



# The effects of cognitive load during an investigative interviewing task on mock interviewers' recall of information

Pamela Hanway\*, Lucy Akehurst, Zarah Vernham and Lorraine Hope

Department of Psychology, University of Portsmouth, Portsmouth, UK

**Purpose.** Although investigative interviewers receive training in interviewing techniques, they often fail to comply with recommended practices. Interviewers are required to actively listen, accurately remember information, think of questions to ask, make judgements, and seek clarification, whilst conducting interviews with witnesses, victims, or suspects. The current study examined the impact of increased cognitive load on mock interviewers' recall of a witness's account.

**Method.** Participants took the role of an investigative interviewer in one of three conditions, high cognitive load (HCL), moderate cognitive load (MCL), or no cognitive load (NCL). Participants watched a video-recorded free narrative of a child witness during which they followed condition-relevant task instructions. Each participant rated their perceived cognitive load during their task and then recalled (free and cued recall) the content of the witness's account.

**Results.** Participants in the HCL and MCL conditions perceived higher cognitive load and demonstrated poorer performance on the free recall task than those in the NCL condition. Participants in the HCL condition demonstrated poorer performance on the cued recall task compared to participants in the NCL condition.

**Conclusions.** The cognitive demands required to complete an investigative interview task led to an increased perceived cognitive load and had a negative impact on recall performance for mock interviewers. Accurately recalling what has been reported by a witness is vital during an investigation. Inaccurate recall can impact on interviewers' questioning and their compliance with recommended interviewing practices. Developing and practising interview techniques may help interviewers to better cope with the high cognitive demands of investigative interviewing.

Despite having knowledge of, and receiving training in, recommended interviewing techniques, interviewers in real-world settings do not always follow best practice guidelines (Criminal Justice Joint Inspectorate, 2014; Powell & Barnett, 2015; Schreiber-Compo *et al.*, 2012). Such guidelines generally recommend a range of interviewing techniques that have been developed from decades of international research for use in

---

*This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.*

\*Correspondence should be addressed to Pamela Hanway, Department of Psychology, University of Portsmouth, King Henry I Building, King Henry I Street, Portsmouth PO2 1DY, UK (email: [pamela.hanway@port.ac.uk](mailto:pamela.hanway@port.ac.uk)).

criminal investigations, child protection enquiries, and intelligence-gathering settings. For example, the PEACE model is recommended for interviewing suspects and witnesses (Bull & Soukara, 2009; Kassin *et al.*, 2010; Milne & Bull, 1999). Similarly, the Achieving Best Evidence guidelines (ABE; Ministry of Justice [MoJ], 2011) and the National Institute for Child Health and Human Development protocol (NICHD; Lamb *et al.*, 2018; Orbach *et al.*, 2000) have been developed, and are recommended for interviewing vulnerable witnesses<sup>1</sup>.

There are, therefore, an abundance of guidelines, which provide advice to practitioners for the optimal approach to obtaining precise and complete statements from interviewees (Bull, 2010; Hershkowitz, 2011; Oxburgh *et al.*, 2015). However, adhering to these guidelines remains a challenge for investigative interviewers (Lamb, 2016; Schreiber-Compo *et al.*, 2012). This may be because interviewing is a complex cognitive task for the interviewer (Lafontaine & Cyr, 2016; Powell, 2002). In an exploratory study, the cognitive load experienced by interviewers was identified as a possible barrier to compliance with recommended techniques (Hanway & Akehurst, 2018). Contrary to recommendations, interviewers' cognitive burden may result in them interrupting the witness or asking questions that have already been answered (Schreiber-Compo *et al.*, 2012). However, as noted by Kleider-Offutt *et al.* (2016), the impact of multiple cognitive demands for investigative interviewers has not been empirically examined. The current study explored the cognitive demands of a mock interview task and tested the effects of cognitive load on the recall of a witness's account.

## Cognitive load and task performance

Cognitive load is the mental workload placed on individuals when they are required to undertake activities (Hart & Staveland, 1988; Van Acker *et al.*, 2018). It signifies working memory use and the demands placed on cognitive resources when carrying out multiple and competing tasks (Dias *et al.*, 2018; Engström *et al.*, 2013). The capacity limitations of working memory mean that without the rehearsal of received sensory information, the processing of information is restricted (van Merriënboer & Sweller, 2010). This can lead to an attentional bottleneck where attending to one element of information causes other cognitive processes, and the associated information, to be neglected (Strayer & Drews, 2007).

Controlled processing is needed to complete cognitive tasks that require attention and the management of information (Bargh, 1984). However, this type of processing is slow and effortful and relies on our limited attention capacity (Strayer & Drews, 2007). High levels of focused attention can be accomplished with effort (Bargh, 1984; Schneider & Shiffrin, 1977), but errors occur if an individual cannot meet the mental demands required to effectively complete the tasks (Paas & van Merriënboer, 1993). Additionally, during complex tasks, there is an increase in cognitive demand; thus, the amount of mental effort required also increases (Kleider-Offutt *et al.*, 2016). The attentional demands required to perform complex tasks may lead to cognitive load and errors, or a reduction in performance (Engle & Kane, 2004; O'Donnell & Eggemeier, 1986).

Cognitive load theory (CLT) identifies three types of load (Sweller, 1988, 1994; Sweller *et al.*, 1998) that are relevant in a variety of applied settings (Galy *et al.*, 2018). The first

---

<sup>1</sup> Vulnerable witnesses for the purpose of this paper include witnesses or victims who may be vulnerable due to their age, intellectual or communication difficulties, or intimidated witnesses (MoJ, 2011).

type, intrinsic load, relates to the load imposed by the fundamental nature of the information being processed and the natural complexity of the task (Schnotz & Kurschner, 2007). The second, extraneous load, is induced by other external factors, such as time pressure (Galy *et al.*, 2012). The third type of load described within CLT is germane load, which is the load used for learning, the development of skills, and the application of skills in a novel situation (Paas, *et al.*, 2004). Notably, germane load is required for the construction and automation of schemas for a particular task (Galy, *et al.*, 2018).

