

Speaking is silver, writing is golden? The role of cognitive and social factors in written versus spoken witness accounts

Melanie Sauerland · Alana C. Krix · Nikki van Kan · Sarah Glunz · Annabel Sak

Published online: 7 March 2014
© Psychonomic Society, Inc. 2014

Abstract Contradictory empirical findings and theoretical accounts exist that are in favor of either a written or a spoken superiority effect. In this article, we present two experiments that put the recall modality effect in the context of eyewitness reports to another test. More specifically, we investigated the role of cognitive and social factors in the effect. In both experiments, participants watched a videotaped staged crime and then gave spoken or written accounts of the event and the people involved. In Experiment 1, 135 participants were assigned to written, spoken-videotaped, spoken-distracted, or spoken-voice recorded conditions to test for the impact of cognitive demand and social factors in the form of interviewer presence. Experiment 2 ($N = 124$) tested the idea that instruction comprehensiveness differentially impacts recall performance in written versus spoken accounts. While there was no evidence for a spoken superiority effect, we found some support for a written superiority effect for description quantity, but not accuracy. Furthermore, any differences found in description quantity as a function of recall modality could be traced back to participants' free reports. Following up with cued open-ended questions compensated for this effect, although at the expense of description accuracy. This suggests that current police practice of arbitrarily obtaining written or spoken accounts is mostly unproblematic.

Keywords Eyewitness testimony · Recall modality · Recall instructions · Interviewer presence · Cognitive load

While silence is certainly not golden when it comes to obtaining eyewitness evidence, the question of whether a witness should preferably speak or write when testifying is a more difficult one. When the police obtain information from eyewitnesses, they can ask for a written description of the course of events and the perpetrators, or they can conduct a personal investigative interview. It appears that the modality of an eyewitness report depends on the seriousness of the crime and the importance of a witness for the case, with proceedings in civil cases often requiring written accounts and more serious crimes predominantly involving oral police interviews (Sauerland & Sporer, 2011). The literature on modality effects in various fields suggests, however, that whether an eyewitness report is given in writing or orally may have a significant effect on the amount and accuracy of the information obtained.

In an early study on the impact of modality on speech production, participants had to discuss one of two topics either in writing or orally (Horowitz & Newman, 1964). The results showed that spoken expression was more productive than written expression in terms of expressed ideas and expansion of previously stated ideas, but also in terms of irrelevant ideas, indicating that speaking was more productive but somewhat less efficient than writing. Similarly, Kellogg (2007) found that spoken renarration of a story was more complete and more accurate (proportion correct) but elicited more distortions than did written renarration. In a survey of medical history, Bergmann, Jacobs, Hoffmann, and Boeing (2004) found that participants underreported some diseases in a written questionnaire they had previously reported in a personal interview.

M. Sauerland · A. C. Krix · N. van Kan · S. Glunz · A. Sak
Department of Clinical Psychological Science, Section Forensic
Psychology, Maastricht University, Maastricht, The Netherlands

M. Sauerland (✉)
Faculty of Psychology and Neuroscience, Maastricht University,
P.O. Box 616, 6200 MD Maastricht, The Netherlands
e-mail: Melanie.Sauerland@maastrichtuniversity.nl

In the eyewitness field, only two studies have addressed the modality issue so far.¹ The first one (Bekerian & Dennett, 1990) presented participants with 16 color slides that depicted a visual narrative of a car accident. Consistent with the literature discussed above, spoken reports were more detailed and more accurate than written reports. More recently, Sauerland and Sporer (2011) tested participants' memory of a staged theft presented to them in a video fragment. Analyses of participants' crime and thief descriptions revealed a clear advantage of spoken descriptions, in terms of both description quantity and accuracy.

This spoken superiority effect can be explained by means of physical, cognitive, and social factors. As compared with writing, speaking demands less muscular energy, is acquired earlier in life (Horowitz & Newman, 1964), is more practiced, and does not require the activation of graphemic representations for spelling words (Kellogg, 2007). Furthermore, speaking puts fewer demands on working memory in the sense that it is faster, and thus ideas have to be stored in memory for a shorter period of time before they are expressed. Finally, the speaking conditions in the described studies involved the presence of an interviewer, which may have a facilitating effect by improving motivation to perform and by providing prompts and encouraging nonverbal cues (Bergmann et al., 2004; Rosenthal, 2002).

A different line of research, however, challenges the idea of a *spoken superiority* effect and rather suggests a *written superiority* effect. Contrary to Kellogg (2007), Grabowski (2007) argued that speaking, not writing, puts higher demands on working memory, such that it impedes output monitoring, lacks self-pacing, and is associated with a larger ratio of produced information units per time interval, thus increasing cognitive load. These assumptions were tested in a series of experiments with different stimulus materials. In an attempt to test the impact of each of the three factors, four conditions were employed—namely, written, invisible written, spoken-voice recorded, and spoken. While writing allows for monitoring previously produced information and control over the produced information units per time interval, as well as self-pacing, invisible writing disables monitoring. Voice recording additionally disables control over time per unit, and normal speaking also disables self-pacing. Note that participants in the voice recording condition were allowed to pause recording but not to rewind and revise. No differences across conditions

were found when European states and capitals were recalled (Experiment 1), while the two written conditions outperformed the two spoken conditions when participants recalled 40 objects they had studied earlier (Experiment 2). The results of a third experiment in which participants watched a filmed theft somewhat resembled those of Horowitz and Newman (1964), with speaking participants reporting more episodes in total but also more repetitions. When repetitions were eliminated, the difference between the two groups in terms of number of episodes recalled disappeared. Writing, however, elicited fewer errors than did speaking. Again, no differences occurred within the two written and the two spoken conditions, indicating that the smaller ratio of produced information units per time interval might be the most crucial factor driving the written superiority effect found here (Grabowski, 2007). Taken together, these results speak to the idea that writing imposes less cognitive load than speaking.

Note, however, that this account remains silent as to the possible facilitating or inhibiting influences of interviewer presence. While Bergmann et al. (2004) emphasized possible positive effects, Wagstaff et al. (2008) focused on potential negative effects. Specifically, they proposed a cognitive-neuropsychological model of social inhibition that postulates that the presence of others places demands on the frontal and executive systems, including working memory, source monitoring, and cognitive inhibition (Kane & Engle, 2002; Mitchell, Johnson, Raye, & Greene, 2004). These systems determine recall success (Engle & Kane, 2004). If the mere presence of others increases cognitive load, this would result in decreased processing capacity for recall. Wagstaff et al. tested this idea in an experiment implementing three different conditions: interviewer only (control), interviewer and one observer, and interviewer and two observers. In agreement with the postulated model, the number of correct responses was significantly smaller in the two experimental conditions than in the control condition, and participants in the two-observer condition performed worse than participants in the one-observer condition. Although Wagstaff et al. did not include a no-interviewer condition, such a condition should be superior to conditions with interviewers, according to the model.

