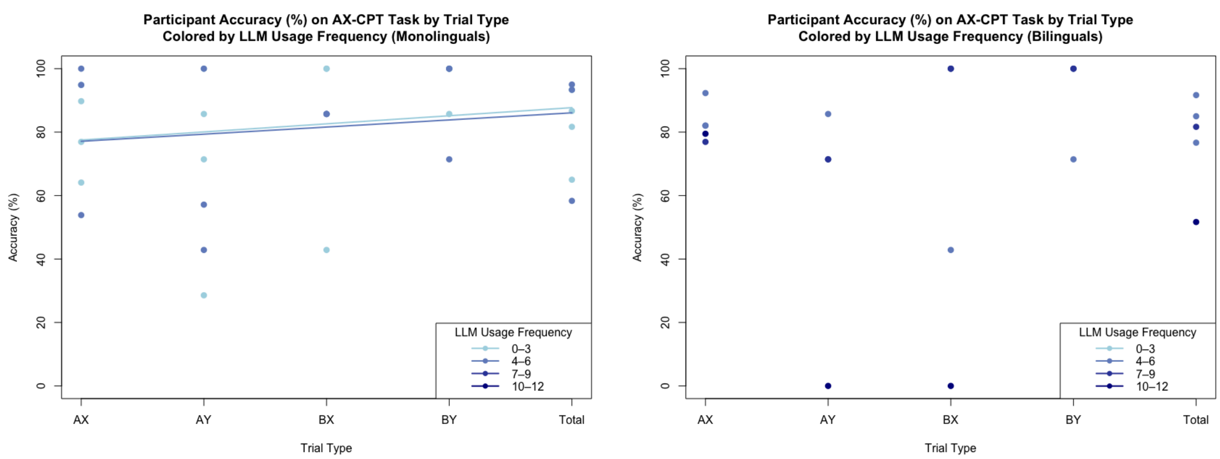

Figure 49: Graph illustrating Participant Accuracy (%) on AX-CPT Task by Trial Type Colored by LLM Usage Frequency for monolinguals and bilinguals.

Trial Type	Cohen's d (Monolinguals)	Cohen's d (Bilinguals)
	0-3 vs 4-6	N/A
LEFTleft	-0.073	N/A
LEFTright	0.157	N/A
RIGHTleft	-0.187	N/A

RIGHTright	0.468	N/A
Total	-0.038	N/A

Figure 50: Table illustrating Cohen's d effect sizes for differences in Participant Accuracy (%) on AX-CPT Task by Trial Type across LLM Usage Frequency for monolinguals and bilinguals.

DISCUSSION

In terms of LLM use by language background, trilinguals reported significantly greater usage of LLMs than bilinguals on both the 0-100 Likert-type scale (*Figure 3*) and the usage frequency measure ranging from *never* (0) to *every 1 minute* (12) (*Figure 4*). Bilinguals, in turn, reported higher usage than monolinguals. Trilinguals and bilinguals also reported using LLMs to support a broader range of cognitive functions than monolinguals (*Figures 5-6*). This trend is further corroborated by the observation that bilinguals who leaned more toward monolingualism, indicated by a greater proficiency disparity between their first and second languages, showed lower levels of LLM use compared to those who leaned more toward bilingualism (*Figures 13-15*).

While these results align with the hypothesis that bilinguals are more active LLMs users — potentially due to the high cognitive demands of managing multiple linguistic systems — the robustness of these findings rests on the adequacy of a 35-participant sample and the uneven distribution of language backgrounds (*Figure 1*). Furthermore, internal validity hinges on self-reported data. Although inconsistencies in self-identification of language background were addressed by reclassifying participants who rated their proficiency at or below 25 on a 0-100 proficiency scale, future studies should improve the categorization of language proficiency like combining self-identification with additional proficiency measures as demonstrated in this study: gathering data on participants' age and context of second and/or third language acquisition, proficiency levels on a 0-100 scale, and patterns of language use — whether integrated, separate, or varied.

Consistent with prior research, this study found that increased LLM use was associated with reduced cognitive performance. Specifically, increased LLM reliance on the grounds of reported reliance on the 0-100 Likert-type scale, usage frequency, and number of cognitive purposes offloaded onto LLMs linked to lower accuracy (*Figures 23, 25, 27, 29, 31, 33*) and faster reaction times, suggesting heightened impatience (*Figures 24, 26, 28, 30, 32, 34*) across both the Simon and AX-CPT tasks.