### **Cognitive load in investigative interviews**

For investigative interviewers, there are several inherent (i.e., intrinsic) features of interviewing that may contribute to a cognitive load, including the generation of questions, identifying topics to pursue, and seeking clarification from interviewees. Interviewers are required to actively listen to, and accurately remember, what interviewees are saying (Fisher *et al.*, 2014). They may also be required to take notes and formulate hypotheses to account for the events described. As such, interviewers must attend to multiple cognitive processes (Kleider-Offutt *et al.*, 2016). At the same time, they are required to adhere to best practice guidance, such as building rapport and forming appropriate questions (Hanway & Akehurst, 2018).

Open questions typically lead to detailed, free narrative responses from interviewees (Dale *et al.*, 1978; Hershkowitz, 2001). Hence, asking open questions is an important feature of an investigative interview (Danby *et al.*, 2017). Interviewers must then accurately remember the often-numerous details provided by interviewees but interviewers' recall of information may be limited and inaccurate (Hyman-Gregory, 2009). The interviewer may introduce this erroneous information to the witness, which may have an impact on the subsequent accuracy and reliability of the witness's testimony (Gudjonsson, 2010; Loftus & Pickrell, 1995). In doing this, interviewers can affect the amount and quality of evidence provided by witnesses (Brown & Lamb, 2015; Gudjonsson, 2010).

In sum, obtaining accurate and detailed accounts from witnesses during investigative interviews can be difficult (Hope & Gabbert, 2019; La Rooy & Dando, 2010). Interviewers hold information provided by witnesses in their memory, whilst at the same time assessing that information, thinking of questions to ask, and identifying the correct order in which to ask those questions (i.e., which topic to ask questions about first; Hanway & Akehurst, 2018). The complex cognitive functions required to complete these tasks are likely to have an impact on interviewers' performance and their judgements (Ask & Landstrom, 2010; Nordstrom *et al.*, 1996).

### **The current research**

The current research examined the effect of increased cognitive demands on participants' perceived cognitive load during a mock interview task. The tasks for each condition were designed to replicate the cognitive demands present during an investigative interview (i.e., to listen to the witness, remember information, judge information, and think of questions to ask; Fisher *et al.*, 2014; Hanway & Akehurst, 2018). We explored the effect of increased cognitive demands on the amount and accuracy of information recalled from a witness's statement by participants who took on the role of interviewers.

Based on previous cognitive load research (e.g., Dias *et al.*, 2018; Nordstrom *et al.*, 1996), we hypothesized that during the interview and recall tasks, participants in a high cognitive load (HCL) condition would report higher perceived cognitive load (PCL) compared to those in a moderate cognitive load (MCL) condition, who would report higher PCL than those in a no cognitive load (NCL) condition. Second, we hypothesized that participants in the HCL condition would recall fewer details, and would have a lower accuracy rate for their free recall of a witness's statement, than those in the MCL condition, who would recall fewer details and have a lower accuracy rate than those in the NCL condition. Third, we predicted that participants in the HCL condition would have a lower percentage accuracy score when answering questions about a witness's statement than those in the MCL condition, who would have lower percentage accuracy score when answering questions about a witness's statement than those in the NCL condition.

## Method

### Design

For this independent-groups study, there was one between-subjects factor, cognitive load, with three levels: high cognitive load (HCL); moderate cognitive load (MCL); and no cognitive load (NCL; control). The dependent variables were perceived cognitive load (PCL), the amount and accuracy of statement details provided by participants during free recall, and the accuracy of their cued recall.

### Participants

A priori G\*power analysis (Faul *et al.*, 2009) for an omnibus one-way ANOVA with three groups indicated that a sample size of 102 participants was required. This was based on power = 0.95, a large effect size of  $f = 0.40$ , and the traditional alpha = .05. A large effect on recall accuracy was predicted on the basis of research showing large effects of working memory capacity on memory accuracy (e.g., Jarrold *et al.*, 2011) and large effects of cognitive load on recall accuracy for the spoken word (e.g., Hunter & Pisoni, 2018).

102 participants, staff and students, were recruited via a university participant pool and workplace advertisements at the university. Participants were invited to take part in a study that examined what it is like to be an investigative interviewer. No monetary incentives were offered to participants, but first year undergraduate psychology students were offered one course credit for their participation. Participants attended for one test session, which lasted approximately 45 minutes. Only adults with English as a first or primary language were recruited. The aim of the study was to assess participants' recall of information provided by a witness, when under varying degrees of cognitive load. Therefore, as experience can have an impact on task performance when under cognitive load (Paas, *et al.*, 2004), prior investigative interviewing experience was an exclusion criterion.

The sample comprised 68 females and 34 males. Participants were aged 18 to 71 years ( $M_{\text{age}} = 25.95$  years,  $SD = 10.02$ , the median age was 22 years). To ensure equal numbers of participants ( $N = 34$ ) in each condition, they were pseudo-randomly allocated to one of the three conditions (HCL, MCL, NCL). Data from one participant were removed from the analysis as their responses suggested a poor understanding of the task and a z-score for accuracy rate of the witness's account was an outlier at  $-3.41$  (Field, 2013). Data from two further participants were removed due to recording equipment

failure. The final sample, therefore, comprised 99 participants who were aged 18 to 71 years ( $M_{\text{age}} = 26.03$  years,  $SD = 10.09$ , median age = 22 years)<sup>2</sup>. There were 67 females and 32 males. For the final analyses, there were 34 participants in the high cognitive load (HCL) condition, 33 in the moderate cognitive load (MCL) condition, and 32 in the no cognitive load (NCL) condition.

## Materials

### Stimulus event

To enable an accurate reflection of a real-world interview, the interview room setting, interview procedure, and recording of the interview were designed to correspond with published guidance for interviewing child witnesses (MoJ, 2011). An eight-year-old child witness was interviewed about an event she had experienced (a recent birthday party). The witness was given an open prompt by the interviewer (i.e., '*Please tell me everything you can remember about the party you went to*'). This question and the witness's subsequent free recall were digitally recorded. The recording of the interview captured a head and shoulders view of the witness. The child's recorded free recall account lasted for 6 minutes and 30 seconds.

### Perceived cognitive load measure

To measure participants' perceived cognitive load, the National Aeronautics and Space Administration, Task Load Index (NASA-TLX) was used. This questionnaire combines information about the magnitude and source of six related factors to derive a sensitive and reliable estimate of workload (Hart & Staveland, 1988).

