Can you accurately monitor your behaviors while multitasking? The effect of multitasking on metacognition

Ruiqiao Guo[a] Yan Liu[a*] Hui Jing Lu[b] Annan Jing[a]

[a]*School of Psychology, Liaoning Normal University, 116029, Dalian, P.R. China*

[b]*Department of Applied Social Sciences, The Hong Kong Polytechnic University, Hong Kong, P.R. China*

[*]Corresponding author:

Yan Liu, PhD, School of Psychology, Liaoning Normal University, Dalian 116029, P.R. China

E-mail: psymaria@pku.edu.cn

Can you accurately monitor your behaviors while multitasking? The effect of multitasking on metacognition

ABSTRACT

With the pace of life accelerating, multitasking has become the norm in daily life. According to research, multiple cognitive processes, including numerical reasoning, comprehension, and writing, are negatively affected by multitasking. However, only a few studies have investigated the relationship between multitasking and metacognition. In this study, the effect of multitasking on metacognition was examined using a prospective monitoring paradigm (prediction of subsequent recall performance). In Experiment 1, the participants simultaneously studied word pairs (primary task) and differentiated between different sound pitches (secondary task) and then predicted their performance in a subsequent memory test for the studied word pairs (prospective metacognitive monitoring). The accuracy of metacognitive evaluation with multitasking was then compared with that without multitasking. In Experiment 2, sounds and icons of real-life applications were used to improve the ecological validity of the experiment in the secondary task. The results indicated that multitasking impaired metacognition in both artificial and real-life simulated scenarios. In addition, the participants who engaged in more media multitasking in their daily lives exhibited poorer metacognitive monitoring abilities in single tasks.

Keywords: Multitasking, Metacognitive accuracy, Relatedness, Ecological validity

1. Introduction

With the pace of modern life accelerating, people often resort to multitasking, a cognitive process that involves the simultaneous or sequential performance of two or more tasks, including dual tasks and task switching (Koch, 2018). With the advent of the internet and the proliferation of mobile media, multitasking has become increasingly prevalent. Media multitasking refers to engaging in two or more tasks simultaneously or switching between different tasks, with at least one of these tasks involving media (Lopez et al., 2018; Murphy et al., 2017). Examples of media multitasking are listening to music while studying, texting on mobile phone while watching TV, and answering a phone call while driving.

Despite extensive research highlighting the detrimental effects of multitasking on various cognitive processes, such as numerical reasoning (Adler & Benbunan-Fich, 2012, 2015), comprehension (Jeong & Hwang, 2012), writing (Lottridge et al., 2015), and memory (Dindar & Akbulut, 2016; Jamet et al., 2020; Rubenking, 2017), the effect of multitasking on metacognition remains unclear. Metacognition refers to the ability of an individual to represent, monitor, and control ongoing cognitive processes (Heyes et al., 2020). It is a second-order process of first-order representations, such as perception, memory, and other cognitive domains (Lehmann, Hangen, & Ettinger, 2022). Metacognition plays a vital role in behavior regulation (Aguilar-Lleyda et al., 2020), learning (Hainguerlot et al., 2018), decision-making (van den Berg et al., 2016), and problem-solving (Anggo & Arapu, 2018). Poor metacognitive abilities are associated with various addictive behaviors (Spada et al., 2015), such as gambling (Jauregui et al., 2016), alcohol use (Spada et al., 2009), nicotine use (Nikčević & Spada, 2008), and problematic technology use (Akbari, 2017).

According to the metacognition model developed by Nelson and Narens (1990), cognitive processes consist of two interdependent levels: an object level and a meta level. The object level involves the processing of various stimuli, such as images and sounds. The meta level involves metacognitive monitoring and control, which allow people to monitor and regulate their cognitive processes at the object level. Both levels require cognitive resources (Kanfer & Ackerman, 1989; Winne & Hadwin, 1998). During multitasking, both object- and meta-level processes share the same cognitive resources (Brasel & Gips, 2011). Typically, a person's cognitive resources are limited at any given time (Lang, 2000). However, compared with single-task processing, multitasking requires more cognitive resources at the object level (Aagaard, 2018; Wiradhany et al., 2021). Therefore, compared with single-task processing, during multitasking, fewer cognitive resources are available at the meta level. We hypothesize that multitasking would impair meta-level processing.

As a core element of metacognition, metacognitive monitoring refers to judgments or evaluations of the object level through the meta level (Nelson & Narens, 1990). In laboratory studies, metacognitive monitoring is typically measured using the participants' rating of how confident they feel regarding accurately completing a task, either before (prospective monitoring) or after (retrospective monitoring) task completion (Lehmann, Hangen, & Ettinger, 2022). In some studies, participants have been asked to make aggregate-item judgments (i.e., global recall estimate based on all items on a list; Hanczakowski et al., 2018). In other studies, participants were asked to make item-by-item judgments (i.e., evaluation of items one by one; Konishi et al., 2021; Konishi et al., 2020; Peng & Tullis, 2021).

In previous studies, researchers have typically used aggregate-item judgments before or after multitasking to explore the effects of multitasking on metacognition. For example, in a study by Finley et al. (2014), the participants were asked to complete a single task first and then a dual task (within-subject design). However, before the dual task, in which the participants simultaneously completed a tracking task and an *n*-back task (which was manipulated at different difficulty levels), the participants were asked to predict their performance on the tracking task. According to the results, as the number of *N* in the *n*-back tasks increased, the participants were able to lower their prospective monitoring judgments of tracking performance, suggesting that they were able to anticipate the costs of multitasking. Hanczakowski et al. (2018) also reported that their participants were able to correctly predict their own performance on a memory task. In this study, the participants were asked to first study 15 words and then recall these words. Between the study phase and the recall phase, the participants were asked to make a prospective monitoring judgment of their recall performance. They were asked to study three 15-word lists under three conditions (within-subject design): (a) no auditory distraction, (b) semantically related auditory distraction, and (c) semantically unrelated auditory distraction. After each study phase, the participants were asked to make an aggregate-item judgment. According to the results, the participants were able to correctly predict whether auditory interference would harm their performance.

According to the aforementioned studies, despite engaging in multitasking, people can still correctly perform metacognitive monitoring when the monitoring judgment is prospective and involves a one-off aggregate response. However, when the metacognitive judgment is retrospective, the research results become mixed. Ralph et al. (2018) conducted a study with a within-subject design, in which the participants were asked to complete a 2-back task under two conditions: a video condition, with a video clip played in the background as a distraction, and a no-video condition, with no distraction. Following task completion, the participants were asked to estimate their accuracy rate. The purpose of this study was to examine the extent to which the participants were aware of their own performance during media multitasking. According to the results, the participants in the video group were able to accurately evaluate their ongoing task performance, indicating a comparable level of metacognitive awareness to that of the participants in the no-video group. These results suggested that the participants were capable of recognizing the effect of media multitasking on their performance. In these aforementioned studies, although the participants were typically aware of the negative effects of multitasking when making aggregate-item judgments in studies with within-subject designs (Finley et al., 2014; Hanczakowski et al., 2018; Ralph et al., 2018), a reality simulation study with a between-subject design yielded a different outcome. For example, in a study by Sanbonmatsu et al. (2016), the participants were asked to play a virtual driving game while talking (cell-phone group) and not talking (control group) on a hands-free cell phone (between-subject design). After completing the driving game, the participants were asked to tabulate their driving errors on a checklist and evaluate their own driving safety levels. According to the results, the cell-phone group made more serious driving errors and reported less safe driving behavior compared with the control group. However, the cell-phone group did not recall making more errors than those made by the control group. These results indicated that talking on a cell phone while driving lowers people's awareness and driving safety and reduces the accuracy of judgment on one's own driving behavior, thereby increasing the risk of accidents.

Of the four studies summarized earlier, the three studies that adopted a within-subject design revealed that people are typically aware of the adverse effects of multitasking, suggesting that multitasking does not affect people's metacognitive judgments of their own task performance. However, the study that adopted a between-subject design revealed that multitasking reduced the metacognitive judgment accuracy of individuals. These contradictory results may be due to the research design. In other words, in the between-subject design, each participant was allowed to complete a metacognitive judgment task under only one condition, whereas in the within-subject design, each participant was allowed to complete a task under different experimental conditions and thus adjusted their own metacognitive judgments accordingly.