Evidence from trial-specific performance suggests that LLM usage may disproportionately impair reactive control. The pattern of poorer AY performance relative to BX performance (*Figures 36, 38, 40*) suggests that participants who relied on LLMs tended to exhibit greater impairments in reactive control than in proactive control. One explanation is that most LLMs are fundamentally reactive systems, optimized for responding to input rather than maintaining goals or anticipating future states (Lu et al., 2024). However, recent developments in agentic LLM-based systems indicate a shift toward proactive behavior, in which models anticipate user needs and initiate actions autonomously. As such systems become more common, users may increasingly externalize proactive control as well. If this trend continues, future studies may observe greater deficits on BX trials, reflecting reduced user engagement in goal maintenance and forward planning as these functions are increasingly delegated to proactive AI agents (Pasternak et al., 2025).

The validity of these findings depends on the assumption that the Simon and AX-CPT tasks served as appropriate proxies for measuring cognitive performance. This assumption is largely supported by the data, as there were marginal to significant accuracy and reaction-time differences in the AX-CPT trials, but could also reflect general performance deficits or difficulty following experimental instructions (Ophir et al., 2009). Future research utilizing neurophysiological assessments like EEG, as employed in the MIT study (Kosmyna et al., 2025), would be necessary to confirm if LLM use correlate with neural under-engagement. Alternatively, other tasks used to measure executive function such as the Flanker, the Stroop task (Donnelly et al., 2019), or the Abbreviated Torrance Test for Adults (Kharkhurin, 2008), on which bilinguals have shown advantages could also be explored in future work.

Regarding the hypothesis that bilingualism serves as a potential buffer against cognitive decline associated with LLM use, the results provided a statistically varied picture of cognitive resilience. While bilinguals outperformed monolinguals on both the Simon and AX-CPT tasks (*Figures 41-42*), consistent with prior evidence of a bilingual advantage, the evidence for a protective buffer against LLM-induced decline was task-dependent and inconclusive. On the Simon task, bilingual participants showed relatively stable performance across trial types with minimal variation, regardless of self-reported LLM reliance (*Figures 43-44)*, aligning with expectations of uniform performance. Monolinguals, by contrast, varied more with LLM engagement. However, this pattern did not hold for the AX-CPT task (*Figures 45-46*), and analyses by LLM usage frequency were inconclusive due to limited participant representation (*Figures 47-50*). These inconsistencies highlight the preliminary nature of this framework and the need to account for task-specific variability and sample size constraints when interpreting the bilingual advantage in LLM use.

The study also supported the link between higher educational attainment and stronger critical thinking (Gerlich, 2025), with education levels positively associated with better cognitive performance (*Figure 52*).

CONCLUSION

This study highlighted a clear hierarchy of LLM engagement based on language background, with trilinguals and bilinguals reporting greater LLM use as compared to monolinguals. There was a negative correlation between LLM use and performance on cognitive tasks. Bilinguals generally performed better than monolinguals on these tasks. Yet, the findings did not provide definitive evidence that the cognitive resilience afforded by bilingualism offsets the cognitive risks of LLM use. Hence, bilingualism had no protective effect on the cognitive consequences of LLM usage. However, given that there were several limitations in sample size and validity of self-reported data and cognitive tasks as proxies for LLM use, language background, and executive functioning, further research is required to substantiate whether the executive control developed through bilingualism buffers against LLM-induced cognitive deficits.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my mentor Eloise Freitag for the continuous guidance she has provided me throughout the course of my research.

REFERENCES

Abhang, P. A., Gawali, B. W., & Mehrotra, S. C. (2016). *Introduction to EEG- and Speech-Based Emotion Recognition* (1st ed.). Academic Press.

Abutalebi, J., & Green, D. (2007, May). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics*, *20*(3), 242-275. 10.1016/j.jneuroling.2006.10.003

Abutalebi, J., & Green, D. W. (2016, April 8). Neuroimaging of language control in bilinguals: neural adaptation and reserve. *Bilingualism: Language and Cognition*, *19*(4), 689-698. 10.1017/s1366728916000225

Arain, M., Haque, M., Johal, L., Mathur, P., Nel, W., Rais, A., Sandhu, R., & Sharma, S. (2013). Maturation of the adolescent brain. *Neuropsychiatr Dis Treat*, *9*, 449-461. 10.2147/ndt.s39776