The NASA-TLX uses a multi-dimensional rating scale questionnaire to evaluate participants' subjective ratings of mental workload; the scale items are mental demand, physical demand, temporal demand, performance, effort, and frustration. These items were selected following analysis of the primary factors that do (and do not) define a subjective experience of workload (Hart, 2006). Each item is measured on a 20-point scale from low to high (except for performance which is measured on a scale from good to poor). A weighted score is obtained by completing 15 pairwise comparisons of the six scale items. For each pair, one item is selected that is more relevant for the participant when completing the task (Hart & Staveland, 1988). For this study, and following the scoring procedure devised by Hart and Staveland (1988), a PCL score out of 100 was calculated by multiplying each scale item score (rating score) by the number of times that item was selected in the pairwise comparisons (adjusted score); the six weighted item scores were then totalled and divided by 15 to obtain an overall PCL score. The NASA-TLX was designed to be used during, or immediately after, a task and has been widely used in a variety of settings to measure the cognitive load perceived by participants when they complete a task (e.g., Hart, 2006; Rizzo *et al.*, 2016).

---

<sup>2</sup> Two participants in the study were aged 71 years. All other participants were aged 18 to 54 years. There was no significant difference in age between the three conditions,  $F(2, 96) = 1.25$ ,  $p = .293$ ,  $\eta^2_p = .03$ . Analyses were conducted with and without these two participants' data, which revealed no differences in the results. For completeness, the data of all participants, including those aged 71 years, were included in the analyses and reported results.

### Procedure

After reading the information sheet and providing written informed consent, participants were allocated to one of the three conditions: HCL, MCL, or NCL. The lead author conducted the research and followed written instructions for all conditions. The experimenter was aware of each participant's condition. To reduce experimenter effects, instructions for each condition were read out verbatim from a written script and all questions were asked verbatim from a prepared script. All participants were instructed to take the role of a police interviewer and were informed that a child had witnessed an event, which the participant needed to investigate. Participants were asked to watch and listen to the witness's recorded interview and were informed that they would be asked some questions after they had watched the interview. In the HCL condition, participants were given the following additional instructions, '*Whilst watching the interview, I would like you to consider carefully what the witness is telling you so that you clearly understand the witness's experience of the event she is describing. Your other task is to identify follow-up questions to ask the witness once she has given her statement. So, whilst you are listening to the child, please think about the wording of your questions and in what order the questions should be asked*'. In the MCL condition, participants were given the following additional instructions, '*Whilst watching the interview, I would like you to consider carefully what the witness is telling you, so that you clearly understand the witness's experience of the event she is describing*'. In the NCL (control) condition, no further instructions were given to participants.

After receiving their specific instructions, all participants watched the recorded interview on a computer screen wearing headphones to reduce distractions. Immediately after watching the interview with the child witness, all participants completed the first PCL measure (i.e., they recorded their perceived cognitive load during the interview task, using the NASA-TLX scale presented via an android tablet application). Participants then carried out a 15-minute distraction task, which required them to work through some unrelated number puzzles.

Following the distraction task, participants were asked to recall as much information, in as much detail as they could, from the witness's recorded statement. After participants finished their free recall, they were asked if there was anything further they could recall about the interview. Once participants had completed the free recall task, they were asked 40 cued recall questions about the content of the witness's interview (e.g., '*What did the witness say was 'quite tricky'?*'; '*Who drove the witness home?*'). The order of these questions was randomized across participants. All participants were audio-recorded whilst they gave their free narrative and answered the cued recall questions.

Participants then completed a second self-report of their PCL for the recall task (i.e., their perceived cognitive load when they were recalling the child's statement and answering the 40 questions). This was again completed using the NASA-TLX scales.

For completeness, as participants in the HCL condition had been asked to think about questions to ask the witness, we then asked them to write down 10 follow-up questions they would ask the witness if they were the investigator in the case. To ensure all participants completed the same tasks, those in the MCL and NCL conditions were also asked to write down 10 questions they would like to ask the witness<sup>3</sup>.

---

<sup>3</sup> Mean time (in seconds) for writing down 10 follow-up questions; HCL,  $M = 256.03$  ( $SD = 72.64$ ); MCL,  $M = 260.81$  ( $SD = 72.55$ ); NCL,  $M = 285.18$  ( $SD = 65.39$ ). The differences were not significant  $F(2, 96) = 1.61$ ,  $p = .205$ ,  $\eta^2_p = .03$ .

Finally, participants were asked to rate, using 7-point scales their confidence in their memory accuracy, from [1] not at all confident to [7] extremely confident; the extent to which they felt motivated to remember the content of the child's interview, from [1] not at all motivated to [7] extremely motivated; the extent to which they found remembering the child's statement easy or difficult, from [1] very easy to [7] very difficult; and the extent to which they found coming up with questions easy or difficult, from [1] very easy to [7] very difficult. Participants in the HCL condition were also asked to rate how motivated they were to think about questions whilst they were listening to the child's statement, from [1] not at all motivated to [7] extremely motivated.

As a manipulation check, participants were then asked to write down the instructions they were given by the researcher before they watched the child's account. Demographic details including age and gender were also recorded. A verbal debrief was provided for all participants and they were thanked for their time and effort.

## Coding

### *Free recall coding*

Verbatim transcripts of the participants' audio-recorded free recall of the witness's statement were coded for quantity and accuracy of details reported. Details were coded as person, action, object, setting, or temporal details. For example, participant accounts were coded as follows 'Amelia (1-person) trotted (1-action) on her horse (1-object) in the stables (1-setting)'. If the participant mentioned a detail relating to time (e.g., 'at the end of the day'), it was coded as a temporal detail. Consistent with previous research and to facilitate assessment of overall accuracy, details were coded as correct, incorrect, or confabulations (Wright & Holliday, 2007). A detail was deemed (1) correct, if it was present in the witness's account and was correctly reported by the participant (e.g., 'she was called Amelia'); (2) incorrect, if a reported detail was discrepant from the witness's account (e.g., participant recalls 'pull the reins back to go' but the witness actually said 'pull the reins back to stop'); and (3) confabulated, if a reported detail was mentioned in the participant's account which was not mentioned at all by the witness (e.g., the participant reported 'they got into a car' but the witness did not mention a car at all during her account). Accuracy rate for the free recall accounts was calculated by dividing the total number of correct details reported by the total number of details reported (i.e., correct plus incorrect plus confabulations). Additionally, to assess indicators of uncertainty in participants' recall of the witness's account, ambiguities were coded (e.g., 'I'm not sure, it was something like...').

