Educational Psychology Review (2024) 36:66
https://doi.org/10.1007/s10648-024-09903-z

META-ANALYSIS

The Relation Between Perceived Mental Effort, Monitoring Judgments, and Learning Outcomes: A Meta-Analysis

Louise David[1] · Felicitas Biwer[1] · Martine Baars[2] · Lisette Wijnia[3] · Fred Paas[2,4] · Anique de Bruin[1]

Accepted: 28 May 2024

Abstract

Accurately monitoring one's learning processes during self-regulated learning depends on using the right cues, one of which could be perceived mental effort. A meta-analysis by Baars et al. (2020) found a negative association between mental effort and monitoring judgments (r=-.35), suggesting that the amount of mental effort experienced during a learning task is usually negatively correlated with learners' perception of learning. However, it is unclear how monitoring judgments and perceptions of mental effort relate to learning outcomes. To examine if perceived mental effort is a diagnostic cue for learning outcomes, and whether monitoring judgments mediate this relationship, we employed a meta-analytic structural equation model. Results indicated a negative, moderate association between perceived mental effort and monitoring judgments (β=-.19), a positive, large association between monitoring judgments and learning outcomes (β=.29), and a negative, moderate indirect association between perceived mental effort and learning outcomes (β=-.05), which was mediated by monitoring judgments. Our subgroup analysis did not reveal any significant differences across moderators potentially due to the limited number of studies included per moderator category. Findings suggest that when learners perceive higher levels of mental effort, they exhibit lower learning (confidence) judgments, which relates to lower actual learning outcomes. Thus, learners seem to use perceived mental effort as a cue to judge their learning while perceived mental effort only indirectly relates to actual learning outcomes.

Keywords Mental effort · Monitoring · Performance · Metacognition · Meta-analysis · Cue-utilization

Knowing how to study and steer, or self-regulate, one's learning process effectively is a key skill in education (Broadbent & Poon, 2015). Even if not being formally taught how, learners often have to make decisions when it comes to their learning

Extended author information available on the last page of the article

Published online: 26 June 2024

behavior. For example, they have to decide whether to re-read a text, which learning task to do next, or which learning strategy to use when memorizing content. To optimize learning outcomes, accurately monitoring learning and—based on this monitoring—accurately regulating one's learning process is essential (De Bruin & Van Merriënboer, 2017; Dunlosky & Rawson, 2012). However, many learners struggle to accurately monitor their learning process, leading them to make suboptimal study decisions (e.g., Baars et al., 2013, 2014; Bjork et al., 2013; Butler & Winne, 1995; Cavalcanti & Sibbald, 2014; Dinsmore & Parkinson, 2013; Kostons et al., 2012), which hampers learning and, consequently, academic outcomes (Dunlosky & Metcalfe, 2008; Dunlosky & Rawson, 2012; Kanfer & Ackerman, 1989). In the current manuscript, we use the term learning outcomes as quantifiable achievements or performances displayed by learners following a learning phase. These outcomes usually refer to knowledge acquisition, which is demonstrated shortly after the learning phase on a near-transfer task.

Given that learners often cannot access information about their learning outcomes directly, they usually monitor their learning outcomes based on indirect information or cues (i.e., cue-utilization framework; Koriat, 1997). More specifically, learners use various cues, such as processing fluency or perceived mental effort, to inform their metacognitive monitoring (Baars et al., 2020; Koriat & Ma'ayan, 2005; Koriat et al., 2006; Koriat et al., 2014a, b; Onan et al., 2022; Undorf & Erdfelder, 2011). Using a meta-analytic approach, Baars and colleagues (2020) found a negative, medium-sized correlation (r = -0.35) between effort and monitoring judgments. This negative correlation suggests that greater perceived effort is often linked to lower estimates of learning, whereas lower perceived mental effort tends to correspond with higher estimations of learning. This pattern implies that learners might use the amount of effort they expend as a cue to guide their monitoring of learning judgments, which we will refer to as "monitoring judgments" going forward. More specifically, if learners experience high mental effort when working on a task, they could interpret this as a sign of poor learning. However, it is unclear to what extent perceived mental effort is associated with actual learning outcomes and, therefore, it is unclear to what extent perceived mental effort is *diagnostic* for actual learning outcomes. If indeed perceived mental effort is used as a cue for making monitoring judgments, but perceived mental effort is not associated with learning outcomes, this could lead to inaccurate monitoring judgments. Subsequently, this could influence learners to take ineffective learning decisions, decreasing their learning and academic outcomes. Therefore, it is crucial to investigate the relationship between perceived mental effort, monitoring judgments, and learning outcomes. Furthermore, investigating if perceived mental effort acts as a diagnostic cue for learning necessitates examining its relationship with learning outcomes. Following the cue-utilization framework as conceptualized by Koriat (1997) and interpreted by De Bruin & Van Merriënboer (2017), it is crucial to determine whether monitoring judgments mediate this relationship. In case of a negative indirect effect between perceived mental effort and learning outcomes, we would interpret this as perceived mental effort being used as a negative cue for perceived learning, or confidence herein, which in turn are positively related to learning outcomes. Hence, using a meta-analytic approach, the current study investigated the relationship between learners' perceived mental effort, monitoring judgments, and learning outcomes.

Importance of Accurate Monitoring in Self-Regulated Learning

Self-regulated learning is a crucial skill in education. Learners are exposed to learning environments where they must plan, monitor, and regulate their learning process. They have to make regulatory decisions according to their learning outcomes and comprehension and, for example, decide whether to re-read a text or choose which learning strategy to use for a specific task. For optimal behavioral regulation and adaptation in accordance with learning, it is imperative that learning progress is accurately monitored (De Bruin & Van Merriënboer, 2017; Dunlosky & Rawson, 2012).

Learners often cannot access their learning outcomes and knowledge directly and, therefore, usually make inferences about their learning outcomes based on indirect information or cues (Koriat, 1993, 1997). According to the cue-utilization framework (Koriat, 1997), three elements are involved in self-regulated learning: cues, monitoring judgments, and actual learning or performance. Learners utilize cues arising from their study experiences or materials to formulate a monitoring judgment, which is called cue-utilization (Koriat, 1997). The extent to which cues predict learning outcomes and are, therefore, diagnostic is termed cue diagnosticity (Brunswik, 1956). The extent to which a monitoring judgment accurately predicts learning outcomes is often coined monitoring accuracy (Butler & Winne, 1995; Nelson, 1984; Schraw, 2009). Often, learners utilize cues to self-regulate their learning process, which are not indicative of actual learning outcomes, meaning that the cues are not diagnostic. This use of non-diagnostic cues thus leads to inaccurate monitoring (De Bruin et al., 2017; Koriat, 1993, 1997; Thiede et al., 2010).

Perceived Mental Effort as a Cue for Monitoring

Learners use various cues such as processing fluency of the learning process, previous task-specific experiences, or perceived mental effort experienced during learning to inform their metacognitive monitoring (e.g., Baars et al., 2020; Koriat & Ma'ayan, 2005; Koriat et al., 2014a, b; Onan et al., 2022; Undorf & Erdfelder, 2011). In the memorizing effort heuristic, Koriat et al. (2006) suggested that perceived mental effort, or invested study time, is often utilized as a cue for metacognitive judgments. That is, learners operate under the assumption that the items they learned with less mental effort exertion are easier to recall compared to those that required more effort (the "easily-learned-easily-remembered" heuristic; Koriat et al., 2006). Here, interpretation is driven by learners' own perception of effort, which is also referred to as a data-driven interpretation of effort (Koriat et al., 2014a, b). This results in a negative relationship between perceived mental effort and monitoring of learning outcomes (e.g., Baars et al., 2013, 2018a, b, 2020; Koriat et al., 2009; Undorf & Erdfelder, 2011).

Interestingly, research indicates that the correlation between mental effort and monitoring does not necessarily need to be negative (Baars et al., 2020; Koriat, 2008; Koriat et al., 2006, 2014b). For example, if learners value learning about a topic or have a salient goal to achieve when investing mental effort, there is a positive or zero-order relationship between mental effort and monitoring judgments, meaning that more mental effort is related to higher perceived learning or certainty of knowing (in case of a positive association). Here, the metacognitive judgments are regarded as goal-driven, and learners invest mental effort in accordance with the importance or interest of the to-be-learned materials (Koriat et al., 2014a, b). So far, goal-driven self-regulation has been manipulated by inducing time pressure, providing incentives (e.g., Ackerman, 2014; Koriat et al., 2006), or by increasing learners' sense of agency by changing the wording of the mental effort self-rating scale such as asking about the invested mental effort instead of the required mental effort (Klepsch & Seufert, 2021; Koriat et al., 2014a, b).

The finding that the correlation between mental effort and monitoring does not necessarily need to be negative is supported by the meta-analysis by Baars and colleagues (2020). While they found an overall negative, medium-sized correlation between effort and monitoring judgments (r = -0.35), their moderator analysis results showed that this association was no longer significant for studies employing goal-driven manipulations, such as incentives, time pressure, or promoting feelings of self-agency (Baars et al., 2020). While the response to goal-driven manipulations and improvement in cue-utilization has been shown to develop with increasing age (Hoffmann-Biencourt et al., 2010; Koriat et al., 2009; Koriat et al., 2014b), Baars and colleagues (2020) did not identify learners' age or level of education as significantly moderating the relationship between effort and monitoring judgments. Contrastingly, Baars and colleagues (2020) identified type of task as a significant moderator with a higher negative correlation in studies using problem-solving tasks compared to other tasks such as word learning or paired associates.

An additional factor potentially influencing the relationship between perceived mental effort and monitoring judgments could be task difficulty (Seufert, 2018). During tasks that are more difficult learners might have fewer available cognitive resources to accurately monitor their learning process compared to easier tasks (Seufert, 2018). It is therefore interesting to explore whether task difficulty moderates the relationship between perceived mental effort and monitoring. Collectively, these studies suggest that perceived mental effort might be used as a cue to self-regulate one's learning. Depending on students' effort interpretation (data- or goal-driven) perceived mental effort may have positive or negative effects on students' monitoring of learning, and thus on their learning outcomes. To unravel the potential relationships between perceived mental effort, monitoring, and learning outcomes, the current study aimed to provide more insight into how perceived mental effort is used as a cue to monitor learning, how that, in turn, affects learning outcomes, and what factors moderate these relationships.

Monitoring Judgments and Learning Outcomes

Learners often face challenges in accurately monitoring and regulating their learning without supplementary aid (e.g., Baars et al., 2013, 2014; De Bruin et al., 2011; Dunlosky & Lipko, 2007; Prinz et al., 2020a; Van Gog et al., 2020), primarily due to a tendency to overestimate their own learning outcomes (Baars & Wijnia, 2018; Baars et al., 2013, 2014). This is potentially due to the additional mental effort that is required to simultaneously fulfill the task demands *and* monitor learning outcomes (Van Gog et al., 2011a, 2011b, 2020) as well as using the wrong cues to infer learning (Koriat, 1997). Especially in complex tasks, learners tend to inaccurately monitor their learning outcomes compared to less complex tasks (Baars et al., 2018a, b). Several generative strategies have been shown to increase learners' monitoring accuracy. In particular, activities such as generating keywords, summaries, or diagrams when reading a text, have been indicated to improve learners' monitoring accuracy because they aid learners in gaining access to more diagnostic cues (for a review, see Prinz et al., 2020b).

To measure the accuracy of a monitoring judgment in relation to learning outcomes, several measures of relative or absolute accuracy can be used (see Schraw, 2009). Relative monitoring accuracy indicates the degree to which learners can discriminate between well-learned and less well-learned materials (Maki & Berry, 1984; Nelson, 1984) and is the relationship of interest in this manuscript. Relative accuracy is typically calculated by a correlational measure (Schraw, 2009). The higher the positive correlation between a learner's estimation of learning outcomes and their actual learning outcomes, the higher their monitoring accuracy. A low positive correlation coefficient indicates low monitoring accuracy. A negative relationship between a learner's monitoring judgment and their actual learning outcomes indicates inaccurate monitoring.

Absolute accuracy refers to the discrepancy between a monitoring judgment and actual learning outcomes and indicates the extent to which learners are able to accurately estimate their absolute level of learning outcomes (Schraw, 2009). If a learner estimates their learning outcomes higher than their actual learning outcomes, overconfidence is observed. Underconfidence is observed when learning outcomes exceed learners' estimation of their learning outcomes.

