A Meta-Analytic Review of the Self-Administered Interview©: Quantity and Accuracy of Details Reported on Initial and Subsequent Retrieval Attempts

Ruth Horry*, Swansea University

Chelsea Hughes, Swansea University

Anagha Sharma, Swansea University

Fiona Gabbert, Goldsmiths, University of London

Lorraine Hope, University of Portsmouth

*Corresponding author. Dr Ruth Horry, Department of Psychology, College of Human and Health Sciences, Swansea University, SA2 8PP, UK. Email: r.horry@swansea.ac.uk.

This paper is accepted for publication at *Applied Cognitive Psychology*. This pre-print may not exactly duplicate the final published version.

This research was supported by funding from the Road Safety Trust, grant number RST 51_3_17, awarded to R. Horry, F. Gabbert, and L. Hope, and S. Davies.

Data availability statement: This meta-analysis was pre-registered on the Open Science Framework: https://osf.io/eq4sz. Data and analysis scripts are also available on the OSF project page: https://osf.io/3s5k9/.

Conflict of interest statement: F. Gabbert and L. Hope developed, and own the copyright for the SAI© [along with R. Fisher]. The SAI© is available to researchers and law enforcement practitioners at no cost, so there are no financial conflicts of interest to declare.

Abstract

The Self-Administered Interview (SAI©) is designed to elicit detailed witness reports in the aftermath of incidents. In two sets of meta-analyses, we compared the number of correct details reported, the number of incorrect details reported, and the accuracy of reports provided by witnesses in initial reports (SAI© vs. other reporting formats) and in subsequent accounts (initial SAI© vs. no initial SAI©). The number of comparisons ranged from 15 to 19, (N = 722 to 977). For initial accounts, the SAI© was associated with more correct details and more incorrect details than other reporting formats; accuracy was slightly lower for the SAI© than for other reporting formats. Subsequent accounts were more detailed and accurate for witnesses who had completed an initial SAI© than for those who had not. The SAI© is an effective tool for capturing detailed initial accounts and for preserving witness memory until a formal interview can be conducted.

Keywords: *Self-Administered Interview, meta-analysis, eyewitness memory, investigative interviewing, eyewitness testimony*

The Self-Administered Interview© (SAI©) was developed by Gabbert et al. (2009) to capture detailed eyewitness accounts following a witnessed event. The SAI© was designed for situations in which witnesses cannot be individually interviewed in a timely manner because of resource limitations—for example, at critical incidents involving large numbers of witnesses. In such cases, the delay between the incident and a formal witness interview can be weeks or months. During such delays, lines of enquiry can be lost, and witnesses' memories will be subject to decay and distortion (e.g., Horry et al., 2014; Paz-Alonso & Goodman, 2008). Engaging in a thorough retrieval attempt shortly after encoding can consolidate the memory trace, mitigating against distortion and decay. Securing high-quality early accounts is therefore vital for ensuring the efficient progress of an investigation and for protecting witnesses' memories until they can be formally interviewed. The SAI© was initially field tested in collaboration with police forces in the UK (Hope & Gabbert, 2011). The SAI© has since been translated into multiple languages, and has been adopted by police forces in multiple countries including Norway, the Netherlands, Sweden and Germany. The College of Policing (2019), the professional body for the police service in England and Wales, has recommended the SAI© for use in incidents that involve high numbers of witnesses, and in which it would be possible for those witnesses to complete the SAI© while still at the scene. The SAI© has also been recommended for use by the College of Policing during the COVID-19 pandemic to facilitate obtaining timely initial accounts and supporting delayed interviews.

This meta-analytic review will address two research questions. First, in the period shortly following a witnessed incident, does the SAI© produce more detailed and accurate accounts than other retrieval formats such as cued recall and free recall tests? Second, what is the impact of completing an initial SAI© on the quantity and accuracy of information reported on a later retrieval attempt?

The Self-Administered Interview

The SAI© is designed to support comprehensive retrieval attempts without the need for an interviewer. The witness works their way through a structured recall attempt, following detailed written instructions designed to provide retrieval support and to encourage metacognitive monitoring and control (Koriat & Goldsmith, 1996). The SAI© incorporates the Report Everything and Mental Context Reinstatement components of the Cognitive Interview (CI; Fisher & Geiselman, 1992), which are the most widely used and most effective components of the CI (Dando et al., 2011). After providing a free recall account of the event, the witness is asked to report any additional information about the people involved in the incident, is encouraged to sketch the scene, and is prompted to report any other details about the incident that they may not have thought to include in their account (e.g., quality of view, weather conditions).

Multiple studies have found that the quantity of correct information reported in an initial SAI© is higher than in other reporting formats such as free recall (e.g., Gabbert et al., 2009; Gawrylowicz et al., 2014a; Hope et al., 2014; Krix et al., 2016). Further research has probed the boundary conditions of this effect, showing, for example, that the magnitude of the SAI© benefit on correct details is reduced if the SAI© is administered following a delay of 24 hours or more (Mackay & Paterson, 2015; Paterson et al., 2015), or if attention was divided during encoding (Krix et al., 2014). There is some evidence, however, that the SAI© produces similar benefits under high and low stress (Gittins et al., 2015; Krix et al., 2016). The vast majority of research with the SAI© has focused on neurotypical young and middle-aged adults, though some studies have investigated other populations such as older adults (Dando et al., 2020; Gawrylowicz et al., 2014b) and adults with Autism Spectrum Disorder (Maras et al. 2014).

The increased quantity of correct details reported in the SAI© is typically accompanied by an increase in the quantity of incorrect details (e.g., Gabbert et al., 2009; Gawrylowicz et al., 2014b; Maras et al., 2014). This pattern is also evident in the CI, upon which the SAI© was based (Memon et al., 2010). This increase in both correct and incorrect details in suggestive of a criterion shift, whereby participants relax their threshold for reporting (versus withholding) recalled details (Goldsmith et al., 2002; Koriat & Goldsmith, 1996). In this meta-analysis, we will estimate meta-analytic effect sizes for incorrect details and accuracy as well as for correct details, thus providing a rounded assessment of the impact of the SAI© on report quality, as well as information quantity.

Effects on subsequent retrieval attempts

The SAI© has the potential to protect witnesses' memories from decay and distortion during the retention interval between an event and a face-to-face interview (Gabbert et al., 2009). Gabbert et al. (2009) found that participants who had initially completed the SAI© reported significantly more correct details on a subsequent retrieval attempt than the participants who had not initially completed the SAI©. Subsequent research has replicated the protective effect of SAI© completion on later reports across a range of reporting formats including cued recall (Gittins et al., 2015; Kraus et al., 2017; Mackay & Paterson, 2015), free recall (Gabbert et al., 2012; Gawrylowicz et al., 2014a; Paterson et al., 2015), the CI (Hope et al., 2014), and a second SAI© (Krix et al., 2014). This protective effect of initial SAI© completion on the quantity of correct details reported after a delay appears to be robust even following exposure to misleading information (Gabbert et al., 2012; Gittins et al., 2015; Paterson et al., 2015; Mackay & Paterson, 2015).

Most studies that have examined errors and accuracy on a delayed report have found no significant differences between participants who had and had not completed an initial SAI© (e.g., Gabbert et al., 2012, Experiment 1; Hope et al., 2014; Maras et al., 2014; Paterson et al., 2015). However, it is possible that such effects do exist, but are too small to be reliably detected by the average study. By estimating meta-analytic effect sizes for subsequently reported incorrect details and accuracy rates, we will have a clearer view of the range of plausible effect sizes for these variables.

Aims and scope of meta-analysis

Our first research question focuses on the quantity and accuracy of information provided in initial accounts captured shortly after a witnessed event. We synthesize studies that have compared initial accounts obtained via the SAI© and initial accounts obtained via other means (e.g., cued recall or free recall). Our second research question focuses on the impact of the initial SAI© completion on delayed memory reports. We synthesize studies that have compared the delayed memory reports of participants who had completed an initial SAI© with those who had not completed an initial SAI©. For each research question, we estimate meta-analytic effect sizes for three dependent measures: a) number correct details reported; b) number incorrect details reported; and c) accuracy of information reported.

A secondary aim of this meta-analysis was to identify any potential moderators of the effectiveness of the SAI©. We considered *estimator variables* (Wells, 1978) that might alter the effectiveness of the SAI©, including the participant's age and situational factors such as alcohol intoxication, stress, and violence. These types of variables are important for identifying the types of situations under which the SAI© might be more or less effective. We also considered *system variables* such as the retention interval between an initial SAI© and a subsequent account. System variables can provide important information to practitioners about how a tool such as the SAI© can be optimally used. Finally, we considered *methodological variables* that characterised key ways in which study designs could differ (e.g., the type of control test against which the SAI© was compared, the modality of the in which the critical event was encoded, and the population from which the participants were recruited). These methodological variables could provide theoretical insights, as well as informing assessments of ecological validity and generalizability.

Methods

Protocol and registration

The protocol was pre-registered, following the PRISMA guidelines for transparent meta-analyses (www.prisma.org). The protocol, data, and analysis scripts are available at https://osf.io/3s5k9/?view_only=56f7f5e896304625a95a64c9f88d4cde.

Eligibility criteria

The following inclusion criteria were applied: (i) Participants were randomly allocated to complete the (original, unmodified) SAI© or to a comparison condition. Modified versions of the SAI© such as the SAW-IT© (MacLean et al., 2019) and ad-hoc structured recall questionnaires (e.g., Roos af Hjelmsäter et al., 2012) were not eligible; (ii) Participants observed a staged event (filmed or live) with known ‘ground truth’, allowing for scoring of correct and incorrect details; (iii) The SAI© was completed in the same experimental session as the witnessed event (i.e., within one hour). This mimics the way the SAI© was designed to be used in practice, and in which it has been recommended for use by policing bodies (College of Policing, 2019); (iv) The study reported at least one of three dependent measures (number of correct details, number of incorrect details, accuracy of reported details); and (v) the results were reported in English.

Further, separate criteria applied for research question 1 (initial reports) and research question 2 (subsequent reports). For immediate recall, studies must have included a control group who completed a retrieval attempt (e.g., free recall, cued recall) in the same experimental session as the witnessed event. For delayed recall, studies must have included a control group who had either had no retrieval attempt immediately following the event, or who had completed a different type of retrieval attempt (e.g., free recall, cued recall). In addition, the delayed report had to be in the same format for participants in the initial-SAI© and comparison conditions.

Studies were excluded if they did not meet the above criteria or if they compared the SAI© with the CI. This is because the SAI© is not intended to be a replacement for the CI, but is designed to be used in situations where a CI is not feasible. Thus, comparing the SAI© to the CI is not likely to be informative for policy makers or practitioners.

Search strategy

A systematic search for published articles was performed using Google Scholar, PsychINFO, PsycArticles, and Web of Science. We used the following Boolean search strings: (i) Self-Administered Interview AND memory; (ii) Self-Administered Interview AND *witness. Searches were limited to articles published from 2009, as this marks the first publication of the SAI© (Gabbert et al., 2009). These searches were performed from the 22nd January 2020 to the 6th February 2020, and they returned 522 unique records.

We also searched for unpublished studies. We emailed authors who had previously published research on the SAI©, and we circulated requests for unpublished data through the Society for Applied Research in Memory and Cognition (SARMAC), the International Investigative Interviewing Research Group (iIIRG), the European Association of Psychology and Law (EAPL), and the South East Eyewitness Network (SEEN). We received 11 unpublished data sets.

Study screening

First, the titles and abstracts of the 522 unique records were screened. Articles which could definitively be ruled ineligible were screened (n = 500). To ensure consistency and reliability, the abstracts and titles of the first 50 records were screened independently by the second and third authors. Disagreements were discussed by the research team until consensus was reached. This process was repeated for the next group of 50 articles, until 100%

agreement was reached. In total, 150 articles were screened by two researchers. The remaining records were screened by the second author. The 11 unpublished records were initially screen for relevance prior to escalating to full-text review. Two records were screened out at this point, with the remaining nine undergoing full-text review.