Inter-coder reliability for the free recall accounts was assessed by selecting 20 interview transcripts (20%), which were coded by an independent scorer. Intra-class correlation coefficients (ICC) using absolute agreement were computed for the following measures: total details [ $r(19) .97, p < .001$ ]; correct details [ $r(19) .95, p < .001$ ]; incorrect details [ $r(19) .83, p < .001$ ]; confabulations [ $r(19) .90, p < .001$ ]; ambiguities [ $r(19) .84, p < .001$ ]. This analysis indicated that the inter-coder reliability was 'good' for the coding of incorrect details and ambiguities, and 'excellent' for the coding of total details, confabulations, and correct details (Koo & Li, 2016).

### Cued recall coding

Answers to 40 cued recall questions were scored as fully correct (e.g., in relation to the location of the event, 'Pink Mead Farm': 2 points), partially correct (e.g., 'Mead stables': 1 point), don't know response (0 points), and incorrect (e.g., 'Crofton stables': -1 point). Total accuracy could therefore range from -40 (all questions answered incorrectly) to 80 (all answers fully correct). The scores were added, and a percentage accuracy score for each participant was calculated.

## Results

### Manipulation check

All 99 participants passed the manipulation check and accurately reported their instructions. As per their instructions, participants in the NCL condition confirmed they were required to watch the interview carefully and participants in the MCL condition confirmed they were to watch the interview and consider what the witness was saying. Participants in the HCL condition confirmed that they were asked to think of questions to ask the witness, as if they were the interviewer in the case, and to watch the interview carefully.

### Hypothesis testing

To examine our hypotheses, we conducted a series of between-groups ANOVAs.

#### Perceived cognitive load

For the 'encoding of interview' task that the participants were first asked to undertake, Levene's test indicated that the assumption of homogeneity of variance for PCL scores had been violated,  $F(2, 96) = 3.94, p = .023$ . Therefore, the more robust Welch equality of means test was examined. As predicted, there was a significant difference in PCL scores between the three conditions;  $F(2, 62.10) = 7.70, p = .001$ , with a large effect size,  $\eta^2_p = .20$  (see Table 1). Tukey HSD post-hoc comparisons showed there was no significant difference between PCL scores for participants in the HCL and MCL conditions ( $p = .209$ ). However, participants in the HCL and MCL conditions scored higher for PCL than those in the NCL condition (HCL,  $p < .001$ ; MCL,  $p = .033$ ). For the 'recall' task, there was no significant difference between the three conditions in terms of PCL scores,  $F(2, 96) = 1.21, p = .304, \eta^2_p = .02$  (see Table 1).

**Table 1.** Mean PCL scores during the 'encoding of interview' and 'recall' tasks for each condition

Condition	PCL for 'encoding the interview' task		PCL for 'recall' task	
	M (SD)	95% CI	M (SD)	95% CI
HCL	53.08 (14.89)	[47.88, 58.27]	68.26 (12.71)	[63.35, 68.35]
MCL	46.21 (13.73)	[41.34, 51.08]	65.64 (12.11)	[61.34, 69.93]
NCL	35.77 (20.28)	[28.46, 43.08]	63.50 (12.66)	[58.94, 68.07]

*Free recall*

With respect to the total number of free recall details reported about the witness's statement, there were no significant differences between the three experimental conditions,  $F(2, 96) = 2.20, p = .117, \eta^2_p = .04$  (see Table 2). In terms of accuracy rate of the details recalled, there was a difference between the three conditions with a large effect size,  $F(2, 96) = 8.54, p < .001, \eta^2_p = .15$ . Post-hoc comparisons of percentage accuracy indicated that there was no significant difference in percentage accuracy for participants in the HCL condition compared with those in the MCL condition ( $p = .476$ ). However, percentage accuracy for participants in the HCL condition was lower than for those in the NCL condition, ( $p < .001$ ). Accuracy was also lower for those in the MCL condition compared with those in the NCL condition ( $p = .015$ ), as shown in Table 2. For details of mean scores for correct details, incorrect details, confabulations, and ambiguity, see the Supplementary Materials.

*Cued recall questions*

For the accuracy of cued recall question responses, there was a difference between the three conditions for percentage accuracy score, with a large effect size,  $F(2, 96) = 7.87, p = .001, \eta^2_p = .14$ . Tukey HSD post-hoc comparisons indicated that percentage accuracy score for participants in the HCL condition was not significantly different from those in the MCL condition ( $p = .114$ ). The percentage accuracy score for participants in the MCL condition was also not significantly different from those in the NCL condition ( $p = .130$ ). However, percentage accuracy score for participants in the HCL condition was significantly lower than for those in the NCL condition ( $p < .001$ ; see Table 3). For details of mean scores for correct, partially correct, incorrect, and don't know responses, see the Supplementary Materials.

*Motivation, confidence, and task difficulty*

A series of Pearson's correlations were calculated to determine whether the dependent variables of motivation, confidence, and task difficulty were correlated with each other. There were significant, but moderate, correlations between the majority of variables (see the Supplementary Materials). Therefore, the assumption of an absence of multicollinearity was met, and to reduce type 1 error, a one-way between-groups MANOVA was conducted to investigate differences between the conditions for participants' motivation, confidence, and how difficult they found the tasks. The MANOVA indicated that there was no significant multivariate effect: Wilks'  $\lambda = .95, F(8, 186) = .62, p = .764, \eta^2_p = .03$  (for

**Table 2.** Total number of details recalled and accuracy rate for each condition during the free recall task

Condition	Total details recalled		Accuracy rate	
	M (SD)	95% CI	M (SD)	95% CI
HCL	116.5 (44.19)	[101.4, 131.6]	0.91 (0.39)	[0.90, 0.93]
MCL	137.3 (32.77)	[125.7, 148.9]	0.92 (0.42)	[0.91, 0.94]
NCL	134.0 (52.90)	[115.0, 153.1]	0.95 (0.37)	[0.94, 0.97]

**Table 3.** Percentage accuracy scores for the question task across the three conditions

Condition	<i>M (SD)</i>	95% CI
HCL	48.90 (14.59)	[43.81, 53.99]
MCL	55.91 (13.62)	[51.08, 60.74]
NCL	62.81 (14.52)	[57.58, 68.05]

details of scores across each of the dependent variables for each condition, see the Supplementary Materials). There were no significant differences at the univariate level.