To summarize, different explanatory accounts exist in favor of both a spoken and a written superiority effect, and both have been supported by empirical evidence. It is the aim of the present study to put the recall modality effect in the context of eyewitness testimony to another test and to investigate possible underlying mechanisms. In Experiment 1, we investigated the role of cognitive demand and interviewer presence in written, spoken-voice recorded, spoken-distracted, and spoken-videotaped conditions. Following Grabowski (2007), cognitive load should be lowest in the written condition (output monitoring, self-pacing, and control over output

¹ Another early witness study on the modality effect exists that found no effect (Lipton, 1977). However, neither effect sizes nor means and standard deviations to establish them were reported to assess the statistical power of the tests conducted. Furthermore, tests of interactions with the other independent variables investigated (sex, retention interval between encoding and recall, structure of recall) were missing. Hence, no conclusions can be drawn from this study (cf. Sauerland & Sporer, 2011).

production per time interval), followed by the spoken-voice recorded (only self-pacing), the spoken-videotaped, and the spoken-distracted conditions (additional cognitive load). Accordingly, recall performance should be expected to decrease in this order.

If, however, writing puts higher demands on working memory than does speaking (Kellogg, 2007), the written condition should be inferior to the spoken-videotaped condition, while no clear predictions can be made for the relation between the written and spoken-distracted condition. In this account, performance in the spoken-voice recorded condition should be comparable to that in the spoken-videotaped condition. Taking into account the possible role of an interviewer (Wagstaff et al., 2008), however, the spoken-voice recorded condition should outperform the spoken-videotaped condition. More specifically, if the presence of others increases cognitive load, the written and the spoken-voice recorded conditions should outperform the spoken-videotaped and spoken-distracted conditions.

We also tested the impact of different levels of executive functioning (i.e., an individual threshold that marks the point when cognitive load is experienced as high) on performance in written relative to spoken statements. For this purpose, we administered several tests that measure different aspects of executive functioning, such as working memory capacity, source memory, and cognitive inhibition. We expected that executive functioning should predict recall performance to a greater extent in conditions imposing higher, rather than lower, cognitive load, because when tasks are less demanding, participants can fully allocate resources to the recall task at hand. In contrast, when performing a recall task with high cognitive load, individuals with higher, relative to lower, executive functioning should be at an advantage.

Experiment 1

Method

Participants

One hundred thirty-five women² ($M_{\text{age}} = 21.0$ years, $SD_{\text{age}} = 2.3$; range, 18–28) participated in return to course credit or a €10 voucher. One participant was excluded because it was unclear which film condition she had been assigned to. Participants were bachelor (84.4 %) and masters (14.1 %) students or were members of the general public (1.5 %). All participants were tested in their mother tongue (Dutch, $n = 87$; German, $n = 48$). The study was approved by the local ethical committee.

² Only females were included for testing an unrelated research question we wished to test.

Design

Participants were randomly assigned within a 4 (interviewing condition: written vs. spoken-voice recorded vs. spoken-videotaped vs. spoken-distracted) \times 4 (film version: 1 through 4) between-participants design. The number of Dutch and Germans was counterbalanced across conditions.

Materials

Stimulus films To avoid a possible influence of features innate to the actor in the perpetrator role (e.g., facial distinctiveness, nature of clothing, or number of clothing items), we created four different film versions using the same four female actors, rotating their roles. Each film version was edited to last approximately 3:20 min. The action in all films can be described as follows:

Two women (the later thief and accomplice) meet in a bar and order drinks. Then the later victim enters and takes a seat at the bar. While talking to the barkeeper, the thief inspects the bag of the victim and pushes it off a stool. Either the thief or the victim picks it up, depending on the film version, and puts it back onto the stool. The thief walks back to the table. Eventually, the thief and the accomplice approach the bar to pay their drinks. While the accomplice is distracting the victim by paying the bill, the thief steals the victim's wallet from her bag. The thief and the accomplice leave. When the victim wants to pay, she cannot find her wallet.

Tasks measuring executive functioning The *operation span* (Ospan) task is a complex span task to measure working memory capacity (Engle, Cantor, & Carullo, 1992). That is, participants are required to pursue a secondary task (solve arithmetic problems), while remembering words. Specifically, participants are presented with equation–word pairs (e.g., “Is $(10/5) - 3 = 2$? PAINT”). After reading out the equation and determining its accuracy, participants read out the to-be-remembered word. Following each trial (ranging from two to five equation–word pairs), participants are prompted to write down the to-be-remembered words. In total, there are 12 trials. According to the partial-credit unit scoring (Conway et al., 2005), a correctly recalled word is considered a correct response, irrespective of whether the word was recalled in the correct order. Then the accuracy across trials is calculated.

We used two different *source monitoring* tests (see Unsworth & Brewer, 2010a), which were taken from another study (Krix, Sauerland, Merckelbach, Gabbert, & Hope, 2013). In the picture source recognition test, participants are shown 30 pictures that appear for 1 s one at a time in one of four quadrants on screen. At test, participants are presented with 30 old and 30 new pictures. They indicate whether a

picture is old or new. If considered old, they are asked in which quadrant it appeared.

In the gender source recognition test, participants hear 30 English one-syllable nouns, which are spoken by a female or a male speaker. At test, participants are presented with 30 old and 30 new words. Participants indicate whether a word is old or new. If considered old, they specify whether it was spoken by the male or the female speaker. The scores are the proportions of correct responses.

In the *random number generation* (RNG) task, which measures cognitive inhibition (Ginsburg & Karpiuk, 1994, 1995), participants randomly generate numbers ranging from 0 to 9 at a pace of one number per second, as indicated by a metronome. Scoring was based on the indices described by Peters, Giesbrecht, Jelicic, and Merckelbach (2007). We will focus on repetition (i.e., identical pairs; e.g., 2, 2), seriation (i.e., consecutive digrams; e.g., 1, 2), poker (i.e., repetitions within 20 sequences of 5 successive responses), and variance of digits.

As simple span measures that do not include a secondary task (Unsworth & Engle, 2007), we employed the *forward and backward digit span* tasks (Wechsler, 1997). Here, strings of digits are read out by the experimenter. The string length increases, ranging from three to eight and from two to seven in the forward and backward versions, respectively. Participants have to repeat the strings, either in the same (forward span) or in the reversed order (backward span). The tasks are stopped when two consecutive errors occur. The length of the last recalled string constitutes the score.

Distractor task A distractor task was used to increase cognitive load during retrieval in the spoken-distracted condition. Participants were shown a presentation on a screen that consisted of 90 % green and 10 % red balls. They were instructed to put up their hand every time a red ball was presented. When they forgot to raise their hand, the experimenter reminded them to do so.

Procedure

To ensure that participants were unaware of the fact that they were expected to act as witnesses, they were told that it would be their task to judge social situations. After giving consent, participants watched one of four stimulus films and were instructed to pay close attention, since they would be questioned about it afterward. If participants recognized one of the actors, they were debriefed and excluded. Otherwise, participants proceeded with the executive functioning tasks, followed by the free recall (FR) instructions regarding the sequence of events. Specifically, participants were asked to report everything they could remember about the actions and surroundings as completely, in as much detail, and as accurately as possible. Participants were discouraged from guessing. Thirteen cued open-ended questions (CQs) followed (see

Appendix 1). Next, participants described all persons (thief, accomplice, victim, and barkeeper). For each person, participants received FR instructions first. Again, participants were asked to be as complete, detailed, and accurate as possible. The description should be so specific that the described person could be recognized in a crowd. Again, guessing was discouraged. After all four FRs had been completed, participants answered 12 more CQs about the appearance of each actor (see Appendix 2). Finally, participants were thanked for their participation and debriefed via e-mail after testing was concluded.