Whether metacognitive judgments are made in an aggregated format or in an item-by-item format may affect people's metacognitive judgments while multitasking. When people are asked to provide an overall assessment before or after multitasking, they become more likely to pay attention to the core differences between the multitasking and nonmultitasking conditions and thus adjust their metacognitive judgments accordingly. Global judgment tends to be affected by motivational and personality variables (Händel et al., 2020). If people are required to provide item-by-item metacognitive judgments during an ongoing primary task without having the chance to thoroughly compare different conditions, their metacognition may become less accurate. However, compared with aggregate-item judgments, item-by-item judgments should more precisely indicate people's metacognitive states during task execution.

To our knowledge, only three studies have examined the effects of multitasking on metacognition through item-by-item judgments (Konishi et al., 2021; Konishi et al., 2020; Peng & Tullis, 2021). Konishi et al. (2020) examined the effects of multitasking on metacognition by using a dual-task paradigm. They used confidence judgments as an indicator of metacognitive monitoring. In their study, the participants were asked to complete a single-color block test, a single-motion block test, and a dual-task block test. For the single-color block, the participants were asked to compare the number of red and blue dots in a circle. For the single-motion block, the participants were asked to judge whether a white dot was moving up or down. For the dual-task block, the participants were asked to simultaneously judge both color and motion. After each trial, the participants were asked to judge their confidence regarding the accuracy of their answers on a vertical scale ranging from 100 (*I am sure I made the correct decision*) to 50 (*I responded randomly*). The results indicated that the confidence judgments made under the dual-task condition were not less accurate than those made under the single-task condition, partly because the dual task was not excessively difficult, leaving relatively adequate resources for the metacognitive level processing. Therefore, in a later study, Konishi et al. (2021) designed a multimodal triple-task paradigm including a sensorimotor tracking task, an auditory 2-back task, and a visual discrimination task. In this study, the participants were asked to complete these tasks simultaneously or separately and then provide confidence judgments regarding their performance at the end of each trial. The results indicated that the participants' confidence judgments were the least accurate under the triple-task condition. To our knowledge, this study was the first to report that multitasking negatively affects metacognition.

However, in a study using judgment of learning (JOL) as an indicator, different findings were reported (Peng & Tullis, 2021). JOL is a type of prospective judgment of the memory performance of studied items. In this study, the participants were asked to learn word pairs and evaluate their own learning outcomes by indicating their level of confidence (from 0 to 100) in recalling each word pair during a subsequent memory test. Half of the participants were asked to engage in multitasking by simultaneously studying a word pair and responding to an auditory stimulus, whereas the other half were asked to focus on only studying a word pair. The results indicated that the JOL accuracy under the multitasking condition was similar to that under the single-task condition. As a prospective monitoring indicator (i.e., prediction of performance in a subsequent recall task), JOL was assumed to be cognitively more complex than retrospective monitoring (i.e., confidence judgment of performance in a completed task; Liu et al., 2018). Therefore, compared with confidence judgments influenced by multitasking (Konishi et al., 2021), JOLs should require more cognitive resources and be more negatively affected by multitasking. However, the findings of Peng and Tullis (2021) do not support this hypothesis. One possible explanation is that the secondary task (pitch discrimination) in the study of Peng and Tullis (2021) was too simple. In other words, the participants were not required to allocate substantial cognitive resources to the secondary task. Thus, adequate resources remained for metacognitive evaluation.

In this study, by manipulating the difficulty of multitasking, we examined the effects of multitasking on metacognitive monitoring by using item-by-item JOLs. We conducted two experiments: Experiment 1 and Experiment 2. In Experiment 1, we asked the participants to study word pairs with and without multitasking involving the identification of the pitch of an auditory stimulus and to perform item-by-item JOL for each word pairs. At last, we asked them to take a memory test for these word pairs following a filler task (i.e., a minus-three task). In general, according to the 'PDR' model (describing the relation between performance (P), objective task difficulty (D), and resources (R)), if the resources invested remain constant, then performance is inversely proportional to objective task difficulty (Wickens, 2020). In this study, we manipulated the difficulty of the pitch discrimination task (single and double identification) while the participants were studying the word pairs. We hypothesized that the JOLs under the multitasking condition would be less accurate than those under the single-task condition.

In Experiment 1, although both the primary and the secondary tasks were executed on a computer, which served as a medium, neither task entailed a typical media activity. According to a review by Kihlstrom (2021), if the study design considerably deviates from the natural environment, the results become less likely to be generalizable to real-life scenarios. Therefore, to increase the ecological validity of this study, we designed Experiment 2 with two modifications. First, to simulate real-life scenarios in which people respond to instant messaging notifications while executing a task (multitasking condition), we used actual notification sounds from instant messaging applications such as QQ and WeChat rather than artificial sounds. Second, unlike in Experiment 1, in which the participants responded to sound stimuli by pressing a key on a keyboard, in Experiment 2, they clicked on the application icon to respond, mimicking the real-life scenario in which people click or tap on instant messaging icons. In accordance with Kihlstrom's (2021) recommendations regarding the alignment of the study design with the natural environment, the goal of Experiment 2 was to enhance the ecological validity of the study and closely simulate media multitasking in order to determine whether multitasking affects metacognitive accuracy.

In this study, we examined the potential link between daily media multitasking and metacognitive accuracy. Multiple studies have investigated the relationships between media multitasking and other aspects of cognitive functioning, such as learning (May & Elder, 2018), attention (Loh et al., 2016), and memory (Uncapher et al., 2016). Notably, Ophir et al. (2009) discovered that individuals who engage in heavy media multitasking may have difficulty inhibiting irrelevant information during a stop-signal task, presumably because of increased mind wandering compared with those who engage in less media multitasking (Kong et al., 2023; Parry & Le Roux, 2021; Wiradhany & Nieuwenstein, 2017). Because metacognitive processing requires active monitoring and regulation, mind wandering may hinder the effectiveness of metacognition. Therefore, in this study, we investigated the relationship between media multitasking behavior and metacognition. We used a media use questionnaire developed by Ophir et al. (2009) to both evaluate the media multitasking behaviors of people in their daily lives and examine the relationship between media multitasking and metacognition.

2. Experiment 1

2.1 Method

2.1.1 Participants

Calculation in G*power revealed that a minimum sample size of 84 was required to obtain a medium effect size of 0.25 with a power of 0.80 at an α level of 0.05 in the within-between repeated measures of 3 × 3 analysis of variance (ANOVA). A total of 115 participants (29 male, $M_{\text{age}} = 21.56$, $SD = 2.53$) were recruited and randomly assigned into three groups: 38 in a single-task group, 38 in an easy multitasking group, and 39 in a difficult multitasking group. All participants received monetary compensation.

2.1.2 Materials

Word Pairs for the Primary Task. A total of 46 Chinese word pairs were used: 10 practice word pairs, 12 strongly related word pairs (e.g., church-priest), 12 somewhat related word pairs (e.g., glass-metal), and 12 unrelated word pairs (e.g., fat-currency). These words have been used in other studies (Chen & Fu, 2010; Nelson et al., 2004) and have been rated in terms of relatedness by an independent group of raters on a 7-point Likert-type scale with endpoints ranging from 0 (*totally unrelated*) to 6 (*strongly related*). The rating scores ranged between 4.58 and 5.96 (M = 5.36, SD = 0.27) for the 12 strongly related word pairs, between 2.11 and 3.80 (M = 3.28, SD = 0.608) for the somewhat related word pairs, and between 0.50 and 1.88 (M = 1.33, SD = 0.35) for the unrelated word pairs.

Acoustic Stimuli for the Secondary Task. Three types of acoustic stimuli that sounding similar to "doo" were used at different frequencies (i.e., 200, 700, and 1600 Hz) for 200 ms. After the volume was adjusted to 50% on the computer, the participants were asked to press keys 1, 2, and 3 in response to sounds with frequencies of 200, 700, and 1600 Hz, respectively.

Media Use Questionnaire. The media use questionnaire was developed by Ophir et al. (2009) to assess the extent to which an individual engages in daily media multitasking. The Chinese version of the questionnaire (Yang & Zhu, 2014) contains 10 different media types, namely print media, television, music, video or computer games, instant messaging, text messaging, face-to-face conversation, texting site, video chat, and homework. In this study, the participants were asked to report the total number of hours that they spent every day on each medium and whether they concurrently used each of the other media ("most of the time," "some of the time," "rarely," or "never") while engaging with the primary medium. The media multitasking index (MMI) was then calculated by summing the number of secondary media used while engaging with the primary medium, weighted by the percentage of time spent on each primary medium. In general, the MMI is used to indicate the level of media multitasking that an individual engages in during a typical media consumption hour.