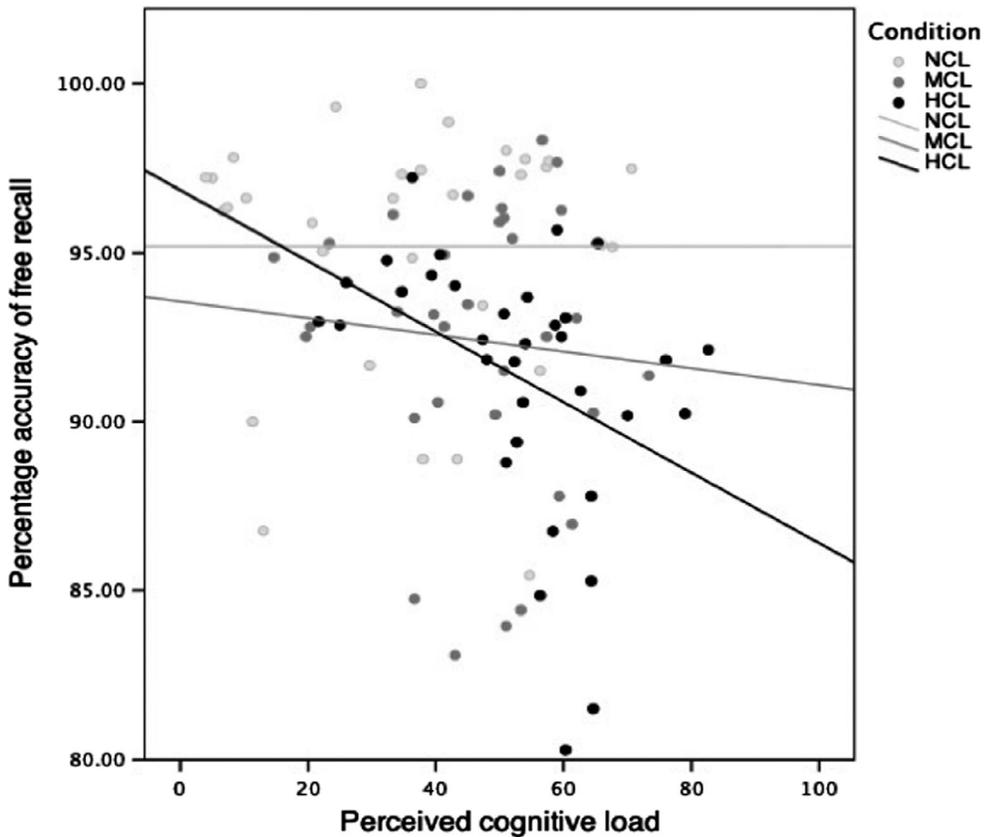
### Exploratory analysis

As our confirmatory analysis showed that increased cognitive demand for participants in the HCL and MCL conditions was associated with increased perceived cognitive load during the ‘encoding the interview’ task and also a lower recall accuracy for the free recall and question tasks, we conducted further exploratory analyses. A Pearson’s correlation showed that there was a relationship between PCL and accuracy of free recall,  $r = -.279$ ,  $p = .003$ . When the sample was split by condition, a linear regression analysis indicated that in the HCL condition, PCL was a predictor of participants’ free recall accuracy rate ( $\beta = -.40$ ,  $p = .018$ ) accounting for 16% of the variance. However, PCL was not a predictor of free recall accuracy for participants in the MCL ( $\beta = -.08$ ,  $p = .653$ ) or NCL conditions ( $\beta < .001$ ,  $p = .1.00$ ) (see Figure 1). PCL was also not a predictor of cued recall percentage accuracy scores across any of the conditions (HCL,  $\beta = -.042$ ,  $p = .815$ ; MCL,  $\beta = -.121$ ,  $p = .502$ ; NCL,  $\beta = -.047$ ,  $p = .797$ ).

### Discussion

We examined the effects of increased cognitive demands on perceived cognitive load and subsequent recall of an interviewee’s account in a mock investigative interviewing task. As predicted, participants who were required to complete tasks that are intrinsic to investigative interviewing (i.e., listening, remembering, judging the information provided, and generating follow-up questions to ask) perceived a higher cognitive load than did participants who were required to complete tasks with fewer cognitive demands (i.e., merely watching and listening to a witness’s statement). Participants who were asked to complete more cognitively demanding tasks were less accurate, when freely recalling information provided by the witness, than those who were asked to perform less cognitively demanding tasks. Additionally, when asked cued questions about the witness’s account, interviewees who completed more demanding cognitive tasks than those asked to perform fewer cognitively demanding tasks whilst watching the interview, provided less accurate responses. Taken together, these results suggest that the demands placed on the participants’ cognitive resources when carrying out the multiple tasks of an investigative interview resulted in a reduction in performance on the tasks.

In exploratory analyses, we found a relationship between PCL and recall accuracy rate. When participants’ scores for the three conditions were examined separately, we found the relationship was moderated by the tasks undertaken by participants (i.e., for the HCL condition, higher levels of perceived cognitive load predicted performance in terms of



**Figure 1.** Linear regression analysis with PCL as a predictor of free recall accuracy rates for each condition (HCL; MCL; NCL). Data points for the three groups are indicated with their associated line of best-fit plot.

free recall accuracy). When more controlled and focused attention was required for the task of generating questions to ask, there was an increase in perceived cognitive load and a reduction in performance. The reduction in recall performance may have been due to a limited capacity to carry out multiple cognitive tasks in working memory (Kahneman, 1973; Reisberg, 2007). However, more automatic processes (i.e., listening and watching the witness) were less affected by cognitive load (Schneider & Shiffrin, 1977). This research provides the first empirical evidence that increased cognitive demands inherent in an investigative interviewing task result in higher perceived cognitive load as well as reduced recall performance for participants.

For the current experimental task, which was designed to reflect real-world interviewing procedures, participants were asked to focus on certain intrinsic features of interviewing, including listening, remembering information, and thinking of questions to ask. Whilst our experimental design included a manipulation of cognitive load based on realistic processes for interviewers, we recognize that investigative interviewing in the field is a complex task and likely requires more cognitive processing than was required for our participants. In practice, interviewers are required to build rapport, interact with the witness, and consider other aspects of the case (Schreiber-Compo *et al.*, 2012).