The remaining 22 published articles, along with the nine unpublished studies, were escalated to full text review. Each study was independently screened by the second and third authors against the inclusion and exclusion criteria. The researchers agreed on all but one study, which was then discussed by the lead author and the two researchers. Eight published and four unpublished studies were excluded from the analysis (for details of exclusion justifications, see the Table S1 in the Supplementary Materials).[1]

Figure 1 shows the number of records that were screened out at each step of the process. In total, 19 studies were included in the meta-analysis. Where there was more than one independent effect size reported in an article (e.g., from separate experiments, or from independent samples of participants within the same experiment), these were extracted separately. Where there were multiple non-independent effect sizes (e.g., a three-group design comparing the SAI© to two comparison conditions), one effect size was identified for extraction. This effect size was chosen to either produce the most conservative test of the SAI© (e.g., comparing the SAI© to a free recall comparison group rather than a no-retrieval comparison group) or to reduce confounds (e.g., such as differences in modality or language between the SAI© and comparison group); details of selected comparisons are provided in the Supplementary Materials (Table S2).

[1] One study (Dando et al., 2020) was under review when we released our call for unpublished studies, and was screened as an unpublished study. However, prior to completion of the meta-analysis, this study was published. It is therefore counted after this point as a published study.

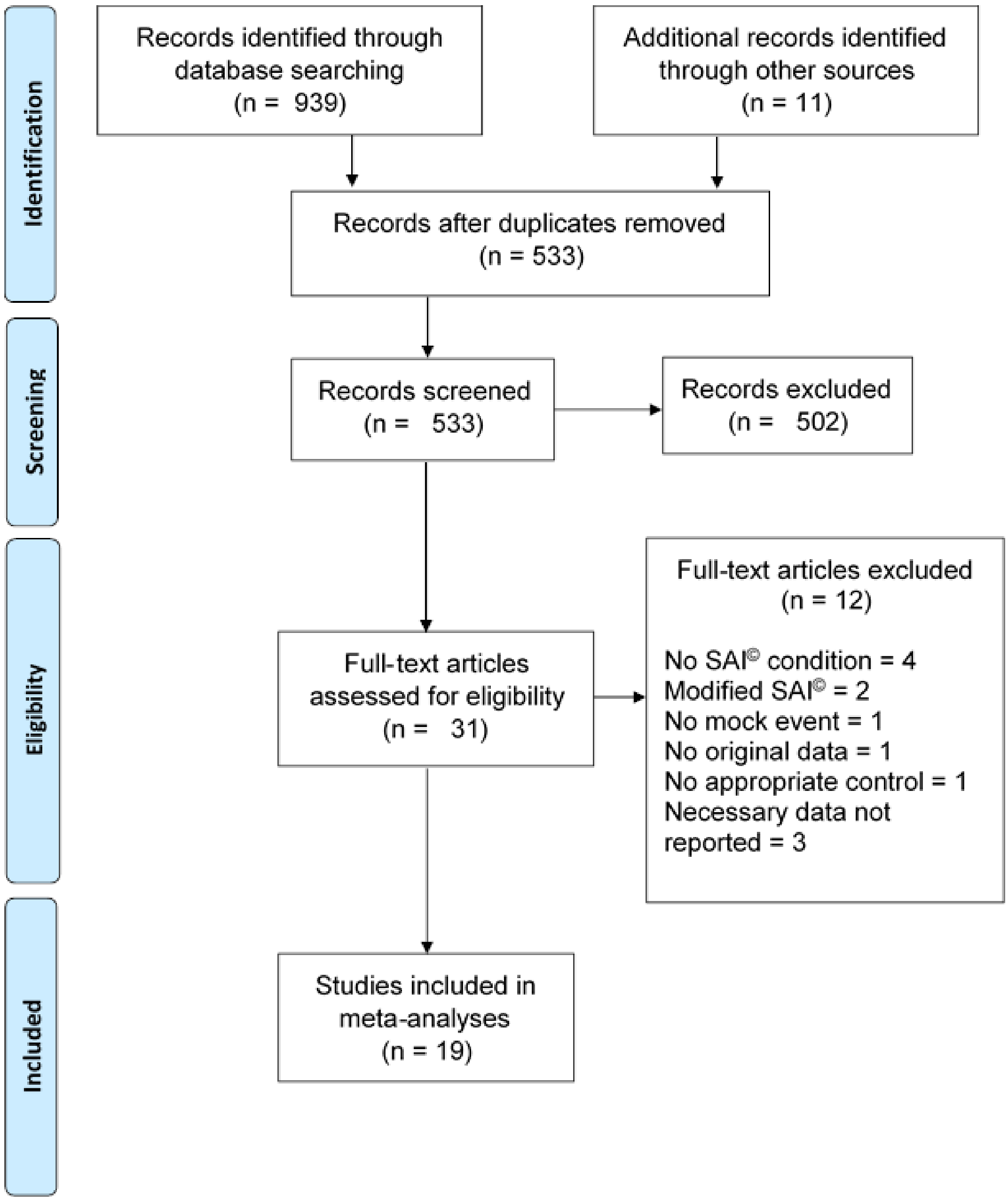


Figure 1. Flow diagram representing the process of identifying and screening studies for inclusion in the meta-analysis

Data coding and extraction

For each study, details of the experimental design, cell sizes, and values for the dependent variables (number of correct details, number of incorrect details, accuracy) were extracted. In all cases where the calculation of "accuracy" was explicitly described, it was calculated as the number of correct details divided by the total number of details reported. Where cell sizes were not reported and the authors were unable to provide the data, we estimated cell sizes by dividing the total reported sample size by the number of between-subjects cells in the study design.

For each effect size, we coded for the following estimator variables: participant age group (Children under 13, Adolescents 13-17, Younger Adults 18-39, Middle-aged Adults, 40-59, Older Adults 60+), exposure to post-event information (absent, present through leading questions or post-event narrative, present through co-witness discussion), alcohol intoxication (present, absent), stress during witnessed event (high, low), and type of event (non-violent crime, violent crime, accident, or other event). We coded for the following system variable: delay between initial and subsequent retrieval attempts (less than one day, one day, two days to one week, more than one week). We coded for the following methodological variables: participant population (students, general population, mixed students and general population, special population), comparison recall test used for initial retrieval attempt (free recall, cued recall, structured interview, no recall attempt), type of memory test used for subsequent report (free recall, cued recall, structured interview, CI), and event modality (live, video). Finally, we coded for publication status (published, unpublished, and whether the any of the original authors of the SAI© (Gabbert, Hope, or Fisher) were a named author on the study (yes or no)[2].

[2] We did not pre-register this moderator variable, but incorporated it at the suggestion of an anonymous reviewer.

Risk of bias in individual studies

We adapted the Cochrane Risk of Bias Tool (Higgins et al., 2011) to make it more suitable for behavioural research (see Table S3 in the Supplementary Materials for details). This tool evaluates selection bias (random allocation of participants to conditions), reporting bias (selective reporting of data), performance bias (blinding of participants and experimenters), detection bias (blinding of coders), and attrition bias (incomplete outcome data), and other forms of bias (including lack of pre-registered hypotheses and analysis plans). For each item, risk was coded as High, Low, or Unclear. Each study was coded independently by the second and third authors, and disagreements were resolved through discussion between the first, second, and third authors.

Summary measures

Random-effects models with Restricted Maximum-Likelihood Estimation were used to estimate meta-analytic effect sizes (Cohen's *d*) with 95% Confidence Intervals (CIs) using the metafor package for R (Viechtbauer, 2010; R Core Team, 2019). The Knapp and Hartung (2003) adjustment was applied to test for statistical significance, which controls the Type I error rate (Viechtbauer et al., 2015). The heterogeneity of effect sizes was assessed using the I^2 statistic, and by examining *Q* and its associated *p* value. If there was evidence of significant heterogeneity ($p < .05$), a metaregression was performed. Moderators were included if there was sufficient variability across studies to allow for a meaningful comparison. Where necessary, the codes used for data extraction were collapsed into fewer categories. Full regression models are reported in the Supplementary Materials (Tables S10 to S14).

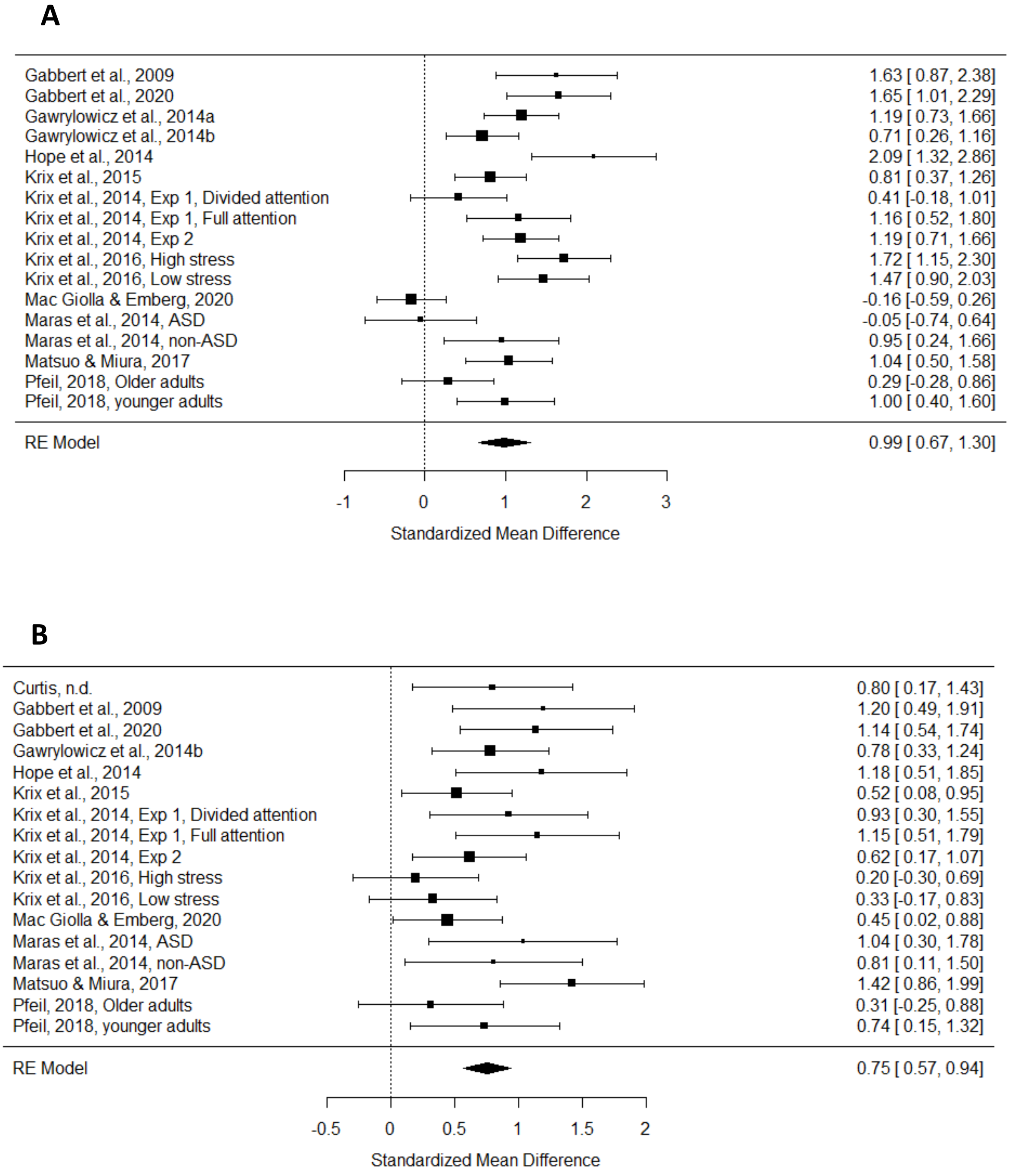
For the published studies in each analysis, publication bias was assessed via visual inspect of funnel plots and a regression test for funnel plot asymmetry. Finally, we used a metaregression approach to predict effect size from total Risk of Bias scores.

Results

Initial accounts

Our first set of analyses compared the quantity and quality of information reported in the SAI© with the quantity and quality of information reporting in other initial reporting formats, such as cued and free recall. The comparisons included in these analyses are shown in Table 1. As not all dependent measures were reported in all studies, the subset of comparisons varied slightly across the three dependent measures: number of correct details ($k = 17$, $N = 977$), number of incorrect details ($k = 17$, $N = 935$), and accuracy ($k = 17$, $N = 977$). Forest plots are shown in Figure 2. Funnel plots can be found in the Supplementary Materials (Figure S1).