Coding of descriptions

For coding the quantity (sum of correct, incorrect, and confabulated details) and accuracy (number of correct details divided by quantity) of participants' descriptions, different coding schemes were developed for each film version. For each of the four films, two of three trained coders (ABC) independently coded all details reported by 10 participants (i.e., 40 statements in total). Specifically, 10 statements each were coded by two coders referring to film 1 (coders A and B), 2 (AC), 3 (BC), and 4 (AC).³

Details were coded as correct or incorrect if they did or did not match the content of the stimulus film, respectively. Details were considered confabulated when they were both incorrect and nonexistent (e.g., describing a hat in the absence of head gear; see Dando, Wilcock, & Milne, 2009, for a similar approach). A statement such as “the woman (1) wore a black (2) t-shirt (3)” yielded three details. Although we did not explicitly code for the precision of details, our coding scheme was so extensive that it accommodated for responses of varying levels of precision (e.g., dark vs. navy blue). Subjective details (e.g., “beautiful shirt”) were not coded. If participants refrained from responding to a cued question (i.e., a “don't know” response), this was accepted as such (after all, participants were instructed not to guess) and regarded as a sign that the participants had exerted their report option (i.e., the freedom to withhold details; Koriat & Goldsmith, 1996). Thus, it was not considered an omission error. To code the accuracy of age, height, and weight estimates, we accepted deviations of 2 years, 4 cm, or 3 kg from the true value. If an interval was given (mostly in FRs) and the range was smaller

³ In hindsight, we realized that the coders were not formally blind to the conditions in this experiment, since they could infer the spoken-videotaped or spoken-distracted conditions, if an interviewer remark was transcribed or if the transcript contained fragments that could be attributed to spoken language. The written versus spoken-voice recorded conditions and the spoken videotaped versus spoken distracted conditions were much less distinguishable. We changed our procedure for Experiment 2 to ensure coders' blindness. Although the procedure in Experiment 1 was suboptimal, the replication of the general pattern of results in Experiment 2 seems to suggest that there is no reason to believe that the procedure employed in Experiment 1 compromised the results.

or analogous to the ranges we used (4 years, 8 cm, or 6 kg), the answer was coded as correct. If the range was larger, however, the answer was coded as incorrect. We are aware that this approach is not in line with grain size theory (Goldsmith, Koriat, & Pansky, 2002), according to which any interval, irrespective of its width, would be considered correct as long as it contains the true value. We decided for this approach for the following reasons. First, when a very wide interval is given, this may be uninformative at best, but adverse at worst, since it entails that a wider range of innocent people could become the focus of the investigations. Second, our coding procedure serves to achieve higher consistency when scoring point versus interval estimates. If, for example, the true age is 30, a point estimate of 20 would be coded as incorrect. It is not reasonable that an interval estimate of 20 to 40 should be coded correct. Third, since the instructions required participants to provide point values, interval estimates were provided very seldom. As a result, the precise nature of the coding procedure of the interval estimates is unlikely to influence our overall results.

Interrater reliability was found to be *substantial* to *almost perfect* (Landis & Koch, 1977). Specifically, for correct recall, Cohen's κ was .89, .83, .89, and .87 for films 1–4, $ps < .001$, respectively. For incorrect recall, κ coefficients were .89, .77, .93, and .75, $ps < .001$, respectively.

Results

For both experiments, we report Cohen's d (Cohen, 1988) for main effects, with $df = 1$ in the numerator, and η_p^2 for main effects, with $df > 1$ in the numerator, and interaction effects (see Sporer & Cohn, 2011). To investigate recall performance as a function of modality (written vs. spoken-voice recorded vs. spoken-videotaped vs. spoken-distracted) and film version (1 through 4), we calculated two-way ANOVAs. Since there were no significant interactions between film version and modality, we collapsed data across film versions for the following analyses, $F_s \leq 1.51$, $ps \geq .153$, $\eta_p^2s \leq .10$. Thus, we computed one-factorial ANOVAs with modality as independent variable and description quantity and accuracy as dependent variables.

An alpha level of .05 was applied for the statistical tests of main effects and interactions. For pairwise comparisons of the modality groups, we made use of an adjusted Bonferroni correction (Shaffer, 1986) and set the critical $\alpha = .017$. Note that post hoc contrasts that are not mentioned were nonsignificant. Results are reported for FRs and the combination of FR and CQ (FR–CQ). Table 1 displays the mean description quantity and accuracy observed for FRs and FR–CQs.

Quantity and accuracy of event descriptions

We found a significant main effect of modality on FR quantity, $F(3, 130) = 3.67$, $p = .014$, $\eta_p^2 = .08$. Post hoc test analyses

revealed that written FRs were more detailed ($M = 88.88$) than spoken-voice recorded statements ($M = 69.50$), $p = .002$, $d = 0.80$, and tended to be more detailed than spoken-distracted statements ($M = 76.06$), $p = .037$, $d = 0.54$. Spoken-videotaped statements also tended to contain more details ($M = 81.73$) than spoken-voice recorded statements, $p = .047$, $d = 0.47$.

For FR–CQs, the main effect of modality on quantity was marginally significant, $F(3, 130) = 2.48$, $p = .064$, $\eta_p^2 = .05$. As for FRs, written statements were more detailed ($M = 101.52$) than spoken-voice recorded statements ($M = 85.53$), $p = .011$, $d = 0.65$.

Modality had no impact on the accuracy of FRs and FR–CQs describing the event, $F_s(3, 129) \leq 0.62$, $ps \geq .324$, $\eta_p^2s \leq .01$.

Quantity and accuracy of person descriptions

Modality had no impact on the quantity or accuracy of FRs and FR–CQs referring to person descriptors, $F_s(3, 133) \leq 1.77$, $ps \geq .156$, $\eta_p^2s \leq .04$.

Executive functioning and interview performance

To reduce the number of predictors to be entered into the regression analyses, we ran a factor analysis (rotation method: Varimax). For this purpose, repetition, seriation, and poker scores were inverted so that higher scores would be associated with higher executive functioning. Using the Kaiser criterion, this yielded a four factor solution that explained 65.75 % of the variance. The factor loadings can be found in Table 2.