Interviews, therefore, occur in a social context, whereby interviewers also perceive witnesses' actions and make judgements about their credibility, reliability, and well-being (Ask & Landstrom, 2010; Hanway & Akehurst, 2018). These extraneous factors, and that of time pressure (i.e., temporal demand), were not present during the current study. However, cognitive load is additive (Leppink *et al.*, 2015). Therefore, the additional factors identified as present when conducting investigative interviews will likely contribute to a higher cognitive load for interviewers in practice (Hanway & Akehurst, 2018; Nordstrom *et al.*, 1996).

Cognitive load theory suggests that automatic processing relies on schemas to reduce effort (Paas *et al.*, 2004). With training, and skill development, more schemas are potentially built. However, if a task is cognitively demanding, and the intrinsic and extraneous load exceeds capacity, then there is little opportunity to form these schemas (Schnotz & Kurchner, 2007). Cognitive load, therefore, may also have an impact on interviewers' skill development. It may be that, despite their training and knowledge of best practice guidance, the intrinsic and extraneous cognitive demands imposed on investigative interviewers each time they conduct a unique interview leaves little capacity for building schemas. Consequently, interviewers are not afforded the opportunity to rely on more automatic processing and they experience significant cognitive load. Thus, interviewers do not always comply with their training (CJJI, 2014; Cross & Hershkowitz, 2017; Powell & Barnett, 2015).

For this study, our aim was to examine the effect of holding information in mind whilst judging that information and thinking of questions to ask a child witness. We also aimed to reduce extraneous load not directly related to the task. Note-taking can be cognitively demanding in itself and may divide attention between listening to the witness, formulating questions, and recording information (Piolat *et al.*, 2005; Schreiber-Compo *et al.*, 2012). Therefore, in the HCL condition, participants were not permitted to note down the questions they were thinking about whilst they were listening to the child. An inevitable limitation of this design was that we could not be sure what participants were thinking during their task. To mitigate this limitation, and to ensure participants had understood their instructions, we included a manipulation check after the recall phase to check participants' understanding of what they had been asked to do. Future research might examine the effects of note-taking for the interviewer.

Whilst the design of this study replicated the cognitive demands experienced by interviewers during real-world interviews, a limitation is that our participants were novice interviewers, who had not received any training in investigative interviewing. As such, the current findings may have limited generalizability to trained or experienced interviewers. However, interviewers in the real world are also required to think about, and comply with, their training when undertaking interviews, which may increase their cognitive load (Hanway & Akehurst, 2018; Schreiber Compo *et al.*, 2012). Considering this, and the additional intrinsic and extraneous factors, it is possible that interviewers in the field will experience more cognitive load than the novice participants in our study. In turn, interviewers' performance in the field may be impacted to a greater extent than was the case for participants in the current experiment. Further research should focus on aspects of investigative interviewing in context. It would be interesting to explore the impact that training and experience have on interviewers' cognitive load as well as the effects of cognitive load on other aspects of interviewer performance, such as the types of questions asked. As some of the variation seen in the current study may be accounted for by individual differences in cognitive ability, this may also be an interesting area for further research, for example, individual differences in working memory capacity (Engle, 2002).

Finally, the sample size estimation may also be a limitation for this study. The sample size was based on a predicted large effect size, which has practical relevance in an applied setting. We considered the approach to be appropriate and in line with similar research in the investigative interviewing literature (e.g., Hoogesteyn *et al.*, 2020; Kontogianni *et al.*, 2018). However, given the sample sizes in each condition ( $N = 32, 33$  and  $34$ ), a larger sample would be needed to detect smaller effects, and significant differences between conditions, in the post-hoc analyses.

The current findings suggest that the cognitive demands required to complete an investigative interview can lead to an increased cognitive load and a reduction in recall accuracy of what was said by an interviewee, which may have an impact on interviewers' questioning and compliance with recommended interviewing practices. Providing interviewers with the opportunity to develop and practise their techniques, so that skills relating to interviewing become more automatic, along with better management of factors which may contribute to additional cognitive load, such as time pressure, may help interviewers to better cope with the high cognitive demands of investigative interviewing.

## Acknowledgements

We thank Charlotte Hudson for her contribution in assessing inter-rater reliability. This research forms part of the doctoral work of the first author who is in receipt of a doctoral studentship funded by the ESRC, South Coast Doctoral Training Partnership (Award: ES/P000673/1). The funding source had no involvement in the design of the study, the conduct of the research, preparation of this article, or the decision to submit this article for publication.

## Conflicts of interest

The authors declare no conflict of interest.

## Author contributions

First and second authors (conception of the research idea). First and second authors (design of the research with feedback and reviews from the third and fourth authors). First author (conduction of the research, analysis and interpretation of the data, and writing of the research paper). Second, third, and fourth authors (providing feedback on the research paper).

## Data availability statement

As the PI's research is being funded by UKRI/ESRC, the data that support the findings of this study will be archived on the UK Data Service website with a secure password protected account. The data are available from the corresponding author upon reasonable request.

## References

- Ask, K., & Landström, S. (2010). Why emotions matter: Expectancy violation and affective response mediate the emotional victim effect. *Law and human behavior*, *34*(5), 392–401.