Monitoring judgments are usually assessed using various self-report items with different scales (e.g., from 1 to 5, 1 to 7, or in percentages from 0 to 100%; Dunlosky & Metcalfe, 2008). There can also be variations in the timing of monitoring judgments. They can be examined prospectively, concurrently, or retrospectively (Schraw, 2009). For example, prospective judgments such as ease of learning judgments (EOLs) aim to measure learners' predictions of how easy it will be to learn information, while judgments of learning (JOLs) aim to measure learners' prediction of how well they think they will recall just-learned information at a later time. These judgments are usually made prior to the task that is judged. Concurrent judgments are made whilst performing the task. Commonly, they are administered on an item-by-item level immediately after learners answer an item of the to-be-judged task (Schraw, 2009). Examples are online confidence ratings, ease of learning judgments, or online learning outcome predictions investigating how confident a learner is in their learning outcomes, how easy it was to work on the task, or how accurate they estimate their learning outcomes (Schraw, 2009). Retrospective judgments are usually made after fully completing the task that is judged. They can refer to an item-by-item level but also to a global task level. Examples are similar to the concurrent judgments (i.e., confidence ratings, ease of learning judgments, or learning outcome accuracy judgments) only that they are administered after all items of the to-be-judged task are completed (Schraw, 2009).

While the type and timing of monitoring judgments significantly vary across the field, dictated by the research question, study design, and target population, it has been established that the timing of these judgments directly impacts their accuracy (Leonesio & Nelson, 1990; Siedlecka et al., 2016). When comparing the accuracy of prospective and concurrent[1] confidence judgments, Siedlecka and colleagues (2016) found that the relationship between confidence and learning outcomes was weaker for prospective confidence ratings. A potential reason for this could be that a learner has less available information to base their metacognitive judgment on compared to judgments that are given concurrently (Siedlecka et al., 2016). Similarly, Baars and colleagues (2020) found that both type and timing of monitoring judgments significantly moderated the relationship between effort and monitoring. They found a weaker correlation for studies administering prospective JOLs compared to other judgments and a weaker correlation for studies administering prospective compared to concurrent judgments (Baars et al., 2020).

Mental Effort and Learning Outcomes

Previous work has established that mental effort is related to monitoring and could thus potentially be used as a cue (Baars et al., 2020). Yet, it remained unclear to what extent perceived mental effort might be a diagnostic cue for monitoring. Mental effort is an essential indicator of cognitive load (Paas et al., 2003; Sweller et al., 1998, 2019), and effortful processing is necessary for explicit learning (e.g., Bjork & Bjork, 2011). A positive relationship exists between effortful processing and learning outcomes as long as the learning task requirements are within the limits of working memory capacity and the cognitive effort necessary for processing and integrating information is directly relevant to learning. However, when the effort required to handle the learning task surpasses these boundaries, learning falls short, and the relationship between effortful processing and task performance becomes negative. For tasks within these cognitive limits, and depending on the complexity of the task, various strategies exist to further enhance learning (Chen et al., 2018, 2023). For complex tasks, this can be done by substituting irrelevant information for the learning task with relevant information (Paas & Van Merriënboer, 2020; Sweller

[1] Although Siedlecka et al. (2016) indicated to have used retrospective judgments, in line with Schraw (2009), we would operationalize them as concurrent judgments.

et al., 2019). For less complex tasks, relevant information can be added to the task (i.e., increased intrinsic cognitive load) or learners can be encouraged to invest more effort into available relevant information (i.e., leveraging desirable difficulties; De Bruin et al., 2023).

Next to mental effort, mental load and task performance are also discussed as important indicators of cognitive load in the literature (Paas & Van Merriënboer, 1994). Mental load is commonly seen as learners' cognitive resources, which are required to meet the task affordances in a bottom-up manner and is usually independent of learners' characteristics such as prior knowledge or expertise, increasing with task complexity. This is often also referred to as data-driven effort (Koriat et al., 2006) which learners experience passively (Seufert, 2018; for an overview, see Scheiter et al., 2020). Mental effort involves the active investment of learners' resources to fulfill the demands of a task, and this is shaped by the learner's attributes, the nature of the learning task, the context of the learning environment, as well as the interplay among these factors (Choi et al., 2014; Paas & Van Merriënboer, 1994). Additionally, mental effort can have a motivational component (Feldon et al., 2019; Paas et al., 2005; Scheiter et al., 2020) and is also referred to as goal-driven effort (Koriat et al., 2006). For example, based on learners' estimated success in reaching certain learning outcomes they might be willing to invest more or less effort when working on a task.

Learning outcomes are influenced by the mental load and mental effort a task elicits and learners' individual characteristics such as their prior knowledge or willingness to invest mental resources (Paas & Van Merriënboer, 1994). For example, if task complexity is high, a task imposes high mental load on the learner and more mental effort investment is necessary to perform well on the task, compared to a less complex version of the task. Learners have to take into account their self-assessed learning outcomes to regulate their effort investment accordingly and, for example, invest additional mental effort, switch to a different learning strategy, or ask for external help (De Bruin et al., 2023).

Mental effort can be measured in multiple ways. On the one hand, there are physiological measures such as heart rate variability, pupil dilation, or skin conductance rate (for an overview, see Ayres et al., 2021), or behavioral measures such as response time or time on task. On the other hand, there are subjective measures using self-report scales. In the current study, we are interested in perceived mental effort due to it being a salient cue that learners might experience during their self-regulated learning (Bruin et al., 2023). The subjective character usually represents learners' self-assessed experience of mental effort and thus provides valuable insights into learners' experiences during learning. A commonly used item in educational science is the Paas-scale (Paas, 1992), in which learners are asked to judge the amount of mental effort they invested on a 9-, 7-, or 5-point Likert scale ranging from "very, very low mental effort" to "very, very high mental effort" (for an overview see Paas et al., 2003). This scale, proven to be user-friendly, valid, and reliable (e.g., Paas et al., 1994), exhibits even higher construct validity than physiological measures, as suggested by the recent review by Ayres et al. (2021).

While the timing of administering these ratings varies in the field, it has been indicated that timing influences the ratings and their predictability of learning outcomes

(Schmeck et al., 2015; Van Gog et al., 2012). Compared to an average of mental effort ratings administered multiple times during a learning task, a single delayed mental effort rating at the end of a series of tasks was higher (Schmeck et al., 2015; Van Gog et al., 2012). This suggests that timing or reference level of mental effort ratings influence the relationship between perceived mental effort and learning outcomes and thus the diagnosticity of mental effort as a cue.

Relationship Between Mental Effort, Monitoring, and Learning Outcomes

According to models of self-regulated learning (e.g., Panadero, 2017), monitoring processes inform regulation processes which in turn affect learning outcomes. Earlier research has shown that mental effort is related to monitoring (e.g., Baars et al., 2020) which suggests that mental effort can be used as a cue for monitoring. This means that, theoretically, mental effort can have an indirect effect on learning outcomes via monitoring processes. That is, learners can use their perceived mental effort to make monitoring judgments, which in turn affect regulation of learning and learning outcomes.

Evidence exists that mental effort, monitoring judgments, and learning outcomes are related in learning tasks. Blissett and colleagues (2018) investigated the triangle between mental effort, monitoring, and learning outcomes in medical reasoning. They investigated the cue-utilization of perceived mental effort, how diagnostic mental effort is as a cue for monitoring, and the monitoring accuracy of certainty judgments. Their findings indicated that mental effort is used as a cue for certainty judgments. Furthermore, they identified mental effort as a diagnostic cue: Higher perceived mental effort was associated with lower learning outcomes. Additionally, they found that monitoring was moderately accurate as higher certainty was related to higher learning outcomes. While this study gives a first insight into the relationship between perceived mental effort, monitoring, and learning outcomes, it is unclear what this relationship looks like when combining multiple studies and when taking into account possible moderators. Knowing to what extent perceived mental effort relates to learners' learning outcomes will indicate whether perceived mental effort is a diagnostic cue for learning outcomes and whether and when learners should use their perceived mental effort as a basis to self-regulate their learning. Also, to understand the role of mental effort in self-regulated learning, it is important to investigate the indirect relationship between mental effort and learning outcomes via monitoring. Given the specific relationships of interest described above and the availability of multiple studies investigating these, a meta-analysis is a promising approach to this question.

The Present Study

Using a meta-analytic approach, we investigated the relationship between perceived mental effort, monitoring judgments, and learning outcomes. Our goal was to take a triangular approach to the cue-utilization framework (De Bruin et al., 2017; Koriat, 1997). Therefore, we specified perceived mental effort as a cue, monitoring of learning judgments as monitoring judgments, and learning outcomes as an approximation of actual learning or performance to understand how they collectively contribute to metacognitive processes. We analyzed the relationships simultaneously by including the paths and variables of interest in a comprehensive model. We first tested whether we could replicate a medium, negative relationship between perceived mental effort and monitoring judgments (i.e., cue-utilization; Baars et al., 2020). We expected a negative relationship between mental effort and monitoring judgments (of perceived learning; Hypothesis 1). For example, we expected that if a learner judges their perceived mental effort to be high, they will judge their perceived learning (reflected by the monitoring judgment) as low. Second, we investigated whether there is a positive relationship between monitoring judgments and learning outcomes (i.e., monitoring accuracy). We expected a positive relationship between monitoring judgments and learning outcomes (Hypothesis 2). For example, we expected that if a learner judges their perceived learning (reflected by the monitoring judgment) as high, learning outcomes during the task would also be high. Third, we wanted to explore whether the relationship between perceived mental effort and learning outcomes (i.e., cue diagnosticity) is mediated by monitoring judgments. We expected monitoring judgments to mediate the relationship between mental effort and learning outcomes (Hypothesis 3) due to the central role of monitoring in self-regulated learning. A negative indirect effect (with a negative direct effect between perceived mental effort and monitoring judgments and a positive direct effect between monitoring judgments and learning outcomes) would indicate that when learners experience high mental effort, this is associated with lower feelings of learning or confidence, which in turn would also relate to actual lower learning. Furthermore, as we expected that certain sample and task characteristics would influence the magnitude of the relationship between mental effort, monitoring judgments, and learning outcomes, we conducted various subgroup analyses to investigate moderating effects of learners' level of education, the formulation and timing of mental effort measurements, types and timing of monitoring judgments, type of task, task difficulty, and whether goal-driven manipulations were used (Hypothesis 4).

Method

Transparency and Openness

We pre-registered our hypotheses, procedure, coding scheme, and analysis plan on the Open Science Framework (OSF) website https://osf.io/u4dgt (pre-registration). Our data set can be found on OSF via https://osf.io/hc5fk/.

Inclusion Criteria

1. *Document type*. We included published, peer-reviewed texts as well as unpublished texts that were written in English (i.e., grey literature such as conference proceedings, dissertations, and master theses). In addition, unpublished data were included.
2. *Sample size.* We only included samples with $N > 10$ per independent sample to make sure to include relatively robust studies. If an article reported multiple independent samples, we coded these samples separately.
3. *Participants.* We only included samples of participants who were healthy and had no psychological variations (e.g., ADHD, autism) and were not under the influence of a severe pharmacological or psychological treatment (e.g., drug administration, induction of stress or mental fatigue, sleep deprivation; e.g., Sugden et al., 2012).
4. *Mental Effort.* We focused on perceived mental effort and, therefore, excluded studies that only reported other mental effort indicators such as response time, time on task, or physiological measures such as pupil dilation (e.g., Huh et al., 2019; Koriat & Nussinson, 2009).

Method

We only included studies in which the (cor)relation between perceived mental effort, monitoring judgments, and learning outcomes and the sample size were reported or received after a request via email to the corresponding author. Also, only studies were included in which perceived mental effort, monitoring judgments, and learning outcomes were measured on a quantitative scale in the context of a learning outcome of a learning task. Furthermore, perceived mental effort, monitoring, and learning outcomes had to be measured in one study or experiment in the same trial for the same item or criterion task for a study to be included. Between-subjects conditions were coded separately, but if perceived mental effort, monitoring, and learning outcomes were measured multiple times under varying conditions for the same group of participants (i.e., within-subjects design), we did not code these conditions separately but used the mean correlation in our analysis.

Systematic Search Strategy

We searched Maastricht University's Web of Science (WOS) Core Collection, PubMed, and the CINAHL, ERIC, LISTA, APA PsycInfo, SocINDEX, and OpenDissertations databases within EBSCOhost. The WOS Core Collection included the Science Citation Index Expanded (1988–present), Social Sciences Citation Index (1988–present), Arts & Humanities Citation Index (1988–present), Conference Proceedings Citation Index – Science (1990–present), Conference Proceedings Citation Index – Social Science & Humanities (1990–present), and the Emerging Sources Citation Index (2018–present). We searched the full text of articles and used the following search terms (ALL ("mental effort" OR "perceived effort" OR "subjective effort" OR "experienced effort" OR "cognitive effort" OR "cognitive load" OR "mental load" OR "germane load" OR "intrinsic load" OR "extraneous load" OR "working memory load")) AND (ALL (monitor* OR "judg* of learning" OR "confidence judg*" OR "confidence rating*" OR "metacognit* judg*" OR "latency-confidence" OR "perceived learning")) AND (ALL (learning OR "self regulat*" OR "metacogniti*" OR "accuracy" OR "diagnostic*" OR "perform*" OR "learning outcome" OR "cue utili*")). We restricted the results to only including documents published from 2000 onwards to maintain the relevance of results. Additionally, the influential cue-utilization framework (Koriat, 1997) underlying the majority of relevant studies was published in 1997, which is why we did not expect many relevant studies before 2000. Additionally, by limiting our search to include primarily recent research, we anticipated a greater homogeneity in the methodologies and technologies used across the studies. Our search yielded 1691 total hits, with 735 hits from WOS, 605 hits from EBSCO, and 351 hits from PubMed. More details can be found in a PRISMA flow chart adapted from Page and colleagues (2021) in the supplementary material.