Moderator codes are shown and inter-rater agreement for each code, calculated as the proportion of identical decisions, can be found in the Supplementary Materials (Table S4). For four of our potential moderators (alcohol intoxication, SAI© modality, exposure to post-event information, and imposition of a time restriction), there was no variability in codes across studies; these moderators are not shown the table and were not included in any analyses. We simplified the codes for participant age (Adolescents, Adults, Older Adults), control condition (Free Recall, Structured Recall), and event type (Violent Crime, Non-Violent Crime) to increase cell sizes.


A
Gabbert et al., 2009
Gabbert et al., 2020
Gawrylowicz et al., 2014a
Gawrylowicz et al., 2014b
Hope et al., 2014
Krix et al., 2015
Krix et al., 2014, Exp 1, Divided attention
Krix et al., 2014, Exp 1, Full attention
Krix et al., 2014, Exp 2
Krix et al., 2016, High stress
Krix et al., 2016, Low stress
Mac Giolla & Emberg, 2020
Maras et al., 2014, ASD
Maras et al., 2014, non-ASD
Matsuo & Miura, 2017
Pfeil, 2018, Older adults
Pfeil, 2018, younger adults
RE Model
1.63 [0.87, 2.38]
1.65 [1.01, 2.29]
1.19 [0.73, 1.66]
0.71 [0.26, 1.16]
2.09 [1.32, 2.86]
0.81 [0.37, 1.26]
0.41 [-0.18, 1.01]
1.16 [0.52, 1.80]
1.19 [0.71, 1.66]
1.72 [1.15, 2.30]
1.47 [0.90, 2.03]
-0.16 [-0.59, 0.26]
-0.05 [-0.74, 0.64]
0.95 [0.24, 1.66]
1.04 [0.50, 1.58]
0.29 [-0.28, 0.86]
1.00 [0.40, 1.60]
0.99 [0.67, 1.30]
-1
0
1
2
3
Standardized Mean Difference
B
Curtis, n.d.
Gabbert et al., 2009
Gabbert et al., 2020
Gawrylowicz et al., 2014b
Hope et al., 2014
Krix et al., 2015
Krix et al., 2014, Exp 1, Divided attention
Krix et al., 2014, Exp 1, Full attention
Krix et al., 2014, Exp 2
Krix et al., 2016, High stress
Krix et al., 2016, Low stress
Mac Giolla & Emberg, 2020
Maras et al., 2014, ASD
Maras et al., 2014, non-ASD
Matsuo & Miura, 2017
Pfeil, 2018, Older adults
Pfeil, 2018, younger adults
RE Model
0.80 [0.17, 1.43]
1.20 [0.49, 1.91]
1.14 [0.54, 1.74]
0.78 [0.33, 1.24]
1.18 [0.51, 1.85]
0.52 [0.08, 0.95]
0.93 [0.30, 1.55]
1.15 [0.51, 1.79]
0.62 [0.17, 1.07]
0.20 [-0.30, 0.69]
0.33 [-0.17, 0.83]
0.45 [0.02, 0.88]
1.04 [0.30, 1.78]
0.81 [0.11, 1.50]
1.42 [0.86, 1.99]
0.31 [-0.25, 0.88]
0.74 [0.15, 1.32]
0.75 [0.57, 0.94]
-0.5
0
0.5
1
1.5
2
Standardized Mean Difference

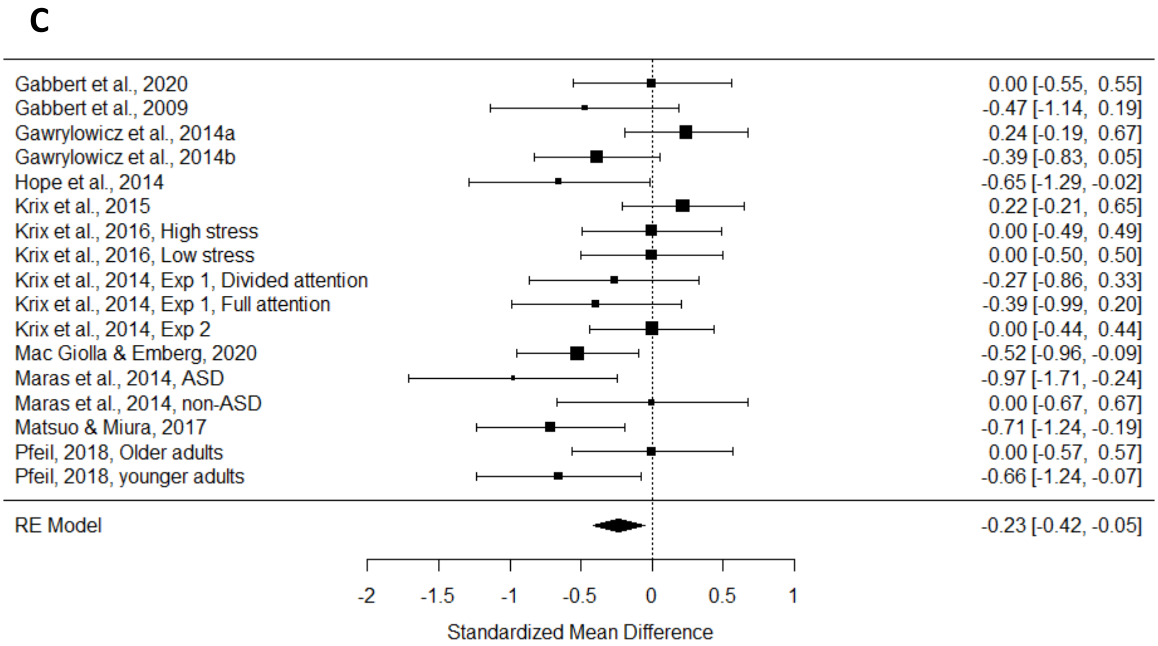


Figure 2. Meta-analytic effect sizes for immediate accounts: SAI© vs. alternative reporting formats. A = Number of correct details; B = Number of incorrect details; C = Accuracy of reported details.

Table 1. Descriptive statistics (cell sizes, means, and standard deviations) for comparisons included in the analyses of initial accounts

	Sample sizes (*n*)		Correct details (*M*, *SD*)		Incorrect details (*M*, *SD*)		Accuracy (*M*, *SD*)	
Study	SAI©	Control	SAI©	Control	SAI©	Control	SAI©	Control
Curtis (n.d.)	21	21	--	--	0.79 (0.57)	0.38 (0.43)	--	--
Gabbert (2009, Exp 1)	18*	18*	70.70 (20.46)	41.50 (14.00)	8.85 (4.85)	4.85 (2.31)	.89 (.05)	.91 (.03)
Gabbert (2020, Exp 2)	26	24	135.82 (40.97)	71.38 (35.43)	12.14 (5.56)	6.33 (4.35)	.92 (.04)	.92 (.04)
Gawrylowicz (2014a)	42*	42*	69.07 (21.40)	45.05 (18.32)	--	--	.96 (.03)	.95 (.05)
Gawrylowicz (2014b)	40*	40*	83.26 (24.95)	66.19 (22.50)	1.51 (1.41)	0.58 (0.89)	.97 (.03)	.98 (.02)
Hope (2014)	20	20	126.45 (35.59)	67.85 (15.54)	9.55 (5.22)	4.35 (3.15)	.93 (.03)	.95 (.03)
Krix (2015)	43	41	112.95 (24.01)	93.10 (24.29)	17.72 (6.88)	13.34 (9.65)	.88 (.05)	.87 (.04)
Krix (2016, High stress)	33	31	21.00 (5.24)	12.61 (4.30)	0.42 (0.71)	0.29 (0.59)	.98 (.03)	.98 (.04)
Krix (2016, Low stress)	32	30	21.59 (4.67)	14.10 (5.42)	0.53 (0.76)	0.30 (0.60)	.98 (.03)	.98 (.04)
Krix (2014, Exp 1, Divided attention)	22	22	49.50 (25.30)	39.00 (24.49)	13.68 (7.77)	7.55 (4.90)	.78 (.07)	.81 (.14)

Krix (2014, Exp 1, Full attention)	22	22	113.41 (32.38)	78.32 (26.64)	13.77 (6.02)	7.55 (4.48)	.89 (.05)	.91 (.05)
Krix (2014, Exp 2)	41	40	119.17 (25.89)	90.05 (22.55)	14.78 (5.60)	11.45 (5.03)	.89 (.04)	.89 (.04)
Mac Giolla (2020)	46	40	19.74 (7.35)	20.99 (7.76)	5.22 (0.76)	4.11 (2.07)	.79 (.08)	.83 (.07)
Maras (2014, ASD)	16*	16*	32.89 (14.35)	33.73 (17.90)	8.56 (5.16)	4.20 (2.62)	.80 (.08)	.88 (.08)
Maras (2014, Non-ASD)	17*	17*	44.59 (12.08)	32.22 (13.37)	6.88 (3.04)	4.44 (2.87)	.87 (.04)	.87 (.08)
Matsuo (2016)	30	30	82.23 (25.53)	54.53 (27.06)	9.03 (5.06)	3.20 (2.67)	.90 (.05)	.94 (.06)
Pfeil (2018, Older adults)	24	24	77.92 (26.34)	71.25 (18.28)	7.42 (4.04)	6.33 (2.63)	.90 (.04)	.90 (.04)
Pfeil (2018, Younger adults)	24	24	112.88 (54.39)	70.13 (24.29)	8.08 (8.08)	3.63 (2.32)	.93 (.03)	.95 (.03)

Note: For studies marked with an asterisk, cell sizes were estimated under the assumption that cell sizes were equal.

Correct details

On average, an additional 23.01 [95% CI, -0.20, 46.22] correct details were reported by participants completing an SAI© in comparison to participants completing a free recall account. For correct details, the meta-analytic effect size was large and significantly greater than zero, $d = 0.99$, 95% CI [0.67, 1.30], $t = 6.57$, $p < .001$, indicating that the SAI© increases the number of correct details reported by approximately one standard deviation. However, there was considerable variability in effect sizes, $I^2 = 78.14\%$, 95% CI [60.26, 90.93], which exceeded the heterogeneity that would be expected by sampling error alone, $Q\ (16) = 72.90$, $p < .001$. We therefore conducted an exploratory moderator analysis.

Inspection of Table S4 revealed that, for many of our extracted moderators, variability was low –for example, only a single effect size was obtained in a high-stress paradigm. Furthermore, some moderators were confounded. For example, only two effect sizes used violent mock event, and these were both from Maras et al. (2014); these two effect sizes were also two of only three that compared the SAI© to an alternative structured recall (as opposed to free recall). We therefore decided to include the following four moderators in the meta-regression: Publication status (published = 1, unpublished = 0), participant age group (adults = 1, older adults = 0), control condition (free recall = 1, structured recall = 0), and whether the study included any of the original authors (yes = 1, no = 0). Collectively, these moderators accounted for 56.07% of the observed heterogeneity in effect sizes, though the residual heterogeneity was statically significant, $QE\ (12) = 29.16$, $p = .004$.

Studies that used a free recall comparison condition produced substantially larger effect sizes than those that used a structured recall comparison, $\beta = 0.95$, 95% CI [0.22, 1.69], $t = 2.83$, $p = .02$. Subgroup analysis showed that the estimated meta-analytic effect size for studies with a free recall comparison condition ($k = 14$) was large and significantly greater than zero, $d = 1.14$, 95% CI [0.85, 1.42], $t = 8.60$, $p < .001$. In contrast, the meta-analytic effect size for studies with a structured recall comparison condition (*k* = 3) was small and not significantly different from zero, *d* = 0.21, 95% CI [-1.28, 1.70], *t* = 0.60, *p* = .61. Neither participant age, *β* = 0.69, 95% CI [-0.09, 1.48], *t* = 1.92, *p* = .08, publication status, *β* = 0.16, 95% CI [-0.46, 0.79], *t* =0.57, *p* = .58, nor original author contributions, *β* = 0.14, 95% CI [-0.42, 0.70], *t* = 0.55, *p* = .60 were statistically significant moderators.