Although all four factors are relevant to eyewitness testimony, we could not enter all of them into the regression equation, since that would have required a much larger sample size (i.e., $N = 186$; Tabachnik & Fidell, 2007). Following theoretical considerations on the contribution of each factor to eyewitness memory, we selected factors 2 and 3, on which source and working memory capacity loaded the highest. Especially the latter has been found to be associated with both correct and incorrect recall (Unsworth & Brewer, 2010b). In contrast, cognitive inhibition (as measured by RNG and mainly loading on factors 1 and 4) is mostly associated with intrusions (but not correct recall; e.g., Peters, Jelicic, Haas, & Merckelbach, 2006). Hence, we considered our factor selection to yield a more complete picture of the relationship between modality, executive functioning, and recall performance than any other possible selection, taking sample size restraints into account. Note, however, that additional tentative analyses with the scores of factors 1 and 4 yielded no significant effects, F -changes ≤ 1.66 , $ps \geq .179$, and $F_s \leq 1.43$, $ps \geq .228$.

To code for modality, we added dummy variables to the regression equation. We calculated the regression equations, once determining the spoken-videotaped group and once the written group as control group. Note that while this approach

Table 1 Mean event and person description quantity and accuracy as a function of modality condition (Experiment 1)

		Written		Spoken-Videotaped		Spoken-Voice Recorded		Spoken-Distracted	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Event Description									
Quantity	FR	88.88 ^{ao}	24.53	81.73 [†]	27.91	69.50 ^{at}	23.85	76.06 ^o	23.20
	FR–CQ	101.52 ^a	24.75	97.37	26.80	85.53 ^a	24.15	92.09	25.76
Accuracy (%)	FR	98.59	2.04	98.72	2.40	98.81	1.83	98.62	1.86
	FR–CQ	96.42	3.40	95.56	3.08	95.38	3.08	95.58	3.60
Person Descriptions									
Quantity	FR	27.67	7.54	25.12	9.41	23.32	7.08	27.03	7.94
	FR–CQ	53.58	9.39	52.82	10.48	51.56	9.12	55.76	9.63
Accuracy (%)	FR	82.22	8.03	87.02	8.23	85.36	10.17	83.21	10.54
	FR–CQ	73.75	5.90	73.75	6.11	72.40	7.45	71.87	7.47

Note. FR = free recall; FR–CQ = combination of free recall and cued questions. Means sharing the same superscript letter within a row differ at $p < .017$. Means sharing the same superscript symbol differ at $p < .05$

does not influence the results of the regression analyses in terms of interaction and main effects, it allows for pairwise comparisons between the interview conditions. The predictors were entered into the regression in two blocks. To analyze the interaction with the source monitoring factor, in the first block, all main effects and the interactions involving the working memory capacity factor were entered, followed by the interaction terms of the source monitoring factor. To analyze the interaction with the working memory capacity factor, this was done the other way around (i.e., for every dependent variable, two regression analyses were conducted). When the interactions were nonsignificant, we reran the analyses with the main effects only. As dependent variables, we again used description quantity and accuracy. Analyses were run separately for event and person descriptions. For the sake of brevity, we report the FR–CQ analyses only.

Event descriptions No significant interaction effects were found between modality and the source monitoring factor, F -changes ≤ 0.67 , $ps \geq .570$. That is, the relationship between source monitoring and recall performance (quantity and accuracy) was not moderated by modality. However, as one would expect, a higher source monitoring factor score was associated with higher event description quantity and accuracy, $R^2 = .12$, $F(5, 120) = 3.30$, $p = .008$, $\beta = .26$, $p = .004$, and $R^2 = .18$, $F(8, 117) = 3.25$, $p = .002$, $\beta = .34$, $p < .001$.

For the working memory capacity factor, there was no significant interaction with modality, F -change = 2.39, $ps \geq .073$, and there was no main effect, $R^2 = .12$, $F(5, 120) = 3.30$, $p = .008$, $\beta = .11$, $p = .192$, for description quantity.

The interaction between modality and the working memory capacity factor score on the accuracy of the reported event details was significant. However, comparisons of all other

Table 2 Factor loadings for factor analysis of measures of executive functioning for Experiments 1 and 2

	Experiment 1				Experiment 2	
	Factor 1	Factor 2 (WMC)	Factor 3 (SM)	Factor 4	Factor 1 (WMC)	Factor 2 (SM)
(A)Ospan	–.09	.65	.05	.42	.59	.09
Forward span	.10	.68	–.01	–.04	.85	.01
Backward span	<.01	.82	.03	–.08	.73	–.01
Picture source	.13	.01	.82	–.07	.07	.82
Gender source	–.14	.03	.79	.12	.01	.81
Repetition (RNG) inverted	.93	–.04	–.02	–.03		
Seriation (RNG) inverted	.08	.23	–.02	.64		
Poker (RNG) inverted	.93	.11	<.01	.02		
Digit variance (RNG)	–.09	–.29	.07	.70		

Note. In Experiment 1, the operation span (Ospan) task was used; in Experiment 2, the automated Ospan (AOspan) task. Factor loadings in bold indicate the highest factor loading for a given measure of executive functioning. WMC = working memory capacity, SM = source monitoring, RNG = random number generation

interviewing conditions with the spoken-videotaped group yielded no significant interaction. When the written group was determined as the control group, the interaction term of the working memory factor score and the spoken-voice recorded group was significant, $\beta = -.28$, $p = .022$. The post hoc analyses indicated a nonsignificant relationship for the written group, $R^2 = .03$, $F(1, 29) = 0.96$, $p = .335$, $\beta = .18$, and a significant negative relationship for the spoken-voice recorded group, indicating that a lower working memory factor score was associated with higher accuracy, $R^2 = .18$, $F(1, 30) = 6.70$, $p = .015$, $\beta = -.43$. Table 3 displays the referring statistics.

Person descriptions For the person descriptions, there were no significant interactions between modality and the source monitoring or the working memory capacity factors, F -changes ≤ 1.05 , $ps \geq .376$. Models without interaction terms also returned no significant results, $F_s \leq 2.11$, $p = .068$.

Discussion

In Experiment 1, we investigated the role of cognitive demand and interviewer presence in the recall modality effect in eyewitness performance. To this end, four different interviewing conditions were tested. Two conclusions can be drawn from

the findings: First, if anything, our results provide support for a written superiority effect (Grabowski, 2007) for description quantity, but not accuracy. Specifically, written FRs describing the event were more detailed than spoken-voice recorded reports and tended to be more detailed than spoken-distracted reports. Additionally, when looking at the means of the remaining nonsignificant comparisons, there was a tendency of the written condition to elicit the most detailed reports, as compared with all spoken conditions. Note, however, that not a single comparison between the standard written and spoken conditions (i.e., written vs. spoken-videotaped) produced a significant result.