2.1.3 Procedure

In this study, a 3 (multitasking: null vs. easy vs. difficult) × 3 (word pair type: strongly related vs. somewhat related vs. unrelated) mixed design was adopted, with multitasking serving as a between-subject variable and word pair type serving as a within-subject variable. The study phase included 36 trials presented in random order: 12 trials with strongly related word pairs, 12 trials with somewhat related word pairs, and 12 trials with unrelated word pairs.

Before the formal experiment was conducted, the participants were asked to execute certain practice tasks. For instance, the participants in the null group underwent 10 trials of word-pair studying and corresponding JOLs to familiarize themselves with the word-pair phase. The multitasking group underwent two practice sessions. In the first session, the participants were asked to practice responding to a secondary task for multitasking purposes. They were asked to listen to a sound with a frequency of 200, 700, or 1600 Hz for 200 ms and to respond to this sound as quickly as possible by pressing the corresponding key on a computer keyboard. They were asked to practice this task 30 times. The results indicated that the accuracy rates of all participants exceeded 75%. In the second session, the participants were asked to undergo the same word-pair presentation as the null group and then complete 10 trials of word-pair tasks combined with the secondary task.

As shown in Fig. 1, the formal experiment comprised three distinct phases: a learning and judgment phase, an interference phase, and a test phase.

Learning and Judgment Phase. During each trial, the participants were initially presented with a 400-ms fixation cross, followed by a cue-target word pair displayed for 2 s and then a screen displaying six numbers (0, 20, 40, 60, 80, and 100). They were then instructed to click on one of these numbers as quickly as possible to indicate their level of confidence in recalling the target word when presented with the cue word alone. The number 0 indicated the lowest confidence level in subsequent retrieval, whereas the number 100 indicated the highest confidence level. The numbers selected by the participants ultimately represented their JOL values for the preceding word pair. For example, following the presentation of a word pair such as "glass-metal," the JOL screen displayed six numbers. Clicking on the number 80 indicated that the participant was able to recall the pair with 80% confidence. After the numbers were responded to, the subsequent trial was started.

Each participant completed a task under only one experimental condition. Under the single-task condition (null multitasking group), the participants studied word pairs without needing to multitask for pitch discrimination. For the easy multitasking group, one sound was presented in the middle of the 2-s period of the word-pair presentation. The participants were asked to study the word pair while responding to a sound with a frequency of 200, 700, or 1600 Hz. For the difficult multitasking group, two sounds were presented in the first 200 ms and from 1001 to 1200 ms during the 2-s word-pair presentation. The participants were asked to respond twice while studying the word pair.

Interference Phase. After 36 trails were completed, a 2-min filler task was executed, in which the participants were asked to count backward from 777 in intervals of three (i.e., 777, 774, 771, 768, etc.).

Test Phase. During this phase, the participants were asked to recall 36 word pairs. For each pair, the cue word (the word on the left of the word pair) was presented, and the participants were asked to input the target word (the word on the right of the word pair) by using ta keyboard. If they were unable to recall the target word, they were asked to leave it empty and to press the "Enter" key to view the next cue word.

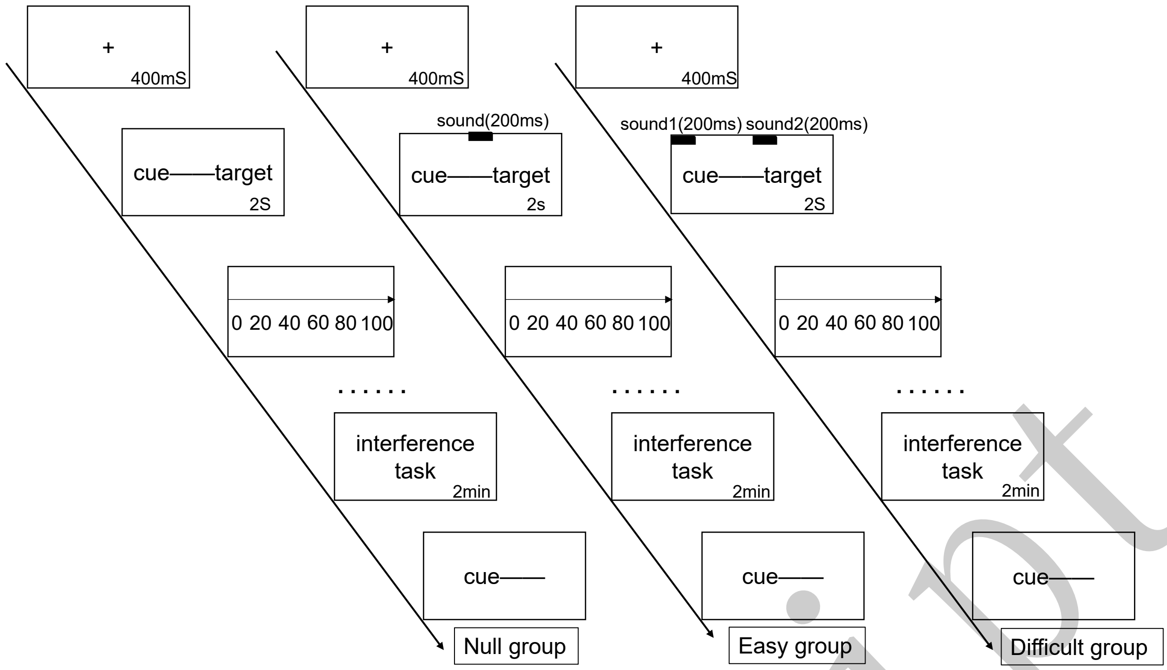

Fig. 1. Visualization of the procedure of Experiment 1. Each trial consisted of a study phase, including a 400-ms fixation cross, followed by a 2-s word-pair presentation and a self-paced JOL task. After the 36-trial phase was completed, a 2-min filler task was administered, and a recall test was conducted. The black bold bar in the easy multitasking condition represents a sound lasting 200 ms in the middle of the 2-s word-pair presentation. The two black bold bars in the difficult multitasking condition represent two sounds during the process of word-pair presentation. Both the first sound and the word pair were simultaneously presented, and the second sound started from 1001 ms during the 2-s word-pair presentation.

2.1.4 Data analysis

JOL accuracy is represented by two indicators: calibration and resolution (Kelemen, 2000). Calibration is a measure of absolute accuracy, and it indicates the extent to which a JOL is close to the actual recall performance. It is calculated by subtracting the percentage of recall performance from the percentage of average JOL magnitude. The closer the calibration value is to zero, the greater is the consistency between the JOL magnitude and the actual memory performance, and thus, the higher is the accuracy of the JOL. Resolution is a measure of relative accuracy, and it indicates how the JOL magnitude differs between recalled and nonrecalled items. It is calculated using the Goodman–Kruskal gamma correlation between the participants' JOL magnitude and their actual recall for each item. The higher the resolution value is, the greater the participants' ability is to assign a higher JOL magnitude to recalled items and a lower JOL magnitude to nonrecalled items. Additionally, following Undorf & Bröder (2019), we assessed participants' use of word pair relatedness in JOLs. We calculated the difference between JOL magnitude for strongly related/unrelated pairs and conducted significance tests. If the difference was >0 and |d|⩾0.2, participants were deemed to utilize word pair relatedness in their JOLs.

2.2 Results and discussion

Table 1 presents the descriptive statistics. We performed 3 (multitasking, between-subject design: null vs. easy vs. difficult) × 3 (word pair type, within-subject design: strongly related vs. somewhat related vs. unrelated) mixed-model ANOVAs for recall performance and JOL magnitude and a one-way ANOVA for calibration and resolution.

2.2.1 Performance of the secondary task

The one-sample *t*-test results indicated that the accuracy rates of the easy task involving multitasking [M = 0.90, SD = 0.21, $t(37)$ = 16.59, p = .001, Cohen's d = 5.53] and the difficult task involving multitasking [M = 0.62, SD = 0.15, $t_1(38)$ = 12.22, p = .001, Cohen's d = 3.91 and M = 0.70, SD = 0.16, $t_2(38)$ = 14.47, p = .001, Cohen's d = 4.63 for responses to the first and second sounds, respectively] were significantly higher than the random level (33.33%).