To assess publication bias, we visually inspected a funnel plot on the subset of published comparisons (*k* = 13); see Figure S1 in the Supplementary Materials. There was no obvious asymmetry in the funnel plot, indicating that publication bias likely did not lead to a systematic inflation of effect sizes in this case. A regression test for funnel plot asymmetry converged on the same conclusion, z = 0.73, *p* = .48.

Incorrect details

On average, an additional 3.33 [95% CI, 0.15, 6.52] incorrect details were reported by participants completing an SAI© in comparison to participants completing a free recall account. The meta-analytic effect size for incorrect details was large and significantly greater than zero, *d* = 0.75, 95% CI [0.57, 0.94], *t* = 8.62, *p* < .001, indicating that the number of incorrect details reported was around three-quarters of a standard deviation higher on an initial SAI© than on other initial reports. The heterogeneity in effect sizes was moderate, I^2 = 38.83%, 95% CI [0.00, 73.15] and not sufficiently large to indicate a moderator analysis, *Q* (16) = 25.62, *p* = .059.

To assess publication bias, we visually inspected a funnel plot on the subset of published comparisons (*k* = 12); see Figure S1 in the Supplementary Materials. A regression test for funnel plot asymmetry indicated that there was some evidence of publication bias, z = 2.66, *p* = .02. To understand this asymmetry, we estimated separate meta-analytic effect sizes for the subset of published studies (*k* = 12) and unpublished studies (*k* = 5). Contrary to what might be expected if findings unfavourable to the SAI© were being selectively withheld, the meta-analytic effect size was actually somewhat higher for the published studies, *d* = 0.80, 95% CI [0.56, 1.04], than for the unpublished studies, *d* = 0.65, 95% CI [0.26, 1.05].

Accuracy

On average, the proportion of correct responses was around 0.02 lower [95% CI, -0.05, 0.02] for participants who completed an SAI© in comparison to participants who completed a free recall account. The SAI© was associated with a small decrease in accuracy, *d* = -0.23, 95% CI [-0.42, -0.05], *t* = 2.75, *p* = .01. Heterogeneity in effect sizes was moderate, I^2 = 41.56%, 95% CI [0.00, 75.04], and somewhat larger than would be expected by sampling error alone, *Q* (16) = 27.38, *p* = .038. We therefore conducted an exploratory moderator analysis.

We included the same four moderators as for correct details: Publication status (published = 1, unpublished = 0), participant age group (adults = 1, older adults = 0), control condition (free recall = 1, structured recall = 0), and contribution of original authors (yes = 1, no = 0). These moderators did not explain a significant amount of the heterogeneity in accuracy, *F*(4, 12) = 0.62, *p* = .66. Unlike for correct details, the type of control condition did not significantly moderate accuracy, *β* = 0.37, 95% CI [-0.22, 0.97], *t* = 1.38, *p* = .19. Neither publication status, *β* = 0.06, 95% CI [-0.43, 0.55], *t* =0.28, *p* = .79, participant age, *β* = -0.004, 95% CI [-0.62, 0.61], *t* = -0.01, *p* = .99, nor contribution of original authors, *β* = -0.19, 95% CI [-0.63, 0.24], *t* = -0.98, *p* = .35, significantly moderated accuracy.

We assessed publication bias via visual inspection of the funnel plot for the subset of unpublished studies (Figure S1 in the Supplementary Materials), and via a regression test for funnel plot asymmetry. This test revealed evidence of asymmetry in the funnel plot, *z* = -2.85, *p* = .02. We calculated separate meta-analytic effect sizes for the subsets of published (*k* = 12) and unpublished (*k* = 5) studies. In both cases, the effect sizes indicated that the SAI© was associated with a decrease in accuracy; however, the effect was somewhat larger for the unpublished studies, *d* = -0.31, 95% CI [-0.85, 0.23], than for the published studies, *d* = -0.21, 95% CI [-0.44, 0.01].

Risk of Bias

Scores for each Risk of Bias item can be found in the Supplementary Materials (Table S5); each item was coded as low risk (0), unclear (1), or high risk (2). Inter-rater agreement is shown in the bottom row of the table, calculated as the proportion of decisions for each criterion that were identical. Note that for Reporting Bias and Attrition Bias, inter-rater agreement was initially quite low. For these items, we reviewed and modified the coding scheme to ensure the criteria were clear. We then discussed each report against those revised criteria to come to an agreement. Justifications for high risk codes can be found in the Supplementary Materials (Table S6). Item codes were summed to create total Risk of Bias (RoB) scores that could vary from 0 to 16. To examine the influence of methodological bias on outcomes, we conducted a series of meta-regressions in which the total RoB score was included as a moderator variable. The SAI© advantage for correct details was significantly moderated by RoB scores, $\beta = 0.20$, 95% CI [0.03, 0.37], $t = 2.47$, $p = .03$, though we note that the *p* value for this effect is relatively large. Finally, RoB did not significantly moderate the effect of the SAI© on accuracy of reported details, $\beta = -0.0004$, 95% CI [-0.12, 0.12], $t = -0.01$, $p = .99$, or on reporting of incorrect details, $\beta = 0.11$, 95% CI [-0.003, 0.22], $t = 2.07$, $p = .056$.

Taken together, we observed some tentative evidence that studies with a higher risk of bias produced larger effect sizes for correct details; the effect size favouring the SAI© was around 0.20 larger for each one-point increase in RoB scores. However, we observed no evidence that studies with higher RoB scores were more favourable to the SAI© in terms of overall accuracy rates.

Subsequent accounts

Our second set of analyses compared the quantity and quality of information reported in a subsequent account between witnesses who had and had not completed an initial SAI©. The comparisons included in the analysis are shown in Table 2, and the extracted moderator codes can be found in the Supplementary Materials, along with inter-rater agreement (Table S7). The number of comparisons varied across the three dependent measures: correct details (k = 19, N = 906), incorrect details (k = 15, N = 722), and accuracy (k = 15, N = 722). Forest plots are shown in Figure 2. Funnel plots can be found in the Supplementary Materials (Figure S2).

Correct details

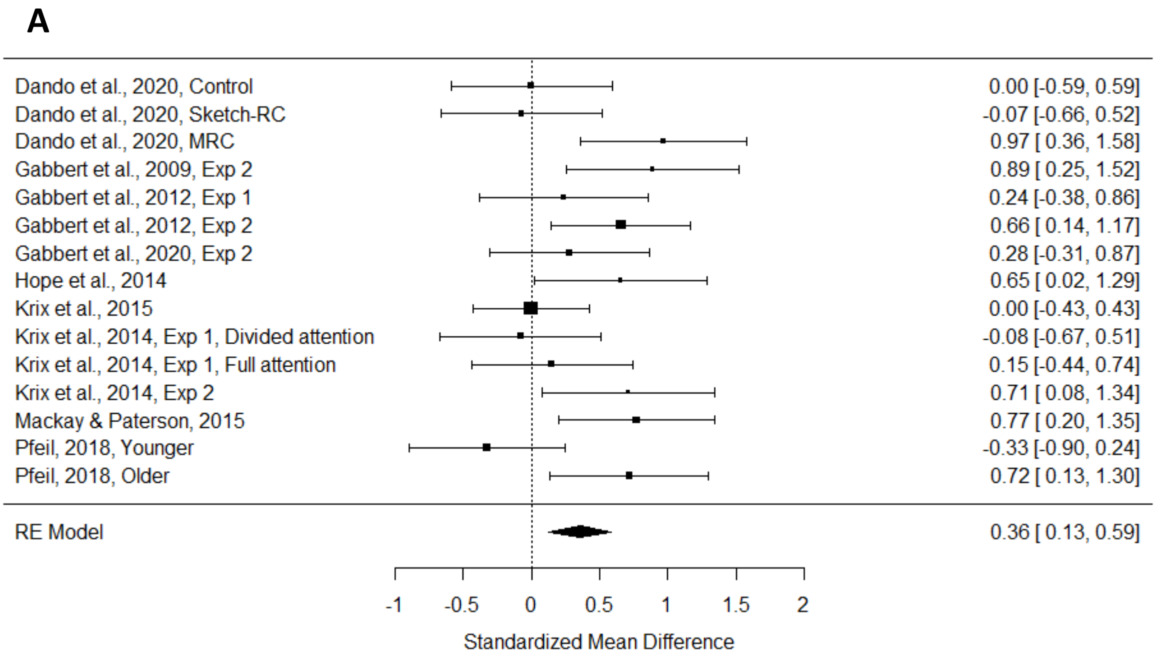
On average, participants who had completed an initial SAI© reported an additional 11.04 correct details [95% CI, -13.48, 35.56] on a subsequent account in comparison to participants who had not. For correct details, the meta-analytic effect size was medium and significantly greater than zero, d = 0.56, 95% CI [0.26, 0.86], t = 3.91, p = .001. The number of correct details reported on a subsequent test was around half a standard deviation higher for participants who had completed an initial SAI© than for those who had not. However, there was considerable heterogeneity in effect sizes, I^2 = 76.11%, 95% CI [58.17, 89.38], which was greater than would be expected by sampling error alone, Q (18) = 74.73, $p < .001$. We therefore conducted an exploratory moderator analysis.

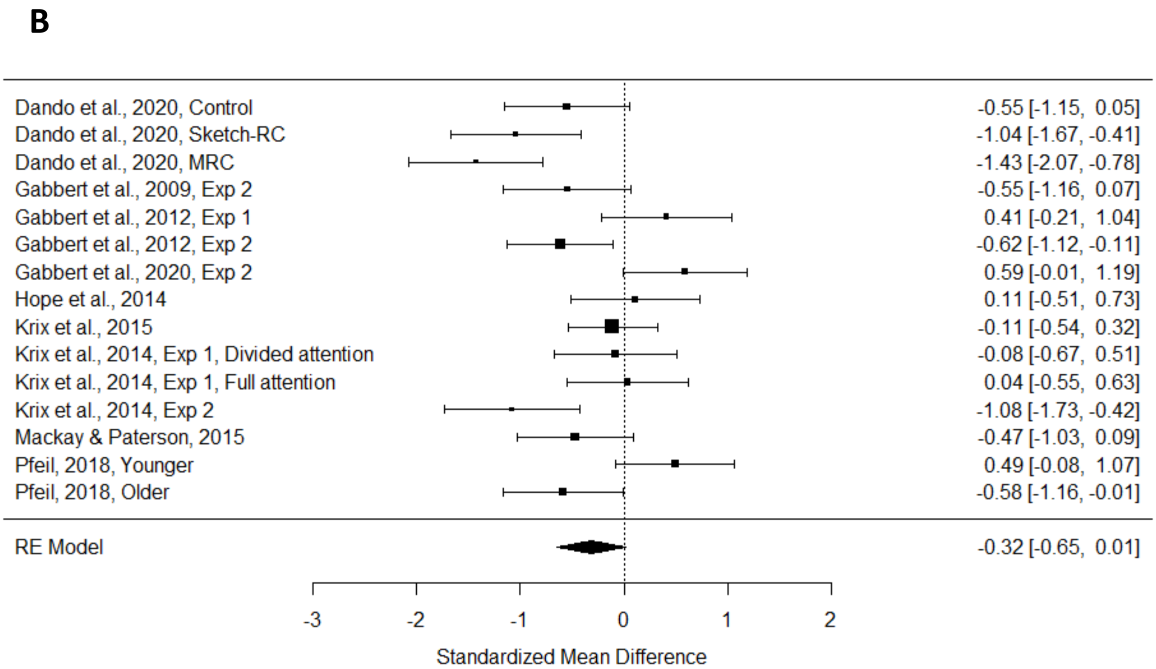
Table 2. Descriptive statistics (cell sizes, means, and standard deviations) for comparisons included in the analyses of delayed accounts