Second, on a descriptive level, the spoken-voice recorded condition consistently elicited the least detailed reports. This contradicts the cognitive-neuropsychological model of social inhibition, which postulates that the presence of an interviewer should adversely affect recall performance (Wagstaff et al., 2008). According to this account, the spoken-voice recorded condition should have outperformed the spoken-videotaped condition. Possible *positive* effects of interviewer presence might be able to explain the results. For example, interviewers may prompt witnesses to make continued efforts to retrieve more details (Bergman et al., 2004; Sauerland & Sporer, 2011). It is difficult, however, to explain the present results

Table 3 Interaction between modality and working memory capacity (WMC) factor score for event descriptions (regression) of Experiment 1

	Event Accuracy	
R^2	.20	
F -change	2.77	
p	.045	
	WMC (control group = spoken-videotaped)	WMC (control group = written)
B	0.01	0.01
$SE B$	0.01	0.01
β	.15	.20
p	.467	.219
	Interaction WMC \times written	Interaction WMC \times spoken-videotaped
B	< .01	< -.01
$SE B$	0.01	0.01
β	.03	-.02
p	.837	.837
	Interaction WMC \times spoken-voice recorded	Interaction WMC \times spoken-voice recorded
B	-0.02	-0.02
$SE B$	0.01	0.01
β	-.25	-.28
p	.068	.022
	Interaction WMC \times spoken-distracted	Interaction WMC \times spoken-distracted
B	<.01	<.01
$SE B$	0.01	0.01
β	.05	.03
p	.680	.815

with a lack of social interaction alone, given that written statements were also made without social interaction. Although speculative, it could also be that students generally lack dictating experience, causing uneasiness while making statements in the spoken-voice recorded condition (Gould, 1978; Gould & Boies, 1978). It would be interesting to include a spoken-voice recorded group that is more practiced in the use of voice recorders to test this notion.

The interactions between executive functioning measures and modality were mostly nonsignificant, making it difficult to conclusively assess the impact of executive functioning as a function of recall modality. Note that the only significant result we obtained in this regard was contrary to earlier findings. Usually, lower working memory capacity is associated with more recall errors (Engle & Kane, 2004). However, we found lower working memory capacity to be associated with higher recall accuracy within the spoken-voice recorded group. Note, however, that we did not find this pattern for any of the other groups. We have no ready explanation for this unanticipated finding and refrain from speculating about it, since it might simply constitute an outlier.

Further exploring the modality effect in witness statements, in Experiment 2, we sought to examine the effect of instruction comprehensiveness on recall performance in written versus spoken accounts. Note that differences in the employed instructions could also serve as an explanation for the contradictory results found in Experiment 1 and by Sauerland and Sporer (2011) regarding the modality effect. Although both studies used FR instructions, Sauerland and Sporer's were less specific. For the event descriptions, they only instructed participants to describe the crime as "detailed as possible." In Experiment 1, however, we gave participants more guidance by asking to give a "complete, detailed, and accurate" description of "actions, events, and the surroundings." Guessing was discouraged. For the description of the perpetrator, the participants of both studies were instructed to give person descriptions that were specific enough to identify the referring persons from a crowd. Nevertheless, in Experiment 1, we added the terms "complete," "detailed," and "accurate" and again discouraged guessing. Possibly, these more specific instructions gave participants the cues they needed to recall more relevant information, thereby eliminating the difference between the two modalities found earlier. Experiment 2 followed up on this idea by varying the supportive nature of the recall instructions for written and spoken-videotaped modality conditions. While the scarce instructions merely included the prompt to describe all details that came to mind about the event and the appearance of the persons involved, the comprehensive instructions contained several components that are known to support retrieval and facilitate recall. First, borrowing elements from rapport-building (e.g., Vallano & Schreiber Comp, 2011), the instructions transferred control to the participants and pointed out that only they had the crucial information that

was needed for solving the crime. Second, using the report everything component, which, like rapport-building, also comes from the Cognitive Interview (Fisher & Geiselman, 1992), the comprehensive instructions invited an account that was as complete and accurate as possible. Third, in order to reduce cognitive load, it was pointed out that recall order did not matter and that details should be recalled as they came to mind. Finally, the comprehensive instructions contained nonleading recall cues, prompting participants to recall details referring to actions, objects, surroundings, facial details, clothing, and so forth (see Gabbert, Hope, & Fisher, 2009). The idea was that these components should facilitate recall and free resources for retrieval. We hypothesized that comprehensive instructions should be especially beneficial under recall conditions that are highly demanding. If the differences in the results of Sauerland and Sporer (2011) and Experiment 1 originate primarily from differences in the instructions, one would expect that comprehensive instructions would be more beneficial for writing participants than for speaking ones. If, however, writing is less demanding than speaking, as suggested by Grabowski (2007) and the results of Experiment 1, the beneficial value of the comprehensive instructions should be greater for speaking than for writing participants. Furthermore, we expected comprehensive instructions and less demanding description conditions to elicit more detailed reports than scarce instructions and more detailed description conditions.

Experiment 2

Method

Participants

One hundred twenty-four participants (95 women; $M_{\text{age}} = 21.5$ years, $SD_{\text{age}} = 4.0$; range, 18–57) took part in exchange for course credit or a €15 voucher. Participants were bachelor (90.3 %) and masters (6.5 %) students, or were members of the general public (3.2 %). All participants were tested in their native language (Dutch, $n = 78$; German, $n = 46$). The study was approved by the local ethical committee. Four participants were excluded from analysis because they were outliers recalling only very few details. This could be attributed to misunderstanding the instructions. Specifically, these participants erroneously understood that their task was to report only the beginning of the film (instead of the complete film). In fact, they were instructed to testify about the "film they had seen in the beginning."

Design

Participants were randomly assigned within a 2 (modality: written vs. spoken) \times 2 (comprehensiveness of instructions:

scarce vs. comprehensive) \times 2 (film version: 1 vs. 2) between-participants design. The number of Dutch and Germans was counterbalanced across conditions.

Procedure

The procedure of Experiment 2 was very similar to that in Experiment 1, with the following exceptions. Participants saw one of two film versions (films 1 and 2 from Experiment 1), rather than one of four. After watching the stimulus film, participants performed the same executive functioning tasks; however, the RNG task was omitted, and the Ospan task was replaced with the automated Ospan (AOspan) task for practical reasons (Unsworth, Heitz, Schrock, & Engle, 2005). Participants received the following instructions for the event description. The instructions provided in the written and spoken conditions were identical.

Scarce: “We would now like to ask you to answer some questions about the film you saw in the beginning. For this purpose, imagine being a witness reporting to the police about the incident shown in the film. If you have any questions, please turn to the experimenter. Please describe all details about the incident that you can remember. Do not guess about details that you cannot remember.”

Comprehensive: “We would now like to ask you to answer some questions about the film you saw in the beginning. For this purpose, imagine being a witness reporting to the police about the incident shown in the film. Note that only you as a witness have valuable first-hand information that can contribute to solving the case. If you have any questions, please turn to the experimenter. Please describe all details about the incident that you can remember. For this purpose, think of all the persons involved, the actions and events, but also of objects and their positions as well as of the surrounding. Mention things as you remember them. It does not matter whether you remember the details in a different order than the one in which they occurred. Your description should be as complete, detailed and precise, but also as accurate as possible. Don’t leave out any details, but do not guess about details that you cannot remember.”

The person descriptions were as follows (depending on who should be described, thief was replaced with accomplice, victim, or barkeeper):

Scarce: “Please describe all details that you can remember about the thief’s appearance and clothing. Do not guess about details that you cannot remember.”