Table 1 Mean and Standard Deviation of recall, calibration, resolution, and JOL magnitude in Experiment 1

Multitask	Word Pair Type	Recall	JOL magnitude	Accuracy of JOL	
				Calibration	Resolution
		M ± *SD*	*M* ± *SD*	*M* ± *SD*	*M* ± *SD*
Null	Strongly related	.40 ± .25	62.20 ± 11.84		
	Somewhat related	.32 ± .21	51.36 ± 13.60	.18 ± .16	.69 ± .26
	Unrelated	.18 ± .15	29.78 ± 16.40		
Easy	Strongly related	.32 ± .17	59.69 ± 18.47		
	Somewhat related	.23 ± .15	42.98 ± 17.99	.30 ± .16	.36 ± .55
	Unrelated	.13 ± .12	42.98 ± 17.99		
Difficult	Strongly related	.21 ± .17	53.42 ± 18.77		
	Somewhat related	.19 ± .17	48.38 ± 20.35	.30 ± .22	.28 ± .73
	Unrelated	.11 ± .10	38.25 ± 21.34		

2.2.2 JOL accuracy

The one-sample *t*-test results indicated that calibrations in the null [$t(37) = 6.9$, $p = .001$, Cohen's $d = 2.24$], easy [$t(35) = 11.4$, $p = .001$, Cohen's $d = 3.8$], and difficult [$t(38) = 8.4$, $p = .001$, Cohen's $d = 2.69$] conditions were significantly greater than zero.

A significant main effect of multitasking was detected for calibration [$F(2, 112) = 4.67$, $p = .01$, $\eta^2 = 0.08$]. The calibration score under the null condition [$M = 0.18$, SD = 0.16] was lower than the calibration score under the easy ($M = 0.3$, $SD = 0.16$, $p = .026$, Cohen's $d = -0.375$) and difficult ($M = 0.3$, $SD = 0.22$, $p = .03$, Cohen's $d = -.326$) conditions (Fig. 2A), indicating that the absolute accuracy of the JOL was lower under the multitasking condition compared with the single-task condition.

The one-sample *t*-test results indicated that the resolution scores under the null [$t(36) = 16.30$, $p = .001$, Cohen's $d = 5.36$] and difficult [$t(35) = 2.28$, $p = .029$, Cohen's $d = 0.76$] conditions were significantly greater than zero, whereas the difference between the resolution of the easy condition ($t(35) = 1.90$, $p = .066$, Cohen's $d = 0.63$) and zero was only marginally significant.

In terms of resolution, the results indicated a significant difference between the groups [$F(2, 108) = 5.59$, $p = .005$, $\eta^2 = 0.099$]. The resolution score of the null group ($M = 0.69$, $SD = 0.26$) was higher than the resolution scores of the easy ($M = 0.36$, $SD = 0.55$, $p = .017$, Cohen's $d = 0.407$) and difficult ($M = 0.28$, $SD = 0.73$, $p = .002$, Cohen's $d = 0.414$) groups (Fig. 2B), indicating that the relative accuracy of the JOL was lower under the multitasking condition compared with the single-task condition.

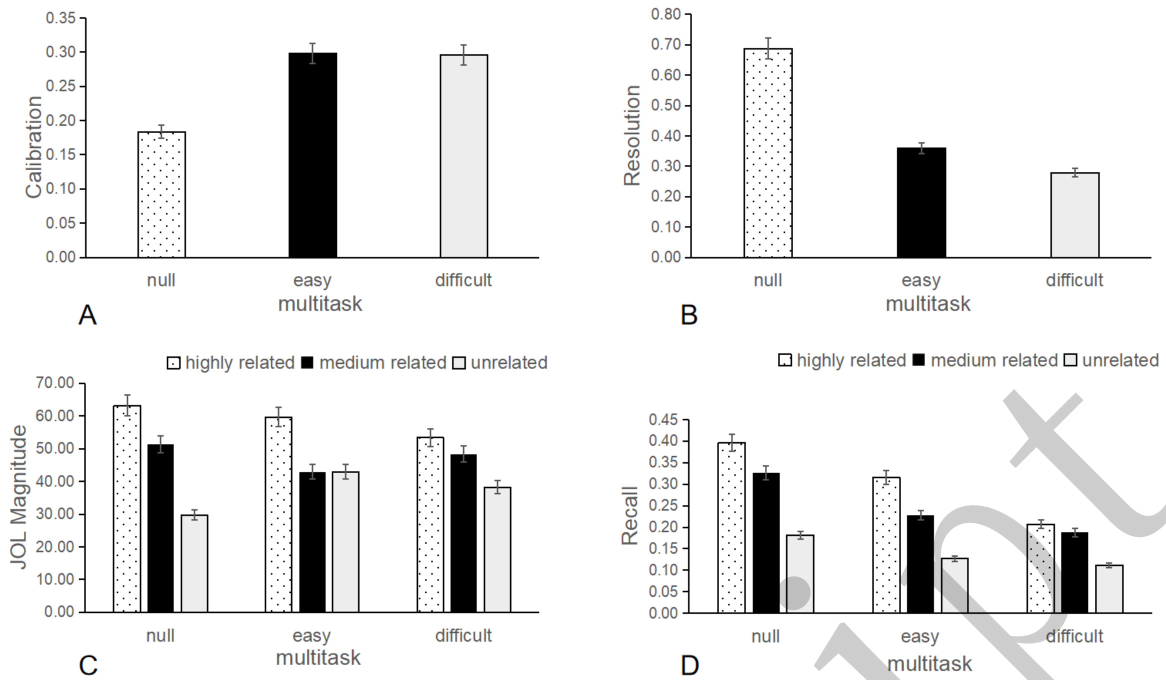

Fig. 2. Visualization of the results of Experiment 1: (A) mean calibration in three groups, (B) mean resolution in three groups, (C) mean JOL magnitude in three groups with different degrees of word-pair relatedness, and (D) mean recall performance in three groups with different degrees of word-pair relatedness. Highly related = strongly related word pair, medium related = somewhat related word pair, unrelated = unrelated word pair. The error bars represent standard errors of the means.

2.2.3 JOL magnitude

Multitasking demonstrated no main effect [$F(2, 112) = 0.14$, $p = .87$, $\eta^2 = 0.002$], indicating that it did not affect the JOL magnitude. By contrast, word pair type demonstrated a main effect [$F(2, 224) = 200.89$, $p = .001$, $\eta^2 = 0.64$]. As shown in Fig. 2C, the JOL magnitude significantly increased with word-pair relatedness. In addition, a significant interaction between word pair type and multitasking was detected [$F(4, 224) = 23.76$, $p = .001$, $\eta^2 = 0.30$]. For strongly related word pairs, $F(2, 112) = 3.4$, $p = .037$, $\eta^2 = 0.057$, the JOL magnitude under the null condition was higher than that under the difficult multitasking condition ($p = .03$). By contrast, for unrelated word pairs, $F(2, 112) = 4.85$, $p = .01$, $\eta^2 = 0.08$, the JOL magnitude under the null condition was lower than that under the easy multitasking condition ($p = .008$). For somewhat related word pairs, $F(2, 112) = 2.22$, $p = .11$, $\eta^2 = 0.038$, the JOL magnitude did not differ between the multitasking conditions. In other words, under the strongly related word-pair condition, the JOL magnitude decreased as the difficulty level increased. By contrast, under the unrelated word-pair condition, the JOL magnitude increased as the difficulty level increased. These results suggested that when the level of difficulty increased, individuals in the unrelated word-pair conditions required additional mental effort to manage information compared with those in the related word-pair conditions. Thus, fewer mental resources were allocated to JOLs, eventually resulting in inaccurate estimations.

2.2.4 Recall

Multitasking demonstrated a significant main effect ($F(2, 112) = 9.0$, $p = .001$, $\eta^2 = 0.14$). The recall performance of the easy and difficult multitasking groups was significantly lower than that of the null multitasking group ($p = .045$, $p = .001$; Fig. 2D). However, no significant difference was observed between the easy and difficult conditions ($p = .25$). Word pair type demonstrated a significant main effect [$F(2, 224) = 52.2$, $p = .001$, $\eta^2 = 0.32$], with a significant decrease in recall performance observed with a decrease in relatedness. A significant interaction was observed between multitasking and word pair type [$F(4, 224) = 2.6$, $p = .037$, $\eta^2 = 0.04$]. For the null [$F(2, 111) = 28.99$, $p = .001$, $\eta^2 = 0.34$] and difficult [$F(2,112) = 6.90$, $p = .002$, $\eta^2 = 0.11$] multitasking conditions, the recall performance of individuals on the strongly related and somewhat related word pairs was higher than their performance on unrelated word pairs. Under the easy multitasking condition [$F(2, 111) = 19.76$, $p = .001$, $\eta^2 = 0.26$], individuals' recall performance for strongly related word pairs was higher than their performance for somewhat related ($p = .013$) and unrelated ($p = .001$) word pairs. In addition, individuals' recall performance for somewhat related word pairs was higher than that for unrelated word pairs ($p = .001$).