	Sample sizes (*n*)		Correct details (*M*, *SD*)		Incorrect details (*M*, *SD*)		Accuracy (*M*, *SD*)	
Study	SAI©	Control	SAI©	Control	SAI©	Control	SAI©	Control
Dando (2020, Control)	22*	22*	21.45 (4.76)	20.36 (5.28)	1.32 (0.89)	1.91 (0.89)	.89 (.20)	.89 (.05)
Dando (2020, S-RC)	22*	22*	38.45 (5.51)	29.26 (5.01)	0.77 (0.68)	1.52 (0.68)	.93 (.19)	.94 (.03)
Dando (2020, MRC)	22*	22*	28.26 (4.99)	24.04 (5.29)	1.35 (0.93)	2.95 (1.25)	.95 (.09)	.81 (.18)
Gabbert (2009 Exp 2)	21	21	62.38 (22.04)	45.90 (24.02)	4.33 (3.09)	7.00 (6.03)	.93 (.05)	.88 (.06)
Gabbert (2012, Exp 1)	20	20	86.25 (23.25)	48.10 (21.97)	3.45 (2.80)	2.40 (2.16)	.96 (.03)	.95 (.05)
Gabbert (2012, Exp 2)	31	31	25.77 (14.49)	11.09 (5.13)	2.87 (2.00)	4.22 (2.32)	.81 (.13)	.73 (.11)
Gabbert (2020, Exp 2)	24	21	96.55 (31.33)	62.81 (33.57)	7.32 (3.15)	5.43 (3.17)	.93 (.03)	.92 (.04)
Gittins (2015, Low stress)	20	18	49.40 (17.73)	39.31 (12.94)	--	--	.--	--
Gittins (2015, High stress)	22	20	43.71 (15.95)	37.65 (15.36)	--	--	--	--
Hope (2014)	20	20	176.95 (41.33)	126.15 (43.45)	11.85 (6.01)	11.25 (4.46))	.94 (.03)	.92 (.03)
Krix (2015)	43	41	102.35 (26.11)	109.59 (26.14)	17.19 (6.39)	18.12 (10.22)	.86 (.05)	.86 (.05)
Krix (2014, Exp 1, Divided attention)	22	22	31.73 (16.73)	38.00 (19.46)	7.72 (5.47)	8.14 (4.99)	.80 (.11)	.81 (.14)

Krix (2014, Exp 1, Full attention)	22	22	79.59 (30.84)	72.00 (30.02)	7.50 (6.22)	7.27 (6.32)	.92 (.06)	.91 (.07)
Krix (2014, Exp 2)	21	20	104.43 (24.83)	117.95 (22.66)	13.57 (5.29)	20.30 (5.29)	.88 (.05)	.85 (.03)
Mackay (2015)	25*	25*	52.60 (21.53)	42.40 (20.72)	7.36 (4.27)	9.32 (3.92)	.88 (.06)	.82 (.09)
Paterson (2015)	26*	26*	48.58 (17.28)	38.96 (12.49)	--	--	--	--
Pfeil (n.d, Younger adults)	24	24	97.21 (34.64)	78.04 (29.26)	6.00 (4.93)	3.92 (3.15)	.94 (.03)	.95 (.03)
Pfeil (2018, Older adults)	24	24	97.17 (28.66)	96.71 (23.32)	7.88 (3.54)	10.49 (5.12)	.92 (.03)	.89 (.05)
Thornhill (2013)	26*	26*	28.32 (9.68)	23.04 (9.84)	--	--	--	--

Note: For studies marked with an asterisk, cell sizes were estimated under the assumption that cell sizes were equal.


A
Dando et al., 2020, Control
Dando et al., 2020, Sketch-RC
Dando et al., 2020, MRC
Gabbert et al., 2009, Exp 2
Gabbert et al., 2012, Exp 1
Gabbert et al., 2012, Exp 2
Gabbert et al., 2020, Exp 2
Hope et al., 2014
Krix et al., 2015
Krix et al., 2014, Exp 1, Divided attention
Krix et al., 2014, Exp 1, Full attention
Krix et al., 2014, Exp 2
Mackay & Paterson, 2015
Pfeil, 2018, Younger
Pfeil, 2018, Older
RE Model
0.00 [-0.59, 0.59]
-0.07 [-0.66, 0.52]
0.97 [0.36, 1.58]
0.89 [0.25, 1.52]
0.24 [-0.38, 0.86]
0.66 [0.14, 1.17]
0.28 [-0.31, 0.87]
0.65 [0.02, 1.29]
0.00 [-0.43, 0.43]
-0.08 [-0.67, 0.51]
0.15 [-0.44, 0.74]
0.71 [0.08, 1.34]
0.77 [0.20, 1.35]
-0.33 [-0.90, 0.24]
0.72 [0.13, 1.30]
0.36 [0.13, 0.59]
-1
-0.5
0
0.5
1
1.5
2
Standardized Mean Difference


B
Dando et al., 2020, Control
Dando et al., 2020, Sketch-RC
Dando et al., 2020, MRC
Gabbert et al., 2009, Exp 2
Gabbert et al., 2012, Exp 1
Gabbert et al., 2012, Exp 2
Gabbert et al., 2020, Exp 2
Hope et al., 2014
Krix et al., 2015
Krix et al., 2014, Exp 1, Divided attention
Krix et al., 2014, Exp 1, Full attention
Krix et al., 2014, Exp 2
Mackay & Paterson, 2015
Pfeil, 2018, Younger
Pfeil, 2018, Older
RE Model
-0.55 [-1.15, 0.05]
-1.04 [-1.67, -0.41]
-1.43 [-2.07, -0.78]
-0.55 [-1.16, 0.07]
0.41 [-0.21, 1.04]
-0.62 [-1.12, -0.11]
0.59 [-0.01, 1.19]
0.11 [-0.51, 0.73]
-0.11 [-0.54, 0.32]
-0.08 [-0.67, 0.51]
0.04 [-0.55, 0.63]
-1.08 [-1.73, -0.42]
-0.47 [-1.03, 0.09]
0.49 [-0.08, 1.07]
-0.58 [-1.16, -0.01]
-0.32 [-0.65, 0.01]
-3
-2
-1
0
1
2
Standardized Mean Difference

C

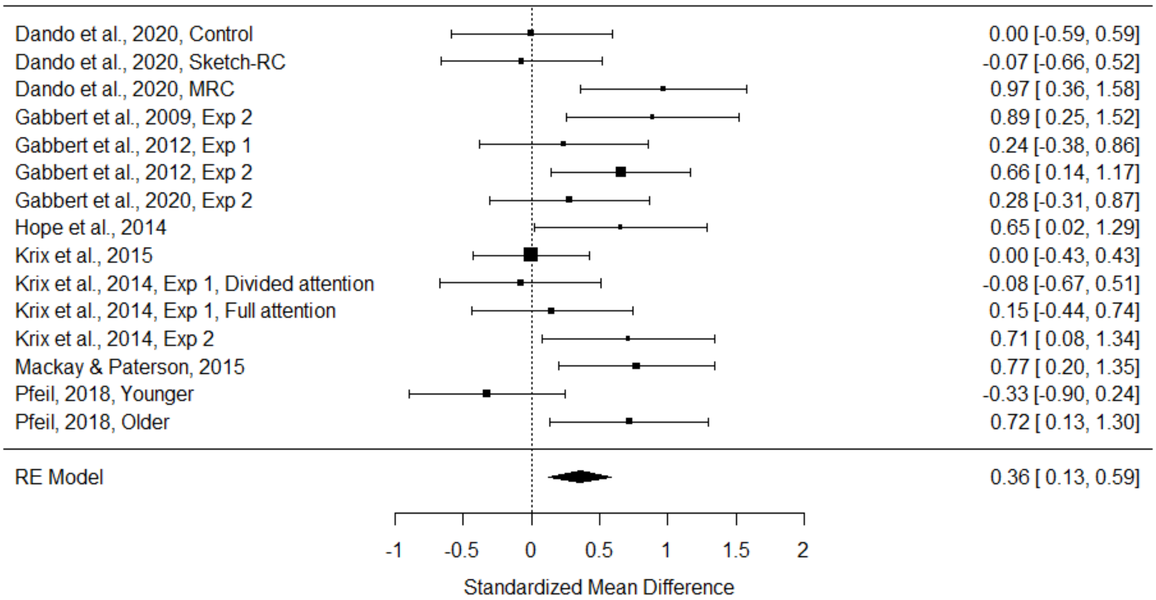


Figure 3. Meta-analytic effect sizes for subsequent accounts, comparing groups who had vs. had not completed an initial SAI©. A = Number of correct details; B = Number of incorrect details; C = Accuracy of reported details.

Inspection of Table S7 allowed us to exclude from consideration several moderators because there were an insufficient number of studies in each category, and/or because they were confounded with other moderators. For example, only two effect sizes were from high-stress paradigms, and both of these used free recall control conditions at time 1, thereby confounding stress with time 1 control condition. Similarly, only three effect sizes used live events, but all of these effect sizes were with older adults. Seven potential moderators were included in the metaregression: Publication status (Published = 1 (k = 15), Unpublished = 0 (k = 4)); Participant age (Adults = 1 (k = 15), Older adults = 0 (k = 4)); Time 1 control group (1 = free recall (k = 9), 0 = no recall (k = 10)); Time 2 report type (1 = free recall (k = 14), 0 =

other recall type (*k* = 5)); Retention interval (1 = more than one week (*k* = 3), 0 = one week or less (*k* = 16)); Exposure to Post-Event Information (1 = present (*k* = 8), 0 = absent (*k* = 11)); and Contribution of Original Authors (1 = yes (*k* = 11), 0 = no (*k* = 8)).

Collectively, the moderators did not explain a statistically significant amount of heterogeneity in effect sizes, $F(7, 11) = 1.82$, $p = .18$, accounting for 35.34% of the variability. The test for residual heterogeneity was statistically significant, $QE = 31.29$, $p = .001$. None of the moderators significantly moderated the meta-analytic effect on correct details: Publication status, $\beta = -0.91$, 95% CI [-2.04 0.21], $t = -1.78$, $p = .10$; Participant age, $\beta = 0.24$, 95% CI [-0.58, 1.06], $t = 0.65$, $p = .53$; Type of control condition at time 1, $\beta = -0.95$, 95% CI [-1.94, 0.05], $t = -2.10$, $p = .060$; Time 2 retrieval type, $\beta = -0.14$, 95% CI [-0.89, 0.61], $t = -0.41$, $p = .69$; Retention interval, $\beta = 0.11$, 95% CI [-0.72, 0.94], $t = 0.29$, $p = .78$; Exposure to post-event information, $\beta = 0.14$, 95% CI [-0.70, 0.98], $t = 0.38$, $p = .71$; Contribution of original authors, $\beta = 0.68$, 95% CI [-0.04, 1.40], $t = 2.07$, $p = .063$.

Visual inspection of the funnel plot on the subset of published studies (*k* = 15) revealed no obvious asymmetry. However, the regression test for funnel plot asymmetry was statistically significant, $t(13) = 2.37$, $p = .03$. To investigate this asymmetry further, we conducted subgroup analyses, estimating separate meta-analytic effect sizes for the published and unpublished comparisons. The effect sizes were similar for published, $d = 0.57$, 95% CI [0.19, 0.95] and unpublished studies, $d = 0.53$, 95% CI [-0.12, 1.18].

Incorrect details

On average, participants who had completed an initial SAI© reported 0.93 fewer incorrect details [95% CI, -4.82, 2.96] on a subsequent account in comparison to participants who had not. The meta-analytic effect size for incorrect details on a subsequent report was not significantly different from zero, $d = -0.32$, 95% CI [-0.65, 0.01], $t = -2.07$, $p = .057$. Heterogeneity of effect sizes was high, $I^2 = 74.25\%$, 95% CI [51.48, 90.06], and above the level that would be expected by sampling error alone, Q (14) = 50.91, $p < .001$. We therefore investigated this heterogeneity with an exploratory moderator analysis. We included the same six moderators as for the analysis of correct details: Publication status, Participant age, Time 1 control condition, Time 2 retrieval type, Retention interval, Exposure to Post-Event Information, and Contribution of original authors.

The moderators accounted for approximately 49.83% of the heterogeneity in effect sizes, though the test for residual heterogeneity was statistically significant, $QE(8) = 16.35$, $p = .02$. However, the omnibus test for the significance of the moderators was not statistically significant, $F(7, 7) = 2.14$, $p = .27$.