- Bargh, J. A. (1984). Automatic and conscious processing of social information. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition*, Vol. 3 (pp. 1–43). Lawrence Erlbaum Associates Publishers.
- Brown, D. A., & Lamb, M. E. (2015). Can children be useful witnesses? It depends how they are questioned. *Child Development Perspectives*, 9(4), 250–5. <https://doi.org/10.1111/cdep.12142>
- Bull, R. (2010). The investigative interviewing of children and other vulnerable witnesses: Psychological research and working/professional practice. *Legal and Criminological Psychology*, 15(1), 5–23. <https://doi.org/10.1348/014466509X440160>
- Bull, R., & Soukara, S. (2009). A set of studies of what really happens in police interviews with suspects. In G. D. Lassiter, & C. A. Meissner (Ed.), *Interrogations and confessions: Research, practice, and policy*. American Psychological Association.
- Criminal Justice Joint Inspectorate (CJJI) (2014). Achieving best evidence in child sexual abuse cases: A joint inspection. [https://www.justiceinspectors.gov.uk/cjji/wp-content/uploads/sites/2/2014/12/CJJI\\_ABE\\_Dec14\\_rpt.pdf](https://www.justiceinspectors.gov.uk/cjji/wp-content/uploads/sites/2/2014/12/CJJI_ABE_Dec14_rpt.pdf)
- Cross, T. P., & Hershkowitz, I. (2017). Psychology and child protection: Promoting widespread improvement in practice. *Psychology, Public Policy, and Law*, 23(4), 503. <https://doi.org/10.1037/law0000141>
- Dale, P. S., Loftus, E. F., & Rathbun, L. (1978). The influence of the form of the question on the eyewitness testimony of preschool children. *Journal of Psycholinguistic Research*, 7(4), 269–77. <https://doi.org/10.1007/BF01068110>
- Danby, M. C., Sharman, S. J., Brubacher, S. P., Powell, M. B., & Roberts, K. P. (2017). Differential effects of general versus cued invitations on children's reports of a repeated event episode. *Psychology, Crime & Law*, 23(8), 794–811. <https://doi.org/10.1080/1068316X.2017.1324028>
- Dias, R. D., Ngo-Howard, M. C., Boskovski, M. T., Zenati, M. A., & Yule, S. J. (2018). Systematic review of measurement tools to assess surgeons' intraoperative cognitive workload. *British Journal of Surgery*, 105(5), 491–501. <https://doi.org/10.1002/bjs.10795>
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11(1), 19–23. <https://doi.org/10.1111/1467-8721.00160>
- Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *Psychology of Learning and Motivation*, 44, 145–200.
- Engstrom, M., Landtblom, A. M., & Karlsson, T. (2013). Brain and effort: brain activation and effort-related working memory in healthy participants and patients with working memory deficits. *Frontiers in Human Neuroscience*, 7, 140. <https://doi.org/10.3389/fnhum.2013.00140>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G\* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149–60. <https://doi.org/10.3758/BRM.41.4.1149>
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.
- Fisher, R. P., Compo, N. S., Rivard, J., & Hirn, D. (2014). Interviewing witnesses. In T. J. Perfect (Ed.), *The Sage Handbook of Applied Memory* (pp. 559–78). Sage Publications Ltd.
- Galy, E., Cariou, M., & Mélan, C. (2012). What is the relationship between mental workload factors and cognitive load types? *International Journal of Psychophysiology*, 83(3), 269–75. <https://doi.org/10.1016/j.ijpsycho.2011.09.023>
- Galy, E., Paxion, J., & Berthelon, C. (2018). Measuring mental workload with the NASA-TLX needs to examine each dimension rather than relying on the global score: an example with driving. *Ergonomics*, 61(4), 517–27. <https://doi.org/10.1080/00140139.2017.1369583>
- Gudjonsson, G. H. (2010). Psychological vulnerabilities during police interviews. Why are they important? *Legal and Criminological Psychology*, 15(2), 161–75. <https://doi.org/10.1348/135532510X500064>
- Hanway, P., & Akehurst, L. (2018). Voices from the front line: police officers' perceptions of real-world interviewing with vulnerable witnesses. *Investigative Interviewing: Research and Practice*, 9(1), 14–33.

- Hart, S. G. (2006, October). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9), 904–8. <https://doi.org/10.1177/154193120605000909>.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, 52(C), 139–83. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- Hershkowitz, I. (2001). Children's responses to open-ended utterances in investigative interviews. *Legal and Criminological Psychology*, 6(1), 49–63. <https://doi.org/10.1348/135532501168190>
- Hershkowitz, I. (2011). Rapport building in investigative interviews of children. In M. E. Lamb, D. La Rooy, C. Malloy, & C. Katz (Eds.), *Children's testimony: A handbook of psychological research and forensic practice* (2nd ed., pp. 109–28). John Wiley & Sons Ltd.
- Hoogesteyn, K., Meijer, E., & Vrij, A. (2020). Examining witness interviewing environments. *Journal of Investigative Psychology and Offender Profiling*, 1–12. <https://doi.org/10.1002/jip.1549>
- Hope, L., & Gabbert, F. (2019). Developments and innovations in evidence-based investigative interviewing. *Evidence-based Investigative Interviewing: Applying Cognitive Principles*, (42–55). Routledge.
- Hunter, C. R., & Pisoni, D. B. (2018). Extrinsic cognitive load impairs spoken word recognition in high- and low-predictability sentences. *Ear and Hearing*, 39(2), 378–89. <https://doi.org/10.1097/AUD.0000000000000493>
- Hyman Gregory, A. (2009). Investigative Interviewing and Memory: How Accurate Are Interviewers' Recollections of Investigative Interviews? (Doctoral dissertation). Florida International University.
- Jarrold, C., Tam, H., Baddeley, A. D., & Harvey, C. E. (2011). How does processing affect storage in working memory tasks? Evidence for both domain-general and domain-specific effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(3), 688–705. <https://doi.org/10.1037/a0022527>
- Kahneman, D. (1973). *Attention and effort*, Vol. 1063. Prentice-Hall.
- Kassin, S. M., Appleby, S. C., & Perillo, J. T. (2010). Interviewing suspects: Practice, science, and future directions. *Legal and Criminological Psychology*, 15(1), 39–55. <https://doi.org/10.1348/135532509X449361>
- Kleider-Offutt, H. M., Clevinger, A. M., & Bond, A. D. (2016). Working memory and cognitive load in the legal system: Influences on police shooting decisions, interrogation and jury decisions. *Journal of Applied Research in Memory and Cognition*, 5(4), 426–33. <https://doi.org/10.1016/j.jarmac.2016.04.008>
- Kontogianni, F., Hope, L., Taylor, P. J., Vrij, A., & Gabbert, F. (2018). The benefits of a self-generated cue mnemonic for timeline interviewing. *Journal of Applied Research in Memory and Cognition*, 7(3), 454–61. <https://doi.org/10.1016/j.jarmac.2018.03.006>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, 15(2), 155–63. <https://doi.org/10.1016/j.jcm.2016.02.012>
- La Rooy, D., & Dando, C. (2010). Witness interviewing. In G. J. Towl & D. A. Crighton (Eds.), *Forensic psychology* (pp. 195–209). BPS and Blackwell Publishing Ltd.
- Lafontaine, J., & Cyr, M. (2016). A study of the relationship between investigators' personal characteristics and adherence to interview best practices in training. *Psychiatry, Psychology and Law*, 23(5), 782–97. <https://doi.org/10.1080/13218719.2016.1152925>
- Lamb, M. E. (2016). Difficulties translating research on forensic interview practices to practitioners: Finding water, leading horses, but can we get them to drink? *American Psychologist*, 71(8), 710. <https://doi.org/10.1037/amp0000039>
- Lamb, M. E., Brown, D. A., Hershkowitz, I., Orbach, Y., & Esplin, P. W. (2018). *Tell me what happened: Questioning children about abuse* (2nd ed.). Wiley-Blackwell.