2.2.5 Correlation between the MMI and task performance

Pearson's correlation analysis of the MMI and task performance indicators was performed under null and multitasking conditions. For the null multitasking group, the MMI was positively correlated with calibration ($r = 0.54$, $p = .001$) but was not correlated with JOL magnitude ($r = 0.21$, $p = .24$), recall performance ($r = -0.29$, $p = .084$), or resolution ($r = 0.07$, $p = .68$). These results indicated that the participants who more frequently engaged in daily multitasking exhibited a lower JOL accuracy in single tasks than those who less frequently engaged in daily multitasking. For the easy multitasking group, no significant correlations were observed between MMI and calibration ($r = -0.013$, $p = .94$), JOL magnitude ($r = 0.02$, $p = .91$), recall performance ($r = 0.044$, $p = .79$), and resolution ($r = 0.008$, $p = .962$). Similarly, for the difficult multitasking group, no significant correlations were observed between MMI and calibration ($r = -0.227$, $p = .17$), JOL magnitude ($r = -0.212$, $p = .2$), recall performance ($r = 0.066$, $p = .695$), and resolution ($r = -0.025$, $p = .88$).

In summary, our hypothesis was supported by the results of Experiment 1. Compared with the single-task group, the multitask groups demonstrated poorer calibration and lower resolution. These results indicated that the participants in the multitasking groups were less accurate than their counterparts in estimating their own performance. According to the correlation analysis results of Experiment 1, in single tasks, the participants who are more frequently engaged in daily media multitasking exhibited a lower accuracy in metacognitive evaluation.

In Experiment 1, the secondary task (i.e., pitch discrimination) was artificial. Therefore, in Experiment 2, we increased the ecological validity of the multitasking design to determine whether the effect of multitasking on metacognitive judgment is observable in a scenario that simulated reality. To achieve this goal, we used actual notification sounds from instant messaging applications as stimuli for the secondary task in Experiment 2.

3. Experiment 2

3.1 Method

3.1.1 Participants

Power analysis indicated that a minimum sample size of 68 was required to obtain an effect size of 0.25 with a power of 0.8 at an α level of 0.05 in the within-between repeated measures of 2 × 3 ANOVAs. A total of 77 college students (47 female, M_{age} = 20.94, *SD* = 2.72) were recruited and randomly assigned into multitasking (*n* = 38) and nonmultitasking (*n* = 39) groups. All participants received monetary compensation.

3.1.2 Materials and procedure

The word pairs used in this experiment were selected from those used in Experiment 1. A 2 (group, between-subject design: single task vs. multiple tasks) × 3 (word pair type, within-subject design: strongly related vs. somewhat related vs. unrelated) mixed design was adopted. The single-task condition was identical to that in Experiment 1. Under the multitasking condition, during the 2-s word-pair presentation, a notification sound from WeChat or QQ was applied in the first second for a duration of 1 s, and the volume was adjusted to 50% on the computer. The participants were asked to click as quickly as possible on the application icon below each word pair corresponding to the sound. Before the word-pair study phase, the participants were trained to familiarize themselves with the notification sound of each application. After the study phase, a filler task and a target word recall task were executed, as in Experiment 1.

Table 2 Mean and Standard Deviation of recall, calibration, resolution, and JOL magnitude in Experiment 2

	Word Pair Type	Recall	Accuracy of JOL

Group			JOL magnitude	Calibration	Resolution
		M ± *SD*	*M* ± *SD*	*M* ± *SD*	*M* ± *SD*
Single task	Strongly related	.57 ± .20	74.62 ± 14.29		
	Somewhat related	.22 ± .17	47.99 ± 18.70	.19± .22	.59 ± .43
	Unrelated	.20 ± .17	33.80 ± 18.61		
multitask	Strongly related	.35 ± .21	58.99 ± 20.11		
	Somewhat related	.08 ± .13	45.39 ± 18.61	.30 ± .20	.47 ± .62
	Unrelated	.07 ± .10	36.05 ± 17.17		

3.2 Results and discussion

Table 2 presents the descriptive statistics. We performed a 2 (group, between-subject design: single task vs. multiple tasks) × 3 (word pair type, within-subject design: strongly related vs. somewhat related vs. unrelated) repeated measures ANOVA for recall performance and JOL magnitude and an independent-samples *t*-test for calibration and resolution.

3.2.1 JOL accuracy

The one-sample *t*-test results indicated that the calibration scores in the single-task ($t(38) = 5.5$, $p = .001$, Cohen's $d = 1.76$) and multitasking ($t(37) = 9.1$, $p = .001$, Cohen's $d = 2.95$) conditions were significantly greater than zero. Notably, the calibration score under the single-task condition ($M = 0.19$, $SD = 0.22$) was lower than that under the multitasking condition ($M = 0.30$, $SD = 0.20$), $t(75) = -2.28$, $p = .025$, Cohen's $d = 0.52$ (Fig. 3A), indicating that the absolute accuracy of the JOL was lower under the multitasking condition than under the single-task condition.

The one-sample *t*-test results also indicated that the resolution score under the single-task ($t(37) = 8.44$, $p = .001$, Cohen's $d = 2.74$) and multitasking ($t(34) = 4.47$, $p = .001$, Cohen's $d = 1.53$) conditions were significantly higher than zero. Although the difference between the single-task ($M = 0.59$, $SD = 0.43$) and multitasking ($M = 0.47$, $SD = 0.62$) groups was not statistically significant in terms of resolution, $t(71) = 0.98$, $p = .061$, Cohen's $d = 0.23$ (Fig. 3B), the relative accuracy of JOL was lower under the multitasking condition than under the single-task condition.

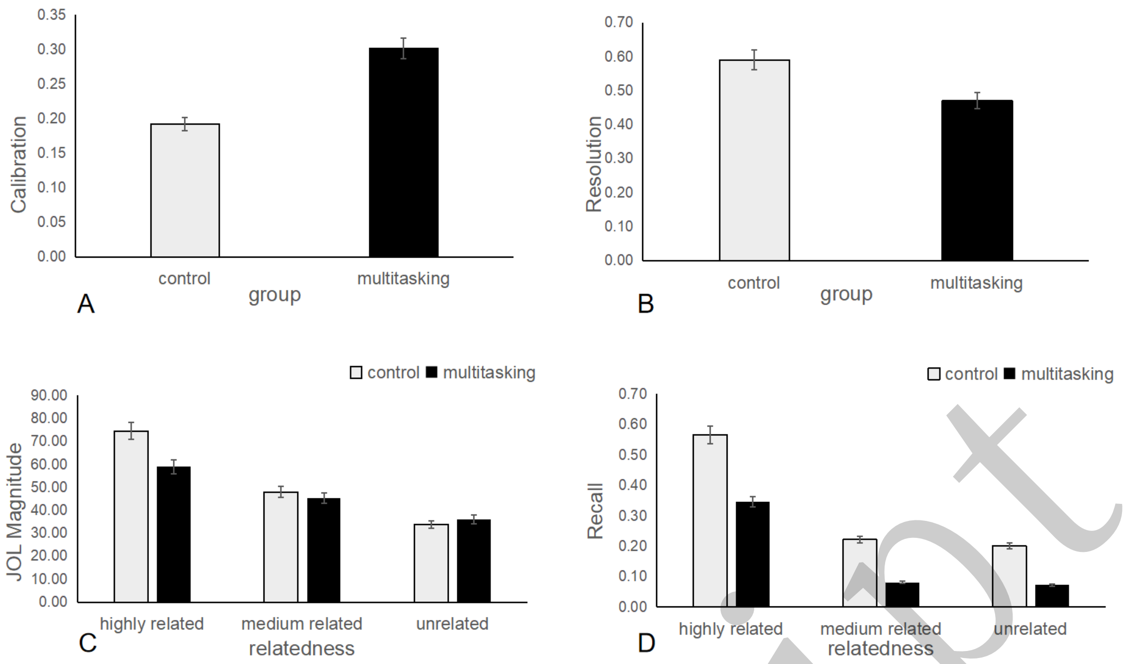

Fig. 3. Visualization of the results of Experiment 2: (A) mean calibration in two groups, (B) mean resolution in two groups, (C) mean JOL magnitude in two groups with different degrees of word-pair relatedness, and (D) mean recall performance in two groups with different degrees of word-pair relatedness. Highly related = strongly related word pair, medium related = somewhat related word pair, unrelated = unrelated word pair. The error bars represent standard errors of the means.