We applied our inclusion criteria in two steps. First, we screened the articles, checking whether the article reported mental effort ratings, monitoring judgments, and a learning outcome measure using the website Rayyan.ai (Ouzzani et al., 2016). In the second step, we checked the remaining inclusion criteria of the articles that measured the three variables.

Included Variables and Effect Sizes

Our key outcome of interest was participants' learning outcome. We operationalized the variable as the actual learning or performance on a learning task, such as the number of correct answers on a test (e.g., Mihalca et al., 2017) or number of correct solution steps in a problem-solving task (e.g., Baars et al., 2014). The main independent variables were perceived mental effort and monitoring judgments. Both variables were measured with self-report items. Perceived mental effort, which usually represents learner's self-assessed experience of mental effort, was usually measured using the Paas-scale (Paas, 1992), in which learners are asked to judge the amount of mental effort they invested on a 9-, 7-, or 5-point Likert scale ranging from "very, very low mental effort" to "very, very high mental effort." Monitoring judgments, which usually are a form of learners' self-assessment of their learning, such as judgments of learning or confidence ratings, are often measured by asking participants to indicate how likely they are to remember studied information on a later test. Often these monitoring judgments, such as JOLs, feeling of knowing, or ease of learning judgments, are answered by indicating one's estimated likelihood to later recall the information on different scales (e.g., from 1 to 5, 1 to 7, or in percentages from 0 to 100%; Dunlosky & Metcalfe, 2008). As mentioned above, the type and timing of monitoring judgments and perceived mental effort ratings can vary.

To perform the meta-analysis, we collected the Pearson's correlation coefficient between perceived mental effort, monitoring judgments, and learning outcomes as reported by other researchers. If a study measured various types of the variables of

Table 1 Overview of Moderator Variables

Moderator		Description (number of studies)
1	Education	Higher & Post Education (165); Other levels of education (55)
2	Mental Effort Measurement	Mental effort rating (167); Other types of effort ratings (69)
3	Mental Effort Wording	Active (166); Other (70)
4	Effort Reference	Item-by-item (146); Global task (90)
5	Type of Monitoring	JOL (169); Other (67)
6	Monitoring Timing	Prospective (157); Concurrent (76)
7	Monitoring Reference	Item-by-item (125); Global task (111)
8	Type of Task	Problem-solving (98); Other tasks (138)
9	Task Difficulty	Not specified (157); Other (79)

The Description column reports all moderator categories that we included in our analysis. To our knowledge, the analytic approach we chose only allows for subgroup analysis, which is why we merged categories with small numbers of effect sizes per moderator to create two groups. Certain moderators did not have enough variability within their categories, which is why we decided against analyzing them. The categories and excluded moderators are not included in this table but are described in the main text. The number in brackets represents the number of effect sizes in the category and is referred to as *k* in the main text.

interest such as ease of learning and JOLs (e.g., Beege et al., 2021), we chose the type that was most commonly used in other studies, which were JOLs. We coded between-subjects conditions as independent samples (e.g., in the study by Baars et al. (2014) the learners in the self-assessment training condition and learners in the no self-assessment training condition were coded as separate samples) and coded all variables on the task level (i.e., each data point reflected a group of people that carried out the same task). If the correlations were not reported in the paper, we contacted the corresponding authors. We extracted or received the Pearson's correlation coefficient and accompanying sample size from 35 papers, with 83 independent samples, and 236 effect sizes. The effect sizes consisted of 83 correlations between mental effort and monitoring, 77 correlations between monitoring and learning outcomes, and 76 correlations between effort and learning outcomes. Based on the correlation coefficient and sample size, we estimated the sampling variances with the rma.mv function of the metafor package (Viechtbauer, 2010) in R (R Core Team, 2023).

Coding

Procedure

Table 1 represents the clustering of moderators we used in our analysis. LD, FB, MB, and LW coded the moderators and control variables of the articles. To facilitate the reproducibility of the coding, we created a detailed coding scheme before data extraction. For calibration purposes and to examine the coding scheme's fit to the articles, all coders coded the same five articles. Coding differences were discussed and resolved, and when necessary, the coding scheme was adjusted or the interpretation of coding categories was further clarified. After this calibration phase, 49 samples (corresponding to 59% of the total data) were coded by two raters to estimate the alignment of the coding. To assess inter-rater reliability, we calculated Cohen's k for categorical variables, which ranged from $k=0.30$ (fair) to $k=1$ (perfect) with a mean of $k=0.77$ (substantial; Landis & Koch, 1977). After discussing the results and coming to an agreement and included additional examples in our coding scheme to improve clarity. Our final coding scheme is available at OSF.

Moderators

Level of Education If reported, we coded the sample's average level of education based on the authors' categorization. The majority of measurements came from samples in higher education ($k=152$), followed by 9th grade ($k=26$), 8th grade ($k=18$), postgraduate education ($k=13$), 3rd grade ($k=6$), and 4th grade ($k=5$). For 13 effect sizes, no level of education was mentioned ($k=13$). If there were participants from multiple levels of education included in the study, then we coded the level of education of the majority of participants. If participants were spread across levels of education, then we coded the sample's mean level of education. For the subgroup analysis, we compared effect sizes from samples within higher or postgraduate education to samples from all other levels of education.

Mental Effort Measurement We coded the type of mental effort measurement. The majority of measures were conducted using a mental effort rating such as the Paas-scale ($k=167$; Paas, 1992) followed by effort ratings ($k=42$), other ratings ($k=22$), and judgments of difficulty ($k=5$). We coded effort ratings as ratings in which learners were not specifically asked about their mental effort but for example their study efforts (e.g., Koriat, 2018). We coded "other" if for example participants were asked how tiring an exercise was (e.g., Kirk-Johnson et al., 2019). For the subgroup analysis, we compared effect sizes from samples measured via mental effort measurements to samples measured by all other types of ratings.

Wording of Mental Effort Judgments We coded the wording of the mental effort measurement. The majority of measurements used active wording ($k=166$). We coded the wording as active if the formulation would suggest that the learner had agency over his effort expenditure (Koriat et al., 2014a, b), such as "investing" or "putting". If the authors mentioned having used the Paas-scale from 1992, this was coded as active as the Paas-scale uses the formulation "invest". Passive ($k=42$) would be words that indicate that the learner did not have agency but that the task requirements were guiding the effort expenditure (Koriat et al., 2014a, b), such as "requiring" or "costing". For the remaining measures ($k=28$), the wording was not specified. For the subgroup analysis, we compared effect sizes from samples measured with active wording compared to samples measured via passive or not specified wording.

Reference of Mental Effort Judgments We coded to which level the mental effort rating referred. The majority of the effort judgments referred to an item ($k=146$), meaning that the judgments were made for different items of a task. In 90 cases, the judgments were made on a global task level ($k=90$).

Types of Monitoring Judgments Based on the authors' categorization, we coded what kind of monitoring judgments were measured. The majority of studies used judgments of learning ($k=169$), followed by confidence ratings ($k=46$), and other types of judgments ($k=21$), such as a learner's self-assessment (e.g., Baars & Wijnia, 2018). For the subgroup analysis, we compared effect sizes measured with JOLs compared to all other types of judgments.

Timing of Monitoring Judgments We coded the timing of monitoring judgments. We coded judgments as prospective if the judgments of learning/performance occurred prior to performing the task that was judged (i.e., predictions; $k=157$). We coded judgments as concurrent if judgments were made during the task that was judged ($k=76$). We coded judgments as retrospective if judgments were made after completing the entire task that was judged ($k=3$). For the subgroup analysis, we compared effect sizes measured with prospective judgments compared to concurrent judgments.

Reference of Monitoring Judgments We coded to which level the monitoring judgment referred. The majority of the monitoring judgments referred to an item ($k=125$), meaning that the judgments were made for different items of a task. In 111 cases, the judgments were made on a global task level ($k=111$).

Type of Task We coded the type of tasks participants had to work on. The majority of tasks were problem-solving tasks ($k=98$), followed by other types of tasks ($k=55$), text comprehension tasks ($k=45$), image learning tasks ($k=30$), and word learning tasks ($k=8$). For the subgroup analysis, we compared effect sizes measured during problem-solving tasks compared to effect sizes measured during all other types of tasks.

Task Difficulty Based on the authors' categorization, we coded the task difficulty. In the majority of cases, task difficulty was not specified ($k=157$), in 56 cases the authors indicated using varying levels of difficulty ($k=56$), followed by difficult tasks ($k=15$), other levels of difficulty ($k=5$), or easy tasks ($k=3$). For the subgroup analysis, we compared effect sizes measured during tasks with not specified task difficulty, compared to effect sizes measured during all other types of task difficulty.

Goal-Driven Manipulations We coded to what extent goal-driven manipulations were used. The majority of cases did not include goal-driven manipulations ($k=220$). In 16 cases, a goal-driven manipulation was present ($k=16$). Examples of goal-driven manipulations were, whether there was a time limit/time pressure to conduct a task (e.g., Koriat, 2018), if researchers tried to promote learner's sense of agency (e.g., Koriat et al., 2014a, b), or tried to promote learner's effort investment (via, for example, instructions; e.g., Onan et al., 2023).

Analytic Approach

We used a meta-analytic structural equation model (MASEM) to analyze our data. All analyses were performed using RStudio (Version 4.2.1; RStudio Team, 2023). As articles reported multiple studies with multiple samples, our data had a nested structure. We considered this dependency by following the WPL approach (Stolwijk et al., 2022; Van den Noortgate et al., 2013; Wilson et al., 2016). Using this approach, we first estimated the synthesized correlation matrix using a three-level hierarchical model. First, to account for dependency amongst effect sizes within each article, we included the article number in our random effects structure. This level assumes that scores within one article can be more similar than scores from other articles. Second, to account for a dependency amongst scores within the same sample, we included the sample ID. This level assumes that scores within one sample are more similar than scores across samples. Third, we added a unique identifier of each task within each independent sample to our random-effects structure. Due to the limited number of articles that reported effect sizes from multiple studies ($k=8$), we did not specify this as an additional level. Thus, our random-effects structure was specified as "~1 | ArticleID/SampleID/MeasureID."

Similar to Stolwijk et al. (2022), we used a random-effects-no-intercept model with maximum likelihood estimation to compute the unadjusted synthesized correlation coefficients for the pooled correlation matrix using the metafor package (Version 4.2.0; Viechtbauer, 2010). We then used this pooled correlation matrix as well as the asymptotic covariance matrix for the Stage 2 analysis.

Based on our hypothesized model, we first specified a full mediation model with mental effort as predictor, monitoring as mediator, and learning outcomes as outcome variable. We then compared the model fit to a partial mediation model. We assessed model fit by comparing the models using the chi-squared difference test ($\Delta\chi 2$), considering a significance level of $\alpha=0.05$ to indicate a substantial difference between the two models. Furthermore, we used additional fit indices like RMSEA and CFI. RMSEA values and their 95% confidence intervals were interpreted following the benchmarks for acceptable (RMSEA ≤ 0.08) and good (RMSEA ≤ 0.05) model fit by Hu & Bentler (1998). CFI values exceeding 0.95 suggest a reasonable fit (Schermelleh-Engel et al., 2003). We conducted our moderator analysis, by using the subgroup analysis approach (Jak & Cheung, 2018) to investigate for each moderator separately whether the parameter estimates differed per subgroup. We did the Stage 2 analysis and the moderator analyses via the metaSEM package (Version 1.3.0; Cheung, 2015).

We interpreted the magnitude of our effect sizes following the rule of thumb by Keith (2015) who established this based on their expertise of effects on learning outcomes. For the direct effects, we considered betas above 0.05 as small, betas above 0.10 as moderate, and betas above 0.25 as large. We interpreted the size of indirect effects by considering that indirect effects are the product of two small, moderate, or large effects. This consideration leads to values of $0.05^2 = 0.003$ (small), $0.10^2 = 0.01$ (moderate), and $0.25^2 = 0.063$ (large) for interpreting the magnitude of the indirect effects.