Nonetheless, examining individual coefficients revealed only one significant moderator – participant age, $\beta = 0.93$, 95% CI [0.09, 1.78], $t = 2.60$, $p = .04$. Subgroup analysis indicated that the meta-analytic effect size for younger adults was small and not significantly below zero, $d = -0.11$, 95% CI [-0.46, 0.23], $t = -0.74$, $p = .48$. The meta-analytic effect size for older adults, however, was large and significantly below zero, $d = -0.88$, 95% CI [-1.54, -0.23], $t = -4.28$, $p = .02$. These findings suggest that the subsequent memory reports of older adults may benefit more from an initial SAI© than the memory reports of younger adults – at least in terms of reducing the number of incorrect details that are reported.

No other moderators were statistically significant: Publication status, $d = -0.33$, 95% CI [-1.60, 0.95], $t = -0.61$, $p = .56$; Time 1 control condition, $d = 0.17$, 95% CI [-0.80, 1.14], $t = 0.41$, $p = .69$; Time 2 report type, $d = 0.39$, 95% CI [-0.61, 1.39], $t = 0.92$, $p = .39$; Retention interval, $d = -0.28$, 95% CI [-1.94, 1.37], $t = -0.41$, $p = .70$; Post-event information, $d = 0.19$, 95% CI [-0.89, 1.26], $t = 0.42$, $p = .69$; Contribution of original authors, $d = 0.03$, 95% CI [-1.06, 1.12], $t = 0.06$, $p = .95$.

Inspection of the funnel plot for the subset of published studies (k = 12) revealed no obvious asymmetry, which was confirmed by a regression test, $t(10) = -0.99$, $p = .35$. We have no evidence, therefore, that studies have been selectively withheld as a function of the effect size for incorrect details on a subsequent test.

Accuracy

On average, the accuracy of subsequent reports was .03 higher [95% CI -.01, .07] for participants who had completed an initial SAI© in comparison to those who had not. Accuracy on a delayed report was higher for participants who had completed an initial SAI© than for participants who had not, $d = 0.36$, 95% CI [0.13, 0.59], $t = 3.35$, $p = .005$. The heterogeneity in effect sizes was moderate, $I^2 = 49.14\%$, 95% CI [4.94, 79.65], and above the level that would be expected by sampling error alone, Q (14) = 27.52, $p = .02$. We therefore investigated this heterogeneity with an explanatory moderator analysis, including the same six moderators as for the analyses of correct and incorrect details: Publication status, Age group, Time 1 control type, Time 2 report type, Retention interval, and Exposure to Post-Event Information.

The moderators explained very little of the heterogeneity in effect sizes. The omnibus test for statistic indicated that the moderators did not explain a statistically significant amount of the heterogeneity, $F(7, 7) = 0.51$, $p = .81$, and that a statistically significant amount of heterogeneity remained, $QE(7) = 17.59$, $p = .014$.

None of the covariates significantly moderated the effect of an initial SAI© on the accuracy of a subsequent account: Publication status, $\beta = -0.39$, 95% CI [-1.69, 0.91], $t = -0.71$, $p = .50$; Participant age, $\beta = -0.17$, 95% CI [-1.03, 0.69], $t = -0.47$, $p = .65$; Time 1 control type, $\beta = -0.31$, 95% CI [-1.30, 0.68], $t = -0.74$, $p = .48$; Time 2 report type, $\beta = -0.66$, 95% CI [-1.68, 0.36], $t = -1.54$, $p = .17$; Retention interval, $\beta = -0.46$, 95% CI [-2.16, 1.24], $t = -0.64$, $p = .54$; Exposure to Post-Event Information, $\beta = 0.17$, 95% CI [-0.93, 1.27], $t =$

0.36, p = .73; Contribution of original authors, β = 0.47, 95% CI [-0.65, 1.58], t = 0.99, p = .35.

To assess the impact of publication bias, we visually inspected the funnel plot (Figure S2) for the subset of published studies (k = 12), which revealed no obvious asymmetry. This lack of asymmetry was confirmed by a regression test, $t(10) = 1.20$, p = .26.

Risk of Bias

RoB codes and total scores for studies included in the analysis of subsequent accounts are shown in the Supplementary Materials (Table S8). Inter-rater agreement is shown in the bottom row. For several variables (Random allocation, Reporting bias, Blinding of Personnel, and Attrition Bias), inter-rater agreement was initially low. Once again, we revised the codes to ensure that they were clear and unambiguous, and we then worked back through each study to ensure that the codes had been applied correctly. Justifications for high risk codes can be found in Table S9 in the Supplementary Materials).

To examine the influence of methodological bias on outcomes, we conducted a series of meta-regressions in which the total Risk of Bias (RoB) score was included as a moderator variable. RoB scores did not significantly moderate the effect of the SAI© on correct details, β = -0.07, 95% CI [-0.25, 0.12], t = -0.74, p = .47, or on accuracy of reported details, β = -0.002, 95% CI [-0.18, 0.18], t = -0.02, p = .98. RoB scores did, however, significantly moderate the effect of the SAI© on incorrect details, β = 0.25, 95% CI [0.03, 0.46], t = 2.46, p = .03. However, we note that the p value for this effect is relatively large, and that the direction of this relationship was positive, meaning that studies with higher RoB scores were less favourable to the SAI© (i.e., the SAI© benefit for reducing incorrect details was smaller) than studies with lower RoB scores. Taken together, these results provide no compelling evidence that the SAI© effects observed in this meta-analysis were being driven by studies that were methodologically biased towards finding an SAI© advantage.

Discussion

We set out to answer two research questions: 1) How, and to what extent, do initial reports obtained via the SAI© differ from initial accounts obtained via alternative recall methods? 2) How, and to what extent, do subsequent accounts differ between witnesses who have and have not completed an initial SAI©? Below, we discuss the findings for each of these two research questions, beginning with initial accounts.

Initial accounts

Compared to alternative recall methods, initial SAI© reports were associated with a large increase in the number of correct details and a slightly smaller increase in the number of incorrect details. The overall accuracy of the initial SAI© reports was slightly lower than the comparison reports. These findings are broadly similar to those from Memon et al.'s (2010) meta-analysis of the CI, who found that the CI was associated with a large increase in correct details (d = 1.20) and a smaller increase in incorrect details (d = 0.24).

The increase in both correct and incorrect details may stem from a criterion shift, in which participants relax their internal threshold for reporting a recalled detail (e.g., Koriat et al., 1996). Details recalled with lower confidence are thereby reported where previously they would have been withheld or perhaps reported in a more coarse-grained way (e.g., Goldsmith et al., 2002). Assuming that the participant is able to monitor the accuracy of recalled details with at least some degree of fidelity, low-confidence details will be less accurate than high-confidence details (see, for example, Brewer et al., 2018). A direct consequence of relaxing the reporting threshold will, therefore, be to increase the number of incorrect details that are reported.

We note that a similar increase in both correct and incorrect details was reported by Memon et al. (2010) in their meta-analysis of the CI. However, the increase in errors observed here is larger than that observed by Memon et al. The SAI© and the CI include similar instructions designed to increase metacognitive monitoring and thereby reduce guessing. Yet perhaps these instructions are more effective when delivered verbally by an interviewer who can dynamically engage with the witness, and who can use eye contact, tone of voice, and body language to increase emphasis on the instruction. An in-person interviewer can also pause, check for comprehension, and repeat instructions if necessary. In written form, witnesses may not fully attend to these instructions, or realize their importance. Future refinements of the SAI© will focus on enhancing the instructions to encourage more effective metacognitive monitoring, thereby reducing errors.

An alternative explanation for the increase in correct details is that the SAI© may alter the granularity of participants' responses. Goldsmith et al. (2002) describe how individuals can choose the level of precision they use to report information – they called this "grain-size". Coarse-grained details ("a light colored car") are more likely to be accurate, but are less informative than fine-grained details ("a light silver car"). Individuals can regulate the accuracy of their reports by adjusting the grain-size of their responses. In the context of eyewitness memory, Weber and Brewer (2008) found that witnesses are more likely to report fine-grained detail when they are confident in that detail; otherwise, they will volunteer coarse-grained information. McCallum et al. (2016), however, found that witnesses have a tendency to report fine-grained information even when the accuracy of that information is low, and when coarse-grained information would therefore be more appropriate. In a recent comparison between an adapted SAI© (in which participants were also asked to rate their confidence for each detail they reported) and a free recall procedure, Brewer et al. (2018) reported that the majority of additional details reported by participants in the SAI© were fine-grained. As fine-grained details are less likely to be correct than coarse-grained details, a tendency to favour fine-grained details is likely to reduce overall accuracy. We could not directly investigate this possibility in this meta-analysis, as information about the precision of reported details was not available to us. However, capturing grain-size in future studies would enable the hypothesis that the accuracy drop in the SAI© stems from a selective increase in fine-grained details to be tested.

There was substantial heterogeneity in the effect sizes for the number of correct details in initial accounts. In an exploratory moderator analysis, we found only one statistically significant moderator of this effect: the type of recall used in the control group. The effect size when comparing the SAI© to a free recall condition was much larger than the effect size when comparing the SAI© to a structured recall condition. The three effect sizes that incorporated a structured recall comparison came from two independent studies (Mac Giolla & Emberg, 2020.; Maras et al., 2014). The control comparison condition used by Mac Giolla & Emberg (2020) was a face-to-face interview which was designed to follow the structure of the SAI©, and to provide similar retrieval support. Maras et al. (2014) presented the participants in their comparison group with a structured recall attempt that followed a similar structure to the SAI©, but without the instructions that are designed to enhance metacognitive monitoring and to increase retrieval support. It is little surprise, then, that these studies showed a much smaller, and non-significant SAI© benefit for correct details than studies that used a much less structured, free recall comparison. Indeed, these studies suggest that the structure provided by the SAI© may play a large role in its effectiveness for eliciting correct information.

No other moderators that we were able to code for explained any of the residual heterogeneity effect sizes for correct details, nor the heterogeneity in effect sizes for accuracy. This suggests that there may be other important moderators that we were unable to capture here.

Subsequent accounts

Rather than supplanting formal face-to-face interviewing, the SAI© is designed to capture initial accounts and preserve the eyewitness's memory until a formal interview can be arranged. It is important, therefore, to consider how witnesses perform on a subsequent retrieval attempt after having completed an earlier SAI©. Here, the results painted a very positive picture. Witnesses who had completed an initial SAI© reported more correct details on a subsequent report than witnesses who did not complete an initial SAI©, and the accuracy of their subsequent reports was also higher overall.

Why might completing an initial SAI© increase both the amount of detail, and the accuracy of a subsequent report? One possibility is that the SAI© may alter witness expectations of what constitutes a comprehensive memory report, and this training effect may carry over to subsequent accounts. Evidence for this possibility comes from Gawrylowicz et al. (2014a), who had witnesses complete an SAI© or a free recall account for a first event. One week later, the witnesses viewed a second event and completed a free recall account. The witnesses who had completed an SAI© for the first event produced more detailed accounts of the second event, indicating that they had developed transferable skills allowing them to produce more comprehensive memory reports. This effect was replicated by Gawrylowicz et al. (2014b) with older adults.

Each of the meta-analytic effect sizes for subsequent accounts showed substantial heterogeneity, yet the moderators that we were able to examine explained very little of this variability. We found some tentative evidence that the SAI© may be more effective for older than younger adults, at least in terms of reducing the number of incorrect details reported on a subsequent account. Memon et al. (2010) also found that the Cognitive Interview provided larger benefits to older adults than younger adults, though this benefit was evident in correct, rather than incorrect, details. Older adults may benefit more from the retrieval support provided by the SAI© than younger adults do, though if that were the case, we would also have expected to see aging effects on initial account quality, and on correct details. Other possibilities are that the protection from forgetting is stronger for older adults, or that older adults benefit more from the training effects that the SAI© may induce (Gawrylowicz et al. 2014b). However, we do urge caution, as our findings are based on a small number of effect sizes with older adults ($k = 4$), three of which came from the same study (Dando et al., 2020). Furthermore, the p value for coefficient was close to .05, and was statistically significant for just one of three of our dependent variables. Clearly the robustness of this finding is an open question, which can only be addressed with further research with older adult samples.