3.2.2 JOL magnitude

Although no main group effect was observed [$F(1, 75) = 2.0$, $p = .16$, $\eta^2 = 0.026$], a main effect of word pair type was observed [$F(2, 150) = 261.96$, $p = .001$, $\eta^2 = 0.78$ (Fig. 3C)]. The JOL magnitude significantly increased with increasing word-pair relatedness. In addition, a significant interaction was observed [$F(2, 150) = 21.54$, $p = .001$, $\eta^2 = 0.22$]. Specifically, when the word pairs were strongly related [$F(1, 72) = 15.24$, $p = .001$, $\eta^2 = 0.175$], the JOL magnitude of the single-task group significantly exceeded that of the multitasking group ($p = .001$, Cohen's $d = 0.454$). However, no difference was observed between the two groups under the somewhat related [$F(1, 72) = 0.539$, $p = .465$, $\eta^2 = 0.007$] and unrelated [$F(1, 72) = 0.182$, $p = .671$, $\eta^2 = 0.003$] word-pair conditions.

3.2.3 Recall

ANOVA revealed a main effect of word pair type [$F(2, 150) = 129.69$, $p = .001$, $\eta^2 = 0.634$ (Fig. 3D)]. The participants recalled significantly more strongly related word pairs compared with somewhat related ($p = .001$) and unrelated ($p = .001$) word pairs. A main group effect was also observed [$F(1, 75) = 32.31$, $p = .001$, $\eta^2 = 0.30$]. The memory performance under the single-task condition was higher than that under the multitasking condition (p = .001). These results indicated that responding to the notification sounds negatively affected the participants' memory performance. However, no interaction was observed between group and relatedness [$F(2, 150) = 2.44$, p = .09, η^2 = 0.03]. The recall rate results were identical to those in Experiment 1, indicating a stable effect of relatedness and multitasking on recall performance.

3.2.4 Correlation analysis

Pearson's correlation analysis of the MMI and task performance indicators was performed for the single-task group. However, no correlation was observed between MMI and calibration (r = 0.29, p = .074), resolution (r = −0.01, p = .93), JOL magnitude (r = 0.08, p = .63), or recall performance (r = −0.27, p = .1). By contrast, in the multitasking group, the MMI was correlated with calibration (r = 0.404, p = .012), resolution (r = 0.45, p = .007), and JOL magnitude (r = 0.355, p = .029) but was not correlated with recall performance (r = −0.22, p = .18).

3.2.5 Meta-analysis

Although Experiments 1 and 2 had the same mixed design, the results of Experiment 2 revealed marginal significance in terms of resolution. Given the consistent trend and size of the p value, in accordance with Ralph et al. (2021), who used a meta-analysis to combine their two studies in a single article, we conducted a meta-analysis of resolution and calibration, including the results of Experiments 1 and 2, to obtain a more robust result. We used the "meta" package of R. According to the results, a small degree of heterogeneity was observed between the two experiments in terms of calibration, Q = 0.46 (p = .5), I^2 = 0.0%, and resolution, Q = 1.5 (p = .22), I^2 = 33.3%. In addition, multitasking affected both calibration (p = .001, Hedges' g = −.42, 95% CI [−.77, −.08]) and resolution (p = .01, Hedges' g = -.42, 95% CI [-.76, -.08]). These results indicated that the participants tended to overestimate their memory performance under the multitasking condition and were unable to effectively differentiate between the items they did or did not memorize well.

Overall, the inconsistency between the correlation results obtained from the two experiments may have been due to the small sample size. Therefore, we combined data from the two experiments to conduct a meta-analysis. We used the "metacor" package of R to conduct a meta-analysis of correlations. The results of the single-task groups revealed a small degree of heterogeneity between the two experiments in terms of calibration [Q = 1.64 (p = .2), I^2 = 38.9%] and recall [Q = 0.01 (p = .91), I^2 = 0.0%]. In addition, the MMI was positively correlated with calibration (r = 0.42, p = .001, 95% CI [.20, .59]) but negatively correlated with recall rate (r = −0.28, p = .016, 95% CI [−.48, −.05]). These results indicated that the participants who frequently multitasked tended to have lower recall performance and to make less accurate metacognitive evaluations compared with their counterparts. In the multitasking group, a large heterogeneity was observed between the two experiments, but no significant correlation was observed with calibration (Q = 7.61 [$p < .001$], I^2 = 86.9%, r = 0.098, p = .76, 95% CI [−.50, .63]), resolution (Q = 3.8 [p = .05], I^2 = 73.7%, r = 0.23, p = .3, 95% CI [−.21, .61]), JOL magnitude (Q = 6.02 [$p < .05$], I^2 = 83.4%, r = 0.078, p = .79, 95% CI [−.45, .57]), or recall performance (Q = 1.49 [p = .22], I^2 = 32.9%, r = −0.08, p = .58, 95% CI [−.27, .62], 95% CI [−.32, .15]).

4. General Discussion

In this study, we examined the effect of multitasking on metacognition. We hypothesized that multitasking would negatively affect the accuracy of metacognitive evaluation in subsequent tasks. The results of Experiment 1 supported this hypothesis, and the results of Experiment 2 further corroborated the findings of Experiment 1, its design had stronger ecological validity. Meta-analysis of the two experiments indicated that multitasking negatively affected both the absolute (i.e., calibration) and the relative (i.e., resolution) accuracy of metacognitive monitoring.

Overall, our results revealed a negative effect of multitasking on the accuracy of prospective metacognitive monitoring. These results are consistent with those of Konishi et al. (2021), who used confidence judgment as an indicator for retrospective monitoring. Compared with previous studies, in which no negative effects were reported from multitasking on metacognitive judgment (Konishi et al., 2020; Peng & Tullis, 2021), in this study, we used more difficult multitasks. We also added a filler task before the memory test to further increase task difficulty. Our results indicated certain competition for cognitive resources between object-level cognition (word-pair learning and sound pitch discrimination) and meta-level processing (metacognitive evaluation of learning). They also indicated that increased task difficulty in object-level cognition requires additional cognitive resources, thereby reducing the resources available for the meta level. Therefore, in Experiment 1, compared with the participants in the single-task group, the participants in the multitasking group made less accurate metacognitive judgments.

Overall, the results of Experiment 2, which exhibited high ecological validity, corroborated those of Experiment 1. To simulate a real-life scenario as closely as possible, in Experiment 2, we replaced the pure tones in Experiment 1 with actual notification sounds from different instant messaging applications. In real-life scenarios, people tend to respond to instant messages from various applications while studying or working. After they hear a notification sound, they use their computer mouse to click on the corresponding application icon. Therefore, in this study, we replaced pressing keys in Experiment 1 to clicking application icons in Experiment 2. The changes made to the secondary task in Experiment 2 improved ecological validity and increased task difficulty. The participants studied the word pairs displayed on the screen while clicking on the corresponding icons after hearing the notification sound, resulting in an intense visual processing load. However, despite the change in difficulty, Experiment 2 yielded similar results to those of Experiment 1. Therefore, both Experiment 1, which was more artificial, and Experiment 2, which was more ecologically valid, demonstrated a negative effect of multitasking on the accuracy of metacognitive evaluation. Our meta-analysis also revealed a low heterogeneity between both experiments and indicated that multitasking had a negative effect on metacognitive monitoring.

As a next step, we explored the mechanisms through which multitasking affects metacognition. We used both absolute accuracy (calibration) and relative accuracy (resolution) to evaluate metacognitive abilities. Calibration refers to the ability to assess the mastery of study items, whereas resolution refers to the ability to distinguish between well-remembered and poorly remembered items. We used the calculation method of Undorf and Bröder (2019). In Experiment 1, 97.37%, 91.89%, and 79.48% of the participants used word-pair relatedness to provide JOLs under the null, easy, and difficult multitasking conditions, respectively. Hence, we speculated that when the participants predicted their performance on memory tests, their limited resources prevented them from effectively using valid cues (i.e., word-pair relatedness) to provide a metacognitive evaluation. However, this speculation requires further validation by empirical studies.