Deviations from Pre-Registration

We initially planned to use a one-stage meta-analytic structural equation (OMASEM) mediation model (Jak et al., 2021) to test the mediation path model. To our knowledge, the OMASEM approach does not allow the inclusion of multiple effect sizes per sample since we can then no longer assume independence between effect sizes. Therefore, in cases of multiple effect sizes of the same sample (i.e., within-subjects design), we planned on aggregating the effect sizes to a mean effect size. In case of separate effect sizes for non-overlapping subsamples (i.e., between-subjects design), we planned to treat them as separate studies in the analysis and thus ignore dependency. However, Stolwijk et al. (2022) proposed a superior method to deal with dependency (i.e., the multilevel approach) which can be used by following a two-stage meta-analytic structural equation mediation model. We thus, followed that approach, which allowed us to take into account the non-overlapping subsamples.

Furthermore, we planned to analyze the moderator goal-driven manipulations. However, it turned out that, in most studies, there was no goal-driven manipulation ($k = 220$). Only 16 measures were taken under a goal-driven manipulation. Therefore, we decided against analyzing this moderator.

Additionally, we coded whether a regulation strategy was used in the included studies and planned to explore the relationship between mental effort, monitoring judgments, and regulation strategies. Using the same analysis approach as for our first research question, we intended to specify mental effort as the predictor, monitoring judgments as the mediator, and instead of learning outcomes, regulation behavior as outcome variable. Due to the low number of studies that measured regulation behavior ($k = 12$), we did not analyze this relationship or the exploratory moderator type of regulation strategy.

Results

Study Descriptives

The current sample consisted of data from 35 manuscripts, with 83 independent samples, 236 effect sizes, and a total sample size of $N = 3973$. Of the studies, 97.1% reported multiple effect sizes. Most studies were conducted in Europe ($k = 154$), followed by Asia ($k = 26$), North America ($k = 23$), online ($k = 21$), and Australia ($k = 12$). The median age of the overall sample was $Mdn = 22.32$, with ages ranging from 9 to 70.8 years.

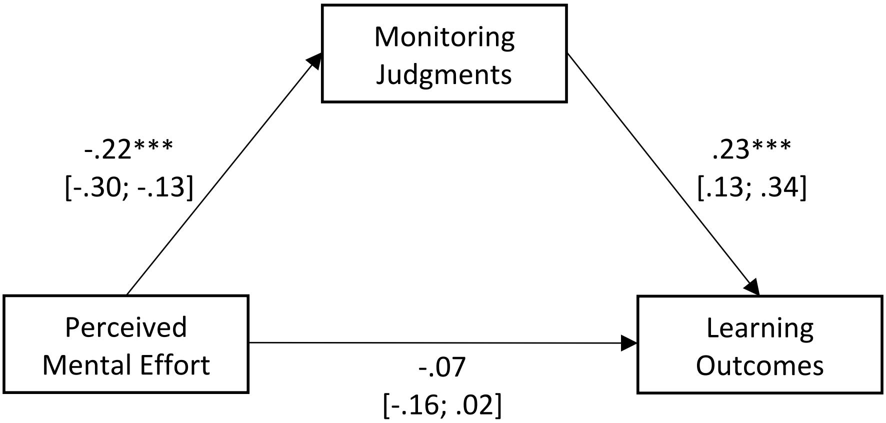

Fig. 1 Partial Mediation Model. The figure presents the partial mediation model, which was not indicated as having a significantly better fit compared to the more parsimonious full mediation model. The standardized regression coefficients of the direct effects specified in the model are indicated

Main Analyses

To investigate our first three research questions, we specified a full mediation model as well as a partial mediation model with perceived mental effort as predictor, monitoring judgments as mediator, and learning outcomes as outcome variable. We specified the direct paths between mental effort and monitoring judgments and monitoring judgments and learning outcomes. Additionally, we specified the indirect effect from mental effort to learning outcomes via monitoring judgments. For the partial mediation model, we added the direct effect between mental effort and learning outcomes (for the differences between the models, please see Figs. 1 and 2). The results of the partial mediation model are included in Fig. 1. The model suggests a negative, statistically significant parameter estimate between mental effort and monitoring judgments $\beta = -0.22$, 95% LBCI [-0.30; -0.13], a positive, statistically significant parameter estimate between monitoring judgments and learning outcomes $\beta = 0.23$, 95% LBCI [0.13; 0.34], and a negative, statistically non-significant parameter estimate between mental effort and learning outcomes $\beta = -0.07$, 95% LBCI [-0.16; 0.02]. The goodness of fit indices of the model indicate that the model is saturated and overfitted and should thus be interpreted with caution, $\chi^2(0) = 0$, $p = 0.00$, with a RMSEA of 0.00, 95% CI [0.00; 0.00], and the CFI of 1. Due to the saturated partial mediation model, we specified a more parsimonious full mediation model, with the direct paths between mental effort and monitoring judgments and monitoring judgments and learning outcomes. Additionally, we specified the indirect effect from mental effort to learning outcomes via monitoring judgments. The χ^2 of the hypothesized full mediation model was statistically non-significant, $\chi^2(1) = 2.57$, $p = 0.1$, so the null hypothesis of exact fit cannot be rejected. Also, the RMSEA of 0.01, 95% CI [0.000; 0.051], and the CFI of 0.98 indicated a close approximate fit of the model.

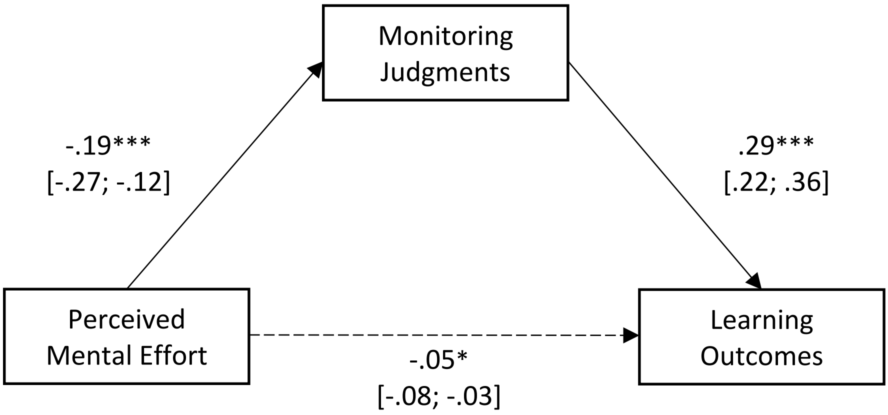

Fig. 2 Full Mediation Model. The figure presents the full mediation model, which was indicated as having good fit indices while not being overfitted. The standardized regression coefficients of the direct effects specified in the model are indicated. Additionally, the indirect effect between mental effort and learning outcomes was specified in the full mediation model and is represented with the dashed arrow. All three paths were significant at $p < 0.05$

Model comparison showed non-significant differences between the full and partial mediation model, suggesting that the saturated partial mediation model was not a significantly better fit than the more parsimonious full mediation model, and we thus continued our analysis with the full mediation model (see Fig. 2). In support of Hypothesis 1, a negative parameter estimate between mental effort and monitoring judgments was found β = -0.19, 95% LBCI [-0.27; -0.12], indicating that with each unit increase in perceived mental effort, monitoring judgments decrease by 0.19 units on average. In support of Hypothesis 2, a positive parameter estimate between monitoring judgments and learning outcomes was found β = 0.29, 95% LBCI [0.22; 0.36], indicating that with each unit increase in monitoring judgments, learning outcomes increased by 0.29 units on average. In support of Hypothesis 3, representing the indirect effect between mental effort and learning outcomes via monitoring judgments, a negative parameter estimate was found β = -0.05, 95% LBCI [-0.08; -0.03]. This indicates that one unit increase in perceived mental effort, on average, was associated with a 0.05 unit decrease in learning outcomes due to the negative effect of perceived mental effort on monitoring judgments, which, in turn, positively related to learners' learning outcomes. Based on the 95% likelihood-based confidence intervals, which did not include 0 for all three coefficients, we concluded that all effects were significantly different from 0 with $p < 0.05$.

Moderator Analysis

We conducted various moderator analyses using subgroup analysis with education (higher and postgraduate education vs other types of education), mental effort measurement (mental effort ratings vs other types of ratings), mental effort wording (active wordings vs other wordings), effort reference (item-by-item level vs global task level), type of monitoring (JOL vs other types of monitoring), monitoring timing (prospective vs concurrent), type of task (problem-solving vs other types of tasks), and task difficulty (not specified vs other difficulty levels) as moderators. We used the full mediation model and investigated whether the subgroup analysis would reveal significant differences in regression coefficients across groups.

Our subgroup analysis revealed no significant differences for any of the moderators at $p<0.05$. The regression coefficients and their corresponding 95% likelihood-based confidence intervals from the full mediation model ran per subgroup can be found in Table 2.

Publication Bias Analysis

We evaluated publication bias using a three-level funnel plot of all effect sizes (Fernández-Castilla et al., 2020). The funnel plot below (Fig. 3) displays the collected effect sizes against their measurement precision, in this case, the standard error. Visual inspections of the plot reveal an asymmetric shape as there are for example few studies with a larger, negative effect size and larger standard error, which indicates that there could potentially be selection or publication bias. We investigated this further by conducting a *p*-curve analysis using the *p*-curve app (Version 4.06; Simonsohn et al., 2014). Via this analysis, it is possible to investigate the likelihood of the findings underlying a true effect or if they are a result of selective reporting. The *p*-curve represents the distribution of statistically significant *p*-values ($p<0.05$). Right-skewed *p*-curves suggest that included studies contain more low (e.g., 0.01 s) than high (e.g., 0.04 s) significant *p*-values and thus provide diagnostic evidence of the presence of evidential value for a true effect. The results of the *p*-curve analysis showed that the distributions of effects underlying the correlations between perceived mental effort, monitoring judgments, and learning outcomes were right-skewed (Fig. 4a; perceived mental effort and monitoring judgements, binomial test $p<0.001$, continuous test: $Z=-19.31$, $p<0.001$; Fig. 4b, monitoring judgements and learning outcomes, binomial test $p<0.001$, continuous test: $Z=-13.74$, $p<0.001$; Fig. 4c, perceived mental effort and learning outcomes, binomial test $p<0.001$, continuous test: $Z=-8.53$, $p<0.001$). The *p*-curves thus indicate evidential value underlying the relationships.

Discussion

In the current study, we aimed to investigate the relationship between perceived mental effort, monitoring judgments, and learning outcomes using a meta-analytic approach. Based on our model comparison, the more parsimonious full mediation model compared to the partial mediation model indicated good fit indices while not being overfitted. Thus, we based our results on the full mediation model. In line with Hypothesis 1, we found a moderate, negative association between perceived mental effort and monitoring judgments ($\beta=-0.19$). This indicates that if a learner judges their perceived mental effort to be high, they will judge their perceived learning

Table 2 Results Moderator Analysis

Moderator	Subgroup	Relationship	β	95% LBCI
Education	Higher/Post Education	Effort-Monitoring	-0.19	[-0.28; -0.10]
		Monitoring-Learning outcome	0.29	[0.21; 0.37]
		Effort-Learning outcome *(Indirect)*	-0.05	[-0.08; -0.03]
	Other Education	Effort-Monitoring	-0.23	[-0.31; -0.15]
		Monitoring-Learning outcome	0.26	[0.17; 0.35]
		Effort-Learning outcome *(Indirect)*	-0.06	[-0.09; -0.03]
Effort Measurement	Mental effort ratings	Effort-Monitoring	-0.19	[-0.28; -0.11]
		Monitoring-Learning outcome	0.29	[0.21; 0.37]
		Effort-Learning outcome *(Indirect)*	-0.05	[-0.08; -0.03]
	Other effort ratings	Effort-Monitoring	-0.19	[-0.34; -0.04]
		Monitoring-Learning outcome	0.29	[0.16; 0.43]
		Effort-Learning outcome *(Indirect)*	-0.05	[-0.10; -0.01]
Effort Wording	Active	Effort-Monitoring	-0.18	[-0.28; -0.08]
		Monitoring-Learning outcome	0.30	[0.21; 0.39]
		Effort-Learning outcome *(Indirect)*	-0.05	[-0.08; -0.03]
	Other wording	Effort-Monitoring	-0.19	[-0.31; -0.07]
		Monitoring-Learning outcome	0.29	[0.19; 0.40]
		Effort-Learning outcome *(Indirect)*	-0.05	[-0.09; -0.02]
Effort Reference	Item-by-item	Effort-Monitoring	-0.20	[-0.29; -0.11]
		Monitoring-Learning outcome	0.29	[0.20; 0.38]
		Effort-Learning outcome *(Indirect)*	-0.06	[-0.08; -0.03]
	Global task	Effort-Monitoring	-0.18	[-0.31; -0.05]
		Monitoring-Learning outcome	0.30	[0.20; 0.41]
		Effort-Learning outcome *(Indirect)*	-0.05	[-0.09; -0.02]
Monitoring Type	JOL	Effort-Monitoring	-0.19	[-0.28; -0.09]
		Monitoring-Learning outcome	0.30	[0.22; 0.38]
		Effort-Learning outcome *(Indirect)*	-0.05	[-0.08; -0.03]
	Other types	Effort-Monitoring	-0.22	[-0.31; -0.12]
		Monitoring-Learning outcome	0.28	[0.19; 0.37]
		Effort-Learning outcome *(Indirect)*	-0.06	[-0.09; -0.03]
Monitoring Timing	Prospective	Effort-Monitoring	-0.19	[-0.29; -0.08]
		Monitoring-Learning outcome	0.30	[0.21; 0.39]
		Effort-Learning outcome *(Indirect)*	-0.05	[-0.08; -0.03]
	Concurrent	Effort-Monitoring	-0.22	[-0.32; -0.12]