- Leppink, J., & van den Heuvel, A. (2015). The evolution of cognitive load theory and its application to medical education. *Perspectives on Medical Education, 4*(3), 119–27. <https://doi.org/10.1007/s40037-015-0192-x>
- Loftus, E. F., & Pickrell, J. E. (1995). The formation of false memories. *Psychiatric Annals, 25*(12), 720–5. <https://doi.org/10.3928/0048-5713-19951201-07>
- Milne, B., & Bull, R. (1999). *Investigative interviewing: Psychology and practice*. Wiley.
- Ministry of Justice (2011). Achieving Best Evidence in Criminal Guidance on interviewing victims and witnesses, and guidance on using special measures, (March), 68–99. <http://www.justice.gov.uk/downloads/victims-and-witnesses/vulnerable-witnesses/achieving-best-evidence-criminal-proceedings.pdf>
- Nordstrom, C. R., Williams, K. B., & LeBreton, J. M. (1996). The effect of cognitive load on the processing of employment selection information. *Basic and Applied Social Psychology, 18*(3), 305–18. [https://doi.org/10.1207/s15324834basps1803\\_4](https://doi.org/10.1207/s15324834basps1803_4)
- O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance, Vol. 2. Cognitive processes and performance* (pp. 1–49). John Wiley & Sons.
- Orbach, Y., Hershkowitz, I., Lamb, M. E., Sternberg, K. J., Esplin, P. W., & Horowitz, D. (2000). Assessing the value of structured protocols for forensic interviews of alleged child abuse victims. *Child Abuse & Neglect, 24*(6), 733–52. [https://doi.org/10.1016/S0145-2134\(00\)00137-X](https://doi.org/10.1016/S0145-2134(00)00137-X)
- Oxburgh, G., Myklebust, T., Grant, T., Milne, R. (Eds.) (2015). *Communication in investigative and legal contexts: Integrated approaches from forensic psychology, linguistics and law enforcement*. John Wiley & Sons.
- Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science, 32*(1), 1–8. <https://doi.org/10.1023/B:TRUC.0000021806.17516.d0>
- Paas, F. G., & Van Merriënboer, J. J. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors, 35*(4), 737–43. <https://doi.org/10.1177/001872089303500412>
- Piolat, A., Olive, T., & Kellogg, R. T. (2005). Cognitive effort during note taking. *Applied Cognitive Psychology, 19*(3), 291–312. <https://doi.org/10.1002/acp.1086>
- Powell, M. B. (2002). Specialist training in investigative and evidential interviewing: Is it having any effect on the behaviour of professionals in the field? *Psychiatry, Psychology and Law, 9*(1), 44–55. <https://doi.org/10.1375/pplt.2002.9.1.44>
- Powell, M. B., & Barnett, M. (2015). Elements underpinning successful implementation of a National Best Practice Child Investigative Interviewing Framework. *Psychiatry, Psychology and Law, 22* (3), 368–77. <https://doi.org/10.1080/13218719.2014.951112>
- Reisberg, D. (2007). *Cognition: exploring the science of the mind*. W.W. Norton.
- Rizzo, L., Dondio, P., Delany, S. J., & Longo, L. (2016, September). Modeling mental workload via rule-based expert system: a comparison with NASA-TLX and workload profile. In IFIP International Conference on Artificial Intelligence Applications and Innovations (pp. 215–29). [https://doi.org/10.1007/978-3-319-44944-9\\_19](https://doi.org/10.1007/978-3-319-44944-9_19)
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*(1), 1. <https://doi.org/10.1037/0033-295X.84.1.1>
- Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review, 19*(4), 469–508. <https://doi.org/10.1007/s10648-007-9053-4>
- Schreiber Compo, N., Hyman Gregory, A., & Fisher, R. (2012). Interviewing behaviors in police investigators: A field study of a current US sample. *Psychology, Crime & Law, 18*(4), 359–75. <https://doi.org/10.1080/1068316X.2010.494604>
- Strayer, D. L., & Drews, F. A. (2007). Cell-phone-induced driver distraction. *Current Directions in Psychological Science, 16*(3), 128–31. <https://doi.org/10.1111/j.1467-8721.2007.00489.x>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12* (2), 257–85. [https://doi.org/10.1016/0364-0213\(88\)90023-7](https://doi.org/10.1016/0364-0213(88)90023-7)

- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295–312. [https://doi.org/10.1016/0959-4752\(94\)90003-5](https://doi.org/10.1016/0959-4752(94)90003-5)
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10(3), 251–96. <https://doi.org/10.1023/a:1022193728205>
- Van Acker, B. B., Parmentier, D. D., Vlerick, P., & Saldien, J. (2018). Understanding mental workload: from a clarifying concept analysis toward an implementable framework. *Cognition, Technology and Work*, 20(3), 351–65. <https://doi.org/10.1007/s10111-018-0481-3>
- Van Merriënboer, J. J., & Sweller, J. (2010). Cognitive load theory in health professional education: design principles and strategies. *Medical Education*, 44(1), 85–93. <https://doi.org/10.1111/j.1365-2923.2009.03498.x>
- Wright, A. M., & Holliday, R. E. (2007). Enhancing the recall of young, young–old and old–old adults with cognitive interviews. *Applied Cognitive Psychology*, 21(1), 19–43. <https://doi.org/10.1002/acp.1260>

Received 16 April 2020; revised version received 21 August 2020

### Supporting Information

The following supporting information may be found in the online edition of the article:

**Table S1.** Mean correct, incorrect, confabulations, and ambiguity, free recall scores for each condition.

**Table S2.** Mean correct, partially correct, incorrect, and don't know, cued-recall scores for each condition.

**Table S3.** Pearson correlations, Means and Standard Deviations associated with confidence motivation and task difficulty.

**Table S4.** Questionnaire scores for each condition.