Exposure to misleading post-event information did not significantly moderate the effect sizes for correct details, incorrect details, or overall accuracy. These results suggest that the SAI© allows witnesses to produce full and accurate accounts, even after they have been exposed to misinformation. However, this conclusion must be viewed as tentative, as we only included non-critical details in our analyses, and therefore we may have missed effects pertaining to critical details. And of course, lack of statistical significance is not evidence of no effect. Indeed, the confidence interval around most of the coefficients in our moderator analyses were wide, indicating a lack of precision in the estimates. We return to this point in the Limitations section.

The type of test used at Time 2 also did not significantly moderate any of the meta-analytic effect sizes. This may indicate that the SAI© provides similar benefits to subsequent retrieval, regardless of the degree of retrieval support provided during that subsequent report. However, our studies were grouped quite crudely, such that we compared studies that used free recall at Time 2 with those that had used any other type of report at Time 2. This “other” category included one study that used a CI (Hope et al., 2014), two that used a second SAI© (Krix et al., 2014; Krix et al., 2015), and two that used cued recall tests (Gabbert et al., 2012, Experiment 2; Mackay & Paterson, 2015). Of course, the retrieval support provided by a cued recall test will depend upon the number of cues that are used – a cued recall test that provides very few prompts will likely produce brief reports that are more comparable to a free recall account than to a structured interview (Fisher et al., 2014). The two cued recall studies we included used 17 and 20 cued recall questions, respectively (Gabbert et al., 2012, Experiment 2; Mackay & Paterson, 2015), so we reasoned that these would provide more retrieval support than a single free-recall prompt. However, inspection of the mean numbers of correct and incorrect details in the study suggests that this assumption may not have been sound, as they were more similar to the free recall studies than to the studies that employed the CI or SAI© at Time 2. Whether the SAI© benefit for subsequent retrieval varies across different types of test, or across different levels of retrieval support, is an open question, ripe for further research.

Risk of Bias

Our Risk of Bias analyses indicated that there is scope for increasing the rigour of studies in this field, thereby minimizing the possibility for study outcomes to be biased towards any particular outcome. First, research protocols could be strengthened by implementing measures to blind research personnel to participants' experimental condition. For example, research materials (e.g., SAI© booklets or free recall booklets) could be handed to participants in opaque folders of equivalent weight such that the experimenter cannot tell which condition the participant is in. Similarly, coders could be blinded to condition if witness reports are first typed out into a separate document and care is taken to eliminate any formatting cues that could break blinding (though this may be difficult to fully achieve in practice). Second, our findings emphasize the need for researchers to be clear and transparent in their reporting. We were often unable to easily determine fairly basic pieces of information, such as cell sizes, participant exclusions, or (in the case of subsequent reports) whether any participants had failed to return for the second session. Non-significant findings were often not reported fully and transparently, and sometimes dependent variables that had been mentioned in the introduction and methods were not reported in the Results. Transparent reporting of methodology and results is crucial for critical evaluation of findings, for replication efforts, and for the accumulation of knowledge. We would encourage researchers in this field to embrace tools such as transparency checklists (e.g., Aczel et al., 2019). Third, the adoption of Open Science practices such as pre-registration and sharing of data, materials, and coding schemes, would bring enormous benefits to the field.

Statistical power of individual studies

The median cell size of studies included in this meta-analysis was 24 for initial accounts (range 16-46) and 22 for subsequent accounts (range 20-43). Justifications for sample sizes were rarely explicitly provided. Where power analyses were conducted, they were based upon large effects (e.g., Krix et al., 2014; Mac Giolla & Emberg, 2020). Using the pwr package for R (Champely, 2018), we conducted sensitivity analyses to explore the statistical power of the average study in this area. For each analysis, we used the median cell sizes (24 for initial reports and 22 for subsequent reports), and we assumed a two-tailed test with an α level of .05. We then calculated power for our six observed meta-analytic effect sizes. The results of these analyses are shown in Table 3.

The average study was well-powered for finding an SAI© effect on correct details on initial accounts. However, for the smaller effects that we observed for incorrect details and accuracy, the power of the average study was low. Furthermore, for subsequent accounts, the average study was underpowered even for correct details. Our meta-analytic effect sizes are likely to be overestimates of the true effect sizes (see, for example, Kvarven et al., 2019); if that is the case, then these sensitivity analyses are *overestimating* statistical power.

Table 3. Sensitivity analyses for hypothetical studies for true effects with the average meta-analytic effect size

	Initial (n = 24)		Subsequent (n = 22)	
	ES	Power	ES	Power
Correct details	0.99	91.88%	0.56	44.24%
Incorrect details	0.75	72.02%	0.32	17.94%
Accuracy	0.23	12.21%	0.36	21.47%

Low power has been noted as a serious concern in other domains of legal psychology (e.g., deception detection; Luke, 2019); investigative interviewing research is no exception. Interviewing research is labour intensive, making large samples expensive and time-consuming to recruit and test. However, within the null hypothesis significance testing paradigm, low statistical power simultaneously increases the Type II error rate and inflates effect sizes for statistically significant effects (see, for example, Yarkoni, 2009). Perhaps researchers should consider adopting networked data collection, pooling resources to conduct larger and more definitive studies than any individual research team could conduct working alone. Such initiatives are becoming more widespread in psychological science and can be an important complement to smaller-scale research programmes (see, for example, the Psychological Science Accelerator; www.psysciacc.org).

Applied Implications

For practitioners, the findings of this meta-analysis are clear: the SAI© produces more detailed initial accounts than less structured types of memory reports, such as free recall. Though SAI© accounts may incorporate some inaccuracies, their overall accuracy is likely to be high, and the absolute number of errors reported is likely to be low. Crucially, a witness who is then interviewed subsequently is likely to produce a fuller and more accurate account if he or she had completed an initial SAI© than if he or she had had no initial retrieval attempt.

From an applied perspective, there is often no real alternative to the SAI©. First responders at active incident scenes have many competing demands on their time and are unlikely to be able to do much more than take very brief verbal accounts from witnesses. The real-world choice is often not between information obtained via SAI© or via an alternative written account or a formal interview, but between the SAI© or nothing. Our meta-analytic effect sizes for initial accounts may *underestimate* the real-world utility of the SAI©, wherein the realistic alternative is either no retrieval attempt or a brief and impoverished verbal account. Indeed, an interesting avenue of future research would be to incorporate comparison conditions that are more closely aligned to the types of brief, spoken interactions that occur between first responders and witnesses at the scene. Such designs would allow estimation of an effect size for immediate reports that may be more informative for practitioners when deciding whether to adopt a tool such as the SAI©.

Importantly, we included studies in which the delay between the witnessed event and the completion of the initial SAI© was less than one hour. The intention, when the SAI© was initially developed, was that it would be deployed at the scene of an incident and used to capture early accounts in the immediate aftermath of the event (Gabbert at al., 2009). Furthermore, it is for this purpose that the SAI© has been recommended for use within the UK by the College of Policing (2019). However, it is quite likely that the delay between an incident and completion of the SAI© will, on average, be longer than one hour in real cases. Frontline officers have many competing demands, and their priorities will often involve securing the scene and ensuring public safety. Witnesses themselves may also need time to calm down if an event was particularly distressing, or they may need to have injuries treated. Even if some witnesses are asked to complete the SAI© while still at the scene, the delay is likely to be greater than an hour. Research has demonstrated that delaying memory reports by days or weeks reduces accuracy and can lead to a shift towards schema-based recall, which also increases the risk that the witness will report schema-consistent incorrect detail (e.g., having seen a gun at a bank robbery; see, for example, Tuckey & Brewer, 2003). There is also evidence that the SAI© decreases in effectiveness with delays of 24 hours or greater (Paterson et al., 2015). Thus, we caution that the results of this meta-analysis can be generalized only to situations wherein the delay between the event and the completion of the SAI© is very short. We also encourage researchers to examine delays of several hours, which may occur in real-world critical incidents with large numbers of witnesses.

Limitations and Future Research Directions

The number of comparisons in this meta-analysis was relatively small, and the confidence intervals around our meta-analytic effect sizes were correspondingly wide. Our conclusions are, of course, subject to revision as further data are accumulated. For some effects (e.g., correct details on initial and subsequent accounts; incorrect details on initial accounts), the lower bound of the confidence interval fell far from zero, suggesting that subsequent research is unlikely to produce qualitative changes in the direction and significance of the effect. Rather, subsequent research is likely to further narrow the uncertainty in the magnitude of the effect. For other effects (e.g., accuracy on initial and subsequent reports, incorrect details), the confidence intervals incorporated values much closer to zero. These findings are more likely to be revised as further data accumulates, and may well change in their significance and/or directionality.

We urge caution against drawing strong conclusions from our moderator analyses, as the number of comparisons included in each category was often small, and the uncertainty around the coefficients was large. Where the number of studies in a meta-analysis is small, power to detect a real moderating effect is low (Cafri et al., 2010; Hedges & Pigott, 2004).

For many moderators, the *p* value was close to .05, and it was never the case that a moderator consistently explained the heterogeneity across initial and subsequent reports, or across all three dependent variables. Furthermore, our moderator categories were often crude, to ensure we could incorporate a reasonable number of comparisons in each category. For example, we categorized retention intervals as one week or less, or more than one week. That we found no significant moderation effect of retention interval does not mean that no effect exists. Perhaps our categories were simply too coarse to detect this source of heterogeneity in the data, or the number of comparisons too few. Indeed, from a theoretical perspective, we would expect retention interval to moderate the effectiveness of an initial SAI© on subsequent reporting, as the original memory trace decays and is subject to distortion. Our moderator analyses should be considered a starting point, which provide suggestions for future lines of enquiry. Non-significant moderators should not be taken to indicate that research is not required, or that the SAI© can be definitively considered to be robust to that particular variable.

Though a formal study space analysis (Malpass et al., 2008) is beyond the scope of this paper, there are important variables that are under-researched. For example, very few studies have manipulated stress, and most have used videos of non-violent mock crimes. Perhaps the next wave of research should seek to recreate the immersive, stressful experience of witnesses more faithfully. How will witnesses respond to the SAI© when they are still experiencing an acute physiological stress reaction from witnessing an unexpected and arousing event, for example? Furthermore, most studies have recruited younger adult populations, with no published research on adolescents and only a handful of studies with older adults. Clearly, there will be a need to recruit more diverse samples of participants to more fully inform policy makers and practitioners about groups for whom the SAI© is more or less likely to be a useful tool.

Conclusions

The SAI© appears to be an effective method for gathering detailed and largely accurate initial accounts following a witnessed event. On average, participants reported an extra 23 additional correct details in an initial SAI© than in an initial free recall account. This relatively large increase in correct details was accompanied by a smaller increase of around three incorrect details, which lowered overall accuracy slightly – by around two per cent. However, despite this small drop in accuracy, the absolute accuracy rate of details reported in the SAI© remained high, at around 90%.

Importantly, memory reports produced at a later time are likely to be more detailed and more accurate if the witness had completed an initial SAI© than if the witness had not engaged in a prior retrieval attempt. On average, participants reported around 11 additional details on a subsequent account if they had completed an initial SAI© than if they had not, and they reported around one less incorrect detail. The net effect of this was to increase the accuracy rate of subsequent accounts by around 3%. These effect sizes were larger when comparing witnesses who had completed an initial SAI© to witnesses who had had no prior retrieval attempt than when comparing them to witnesses who had had some form of initial retrieval attempt, such as a free recall test. In an applied context, where the choice is often between a tool such as the SAI© and a brief, impoverished initial account, the SAI© offers clear investigative benefits.

References

Aczel, B., Szaszi, B., Sarafoglou, A., Kekecs, Z., Kucharský, Š., Benjamin, D., … & Wagenmakers, E.-J. (2019). A consensus-based transparency checklist. *Nature Human Behaviour, 4,* 4-6. doi: 10.1038/s41562-019-0772-6

Boulter, C. R. (2017). Testing two methods to improve eyewitness testimony: Pre-interview instruction techniques and three different types of photo line-ups. Available at: https://search.proquest.com/docview/1989696815

Brewer, N., Vagadia, A. N., Hope, L., & Gabbert, F. (2018). Interviewing witnesses: Eliciting coarse-grain information. *Law and Human Behavior, 42,* 458-471. doi: 10.1037/lhb0000294

Brown, C., Frowd, C. D., & Portch, E. (2017). Tell me again about the face: Using repeated interviewing techniques to improve feature-based facial composite technologies. *Proceedings – 2017 7th International Conference on Emerging Security Technologies, 06-08 Sep 2017, Canterbury, UK.* IEEE, pp. 38-43. doi: doi.org/10.1109/EST.2017.8090396

Cafri, G., Kromrey, J. D., & Brannick, M. T. (2010). A meta-meta-analysis: Empirical review of statistical power, type I error rates, effect sizes, and model selection of meta-analyses published in psychology. *Multivariate Behavioral Research, 45,* 239-270. doi: 10.1080/00273171003680187

Champely, S. (2018). pwr: Basic functions for power analysis. R package 1.2.-2. https://CRAN.R-project.org/package=pwr

College of Policing (2019). *Obtaining initial accounts from victims and witnesses: Guidelines for first responders*. Coventry, UK: College of Policing Limited.