Table 2 (continued)

Moderator	Subgroup	Relationship	β	95% LBCI
		Monitoring-Learning outcome	0.27	[0.18; 0.37]
		Effort-Learning outcome *(Indirect)*	-0.06	[-0.09; -0.03]
Monitoring Reference	Item-by-item	Effort-Monitoring	-0.21	[-0.30; -0.13]
		Monitoring-Learning outcome	0.28	[0.19; 0.37]
		Effort-Learning outcome *(Indirect)*	-0.06	[-0.09; -0.03]
	Global task	Effort-Monitoring	-0.17	[-0.28; -0.05]
		Monitoring-Learning outcome	0.31	[0.21; 0.41]
		Effort-Learning outcome *(Indirect)*	-0.05	[-0.08; -0.02]
Task Type	Problem-solving	Effort-Monitoring	-0.18	[-0.28; -0.08]
		Monitoring-Learning outcome	0.30	[0.22; 0.39]
		Effort-Learning outcome *(Indirect)*	-0.05	[-0.08; -0.02]
	Other task types	Effort-Monitoring	-0.19	[-0.29; -0.10]
		Monitoring-Learning outcome	0.29	[0.21; 0.38]
		Effort-Learning outcome *(Indirect)*	-0.05	[-0.08; -0.03]
Task difficulty	Not specified	Effort-Monitoring	-0.20	[-0.29; -0.10]
		Monitoring-Learning outcome	0.29	[0.21; 0.37]
		Effort-Learning outcome *(Indirect)*	-0.05	[-0.08; -0.03]
	Other levels difficulty	Effort-Monitoring	-0.19	[-0.32; -0.06]
		Monitoring-Learning outcome	0.29	[0.17; 0.42]
		Effort-Learning outcome *(Indirect)*	-0.05	[-0.10; -0.02]

In the table, the standardized regression coefficient and their corresponding 95% likelihood-based confidence interval are displayed. These coefficients were based on the full mediation model that was run for each of the subgroups separately and then compared for statistically significant differences in their effects

(reflected by their monitoring judgment) as low. This indicates that perceived mental effort is potentially used as a cue and aligns with earlier research (Baars et al., 2020). In line with Hypothesis 2, we found a large, positive relationship between monitoring judgments and learning outcomes (β = 0.29). This indicates that if a learner judges their perceived learning (reflected by their monitoring judgment) as high, their learning outcomes during the task would also be higher. Similarly to Blissett et al. (2018), we would still interpret this as low to moderate accuracy when monitoring one's learning due to the beta coefficient. However, this interpretation of magnitude should be regarded with caution as relative metacomprehensive accuracy is typically judged via the Gamma or Pearson correlation coefficients (for a recent meta-analysis on relative metacomprehension accuracy, see Prinz et al., 2020a). In line with Hypothesis 3, we found a significant negative, moderate-sized indirect

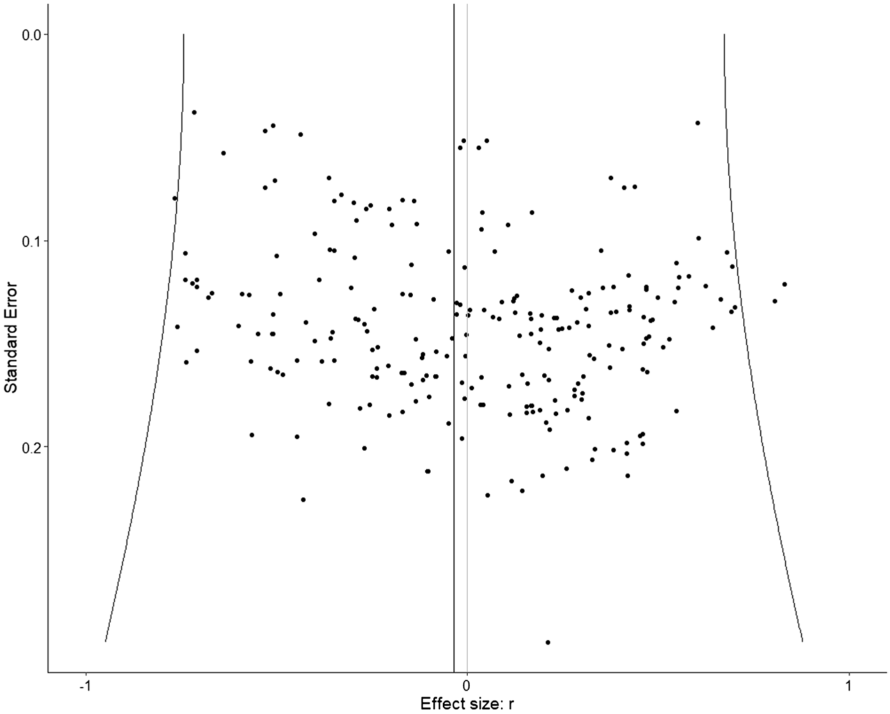

Fig. 3 Funnel Plot of all Effect Sizes. The figure shows a funnel plot displaying the collected effect sizes against their standard error

association between perceived mental effort and learning outcomes (β = -0.05), which was mediated by monitoring judgments. This suggests that when learners perceive higher levels of mental effort, their learning outcomes tend to decrease as mediated by monitoring judgments, indicating the important role of monitoring judgments in explaining the relationship between perceived mental effort and learning outcomes. These findings imply that learners use their perceptions of invested mental effort as a cue when monitoring learning, while the significant indirect negative effect between effort and learning outcomes indicates that mental effort might be indirectly related to actual learning outcomes. More specifically, the negative indirect effect seems to indicate that when learners experience high mental effort, this is associated with lower feelings of learning or confidence, which in turn relates to actual lower learning. When learners experience lower mental effort, this is associated with higher feelings of learning or confidence, which in turn relates to higher actual learning. This suggests that when learners perceive high mental effort, they seem to score lower on the learning outcomes. In line with the cue-utilization perspective, our findings suggest that learners judged their learning outcomes relatively accurately (due to the positive direct effect between monitoring judgments and learning outcomes) to be lower (due to the negative direct effect between perceived mental effort and monitoring judgment) when they experienced high mental effort.

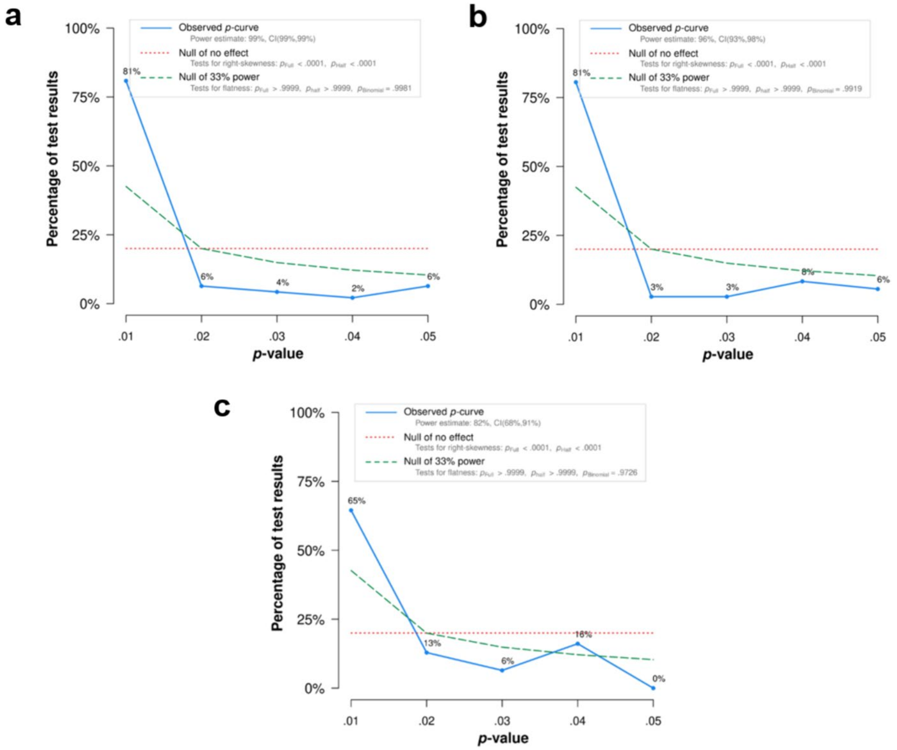

Fig. 4 P-Curve Analysis of the Effect Sizes per Relationship. *P*-curve analysis of studies on the relationship between mental effort, monitoring judgments, and learning outcomes presenting the distribution of significant *p*-values. **a** Analysis of the correlations between mental effort and monitoring judgments. The observed *p*-curve includes 47 statistically significant ($p < 0.05$) results, of which 41 are $p < 0.025$. There were 36 additional results entered but excluded from the *p*-curve because they were $p > 0.05$. **b** Analysis of the correlations between monitoring judgments and learning outcomes. The observed *p*-curve includes 36 statistically significant ($p < 0.05$) results, of which 31 are $p < 0.025$. There were 41 additional results entered but excluded from the *p*-curve because they were $p > 0.05$. **c** Analysis of the correlations between mental effort and learning outcomes. The observed *p*-curve includes 31 statistically significant ($p < 0.05$) results, of which 26 are $p < 0.025$. There were 45 additional results entered but excluded from the *p*-curve because they were $p > 0.05$

The lack of a statistically significant direct relationship between perceived mental effort and learning outcomes could be explained by cognitive load theory (Sweller et al., 1998, 2019). As introduced earlier, the relationship between mental effort and learning outcomes could be positive when task requirements meet the learner's working memory capacity and the cognitive effort necessary for processing and integrating information is directly relevant to learning. When the mental effort required for a learning task exceeds this capacity, learning is hampered and the relationship between mental effort and learning outcomes becomes negative. In addition, it might be that in some instances learners' perceived high mental effort is due to factors such as suboptimal task design, a distracting environment, less effective strategies, or incorrect strategy utilization, which are not conducive to effective learning. When learners experience high mental effort due to such non-learning conducive factors (i.e., high extraneous load; Paas & Van Merriënboer, 1994), we would anticipate the relationship between mental effort and learning outcomes to be different compared to instances in which task design is optimal, the environment is distraction-free, and learners employ correct learning/problem-solving strategies. Future research could investigate to what extent task design and strategy use play a role in the relationships between mental effort, monitoring judgments, and learning outcomes.

Another factor, which could potentially explain the statistically non-significant direct relationship between mental effort and learning outcomes, is that the studies included in this meta-analysis might have been too heterogeneous in their approximation of mental effort and/or load. Although we did not find evidence for this in our subgroup analysis, it could be that the use of different self-report questions and formulations hereof might have caused too much noise to detect a significant relation or by measuring extraneous and germane load at the same time and combining them to indicate load (e.g., Schleinschok et al., 2017).

When investigating the direct relationships between perceived mental effort, certainty as the monitoring judgment, and learning outcomes, Blissett and colleagues (2018) found a significant negative relationship between mental effort and monitoring (β = -0.37) and a significant positive relationship between monitoring and learning outcomes (β = 0.34). The results from our study extend these findings by discovering similar results across different settings, tasks, and populations. Additionally, they found a significant negative relationship between mental effort and learning outcomes (β = -0.17) which they interpreted as mental effort being diagnostic due to its predictive value for learning outcomes. Our model comparison showed non-significant differences between the full and partial mediation model, suggesting that the saturated partial mediation model with a direct effect from mental effort to learning outcomes was not a significantly better fit than the more parsimonious full mediation model. Extending the findings from Blissett and colleagues (2018), our findings suggest evidence for an indirect rather than a direct effect between mental effort and learning outcomes, mediated by monitoring judgments.

Our data seem to support the idea that effort is used as a cue for monitoring even though it is not directly related to learning outcomes, highlighting the importance of metacognition during learners' self-regulated learning. As perceived mental effort seems to be used as a cue, it is important to teach students how to interpret their mental effort experience correctly and use it adequately when monitoring their learning. Our findings show that learners tend to interpret it as negatively related to learning when in reality this is not necessarily the case. Additionally, the mediation pathway between mental effort and learning outcomes through monitoring indicates a potentially important role that monitoring has in linking perceived mental effort to actual learning outcomes. Incorporating and further exploring this mediation mechanism into self-regulated learning models could further enhance our understanding of the on-task experiences learners have and how these relate to their actual learning outcomes.