Comprehensive: “Please describe all details that you can remember about the thief’s appearance and clothing. Your description should be as complete, detailed and

precise, but also as accurate as possible. For this purpose, think of facial details, hair/hairstyle, build, and clothing. Your description should be so specific that reading the description would enable someone to pick the described person from a crowd. Don’t leave out any details, but do not guess about details that you cannot remember.”

Automated Ospan task

The automated version of the Ospan task (Unsworth et al., 2005) is a computerized version of the Ospan task (Engle et al., 1992) which was used in Experiment 1. Here, participants are presented with a set of to-be-memorized letters after solving an arithmetic problem. At test, the letters have to be recalled in the order of presentation by clicking on the appropriate letters. The partial-credit unit scoring was used to obtain participants’ scores.

Coding of descriptions

The coding procedure was analogous to that in Experiment 1. To ensure coders’ blindness to the conditions, the transcriptions were revised by a person uninvolved in the coding process, so that no inferences about the conditions could be made (e.g., deleting clarifications made by the interviewer if the participants had a question or removing hesitation sounds like “uhm”). For establishing interrater reliability, 12 randomly selected statements were coded by three independent coders. According to Landis and Koch (1977), interrater reliability was *almost perfect*. Specifically, Fleiss’s (1971) κ was 0.92 for correct details and 0.84 for incorrect details.

Results

To investigate recall performance as a function of modality (written vs. spoken), instructions (scarce vs. comprehensive), and film version (film 1 vs. film 2), we calculated three-way ANOVAs. When there were no significant interactions between film version and the other variables, we report analyses collapsed across films. Table 4 displays the mean description quantity and accuracy observed for FRs and FR–CQs for event and person descriptions. All nonreported main effects and interactions were nonsignificant.

Quantity and accuracy of event descriptions

For FRs, the main effects of instructions, $F(1, 112) = 8.18, p = .005, d = 0.51$, was significant, indicating that comprehensive instructions elicited more details ($M = 83.57$) than did scarce instructions ($M = 72.97$). The main effect of modality, $F(1, 112) = 11.18, p = .001, d = 0.58$, was qualified by a significant

Table 4 Mean event and person description quantity and accuracy as a function of modality and instruction (Experiment 2)

			Written		Spoken-Videotaped		Total		
			<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Event Description									
Quantity	FR*	Scarce	76.07	22.11	70.06	14.07	72.97 ^a	18.48	
		Comprehensive	92.47	23.01	74.67	19.53	83.57 ^a	22.98	
		Total	84.41 ^b	23.85	72.33 ^b	16.99	78.27	21.44	
	FR–CQ	Scarce	86.07	22.51	86.81	17.88	86.45 ^c	20.08	
		Comprehensive	102.00 ^d	22.50	87.27 ^d	22.70	94.63 ^c	23.60	
		Total	94.17	23.71	87.03	20.22	90.54	22.20	
	Accuracy (%)	FR	Scarce	95.14	4.36	93.65	4.68	94.37	4.55
			Comprehensive	95.25	3.25	95.63	3.25	95.44	3.23
			Total	95.20	3.80	94.62	4.13	94.90	3.97
FR–CQ		Scarce	94.40	4.25	92.92	4.37	93.63	4.34	
		Comprehensive	94.62	3.27	94.16	3.43	94.39	3.33	
		Total	94.51	3.47	93.53	3.96	94.01	3.87	
Person Descriptions									
Quantity	FR	Scarce	27.34	8.75	26.29	7.65	26.80 ^c	8.15	
		Comprehensive	31.47	7.68	30.60	7.91	31.03 ^c	7.74	
		Total	29.44	8.41	28.41	8.02	28.92	8.19	
	FR–CQ	Scarce	50.17	9.16	52.19	9.70	51.22	9.42	
		Comprehensive	51.30	10.13	52.40	10.36	51.85	10.17	
		Total	50.75	9.60	52.30	9.95	51.53	9.77	
	Accuracy (%)	FR	Scarce	84.05	8.57	81.67	10.17	82.82	9.43
			Comprehensive	81.61	9.75	83.72	7.66	82.67	8.76
			Total	82.81	9.19	82.68	9.01	82.74	9.06
FR–CQ		Scarce	70.83	9.01	66.62	10.20	68.66	9.79	
		Comprehensive	71.08	10.38	72.02	7.56	71.55	9.01	
		Total	70.96	9.65	69.28	9.33	70.10	9.48	

Note. Variables marked with an asterisk indicate a significant interaction with the film version. Means sharing the same superscript letter within a row or column indicate significant main effects or simple main effects with $p < .05$. FR = free recall, FR–CQ = combination of free recall and cued questions

modality × film interaction, $F(1, 112) = 4.03, p = .047, \eta_p^2 = .04$. Specifically, in film 1, there was no difference in FR quantity as a function of modality, $F(1, 116) = 0.69, p = .407, d = 0.21$, while in film 2, written FR statements ($M = 90.63$) were more detailed than spoken statements ($M = 71.13$), $F(1, 116) = 14.06, p < .001, d = 1.02$.

For FR–CQ quantity, the main effect of instructions, $F(1, 116) = 4.38, p = .039, d = 0.37$, was qualified by a marginally significant interaction between modality and instructions, $F(1, 116) = 3.90, p = .051, \eta_p^2 = .03$. Simple main effects analyses revealed no modality effect for scarce instructions (written, $M = 86.07$; spoken, $M = 86.81$), $F(1, 116) = 0.02, p = .894, d = -0.04$. With comprehensive instructions, however, writing ($M = 102.00$) elicited more details than did speaking ($M = 87.27$), $F(1, 116) = 7.07, p = .009, d = 0.65$.

Event description accuracy of FRs and FR–CQs did not differ as a function of modality or instructions, and there was no interaction, $F_s \leq 2.11, p_s \geq .149, \eta_p^2_s \leq .01, |d|_s \leq 0.27$.

Quantity and accuracy of person descriptions

For quantity of person descriptions obtained in the FR, only the main effect of instructions became significant, with comprehensive instructions leading to more detailed FRs ($M = 31.03$) than did scarce instructions ($M = 26.80$), $F(1, 116) = 8.33, p = .005, d = 0.53$.

Person description accuracy established for FRs and FR–CQs did not differ as a function of modality or instructions, and there was no interaction, $F_s \leq 2.73, p_s \geq .101, \eta_p^2_s \leq .02, |d|_s \leq 0.30$.

Executive functioning and interview performance

As in Experiment 1, we ran a factor analysis (rotation method: Varimax) on the executive functioning measures in order to reduce the number of predictors that would be entered into a regression analysis. Using the Kaiser criterion, this yielded a two-factor solution (one source monitoring and one working memory capacity factor) that explained 58.79 % of the variance (see Table 2 for the factor loadings). Note that these two factors precisely correspond to the working memory capacity and source memory factors obtained in Experiment 1.

These two-factors, modality, and instruction conditions, as well as the interaction terms, were entered into separate regression equations (enter method) for person and event quantity and accuracy. Again, we report FR–CQ data only. When the interaction terms were nonsignificant, they were removed one at a time, and the analyses were rerun until only main effects or significant interaction terms remained. Table 5 shows the results of the final regression equations, after nonsignificant predictors were removed.