Generally, whether multitasking affects the accuracy of metacognitive assessment depends on the presence of interblock comparisons or reminders. Consistent with previous research (Finley et al., 2014; Hanczakowski et al., 2018; Ralph et al., 2018), a preliminary study with a within-subject design (reported in the supplementary materials) did not identify any effect of multitasking on the accuracy of metacognitive assessment. In the preliminary study, the participants executed both single and dual tasks and were able to spontaneously compare the two conditions while forming JOLs. Comparing these conditions allowed the participants to recognize the negative effects of multitasking, thereby resulting in accurate metacognitive judgments through the prediction of lower memory performance under the multitasking condition compared with the single-task condition. By contrast, in Experiments 1 and 2, the participants executed tasks under only one condition and were unable to receive feedback regarding their learning outcomes under all conditions, thus preventing them from adjusting their own metacognitive assessments. Therefore, the results of both experiments indicated that the JOL magnitude did not differ between the single-task and multitasking conditions. However, the performance of recall was lower under the multitasking condition than under the single-task condition. In summary, the presence of interblock comparisons or reminders is essential in accurately evaluating the effect of multitasking on metacognitive assessments.

We examined the correlation between daily media multitasking behavior and metacognitive accuracy under both single-task and multitasking conditions. Our results indicated that the participants who frequently engaged in multitasking behaviors tended to have less accurate metacognitive evaluations, compared with their counterparts, when performing individual tasks. However, we did not observe a significant correlation between media multitasking behavior and metacognitive accuracy in the multitasking group. These results suggested that an individual's daily media multitasking behavior may be related to their performance in single tasks but not in multiple tasks. This is likely due to the significant decline in task performance caused by multitasking, irrespective of whether individuals are heavy or light multitaskers. Consequently, this effect may obscure individual differences (Minear et al., 2013). However, these findings require further validation because of the relatively small sample size in our study.

Overall, this study contributes to metacognition research and theory by demonstrating that multitasking reduces the accuracy of metacognitive monitoring. According to the theoretical metacognition model of Nelson and Nelson (1990) and the limited capacity model of mediated message processing of Lang (2000), object-level cognition and meta-level processing share the same cognitive resources, with multitasking requiring additional resources at the object level, thereby leaving insufficient resources for meta-level processes. This impairment in meta-level processing reduces the accuracy of metacognitive monitoring. However, none of the previous studies, including those of Hanczakowski et al. (2018), Peng and Tullis (2021), and Ralph et al. (2018), supported this theoretical inference, except for two studies. Sanbonmatsu et al. (2016) discovered that their participants were able to identify the negative effect of multitasking but underestimated their own errors. Konishi et al. (2021) highlighted the negative effects of multitasking on retrospective monitoring through item-by-item judgments. The present study expands these findings by demonstrating the negative effect of multitasking on the accuracy of prospective monitoring, thereby supporting the limited capacity model.

Overall, this study provides practical implications for people’s lives. Multitasking negatively affects task performance and metacognitive evaluation, which prevents people from perceiving problems during multitasking. Therefore, people must resist the distraction of irrelevant media and concentrate on the primary task, especially in dangerous environments, such as on public transportation for drivers, where hands-free phones negatively affect the drivers’ self-awareness and increase their risk of misjudgment (Sanbonmatsu et al., 2016). Within the context of education, teachers should remind their students of the negative effects of multitasking, because the mere presence of a smartphone can reduce their available cognitive resources and result in the overestimation of performance (Ward et al., 2017). Keeping electronic devices out of reach and muting notifications can therefore improve learning concentration and metacognitive processing, thereby increasing learning effectiveness. Media multitasking also has a negative effect on the accuracy of JOLs, indicating that frequent media multitasking may be associated with a decline in self-reflective abilities.

This study has several limitations. First, although the results indicated a negative effect of multitasking on memory and metacognitive monitoring accuracy, we did not investigate the relationship between these three concepts. According to previous research, metacognition mediates the relationship between media multitasking behavior and learning performance (Gandolfi et al., 2021). In addition, impairments in metacognitive monitoring are associated with various psychobehavioral problems, such as addictive behaviors (Spada et al., 2015). Therefore, future studies should examine whether behavioral problems arising from media multitasking are the result of impaired metacognitive monitoring. Second, this study included only college students. Therefore, future studies should investigate adolescents who frequently engage in media multitasking. Third, the ecological validity of this study should be improved by adopting more realistic experimental settings. Finally, while the correlation between media multitasking and poor metacognitive monitoring is intriguing, our small sample size calls for further verification in future studies.

5. Conclusion

Multitasking reduces the accuracy of metacognitive monitoring, whether in the laboratory or in daily life.

Declarations

Competing interests: The authors have no conflicts of interest to declare.

Ethics approval: Research involving Human Participants: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and ethics approval was obtained from the Institutional Review Board of the authors' affiliating university, and parental consent was obtained before the study.

Consent to participate: Informed consent was obtained from all individual participants included in the study.

Author contributions statement: Ruiqiao Guo: Conceptualization (Ruiqiao Guo, Yan Liu), Methodology (Ruiqiao Guo, Yan Liu), Software (Ruiqiao Guo), Validation (Ruiqiao Guo), Formal Analysis (Ruiqiao Guo), Investigation (Ruiqiao Guo), Resources (Ruiqiao Guo), Data Curation (Ruiqiao Guo, Yan Liu, Annan Jing), Writing-Original Draft (Ruiqiao Guo), Writing - Review & Editing (Yan Liu, Hui Jing Lu), Supervision (Yan Liu), Project Administration (Yan Liu), Funding Acquisition(Yan Liu)

Funding: Research grants from Social Science Foundation of Liaoning Province of China (L21BSH007)

Availability of data and materials: The datasets generated during the current study are available in the github (https://github.com/rachelzzy/metacognition)

Reference

Aagaard, J. (2018). Multitasking as distraction: A conceptual analysis of media multitasking research. *Theory & Psychology*, 095935431881576. https://doi.org/10.1177/0959354318815766

Aguilar-Lleyda, D., Lemarchand, M., & De Gardelle, V. (2020). Confidence as a priority signal. *Psychological Science*, *31*(9), 1084-1096.

Akbari, M. (2017). Metacognitions or distress intolerance: The mediating role in the relationship between emotional dysregulation and problematic internet use. *Addictive Behaviors Reports*, *6*, 128-133.

Adler, R. F., & Benbunan-Fich, R. (2012). Juggling on a high wire: Multitasking effects on performance. *International Journal of Human-Computer Studies*, *70*(2), 156-168.

Adler, R. F., & Benbunan-Fich, R. (2015). The effects of task difficulty and multitasking on performance. *Interacting with Computers*, *27*(4), 430-439.

Anggo, M., & Arapu, L. (2018, June). The use of mathematics teaching aids to train metacognition ability of elementary school students. In *Journal of Physics: Conference Series* (Vol. 1028, No. 1, p. 012143). IOP Publishing.

Brasel, S. A., & Gips, J. (2011). Media multitasking behavior: concurrent television and computer usage. *Cyberpsychology, Behavior, and Social Networking*, *14*(9), 527-534.

Chen, G., & Fu, X. (2010, August). The foundation of JOLs and influencing factors. In *2010 Sixth International Conference on Natural Computation* (Vol. 4, pp. 2181-2185). IEEE.

Dindar, M., & Akbulut, Y. (2016). Effects of multitasking on retention and topic interest. *Learning and Instruction*, *41*, 94-105.

Finley, J. R., Benjamin, A. S., & McCarley, J. S. (2014). Metacognition of multitasking: How well do we predict the costs of divided attention? *Journal of Experimental Psychology: Applied*, *20*(2), 158-165. https://doi.org/10.1037/xap0000010

Gandolfi, E., Soyturk, I., & Ferdig, R. E. (2021). Evaluating U.S. gamers' metacognitions about digital entertainment: Validation of metacognition about online gaming scale in the U.S. context. *Journal of Affective Disorders*, *295*, 954-959. https://doi.org/https://doi.org/10.1016/j.jad.2021.08.133

Hainguerlot, M., Vergnaud, J.-C., & De Gardelle, V. (2018). Metacognitive ability predicts learning cue-stimulus associations in the absence of external feedback. *Scientific Reports*, *8*(1), 1-8.

Hanczakowski, M., Beaman, C. P., & Jones, D. M. (2018). Learning through clamor: The allocation and perception of study time in noise. *Journal of Experimental Psychology: General*, *147*(7), 1005-1022. https://doi.org/10.1037/xge0000449

Händel, M., de Bruin, A. B. H., & Dresel, M. (2020). Individual differences in local and global metacognitive judgments. *Metacognition and Learning*, 15(1), 51–75. https://doi.org/10.1007/s11409-020-09220-0

Heyes, C., Bang, D., Shea, N., Frith, C.D., & Fleming, S.M. (2020). Knowing ourselves together: The cultural origins of metacognition. *Trends in Cognitive Sciences, 24*, 349 - 362.