Bawa, A., Khadpe, P., Joshi, P., Bali, K., & Choudhury, M. (2020, May 29). Do Multilingual Users Prefer Chat-bots that Code-mix? Let's Nudge and Find Out! *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW1), 1–23. 10.1145/3392846

Beatty-Martínez, A. L., Navarro-Torres, C. A., Dussias, P. E., Bajo, M. T., Guzzardo Tamargo, R. E., & Kroll, J. F. (2020, June). Interactional context mediates the consequences of bilingualism for language and cognition. *J Exp Psychol Learn Mem Cogn*, *46*(6), 1022-1047. 10.1037/xlm0000770

Bialystok, E. (1992). Attentional control in children's metalinguistic performance and measures of field independence. *Developmental Psychology*, *28*(4), 654-664. 10.1037/0012-1649.28.4.654

Bialystok, E. (2006, March). Effect of bilingualism and computer video game experience on the Simon task. *Canadian Journal of Experimental Psychology*, *60*(1), 68-79. 10.1037/cjep2006008

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004, June). Bilingualism, aging, and cognitive control: evidence from the Simon task. *Psychology and Aging*, *19*(2), 290-303. 10.1037/0882-7974.19.2.290

Boroujeni, K. B., Sigona, M. K., Treuting, R. L., Manuel, T. J., Caskey, C. F., & Womelsdorf, T. (2022, September 6). Anterior cingulate cortex causally supports flexible learning under motivationally challenging and cognitively demanding conditions. *PLoS Biol*, *20*(9), e3001785. 10.1371/journal.pbio.3001785

Braver, T. S. (2012, February). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106-113. 10.1016/j.tics.2011.12.010

Calvo, N., Anderson, J. A. E., Berkes, M., Freedman, M., Craik, F. I. M., & Bialystok, E. (2023, February 23). Gray Matter Volume as Evidence for Cognitive Reserve in Bilinguals With Mild Cognitive Impairment. *Alzheimer Dis Assoc Disord*, *37*(1), 7-12. 10.1097/wad.0000000000000549

Cheng, B., Lin, E., Wunderlich, A., Gramann, K., & Fabrikant, S. I. (2023, February 14). Using spontaneous eye blink-related brain activity to investigate cognitive load during mobile map-assisted navigation (W. Schlee, Ed.). *Frontiers in Neuroscience*, *17*. 10.3389/fnins.2023.1024583

Close encounters of the AI kind. (2025, March 12). Imagining the Digital Future Center. Retrieved July 31, 2025, from https://imaginingthedigitalfuture.org/reports-and-publications/close-encounters-of-the-ai-kind/

Donnelly, S., Brooks, P. J., & Homer, B. D. (2019, February 27). Is there a bilingual advantage on interference-control tasks? A multiverse meta-analysis of global reaction time and interference cost. *Psychonomic Bulletin & Review*, *26*, 112201147. 10.3758/s13423-019-01567-z

D'Souza, D., & D'Souza, H. (2016, May 18). Bilingual Language Control Mechanisms in Anterior Cingulate Cortex and Dorsolateral Prefrontal Cortex: A Developmental Perspective. *Journal of Neuroscience*, *36*(20), 5434-5436. 10.1523/jneurosci.0798-16.2016

Dunn, T. L., & Risko, E. F. (2015, August 26). Toward a Metacognitive Account of Cognitive Offloading. *Cognitive Science*, *40*(5), 1080-1127. 10.1111/cogs.12273

Engel de Abreu, P. M. (2011, July). Working memory in multilingual children: is there a bilingual effect? *Memory*, *19*(5), 529-537. 10.1080/09658211.2011.590504

Gerlich, M. (2025, January 3). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking (L. J. Moore, Ed.). *Societies*, *15*(1), 6. 10.3390/soc15010006

Gonthier, C., Macnamara, B. N., Chow, M., Conway, A. R. A., & Braver, T. S. (2016, November 22). Inducing Proactive Control Shifts in the AX-CPT (A. Vallesi, Ed.). *7*. 10.3389/fpsyg.2016.01822

Goodrich, J. M., & Lonigan, C. J. (2017, August). Language-Independent and Language-Specific Aspects of Early Literacy: An Evaluation of the Common Underlying Proficiency Model. *J Educ Psychol*, *109*(6), 782-793. 10.1037/edu0000179

Gutiérrez-Clellen, V. F., Calderón, J., & Weismer, S. E. (2004, August). Verbal working memory in bilingual children. *Journal of Speech, Language, and Hearing Research*, *47*(4), 863-876. 10.1044/1092-4388(2004/064)