Event descriptions Analogous to the above reported marginally significant interaction between modality and instructions, this interaction was significant (for details, see the ANOVA

reported above). The regression equation regarding accuracy of the event descriptions was nonsignificant.

Person descriptions The regression equation regarding quantity of the person descriptions was not significant. For person description accuracy, a higher source memory factor score was associated with increased accuracy.

Discussion

In Experiment 2, we studied the effect of recall instructions on the recall modality effect in eyewitness accounts. To this end, we included a spoken-videotaped and a written condition and employed either scarce or comprehensive instructions. While Sauerland and Sporer's (2011) results suggested that writing is more demanding than speaking, the results of Experiment 1 and Grabowski (2007) suggest the opposite. We expected the less demanding description condition (i.e., writing or speaking) to produce more detailed accounts and comprehensive instructions to elicit more detailed reports than would scarce instructions. Furthermore, we expected that the beneficial value of comprehensive instructions would be greater under more rather than less demanding conditions (i.e., an interaction).

Table 5 Results of regression analyses predicting recall performance from modality, instructions, and executive functioning factor scores (Experiment 2)

Variable	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>	<i>R</i> ²	<i>F</i>
Quantity of Event Description						.10	2.40*
Modality	2.34	5.65	.05	0.42	.679		
Instruction	17.00	5.72	.39	2.97	.004		
WMC	2.22	2.06	.10	1.08	.284		
SM	0.71	2.04	.03	0.35	.728		
Modality × instruction	−16.38	8.04	−.32	−2.04	.044		
Accuracy of Event Description						.03	0.99
Modality	−0.01	0.01	−.13	−1.38	.171		
Instruction	0.01	0.01	.12	1.33	.188		
WMC	< 0.01	< 0.01	.06	0.59	.559		
SM	< 0.01	< 0.01	−.01	−0.08	.933		
Quantity of Person Descriptions						.03	1.00
Modality	1.90	1.85	.10	1.03	.306		
Instruction	0.77	1.82	.04	0.43	.671		
WMC	1.14	0.93	.11	1.23	.222		
SM	1.26	0.93	.13	1.36	.178		
Accuracy of Person Descriptions						.08	2.61*
Modality	< −0.01	0.02	−.02	−0.21	.832		
Instruction	0.03	0.02	.15	1.66	.099		
WMC	0.01	0.01	.11	1.26	.210		
SM	0.02	0.01	.22	2.40	.018		

Note. WMC = working memory capacity factor score; SM = source monitoring factor score

* $p < .05$

Our results provide only limited support for our hypotheses. First, replicating the well-established finding that recall instructions are an important determinant of recall performance (e.g., Fisher & Geiselman, 1992; Gabbert et al., 2009), we found that comprehensive instructions led to more detailed accounts than did scarce instructions for FRs in general, as well as the event FR–CQs (but not the person FR–CQs). Second, a main effect of modality such that writing was more beneficial than speaking was apparent only for the event FR, but not the person FR or the FR–CQs in general. Third and similarly, an interaction between instructions and modality became significant only for the event FR–CQ (but not the person FR–CQ and FRs in general). Finally, similar to Experiment 1, the level of executive functioning did not interact with the modality to predict recall performance.

Note that our findings lend some support to Grabowski's (2007) hypothesis that writing is less demanding than speaking. They are, however, in direct conflict with the idea that differences in the instructions used in Experiment 1 and by Sauerland and Sporer (2011) might serve as an explanation for the contradictory results found in these two experiments. What is more, the interaction between instructions and modality took a different form than expected. We had expected both written and spoken conditions to profit from the comprehensive instructions, with a greater beneficial value for the more demanding condition. However, speaking participants did not profit from the comprehensive instructions at all, while writing participants did. Thus, the findings are in line neither with the assumption that speaking is more demanding than writing (Grabowski, 2007) nor with the opposing view that writing is more demanding than speaking (Kellogg, 2007). Either way, both conditions should have profited from the comprehensive instructions, only with a greater beneficial value for the one, as compared with the other, condition. Thus, the confusing finding is that speaking participants did not profit from the comprehensive instructions at all when it came to providing event descriptions. One possible explanation for this finding could be that participants might have had difficulties remembering the instructions in the spoken condition. On the other hand, the significant effect of instruction comprehensiveness for person description quantity contradicts this argument.

General discussion

Across two experiments, we investigated the relevance of recall modality for eyewitness performance. First, given that theoretical explanations and supporting empirical data exist for both a possible written and a possible spoken superiority effect (e.g., Grabowski, 2007; Kellogg, 2007; Sauerland & Sporer, 2011), we were interested in the direction of the effect. Second, we wanted to test the mechanisms that drive the effect. To this end, we assessed the role of cognitive demand,

including executive functioning, the presence of an interviewer, and the role of recall instruction comprehensiveness.

Turning to the first question, our data lend no support to the notion of a spoken superiority effect. This is contrary to the idea that speaking demands fewer cognitive resources (Kellogg, 2007) and contradicts earlier findings in the eyewitness literature (Bekerian & Dennett, 1990; Sauerland & Sporer, 2011) and other fields (Bergmann et al., 2004; Horowitz & Newman, 1964). In line with Grabowski (2007), however, the data provide some, albeit limited, evidence for a written superiority effect. Across two experiments, written event FRs were consistently somewhat superior to spoken FRs (note, though, that a significant difference between written and spoken-videotaped conditions was found only in Experiment 2, but not in Experiment 1). However, this effect disappeared when looking at the complete event descriptions (i.e., FR–CQs) and at person descriptions. Furthermore, the accuracy of reports provided in writing or orally did not differ at all. Altogether, our results suggest that any differences in description quantity occur in the FRs and that following up with cued open-ended questions seems to be an effective tool to level possible differences between spoken and written recall. Unfortunately, however, as Tables 1 and 4 indicate, the use of cued questions comes with a decreased level of accuracy, as is known from previous research (Lipton, 1977; Powell, Fisher, & Wright, 2005; Sauerland & Sporer, 2011).

More support for the view that writing might be superior to speaking evolves when taking into the account interviewer presence. In this context, the literature on the effects of the presence of others suggests a facilitative effect on performance for simple tasks but an inhibiting effect for more complex tasks (Bond & Titus, 1983; Zajonc, 1965). Accordingly, performance under more cognitively demanding and, hence, complex recall conditions might deteriorate with the inclusion of an interviewer, while less demanding recall conditions might benefit from it. Contrasting the interviewer-present and interviewer-absent spoken conditions in Experiment 1 yields the conclusion that interviewer presence has a facilitating, rather than an inhibiting, effect on eyewitness recall performance. A direct comparison of the two interviewer-absent conditions in Experiment 1 furthermore suggests that, if anything, participants perform better when writing than speaking. A direct comparison of written and spoken interviewer-present conditions is not possible in the present design, since we did not include a written condition with an interviewer (or rather an observer). On the basis of the data available here, we reason that writing is not more complex than speaking. From this it follows that if the presence of others facilitates spoken witness reports, it should also facilitate written ones. The inclusion of a written condition both with and without an observer present (in addition to interviewer-present and interviewer-absent spoken

conditions) is essential to making an informed judgment on this issue. Ideally, the spoken interviewer-absent condition in such a follow-up study should include participants who are practiced in the use of voice recorders in order to avoid the possible negative effects of operating voice recorders (Gould, 1978; Gould & Boies, 1978).