*Curtis, E. (Unpublished). The efficacy of the Self-Administered Interview in minimising the misinformation effect in adolescents.

*Dando, C. J., Gabbert, F., & Hope, L. (2020). Supporting older eyewitnesses' episodic memory: The self-administered interview and sketch reinstatement of context. *Memory*. doi: 10.1080/09658211.2020.1757718

Dando, C. J., Wilcock, R., Behnkle, C., & Milne, R. (2011). Modifying the cognitive interview: Countenancing forensic application by enhancing practicability. *Psychology, Crime & Law, 17,* 491-511. doi: 10.1080/10683160903334212

*Eggleton, R. (2016). Exploring the potential of free and structured collective interviewing.

Fisher, R. P., & Geiselman, R. E. (1992). Memory enhancing techniques for investigative interviewing: The Cognitive Interview. Springfield, IL: Charles C Thomas Books.

Fisher, R. P., Schreiber Compo, N. S., Rivard, J., & Hirn, D. (2014). Interviewing witnesses. In T. J. Perfect & D. S. Lindsay (Eds.). *The Sage Handbook of Applied Memory* (pp. 559-578). Los Angeles, CA: Sage.

Flowe, H. D., Humphries, J. E., Takarangi, M. K., Zelek, K., Karoğlu, N., Gabbert, F., & Hope, L. (2019). An experimental examination of the effects of alcohol consumption and exposure to misleading postevent information on remembering a hypothetical rape scenario. *Applied Cognitive Psychology, 33,* 393-413. doi: 10.1002/acp.3531

*Gabbert, F., Hope, L., Horry, R., Drain, T., & Hughes, C. (2020). Examining the efficacy of the Self-Administered Interview in online and mobile formats. Manuscript under review.

*Gabbert, F., Hope, L., & Fisher, R. P. (2009). Protecting eyewitness evidence: Examinung the efficacy of a self-administered interview tool. *Law and Human Behavior, 33,* 298-307. doi: 10.1007/s10979-008-9146-8

*Gabbert, F., Hope, L., Fisher, R. P., & Jamieson, K. (2012). Protecting against misleading post-event misinformation with a self-administered interview. *Applied Cognitive Psychology, 28,* 568-575. doi: 10.1002/acp.2828

*Gawrylowicz, J., Memon, A., & Scoboria, A. (2014a). Equipping witnesses with transferable skills: The Self-Administered Interview©. Psychology, Crime & Law, 20, 315-325. doi: 10.1080/1068316X.2016.777961

*Gawrylowicz, J., Memon, A., Scoboria, A., Hope, L., & Gabbert, F. (2014b). Enhancing older adults' eyewitness memory for present and future events with the Self-Administered Interview. *Psychology and Aging, 29,* 885-890. doi: 10.1037/a0038048

*Gittins, C. B., Paterson, H. M., & Sharpe, L. (2015). How does immediate recall of a stressful event affect psychological response to it? *Journal of Behavior Therapy and Experimental Psychiatry, 46,* 19-26. doi: 10.1016/j.jbtep.2014.07.006

Goldsmith, M., Koriat, A., & Weinberg-Eliezer, A. (2002). Strategic regulation of grain size memory reporting. *Journal of Experimental Psychology: General, 131,* 73-95. doi: 10.1037/0096-3445.131.1.73

Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods, 9,* 426-445. doi: 10.1037/1082-989X.9.4.426

Higgins, J. P., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., ... & Sterne, J. A. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ-British Medical Journal*, *343*(7829), 889. doi: 10.1136/bmj.d5928

Hope, L., & Gabbert, F. (2011). Protecting eyewitness evidence: What can the Self Administered Interview (SAI©) contribute to the investigation of road traffic incidents? *Impact, 19*, 15-18.

*Hope, L., Gabbert, F., Fisher, R. P., & Jamieson, K. (2014). Protecting and enhancing eyewitness memory: The impact of an initial recall attempt on performance in an investigative interview. *Applied Cognitive Psychology, 28,* 304-313. doi: 10.1002/acp.2984

Horry, R., Colton, L.-M., & Williamson, P. (2014). Confidence-accuracy resolution in the misinformation paradigm is influenced by the availability of source cues. *Acta Psychologica, 151,* 164-173. doi: 10.1016/j.actpsy.2014.06.006

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine, 22,* 2693-2710. doi: 10.1002/sim.1482

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103,* 490-517. doi: 10.1037/0033-295X.103.3.490

Kraus, U., Zeier, F., Wagner, W., Paelecke, M., & Hewig, J. S. (2017). Comparing the quality of memory reports in different initial eyewitness questioning approaches. *Cogent Psychology, 4,* 1403063. doi: 10.1080/23311908.2017.1403063

*Krix, A. C., Sauerland, N., Gabbert, F., & Hope, L. (2014). Providing eyewitnesses with initial retrieval support: What works at immediate and subsequent recall? *Psychology, Crime & Law, 20,* 1005-1027. doi: 10.1080/1068316X.2014.902456

*Krix, A. C., Sauerland, M., Lorei, C., & Rispens, I. (2015). Consistency across repeated eyewitness interviews: Contrasting police detectives' beliefs with actual eyewitness performance. *PLoS One, 10,* e0118641. doi: 10.1371/journal.pone.0118641

Krix, A. C., Sauerland, M., Merckelbach, H., Gabbert, F., & Hope, L. (2015). How effective is retrieval support for witnesses with different levels of working and source memory? *Journal of Cognitive Psychology, 27,* 335-348. doi: 10.1080/20445911.2014.1003219

*Krix, A. C., Sauerland, M., Raymaekers, L. H. C., Memon, A., Quadflieg, C. W. E. M., & Smeets, T. (2016). Eyewitness evidence obtained with the Self-Administered Interview© is unaffected by stress. *Applied Cognitive Psychology, 30,* 103-112. doi: 10.1002/acp.3173

Kvarven, A., Strømland, E., & Johannesson, M. (2019). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behavior, 4,* 423-434. doi: 10.1038/s41562-019-0787-z

LaPaglia, J. A., & Chan, J. C. K. (2019). Telling a good story: The effects of memory retrieval and context processing on eyewitness suggestibility. *PLoS One, 14,* e0212592. doi: 10.1371/journal.pone.0212592

Luke, T. J. (2019). Lessons from Pinocchio: Cues to deception may be highly exaggerated. *Perspectives on Psychological Science, 14*, 646-671. doi: 10.1177/17456919198385258

*Mackay, T. L., & Paterson, H. M. (2015). How does timing of recall affect eyewitness memory and psychological distress? *Journal of Police and Criminal Psychology, 30,* 242-253. doi: 10.1007/s11896-014-9156-z

*Mac Giolla, E., & Emberg, E. (2020). The self-interview: A way to improve the testimony of witnesses with Swedish as a second language. Unpublished dataset.

MacLean, C. L., Gabbert, F., & Hope, L. (2019). The Self-Administered Witness Interview Tool (SAW-IT): Enhancing witness recall of workplace incidents. *Applied Cognitive Psychology, 33,* 1212-1223. doi: 10.1002/acp.3568

MacLean, C. L., Stinson, V., Kelloway, E. K., & Fisher, R. P. (2011). Improving workplace incident investigations by enhancing memory recall. *International Journal of Workplace Health Management, 4,* 257-273. doi: 10.1108/17538351111172617

Malpass, R. S., Tredoux, C. G., Compo, N. S., McQuiston-Surrett, D., Maclin, O. H., Zimmerman, L. A., & Topp, L. D. (2008). Study space analysis for policy development. *Applied Cognitive Psychology, 22,* 789-801. doi: 10.1002/acp.1483

*Maras, K., L., Mulcahy, S., Memon, A., Picariello, F., & Bowler, D. M. (2014). Evaluating the effectiveness of the Self-Administered Interview© for witnesses with Autism

Spectrum Disorder. *Applied Cognitive Psychology, 28,* 693-701. doi: 10.1002/acp.3055

*Matsuo, K., & Miura, H. (2016). Effectiveness of the Self-Administered Interview and drawing pictures for eliciting eyewitness memories. *Psychology, Psychiatry & Law, 24,* 643-654. doi: 10.1080/13218719.2016.1254587

Matuso, K., & Miura, H. (2019). Effects of writing in the Self-Administered Interview on recalling and retaining eyewitness memories. Paper presented at the conference of the European Association of Psychology and Law, Santiago de Compostela, Spain.

McCallum, N. A., Brewer, N., & Weber, N. (2016). Memorial monitoring and control: How confidence and social and financial consequences affect eyewitnesses' reporting of fine-grain information. *Applied Cognitive Psychology, 30,* 375-386. doi: 10.1002/acp.3212

Memon, A., Meissner, C. A., & Fraser, J. (2010). The Cognitive Interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law, 16,* 340-372. doi: 10.1037/a0020518

*Paterson, H. M., Eijkemans, H., & Kemp, R. I. (2015). Investigating the impact of delayed administration on the efficacy of the Self-Administered Interview. *Psychiatry, Psychology and Law, 22,* 307-317. doi: 10.1080/13218719.2014.947670

Paz-Alonso, P. M., & Goodman, G. S. (2008). Trauma and memory: Effects of post-event misinformation, retrieval order, and retention interval. *Memory, 16,* 58-75. doi: 10.1080/09658210701363164

*Pfeil, K. (2018). The effectiveness of the Self-Administered Interview©: A meta-analytic review and empirical study with older adult witnesses. Unpublished doctoral dissertation, available at https://www.repository.cam.ac.uk/handle/1810/274905

Pinto, L. H., Stein, L. M., Gabbert, F., & Hope, L. (n.d.). Can I tell you what happened instead? Testing an oral modality of the Self-Administered Interview©. Unpublished manuscript.

Pinto, L. H., Stein, L. M., Gabbert, F., & Hope, L. (n.d.). Does reviewing a previous statement improve eyewitness recall? Unpublished manuscript.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Roos af Hjelmsäter, E., Strömwall, L. A., & Granhag, P. A. (2012). The self-administered interview: A means of improving children's eyewitness performance? *Psychology, Crime & Law, 18,* 897-911. doi: 10.1080/1068316X.2011.582844

*Thornhill, E. L. (2013). The Self-Administered Interview: Does type of recall tool and gender affect the investigation relevant information obtained after a criminal incident? Unpublished undergraduate dissertation.

Tuckey, M. R., & Brewer, N. (2003). The influence of schemas, stimulus ambiguity, and interview schedule on eyewitness memory over time. *Journal of Experimental Psychology: Applied, 9,* 101-118. doi: 10.1037/1076-898X.9.2.101

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36,* 1-48. doi: 10.18637/jss.v036.i03ci

Viechtbauer, W., López-López, J. A., Sánchez-Meca, J., & Marín-Martínez, F. (2015). A comparison of procedures to test for moderators in mixed-effects meta-regression models. *Psychological Methods, 20,* 360-374. doi: 10.1037/met0000023

Weber, N., & Brewer, N. (2008). Eyewitness recall: Regulation of grain size and the role of confidence. *Journal of Experimental Psychology: Applied, 14,* 50-60. doi: 10.1037/1076-898X.14.1.50

Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology, 36,* 1546-1557. doi: 10.1037/0022-3514.36.12.1546

Yarkoni, T. (2009). Big correlations in little studies: Inflated fMRI correlations reflect low statistical power – Commentary on Vul et al. (2009). *Perspectives on Psychological Science, 4*, 294-298. doi: 10.1111/j.1745-6924.2009.01127.x