To counteract learners' low metacognitive skills, instructional approaches that support learners in aligning their perceived effort and perceptions of learning with their actual learning, are necessary. Generative strategies such as generating keywords, summaries, or diagrams when learners are required to read a text, have been suggested to facilitate learners' monitoring accuracy by providing learners with more salient access to more diagnostic cues (Prinz et al., 2020b). Educators could for example try to integrate such strategies to facilitate learners' calibrations of their perceived learning. Furthermore, providing learners with feedback on their actual learning outcomes could help them to gain a deeper understanding of their actual learning and the corresponding on-task experiences (De Bruin et al., 2023).

Alternatively, to increase monitoring accuracy, it might be important to refrain from asking learners to rate their mental effort, to reduce the use of effort as a cue for monitoring. While there is unpublished data by Raaijmakers et al., (2023), which indicates that the order of mental effort and monitoring judgment administration does not influence learners' ratings, it might be that the specific combination of perceived mental effort ratings and monitoring judgments could cause changes in monitoring judgments compared to situations in which only monitoring judgments are administered.

Hoch and colleagues (2023) who examined the predictive value of mental effort, task difficulty, and confidence appraisals for performance, found that compared to effort appraisals, confidence, and difficulty appraisals demonstrated stronger associations with performance. They concluded that mental effort appraisals reflect fluency and might thus not serve as reliable predictors of actual performance and instead are misleading (Hoch et al., 2023). These findings suggest that monitoring judgments instead of mental effort might be more reliable predictors of learning success and could thus potentially serve as a more reliable cue when for example taking a regulatory decision.

However, there are also situations during which effort is diagnostic of learning outcomes, but this diagnosticity could still be misleading. For example, in the case of desirable difficulties (i.e., learning conditions that are experienced as effortful while also enhancing learning and long-term retention; Bjork & Bjork, 2011) learners often misinterpret this higher mental effort as a cue for low learning which biases learners to lower judgments of their learning and to choose learning strategies that require less mental effort (Hui et al., 2022; Kirk-Johnson et al., 2019; Onan et al., 2022). Also, when using desirable difficulties, initial learning can be lower compared to other learning strategies. For example, when using retrieval practice compared to re-study, the beneficial effects can become evident only after a few days (Karpicke & Roediger, 2008; Nunes & Karpicke, 2015). This means that immediate learning outcomes can sometimes be lower while effort is higher compared to an ineffective learning strategy. Even though immediate perceptions of effort during the learning task would not be diagnostic in this case, it is important to consider potential delayed learning effects as the cue diagnosticity of mental effort may become apparent after a delay. Learners might benefit from interventions that address this paradox and thus be able to regulate their learning accordingly (De Bruin et al., 2023). Future studies should investigate whether or not mental effort is a diagnostic cue for delayed compared to immediate learning outcomes. This could also be meaningful in order to gain a better understanding of how mental effort and monitoring relate to learning outcomes. In our analysis, we considered the relationships between mental effort, monitoring judgments, and immediate learning outcomes.

Immediate learning outcomes, however, might not be the best approximation for long-term learning, thus limiting the interpretation of our results. More specifically, the fact that monitoring judgments are positively associated with immediate learning outcomes does not necessarily mean that they are also positively related to long-term learning. The studies included in this meta-analysis commonly measured learning outcomes immediately after a learning task or after a short delay. Furthermore, if transfer was included, this was usually a near-transfer task. Due to the lack of temporal changes and far-transfer tasks captured in the included studies, we cannot differentiate between performance and learning as suggested by Soderstom and Bjork (2015) as we do not know whether the changes in behavior/knowledge were permanent. Future research should investigate to what extent perceived mental effort and monitoring judgments relate to long-term learning.

Surprisingly, our moderator analysis revealed no significant differences across subgroups. More specifically, the moderators education, mental effort measurement, mental effort wording, effort reference, type of monitoring, monitoring timing, monitoring reference, and type of task or task difficulty did not significantly moderate the relationships between mental effort, monitoring judgments, and learning outcomes. This finding contrasts with previous findings such as the meta-analysis by Baars and colleagues (2020), who found timing of monitoring judgment, type of task, and goal-driven manipulations significantly moderated the relationship between mental effort and monitoring. While we were not able to assess the moderator goal-driven manipulations due to the small number of goal-driven manipulations in our sample, we did not find a significant difference in the magnitude of effect sizes for timing of monitoring judgment or type of task.

A potential reason for these different findings might be the statistical analysis approach used. Using a multilevel meta-analytic structural equation model, we conducted a subgroup analysis and thus investigated whether the effect sizes in the full mediation model were statistically different amongst samples from two different subgroups. We did not test the moderating influence of the variables on the separate paths. To our knowledge, using a multilevel meta-analytic structural equation model with subgroup analysis is currently the best way to examine the influence of moderating variables in a two-stage model (Jak & Cheung, 2018). Furthermore, another potential reason for this incongruence in findings compared to earlier work could be the limited number of studies we were able to include in our analysis or a lack of variability within our moderators. Additionally, Baars and colleagues (2020) did not focus on perceived mental effort as measured by self-report items only but also took (mental) effort indicators such as response time or time on task into account. This broader perspective on (mental) effort might have caused the difference in results. In the present study, we focused on perceived mental effort as a more subjective representation of effort compared to a more objective operationalization, via for example time on task, for two reasons. During self-regulated learning, learners often need to regulate their effort investment. In order to do so, they need to monitor their effort. As learners usually do not have direct access to the mental effort they invested during a task, they infer the amount of mental effort based on subjective experiences of effort (De Bruin et al., 2023). Therefore, in the context of self-regulated learning, perceived mental effort seemed a more appropriate operationalization compared to a more objective one. A second reason was to avoid ambiguity between learning outcomes and effort measures. While for example time on task is sometimes operationalized as an indication of mental effort, it is also often used as a learning outcome measure. To avoid this ambiguity, we focused on perceived mental effort only. Future studies could investigate this further by testing whether the current findings extend to studies employing other operationalization of (mental) effort such as study time. Such an approach might also be interesting in light of the previously raised point that the specific combination of perceived mental effort ratings and monitoring judgments could potentially cause changes in monitoring judgments compared to situations in which only monitoring judgments are administered. However, unpublished data indicates that the order of administration seems to not influence mental effort ratings or monitoring judgments (Raaijmakers et al., 2023).

Limitations and Future Studies

The current study has various limitations, which should be taken into consideration when interpreting these findings. First, the effect sizes used in our study were purely correlational, thus limiting our ability to establish causal relationships. In line with the cue-utilization framework (Koriat, 1997), we theorized directional paths from mental effort to monitoring, from monitoring to learning outcomes, and from mental effort to learning outcomes. While our study contributes to an increased understanding of the associations across diverse studies, the observed correlations do not confirm the directionality or causality of these relationships. For example, we cannot exclude the fact that the relationships between some of the variables are reversed. More specifically, it might be the case, that in some instances perceived mental effort and monitoring judgments are influenced by learning outcomes. Raaijmakers and colleagues (2017) found that performance feedback influenced learners' perceptions of their invested effort. While we cannot exclude this reversed directionality, the majority of studies included in this meta-analysis did not provide learners with access to their actual learning outcomes, and commonly mental effort judgments and monitoring judgments were asked prior to the learning outcome measure. In order to gain more insight into the causal relationships between mental effort, monitoring, and learning outcomes, future research could manipulate learners' experience of their perceived mental effort or their monitoring which would allow for causal inference and would thus be essential in validating the directionality of the proposed relationships within the cue-utilization framework. Additionally, it is important to acknowledge the current lack of comprehensive understanding regarding optimal strategies to assist learners in effectively monitoring and regulating their learning. We argue that accurate monitoring is a requirement for effective self-regulated learning. However, research suggests, that accurate monitoring does not necessarily mean that learners will actually engage in (beneficial) regulation. For example, prompting learners to monitor their understanding or informing them about the dangers of inaccurate monitoring does not necessarily lead to improved learning outcomes (Berthold et al., 2007; Nückles et al., 2020; Roelle et al., 2017).

This suggests that even if learners are able to accurately monitor their learning, they will not necessarily engage in effective self-regulation due to for example the additional mental effort that is required but not available to regulate effectively (De Bruin et al., 2023).

Furthermore, due to the correlational nature of the data, it might be the case that an underlying third variable such as motivation influences the associations between mental effort, monitoring judgments, and learning outcomes. Research has shown that motivational profiles were related to monitoring accuracy and learning outcomes after self-regulated learning skills training (Baars & Wijnia, 2018; Wijnia & Baars, 2021). Specifically, results showed that learners with higher-quality motivation obtained better scores on monitoring accuracy and problem-solving tasks after the intervention. Results further suggested that participants with lower-quality motivation perceived the learning task as more effortful than students with higher-quality motivation. Therefore, individual differences in learners' motivation could potentially serve as a third variable influencing the observed correlations. Depending on learners' motivation, they might experience increased mental effort and more accurate monitoring of their learning processes. Additionally, motivational profiles might independently contribute to enhance their learning outcomes. As we did not have access to learners' motivation, we cannot rule out the possibility that the observed associations may be confounded by, for example, learners' motivation. Future research could incorporate measures of motivation or manipulate motivation for a more nuanced understanding of the relationships between mental effort, monitoring, and actual learning outcomes. Moreover, in order to gain more insight into the causal relationships between mental effort, monitoring, and learning outcomes, future research could manipulate learners' experience of their perceived mental effort or their monitoring which would allow for causal inference and would thus be essential in validating the directionality of the proposed relationships within the cue-utilization framework.

Second, we included a relatively small number of studies in our analysis. While there is no common cut-off score of minimum number of studies necessary to apply meta-analytic structural equation modeling, it is suggested to include a sufficient number of studies to ensure the reliability and generalizability of the meta-analytic findings. A small number of studies might limit the statistical power and precision of the analysis, making it challenging to draw meaningful conclusions. Additionally, the studies we selected for our meta-analysis stem from a particular intersection of research into metacognition and cognitive load theory. It is possible that this intersection excluded certain research lines or paradigms, where, e.g., measures of effort are not typically used. The conclusions drawn regarding the relationships of interest are situated within these frameworks and while providing valuable insights within this theoretical context, their generalizability to broader educational settings or other fields of research might therefore be limited. Future studies should investigate the relationships further and consolidate our findings.

Third, the role of the moderators that were tested in the current study needs further investigation. We were not able to replicate previous meta-analytic findings, which identify moderating effects of certain types of tasks and timing of judgments.

Potential reasons for this include the lack of variability in our moderators or the limited sample size. Future research is necessary to disentangle the effects moderators might have on the relationships between mental effort, monitoring judgments, and learning outcomes.

Fourth, we have restricted the publication period of studies eligible for our meta-analysis and only included studies published in the year 2000 or thereafter. We decided to do so to maintain the contemporary relevance of results, to increase the likelihood of homogeneity in the studies' employed methodology and technology, and because the influential cue-utilization framework (Koriat, 1997) underlying the majority of relevant studies was published in 1997, which is why we did not expect many relevant studies before 2000. However, in restricting the publication period we might have biased our selected sample and missed influential studies published before the year 2000.

Conclusion

In the current study, we investigated the association between mental effort, monitoring, and learning outcomes using a meta-analytic approach. The results indicate a negative association between mental effort and monitoring judgments, a positive association between monitoring judgments and learning outcomes, and a significant indirect association between mental effort and learning outcomes mediated by monitoring judgments. Surprisingly, our subgroup analysis did not reveal any significant differences across moderators. These findings suggest that mental effort is used as a cue for monitoring judgments and is only indirectly related to immediate learning outcomes via monitoring judgments. We did not find evidence that these relationships were influenced by certain task or learner characteristics.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10648-024-09903-z.

Acknowledgements We would like to thank the members of the EARLI Emerging Field Group 3 "Monitoring and Regulation of Effort" for the ongoing discussions on the topic of effort monitoring and regulation that contributed to this research.

Author Contributions All authors contributed to the study concept and design. Material preparation, data collection, and analysis were performed by LD in close consultation with FB, MB, and LW. All authors contributed to the manuscript.

Data Availability The data used for this study are available via the project's Open Science Framework (OSF) page: https://osf.io/hc5fk/.

Declarations

This work was funded by the Dutch Research Council (NWO) (VIDI grant number VI.Vidi.195.135) to Anique de Bruin.

Conflict of Interest The authors declare no conflict of interest.