Kearney, M., Binns, R., & Gal, Y. (2025, July 22). Language Models Change Facts Based on the Way You Talk. 10.48550/arXiv.2507.14238

Kharkhurin, A. V. (2008, July 1). The effect of linguistic proficiency, age of second language acquisition, and length of exposure to a new cultural environment on bilinguals' divergent thinking. *Bilingualism: Language and Cognition*, *11*(2), 225-243. 10.1017/S1366728908003398

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025, June). Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task. 10.48550/arXiv.2506.08872

Lu, Y., Yang, S., Qian, C., Chen, G., Luo, Q., Wu, Y., Wang, H., Cong, X., Zhang, Z., Lin, Y., Liu, W., Wang, Y., Liu, Z., Liu, F., & Sun, M. (2024, December 3). Proactive Agent: Shifting LLM Agents from Reactive Responses to Active Assistance. 10.48550/arXiv.2410.12361

Namazi, M., & Thordardottir, E. (2010, August 16). A working memory, not bilingual advantage, in controlled attention. *International Journal of Bilingual Education and Bilingualism*, *13*(5), 597-616. 10.1080/13670050.2010.488288

Nicoladis, E. (2007, June 11). The effect of bilingualism on the use of manual gestures. *Applied Psycholinguistics*, *28*(3), 441-454. 10.1017/s0142716407070245

Ophir, E., Nass, C., & Wagner, A. D. (2009, September 15). Cognitive control in media multitaskers (M. I. Posner, Ed.). *PNAS*, *106*(37), 15583-15587. 10.1073/pnas.0903620106

Pasternak, G., Rajagopal, D., White, J., Atreja, D., Thomas, M., Hurn-Maloney, G., & Lewis, A. (2025, October 29). Beyond Reactivity: Measuring Proactive Problem Solving in LLM Agents. 10.48550/arXiv.2510.19771

Peláez-Sánchez, I. C., Velarde-Camaqui, D., & Glasserman-Morales, L. D. (2024, June 14). The impact of large language models on higher education: exploring the connection between AI and Education 4.0 (D. H. Montoya, Ed.). *Frontiers in Education*, *9*. 10.3389/feduc.2024.1392091

Polizzotto, N. R., Hill-Jarrett, T., Walker, C., & Cho, R. Y. (2018, May 31). Normal development of context processing using the AXCPT paradigm. *PLoS ONE*, *13*(5), e0197812. 10.1371/journal.pone.0197812

Ransdell, S., Barbier, M.-L., & Niit, T. (2008, December 22). Metacognitions about Language Skill and Working Memory among Monolingual and Bilingual College Students: When Does Multilingualism Matter? *International Journal of Bilingual Education and Bilingualism*, *9*(6), 728-741. 10.2167/beb390.0

Schroeder, S. R., & Marian, V. (2016, April 1). Cognitive consequences of trilingualism. *SageJournals*, *21*(6). 10.1177/1367006916637288

Shanmugasundaram, M., & Tamilarasu, A. (2023, November 24). The impact of digital technology, social media, and artificial intelligence on cognitive functions: a review (J. Poh, Ed.). *Frontiers in Cognition*, *2*. 10.3389/fcogn.2023.1203077

Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, *51*(3), 300–304. 10.1037/h0020586

Sparrow, B., Liu, J., & Wegner, D. M. (2011, July 14). Google Effects on Memory: Cognitive Consequences of Having Information at Our Fingertips. *Science*, *333*(6043), 776-778. 10.1126/science.1207745

Stadler, M., Bannert, M., & Sailer, M. (2024, 11). Cognitive ease at a cost: LLMS reduce mental effort but compromise depth in student scientific inquiry (N. Nistor, Ed.). *Computers in Human Behavior*, *160*, 108386. 10.1016/j.chb.2024.108386

Sweller, J. (2024, February 21). Cognitive load theory and individual differences. *Learning and Individual Differences*, *110*, 102423. 10.1016/j.lindif.2024.102423

Vorstman, E. L. P., De Swart, H., Ceginskas, V., & Bergh, H. V. D. (2009, July 24). Language learning experience in school context and metacognitive awareness of multilingual children. *International Journal of Multilingualism*, *6*(3), 258-280. 10.1080/14790710902878692

Yang, E. (2017, July 18). Bilinguals' Working Memory (WM) Advantage and Their Dual Language Practices (S. Kroener, Ed.). *Brain Sciences*, *7*(7), 86. 10.3390/brainsci7070086