Jamet, E., Gonthier, C., Cojean, S., Colliot, T., & Erhel, S. (2020). Does multitasking in the classroom affect learning outcomes? A naturalistic study. *Computers in Human Behavior*, *106*, 106264.

Jauregui, P., Urbiola, I., & Estevez, A. (2016). Metacognition in pathological gambling and its relationship with anxious and depressive symptomatology. *Journal of Gambling Studies*, *32*(2), 675-688.

Jeong, S.-H., & Hwang, Y. (2012). Does multitasking increase or decrease persuasion? Effects of multitasking on comprehension and counterarguing. *Journal of Communication*, *62*(4), 571-587.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of applied psychology*, *74*(4), 657.

Kelemen, W. L. (2000). Metamemory cues and monitoring accuracy: Judging what you know and what you will know. *Journal of Educational Psychology*, *92*(4), 800.

Kihlstrom, J. F. (2021). Ecological validity and "ecological validity". *Perspectives on Psychological Science*, *16*(2), 466-471.

Koch, I., Poljac, E., Müller, H., & Kiesel, A. (2018). Cognitive structure, flexibility, and plasticity in human multitasking—An integrative review of dual-task and task-switching research. *Psychological Bulletin, 144*(6), 557–583. https://doi.org/10.1037/bul0000144

Kong, F., Meng, S., Deng, H., Wang, M., & Sun, X. (2023). Cognitive Control in Adolescents and Young Adults with Media Multitasking Experience: a Three-Level Meta-analysis. In *Educational Psychology Review* (Vol. 35, Issue 1). Springer US. https://doi.org/10.1007/s10648-023-09746-0

Konishi, M., Berberian, B., de Gardelle, V., & Sackur, J. (2021). Multitasking costs on metacognition in a triple-task paradigm. *Psychonomic Bulletin & Review*, 1-10. https://doi.org/10.3758/s13423-021-01967-0

Konishi, M., Compain, C., Berberian, B., Sackur, J., & de Gardelle, V. (2020). Resilience of perceptual metacognition in a dual-task paradigm. *Psychonomic Bulletin Review*, *27*(6), 1259-1268. https://doi.org/10.3758/s13423-020-01779-8

Lang, A. (2000). The limited capacity model of mediated message processing. *Journal of Communication*, *50*(1), 46-70.

Lehmann, M., Hagen, J., & Ettinger, U. (2022). Unity and diversity of metacognition. *Journal of Experimental Psychology: General.* Advance online publication. https://doi.org/10.1037/xge0001197

Liu, Y., Su, Y., Xu, G., & Pei, M. (2018). When do you know what you know? The emergence of memory monitoring. *Journal of Experimental Child Psychology*, *166*, 34-48.

Loh, K. K., Tan, B. Z. H., & Lim, S. W. H. (2016). Media multitasking predicts video-recorded lecture learning performance through mind wandering tendencies. *Computers in Human Behavior*, *63*, 943-947. https://doi.org/10.1016/j.chb.2016.06.030

Lopez, R. B., Salinger, J. M., Heatherton, T. F., & Wagner, D. D. (2018). Media multitasking is associated with altered processing of incidental, irrelevant cues during person perception. *BMC Psychology*, *6*(1), 1-7.

Lottridge, D. M., Rosakranse, C., Oh, C. S., Westwood, S. J., Baldoni, K. A., Mann, A. S., & Nass, C. I. (2015). The effects of chronic multitasking on analytical writing. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems,

May, K. E., & Elder, A. D. (2018). Efficient, helpful, or distracting? A literature review of media multitasking in relation to academic performance. *International Journal of Educational Technology in Higher Education*, *15*(1), 13. https://doi.org/10.1186/s41239-018-0096-z

Minear, M., Brasher, F., McCurdy, M., Lewis, J., & Younggren, A. (2013). Working memory, fluid intelligence, and impulsiveness in heavy media multitaskers. *Psychonomic Bulletin & Review*, *20*(6), 1274-1281. https://doi.org/10.3758/s13423-013-0456-6

Murphy, K., McLauchlan, S., & Lee, M. (2017). Is there a link between media-multitasking and the executive functions of filtering and response inhibition? *Computers in Human Behavior*, *75*, 667-677.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402-407. https://doi.org/10.3758/BF03195588

Nelson, T., Narens, L. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning and Motivation*, *26*, 125-173.

Nikčević, A. V., & Spada, M. M. J. B. (2008). Metacognitions across the continuum of smoking dependence. *Behavioural and Cognitive Psychotherapy*, *36*(3), 333-339.

Ophir, E., Nass, C., & Wagner, A. D. (2009). Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences*, *106*(37), 15583-15587. https://doi.org/10.1073/pnas.0903620106

Parry, D. A., & Le Roux, D. B. (2021). “Cognitive Control in Media Multitaskers” Ten Years On: A Meta-Analysis. *Cyberpsychology*, 15(2). https://doi.org/10.5817/CP2021-2-7

Peng, Y., & Tullis, J. G. (2021). Dividing attention impairs metacognitive control more than monitoring. *Psychonomic Bulletin Review*, 1-11.

Ralph, B., Seli, P., Wilson, K., & Smilek, D. (2018). Volitional media multitasking: Awareness of performance costs and modulation of media multitasking as a function of task demand. *Psychological Research*, *84*, 1-20. https://doi.org/10.1007/s00426-018-1056-x

Ralph, B. C., Smith, A. C., Seli, P., & Smilek, D. J. P. r. (2021). The relation between task-unrelated media multitasking and task-related motivation. *85*(1), 408-422.

Rubenking, B. (2017). Boring is bad: Effects of emotional content and multitasking on enjoyment and memory. *Computers in Human Behavior*, *72*, 488-495. https://doi.org/https://doi.org/10.1016/j.chb.2017.03.015

Sanbonmatsu, D. M., Strayer, D. L., Biondi, F., Behrends, A. A., & Moore, S. M. (2016). Cell-phone use diminishes self-awareness of impaired driving. *Psychonomic Bulletin & Review*, *23*(2), 617-623. https://doi.org/10.3758/s13423-015-0922-4

Spada, M. M., Caselli, G., & Wells, A. J. B. R. (2009). Metacognitions as a predictor of drinking status and level of alcohol use following CBT in problem drinkers: A prospective study. *Behaviour Research and Therapy*, *47*(10), 882-886.

Spada, M. M., Giustina, L., Rolandi, S., Fernie, B. A., & Caselli, G. (2015). Profiling metacognition in gambling disorder. *Behavioural and Cognitive Psychotherapy*, *43*(5), 614-622.

Uncapher, M. R., K. Thieu, M., & Wagner, A. D. (2016). Media multitasking and memory: Differences in working memory and long-term memory. *Psychonomic Bulletin & Review*, *23*(2), 483-490.

Undorf, M., & Bröder, A. (2019). Cue integration in metamemory judgements is strategic. *Quarterly Journal of Experimental Psychology*, *73*(4), 629-642. https://doi.org/10.1177/1747021819882308

van den Berg, R., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). Confidence is the bridge between multi-stage decisions. *Current Biology* , *26*(23), 3157-3168.

Ward, A. F., Duke, K., Gneezy, A., & Bos, M. W. (2017). Brain drain: the mere presence of one's own smartphone reduces available cognitive capacity. *Journal of the Association for Consumer Research*, *2*(2), 140-154. https://doi.org/10.1086/691462

Wickens, C. D. (2020). *Processing resources and attention* (pp. 3-34). CRC Press.

Winne, P. H., & Hadwin, A. E. (1998). *Studying as self-regulated learning* (pp. 291-318). Routledge.

Wiradhany, W., Baumgartner, S. E., & de Bruin, A. B. H. (2021). Exploitation – Exploration Model of Media Multitasking. *Journal of Media Psychology*, 33(4), 169–180. https://doi.org/https://doi.org/10.1027/1864-1105/a000303

Wiradhany, W., & Nieuwenstein, M. R. (2017). Cognitive control in media multitaskers: Two replication studies and a meta-Analysis. *Attention, Perception, & Psychophysics*, 79(8), 2620–2641. https://doi.org/10.3758/s13414-017-1408-4

Yang, X., & Zhu, L. (2014). Relationship among media multitasking, personality and negative mood in college students. *Chinese Mental Health Journal*, *28*(4), 277-282.