References

References marked with an asterisk were included in the meta-analysis

Ackerman, R. (2014). The diminishing criterion model for metacognitive regulation of time investment. *Journal of Experimental Psychology: General, 143*(3), 1349–1368. https://doi.org/10.1037/a0035098

*Allen, P. J., Finlay, J., Roberts, L. D., & Baughman, F. D. (2019). An experimental evaluation of StatHand: A free application to guide students' statistical decision making. *Scholarship of Teaching and Learning in Psychology*, *5*(1), 23–36.https://doi.org/10.1037/stl0000132

Ayres, P., Lee, J. Y., Paas, F., & Van Merriënboer, J. J. G. (2021). The validity of physiological measures to identify differences in intrinsic cognitive load. *Frontiers in Psychology, 12*, 702538. https://doi.org/10.3389/fpsyg.2021.702538

Baars, M., Wijnia, L., De Bruin, A., & Paas, F. (2020). The relation between students' effort and monitoring judgments during learning: A meta-analysis. *Educational Psychology Review, 32*(4), 979–1002. https://doi.org/10.1007/s10648-020-09569-3

*Baars, M., & Wijnia, L. (2018). The relation between task-specific motivational profiles and training of self-regulated learning skills. *Learning and Individual Differences*, *64*, 125–137.https://doi.org/10.1016/j.lindif.2018.05.007

*Baars, M., Visser, S., Gog, T. V., Bruin, A. D., & Paas, F. (2013). Completion of partially worked-out examples as a generation strategy for improving monitoring accuracy. *Contemporary Educational Psychology*, *38*(4), 395–406.https://doi.org/10.1016/j.cedpsych.2013.09.001

*Baars, M., Vink, S., Van Gog, T., De Bruin, A., & Paas, F. (2014). Effects of training self-assessment and using assessment standards on retrospective and prospective monitoring of problem-solving. *Learning and Instruction*, *33*, 92–107.https://doi.org/10.1016/j.learninstruc.2014.04.004

*Baars, M., Wijnia, L., & Paas, F. (2017). The association between motivation, affect, and self-regulated learning when solving problems. *Frontiers in Psychology*, *8*, 1346. https://doi.org/10.3389/fpsyg.2017.01346

*Baars, M., Leopold, C., & Paas, F. (2018a). Self-explaining steps in problem-solving tasks to improve self-regulation in secondary education. *Journal of Educational Psychology*, *110*(4), 578–595.https://doi.org/10.1037/edu0000223

*Baars, M., Van Gog, T., De Bruin, A., & Paas, F. (2018b). Accuracy of primary school children's immediate and delayed judgments of learning about problem-solving tasks. *Studies in Educational Evaluation*, *58*, 51–59.https://doi.org/10.1016/j.stueduc.2018.05.010

*Bednall, T. C. (2009). *Effects of self-regulatory aids on autonomous study* [Doctoral dissertation, University New South Wales]. EBSCO OpenDissertations. https://doi.org/10.26190/UNSWORKS/19106

*Beege, M., Krieglstein, F., Schneider, S., Nebel, S., & Rey, G. D. (2021). Is there a (dis-)fluency effect in learning with handwritten instructional texts? Evidence from three studies. *Frontiers in Education*, *6*, 678798. https://doi.org/10.3389/feduc.2021.678798

Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction, 17*(5), 564–577. https://doi.org/10.1016/j.learninstruc.2007.09.007

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the real world: Essays illustrating fundamental contributions to society.* (pp. 56–64).

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*(1), 417–444. https://doi.org/10.1146/annurev-psych-113011-143823

Blissett, S., Sibbald, M., Kok, E., & Van Merriënboer, J. (2018). Optimizing self-regulation of performance: Is mental effort a cue? *Advances in Health Sciences Education, 23*(5), 891–898. https://doi.org/10.1007/s10459-018-9838-x

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education, 27*, 1–13. https://doi.org/10.1016/j.iheduc.2015.04.007

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). University of California Press.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245–281. https://doi.org/10.3102/00346543065003245

Cavalcanti, R. B., & Sibbald, M. (2014). Am I right when I am sure? Data consistency influences the relationship between diagnostic accuracy and certainty. *Academic Medicine, 89*(1), 107–113. https://doi.org/10.1097/ACM.0000000000000074

Chen, O., Castro-Alonso, J. C., Paas, F., & Sweller, J. (2018). Extending cognitive load theory to incorporate working memory resource depletion: Evidence from the spacing effect. *Educational Psychology Review, 30*(2), 483–501. https://doi.org/10.1007/s10648-017-9426-2

Chen, O., Paas, F., & Sweller, J. (2023). A cognitive load theory approach to defining and measuring task complexity through element interactivity. *Educational Psychology Review, 35*(2), 63. https://doi.org/10.1007/s10648-023-09782-w

Cheung, M.W.-L. (2015). metaSEM: An R package for meta-analysis using structural equation modeling. *Frontiers in Psychology, 5*, 521. https://doi.org/10.3389/fpsyg.2014.01521

Choi, H.-H., Van Merriënboer, J. J. G., & Paas, F. (2014). Effects of the physical environment on cognitive load and learning: Towards a new model of cognitive load. *Educational Psychology Review, 26*(2), 225–244. https://doi.org/10.1007/s10648-014-9262-6

*Cuevas, H. M., Fiore, S. M., & Oser, R. L. (2002). Scaffolding cognitive and metacognitive processes in low verbal ability learners: Use of diagrams in computer-based training environments. *Instructional Science*, *30*(6), 433–464.https://doi.org/10.1023/A:1020516301541

*Currie, J., Bond, R. R., McCullagh, P., Black, P., Finlay, D. D., & Peace, A. (2018). Eye tracking the visual attention of nurses interpreting simulated vital signs scenarios: Mining metrics to discriminate between performance level. *IEEE Transactions on Human-Machine Systems*, *48*(2), 113–124.https://doi.org/10.1109/THMS.2017.2754880

De Bruin, A. B. H., & van Merriënboer, J. J. G. (2017). Bridging cognitive load and self-regulated learning research: A complementary approach to contemporary issues in educational research. *Learning and Instruction, 51*, 1–9. https://doi.org/10.1016/j.learninstruc.2017.06.001

De Bruin, A. B. H., Thiede, K. W., Camp, G., & Redford, J. (2011). Generating keywords improves metacomprehension and self-regulation in elementary and middle school children. *Journal of Experimental Child Psychology, 109*(3), 294–310. https://doi.org/10.1016/j.jecp.2011.02.005

De Bruin, A. B. H., Dunlosky, J., & Cavalcanti, R. B. (2017). Monitoring and regulation of learning in medical education: The need for predictive cues. *Medical Education, 51*(6), 575–584. https://doi.org/10.1111/medu.13267

De Bruin, A. B. H., Biwer, F., Hui, L., Onan, E., David, L., & Wiradhany, W. (2023). Worth the effort: The start and stick to desirable difficulties (S2D2) framework. *Educational Psychology Review, 35*(2), 41. https://doi.org/10.1007/s10648-023-09766-w

*Dentakos, S., Saoud, W., Ackerman, R., & Toplak, M. E. (2019). Does domain matter? Monitoring accuracy across domains. *Metacognition and Learning*, *14*(3), 413–436.https://doi.org/10.1007/s11409-019-09198-4

Dinsmore, D. L., & Parkinson, M. M. (2013). What are confidence judgments made of? Students' explanations for their confidence ratings and what that means for calibration. *Learning and Instruction, 24*, 4–14. https://doi.org/10.1016/j.learninstruc.2012.06.001

Dunlosky, J., & Metcalfe, J. (2008). *Metacognition*. Thousand Oaks: Sage Publications.

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science, 16*(4), 228–232. https://doi.org/10.1111/j.1467-8721.2007.00509.x

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Improving Self-Monitoring and Self-Regulation* *of Learning: From Cognitive Psychology to the Classroom, 22*(4), 271–280. https://doi.org/10.1016/j.learninstruc.2011.08.003

Feldon, D. F., Callan, G., Juth, S., & Jeong, S. (2019). Cognitive load as motivational cost. *Educational Psychology Review, 31*(2), 319–337. https://doi.org/10.1007/s10648-019-09464-6

Fernández-Castilla, B., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2020). Visual representation of meta-analyses of multiple outcomes: Extensions to forest plots, funnel plots, and caterpillar plots. *Methodology, 16*(4), 299–315. https://doi.org/10.5964/meth.4013

Hoch, E., Sidi, Y., Ackerman, R., Hoogerheide, V., & Scheiter, K. (2023). Comparing mental effort, difficulty, and confidence appraisals in problem-solving: A metacognitive perspective. *Educational Psychology Review, 35*(2), 61. https://doi.org/10.1007/s10648-023-09779-5

Hoffmann-Biencourt, A., Lockl, K., Schneider, W., Ackerman, R., & Koriat, A. (2010). Self-paced study time as a cue for recall predictions across school age. *British Journal of Developmental Psychology, 28*(4), 767–784. https://doi.org/10.1348/026151009X479042

*Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2014). Comparing the effects of worked examples and modeling examples on learning. *Computers in Human Behavior*, *41*, 80–91.https://doi.org/10.1016/j.chb.2014.09.013

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424.

Huh, D., Kim, J., & Jo, I. (2019). A novel method to monitoring changes in cognitive load in video-based learning. *Journal of Computer Assisted Learning, 35*(6), 721–730. https://doi.org/10.1111/jcal.12378

Hui, L., Bruin, A. B. H., Donkers, J., & Merriënboer, J. J. G. (2022). Why students do (or do not) choose retrieval practice: Their perceptions of mental effort during task performance matter. *Applied Cognitive Psychology, 36*(2), 433–444. https://doi.org/10.1002/acp.3933

*İlïc, U., & Akbulut, Y. (2019). Effect of disfluency on learning outcomes, metacognitive judgments and cognitive load in computer assisted learning environments. *Computers in Human Behavior*, *99*, 310–321.https://doi.org/10.1016/j.chb.2019.06.001

Jak, S., & Cheung, M.W.-L. (2018). Testing moderator hypotheses in meta-analytic structural equation modeling using subgroup analysis. *Behavior Research Methods, 50*(4), 1359–1373. https://doi.org/10.3758/s13428-018-1046-3

Jak, S., Li, H., Kolbe, L., De Jonge, H., & Cheung, M. W. L. (2021). Meta-analytic structural equation modeling made easy: A tutorial and web application for one-stage MASEM. *Research Synthesis Methods, 12*(5), 590–606. https://doi.org/10.1002/jrsm.1498

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), 657–690. https://doi.org/10.1037/0021-9010.74.4.657

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 319*(5865), 966–968. https://doi.org/10.1126/science.1152408

Keith, T. Z. (2015). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling* (2nd ed.). New York, NY: Routledge.

*Kirk-Johnson, A., Galla, B. M., & Fraundorf, S. H. (2019). Perceiving effort as poor learning: The misinterpreted-effort hypothesis of how experienced effort and perceived learning relate to study strategy choice. *Cognitive Psychology*, *115*, 101237. https://doi.org/10.1016/j.cogpsych.2019.101237

Klepsch, M., & Seufert, T. (2021). Making an effort versus experiencing load. *Frontiers in Education, 6*, 645284. https://doi.org/10.3389/feduc.2021.645284

Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language, 52*(4), 478–492. https://doi.org/10.1016/j.jml.2005.01.001

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*(4), 609–639. https://doi.org/10.1037/0033-295X.100.4.609

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Koriat, A., & Nussinson, R. (2009). Attributing study effort to data-driven and goal-driven effects: Implications for metacognitive judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(5), 1338–1343. https://doi.org/10.1037/a0016374

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General, 135*(1), 36–69. https://doi.org/10.1037/0096-3445.135.1.36

Koriat, A., Ackerman, R., Lockl, K., & Schneider, W. (2009). The memorizing effort heuristic in judgments of learning: A developmental perspective. *Journal of Experimental Child Psychology, 102*(3), 265–279. https://doi.org/10.1016/j.jecp.2008.10.005

Koriat, A., Ackerman, R., Adiv, S., Lockl, K., & Schneider, W. (2014a). The effects of goal-driven and data-driven regulation on metacognitive monitoring during learning: A developmental perspective. *Journal of Experimental Psychology: General, 143*(1), 386–403. https://doi.org/10.1037/a0031768

*Koriat, A., Nussinson, R., & Ackerman, R. (2014b). Judgments of learning depend on how learners interpret study effort. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1624–1637.https://doi.org/10.1037/xlm0000009

*Koriat, A. (2018). Agency attributions of mental effort during self-regulated learning. *Memory & Cognition*, *46*(3), 370–383.https://doi.org/10.3758/s13421-017-0771-7

Koriat, A. (2008). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory & Cognition, 36*(2), 416–428. https://doi.org/10.3758/MC.36.2.416

Kostons, D., Van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and Instruction, 22*(2), 121–132. https://doi.org/10.1016/j.learninstruc.2011.08.004

*Kostons, D., & De Koning, B. B. (2017). Does visualization affect monitoring accuracy, restudy choice, and comprehension scores of students in primary education? *Contemporary Educational Psychology*, *51*, 1–10.https://doi.org/10.1016/j.cedpsych.2017.05.001

*Kuhn, J., Van Den Berg, P., Mamede, S., Zwaan, L., Bindels, P., & Van Gog, T. (2022). Improving medical residents' self-assessment of their diagnostic accuracy: Does feedback help? *Advances in Health Sciences Education*, *27*(1), 189–200.https://doi.org/10.1007/s10459-021-10080-9

*Kuhn, J., Mamede, S., Van Den Berg, P., Zwaan, L., Van Peet, P., Bindels, P., & Van Gog, T. (2023). Learning deliberate reflection in medical diagnosis: Does learning-by-teaching help? *Advances in Health Sciences Education*, *28*(1), 13–26.https://doi.org/10.1007/s10459-022-10138-2

*Lachner, A., Backfisch, I., Hoogerheide, V., Van Gog, T., & Renkl, A. (2020). Timing matters! Explaining between study phases enhances students' learning. *Journal of Educational Psychology*, *112*(4), 841–853.https://doi.org/10.1037/edu0000396

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Leonesio, R. J., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 464–470. https://doi.org/10.1037/0278-7393.16.3.464

*Macaluso, J. A., Beuford, R. R., & Fraundorf, S. H. (2022). Familiar strategies feel fluent: The role of study strategy familiarity in the misinterpreted-effort model of self-regulated learning. *Journal of Intelligence*, *10*(4), 83. https://doi.org/10.3390/jintelligence10040083

Maki, R. H., & Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(4), 663–679. https://doi.org/10.1037/0278-7393.10.4.663

*Mihalca, L., & Mengelkamp, C. (2020). Effects of induced levels of prior knowledge on monitoring accuracy and performance when learning from self-regulated problem solving. *Journal of Educational Psychology*, *112*(4), 795–810.https://doi.org/10.1037/edu0000389

*Mihalca, L., Mengelkamp, C., & Schnotz, W. (2017). Accuracy of metacognitive judgments as a moderator of learner control effectiveness in problem-solving tasks. *Metacognition and Learning*, *12*(3), 357–379.https://doi.org/10.1007/s11409-017-9173-2

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*(1), 109–133. https://doi.org/10.1037/0033-2909.95.1.109

Nückles, M., Roelle, J., Glogger-Frey, I., Waldeyer, J., & Renkl, A. (2020). The self-regulation-view in writing-to-learn: Using journal writing to optimize cognitive load in self-regulated learning. *Educational Psychology Review, 32*, 1089–1126. https://doi.org/10.1007/s10648-020-09541-1

Nunes, L. D., & Karpicke, J. D. (2015). Retrieval-based learning: Research at the interface between cognitive science and education. In R. A. Scott & S. M. Kosslyn (Eds.), *Emerging trends in the social and behavioral sciences* (1st ed., pp. 1–16). Wiley. https://doi.org/10.1002/9781118900772.etrds0289

*Onan, E., Wiradhany, W., Biwer, F., Janssen, E. M., & De Bruin, A. B. H. (2022). Growing out of the experience: How subjective experiences of effort and learning influence the use of interleaved practice. *Educational Psychology Review*, *34*(4), 2451–2484.https://doi.org/10.1007/s10648-022-09692-3

*Onan, E., Wiradhany, W., Biwer, F., & De Bruin, A. B. H. (2023). *Instruction meets experience: Using Theory- and Experience-based methods to promote the use of desirable difficulties.* [Manuscript submitted for publication]. Maastricht University.

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—A web and mobile app for systematic reviews. *Systematic Reviews, 5*(1), 210. https://doi.org/10.1186/s13643-016-0384-4

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*(4), 429–434. https://doi.org/10.1037/0022-0663.84.4.429

Paas, F., & Van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*(4), 351–371. https://doi.org/10.1007/BF02213420

Paas, F., & Van Merriënboer, J. J. G. (2020). Cognitive-load theory: Methods to manage working memory load in the learning of complex tasks. *Current Directions in Psychological Science, 29*(4), 394–398. https://doi.org/10.1177/0963721420922183

Paas, F. G. W. C., Van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of Cognitive Load in Instructional Research. *Perceptual and Motor Skills, 79*(1), 419–430. https://doi.org/10.2466/pms.1994.79.1.419

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63–71. https://doi.org/10.1207/S15326985EP3801_8

Paas, F., Tuovinen, J. E., Van Merriënboer, J. J. G., & Darabi, A. A. (2005). A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. *Educational Technology Research and Development, 53*(3), 25–34. https://doi.org/10.1007/BF02504795

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 71*. https://doi.org/10.1136/bmj.n71

*Paik, E. S., & Schraw, G. (2013). Learning with animation and illusions of understanding. *Journal of Educational Psychology*, *105*(2), 278–289.https://doi.org/10.1037/a0030281

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology, 8*, 422.

Prinz, A., Golke, S., & Wittwer, J. (2020a). How accurately can learners discriminate their comprehension of texts? A comprehensive meta-analysis on relative metacomprehension accuracy and influencing factors. *Educational Research Review, 31*, 100358. https://doi.org/10.1016/j.edurev.2020.100358

Prinz, A., Golke, S., & Wittwer, J. (2020b). To what extent do situation-model-approach interventions improve relative metacomprehension accuracy? Meta-analytic insights. *Educational Psychology Review, 32*(4), 917–949. https://doi.org/10.1007/s10648-020-09558-6

Raaijmakers, S. F., Baars, M., Schaap, L., Paas, F., & Van Gog, T. (2017). Effects of performance feedback valence on perceptions of invested mental effort. *Learning and Instruction, 51*, 36–46. https://doi.org/10.1016/j.learninstruc.2016.12.002

Raaijmakers, S. F., Schaap, L., Van Gog, T., & Paas, F. (2023). *Assessing performance before mental effort has no effect on mental effort, but receiving fixed feedback has*. [Unpublished manuscript]. OSF. https://osf.io/ryzqj. Accessed 16 Nov 2023.

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Roelle, J., Schmidt, E. M., Buchau, A., & Berthold, K. (2017). Effects of informing learners about the dangers of making overconfident judgments of learning. *Journal of Educational Psychology, 109*(1), 99–117. https://doi.org/10.1037/edu0000132

RStudio Team. (2023). *RStudio: integrated development environment for R*. RStudio, PBC. http://www.posit.co/

Scheiter, K., Ackerman, R., & Hoogerheide, V. (2020). Looking at mental effort appraisals through a metacognitive lens: Are they biased? *Educational Psychology Review, 32*(4), 1003–1027. https://doi.org/10.1007/s10648-020-09555-9

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*(2), 23–74.

*Schleinschok, K., Eitel, A., & Scheiter, K. (2017). Do drawing tasks improve monitoring and control during learning from text? *Learning and Instruction*, *51*, 10–25.https://doi.org/10.1016/j.learninstruc.2017.02.002

Schmeck, A., Opfermann, M., Van Gog, T., Paas, F., & Leutner, D. (2015). Measuring cognitive load with subjective rating scales during problem solving: Differences between immediate and delayed ratings. *Instructional Science, 43*(1), 93–114. https://doi.org/10.1007/s11251-014-9328-3

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning, 4*(1), 33–45. https://doi.org/10.1007/s11409-008-9031-3

Seufert, T. (2018). The interplay between self-regulation in learning and cognitive load. *Educational Research Review, 24*, 116–129. https://doi.org/10.1016/j.edurev.2018.03.004

Siedlecka, M., Paulewicz, B., & Wierzchoń, M. (2016). But I was so sure! Metacognitive judgments are less accurate given prospectively than retrospectively. *Frontiers in Psychology, 7*, 218. https://doi.org/10.3389/fpsyg.2016.00218

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file drawer. *Journal of Experimental Psychology: General, 143*(2), 534–547. https://doi.org/10.1037/a0033242

*Skulmowski, A., & Rey, G. D. (2018). Realistic details in visualizations require color cues to foster retention. *Computers & Education*, *122*, 23–31.https://doi.org/10.1016/j.compedu.2018.03.012

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: an integrative review. *Perspectives on Psychological Science, 10*(2), 176–199. https://doi.org/10.1177/1745691615569000

*Sondermann, C., & Merkt, M. (2023). Like it or learn from it: Effects of talking heads in educational videos. *Computers & Education*, *193*, 104675. https://doi.org/10.1016/j.compedu.2022.104675

Stolwijk, I., Jak, S., Eichelsheim, V., & Hoeve, M. (2022). Dealing with dependent effect sizes in MASEM: A comparison of different approaches using empirical data. *Zeitschrift Für Psychologie, 230*(1), 16–32. https://doi.org/10.1027/2151-2604/a000485

Sugden, C., Housden, C. R., Aggarwal, R., Sahakian, B. J., & Darzi, A. (2012). Effect of pharmacological enhancement on the cognitive and clinical psychomotor performance of sleep-deprived doctors: A randomized controlled trial. *Annals of Surgery, 255*(2), 222–227. https://doi.org/10.1097/SLA.0b013e3182306c99

Sweller, J., Van Merrienboer, J. J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296.

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Thiede, K. W., Griffin, T. D., Wiley, J., & Anderson, M. C. M. (2010). Poor metacomprehension accuracy as a result of inappropriate cue use. *Discourse Processes, 47*(4), 331–362. https://doi.org/10.1080/01638530902959927

Undorf, M., & Erdfelder, E. (2011). Judgments of learning reflect encoding fluency: Conclusive evidence for the ease-of-processing hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(5), 1264–1269. https://doi.org/10.1037/a0023719

Van Den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods, 45*(2), 576–594. https://doi.org/10.3758/s13428-012-0261-6

Van Gog, T., Kester, L., & Paas, F. (2011a). Effects of concurrent monitoring on cognitive load and performance as a function of task complexity. *Applied Cognitive Psychology, 25*(4), 584–587. https://doi.org/10.1002/acp.1726

Van Gog, T., Kester, L., & Paas, F. (2011b). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology, 36*(3), 212–218. https://doi.org/10.1016/j.cedpsych.2010.10.004

Van Gog, T., Kirschner, F., Kester, L., & Paas, F. (2012). Timing and frequency of mental effort measurement: Evidence in favour of repeated measures. *Applied Cognitive Psychology, 26*(6), 833–839. https://doi.org/10.1002/acp.2883

Van Gog, T., Hoogerheide, V., & Van Harsel, M. (2020). The role of mental effort in fostering self-regulated learning with problem-solving tasks. *Educational Psychology Review, 32*(4), 1055–1072. https://doi.org/10.1007/s10648-020-09544-y

Viechtbauer, W. (2010). Conducting meta-analyses in *R* with the **metafor** package. *Journal of Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

*Wang, J., & Antonenko, P. D. (2017). Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in Human Behavior*, *71*, 79–89.https://doi.org/10.1016/j.chb.2017.01.049

*Wang, J., Antonenko, P., & Dawson, K. (2020). Does visual attention to the instructor in online video affect learning and learner perceptions? An eye-tracking analysis. *Computers & Education*, *146*, 103779. https://doi.org/10.1016/j.compedu.2019.103779

*Weber, S. (2022). Einfluss von Prompts auf die Nutzung von Hinweisreizen beim Monitoring in einer multimedialen Lernumgebung. [Unpublished Manuscript].

*Wesenberg, L., Krieglstein, F., Jansen, S., Rey, G. D., Beege, M., & Schneider, S. (2022). The influence of the order and congruency of correct and erroneous worked examples on learning and (meta-)cognitive load. *Frontiers in Psychology, 13*, 1032003. https://doi.org/10.3389/fpsyg.2022.1032003

*Wijnia, L., & Baars, M. (2021). The role of motivational profiles in learning problem-solving and self-assessment skills with video modeling examples. *Instructional Science*, 49(1), 67–107.https://doi.org/10.1007/s11251-020-09531-4

Wilson, S. J., Polanin, J. R., & Lipsey, M. W. (2016). Fitting meta-analytic structural equation models with complex datasets. *Research Synthesis Methods, 7*(2), 121–139. https://doi.org/10.1002/jrsm.1199

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Louise David[1] · Felicitas Biwer[1] · Martine Baars[2] · Lisette Wijnia[3] · Fred Paas[2,4] · Anique de Bruin[1]

✉ Louise David
l.david@maastrichtuniversity.nl

[1] Department of Educational Development and Research, School of Health Professions Education (SHE), Faculty of Health, Medicine, and Life Sciences, Maastricht University, P.O. Box 616, 6200 MD Maastricht, the Netherlands

[2] Department of Psychology, Education, and Child Studies, Erasmus School of Social and Behavioural Sciences, Erasmus University Rotterdam, Rotterdam, the Netherlands

[3] Department of Conditions for Lifelong Learning, Faculty of Educational Sciences, Open University of the Netherlands, Heerlen, the Netherlands

[4] School of Education/Early Start, University of Wollongong, Wollongong, Australia