Moving to the role of cognitive demand in the recall modality effect, little can be said on the basis of the present data. If, for example, output production per time interval was crucial by itself, as suggested by Grabowski (2007), the written condition should consistently have outperformed the spoken ones. It is possible, however, that this factor interacts with interviewer or observer presence, as discussed above.

As for the influence of recall instructions, comprehensive instructions were mostly beneficial *across* modalities, while fluctuations as a function of recall modality were either absent or inconsistent. Similarly, differential effects of executive functioning on recall according to recall modality were absent or inconsistent. From a legal perspective, it is desirable to facilitate all witness reports, including those of individuals with various degrees of executive functioning. Given, however that no differential effects were found for the written as compared with the spoken-videotaped conditions, which could be considered the standard police interviewing conditions, this point may not need to be a concern in the current police practice.

A limitation of the present study is the fact that we tested our participants for visual material only (i.e., no conversations could be heard in the stimulus film). Hence, it is unclear whether our findings transfer to auditory details, and our conclusions must be limited to visual details. This point also pertains to the use of a visual (instead of a verbal) distraction task in Experiment 1. We do not know whether the visual distraction task we used would have inhibited recall of auditory material, too. Previous research indicates that it might not have, since visual distraction during recall selectively inhibits recall of visual material and auditory distraction selectively inhibits recall of auditory material (Vredeveldt, Hitch, & Baddeley, 2011). However, since we tested both Dutch- and German-speaking participants, we wanted to keep the influence of language to a minimum and, hence, refrained from including auditory material in the stimulus film. On the basis of Vredeveldt et al.'s findings and given the nature of the stimulus film (i.e., containing visual details only), the use of a visual instead of an auditory distraction task seemed most appropriate.

Overall, the present results suggest that writing might be superior to speaking when making a witness statement, at least in highly educated samples. This effect might be even stronger when an observer is introduced to the written condition. Given, however, that the deficit in spoken descriptions that might occur in FRs can be compensated by the use of cued open-ended follow-up questions, a general recommendation

for obtaining written, rather than spoken, testimony is unwarranted on the basis of the present data. This is reassuring insofar as it suggests that current police practice of obtaining written versus spoken accounts as a function of crime seriousness and witness centrality is unproblematic. Two notes of caution are in order, however. First, meaningful differences between recall modalities might arise as a function of (lower) educational background, increased task difficulty, personal preference, or other factors. Second, the present findings directly contradict the findings of a recent eyewitness study in which strong effects in support of a spoken superiority effect in witness accounts were found (Sauerland & Sporer, 2011). Of note, the methodologies used here and by Sauerland and Sporer was highly similar in terms of the sample (mainly university students), experimenters (female masters students), stimulus material (staged crime videos with four actors, depicting the theft of an object: a pair of sunglasses or a wallet, comparable in length), and coding procedure. We can only speculate about the reasons underlying these conflicting findings. Possibly, the effect is subject to the decline effect (Lehrer, 2010; Schooler, 2011), or one of the findings constitutes a false alarm. The chance for such a result is 5 % after all. In either case, these contradictory findings in the light of no obvious differences in methodology give reason for caution and should be cause for further investigation into the recall modality effect in eyewitness testimony. For now, we conclude that both writing and speaking seem to be appropriate tools when obtaining eyewitness reports. Which of them is golden rather than silver, however, cannot be decided on the basis of the data available at present.

Author Note We thank Eva Billen and Bo Hollanders for their help in collecting data and Timm M. Kuhl for preparing the data for analyses. This research was supported by a PhD scholarship from Studienstiftung des deutschen Volkes (German National Academic Foundation) awarded to the second author.

Appendixes

Appendix 1 Cued questions for event description

1. Where did the event take place? Describe the location of the event.
2. How long was the duration of the film?
3. How many persons were at the crime scene?
4. How many persons were involved in the theft?
5. How was the wallet stolen? Describe the manner in which the theft occurred.
6. Did the thief have (an) accomplice(s)?

If yes, continue with question 7.

If no, continue with question 9.

7. What did the accomplice(s) do during the theft?
8. Did the accomplice(s) stay at the crime scene after the wallet was stolen? If yes, for how long?
9. What did the bystander(s) do during the theft of the wallet?
10. How long did the theft of the wallet take?
11. Did the thief stay at the crime scene after the wallet had been stolen? If yes, for how long?
12. Where did the thief put the stolen wallet?
13. How long did it take until the victim realized that her wallet had been stolen?

Appendix 2 Cued questions for person descriptions

1. How old was the thief? Give a specific age (e.g., x years).
2. Describe the hair color of the thief.
3. Describe the hair length of the thief.
4. Describe the hairstyle of the thief.
5. Describe the build of the thief.
6. Describe the body proportions of the thief.
7. Estimate the height of the thief (cm). Give a specific height (e.g., x cm).
8. Estimate the weight of the thief (kg). Give a specific weight (e.g., x kg).
9. Did the thief have any distinctive features? If yes, describe them.
10. Describe the clothing of the thief.
11. Did the thief wear jewelry or accessories? If yes, describe them.
12. Are there additional details that you remember, but that you have not been asked about yet? If yes, please describe them.

The questions referring to the other actors were identical.

References

- Bekerian, D. A., & Dennett, J. L. (1990). Spoken and written recall of visual narratives. *Applied Cognitive Psychology*, *4*, 175–187. doi:10.1002/acp.2350040303
- Bergmann, M. M., Jacobs, E. J., Hoffmann, K., & Boeing, H. (2004). Agreement of self-reported medical history: Comparison of an in-person interview with a self-administered questionnaire. *European Journal of Epidemiology*, *19*, 411–416. doi:10.1023/B:EJEP.0000027350.85974.47
- Bond, C. F., & Titus, L. J. (1983). Social facilitation: A meta-analysis of 241 studies. *Psychological Bulletin*, *94*, 265–292. doi:10.1037/0033-2909.94.2.265
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769–786. doi:10.3758/BF03196772
- Dando, C., Wilcock, R., & Milne, R. (2009). The cognitive interview: The efficacy of a modified mental reinstatement of context procedure for frontline police investigators. *Applied Cognitive Psychology*, *23*, 138–147. doi:10.1002/acp.1451
- Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 972–992. doi:10.1037/0278-7393.18.5.972
- Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 145–199). New York: Elsevier.
- Fisher, R. P., & Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing: The cognitive interview*. Springfield, IL: Charles C. Thomas.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*, 378–382. doi:10.1037/h0031619
- Gabbert, F., Hope, L., & Fisher, R. P. (2009). Protecting eyewitness evidence: Examining the efficacy of a self-administered interview tool. *Law and Human Behavior*, *33*, 298–307. doi:10.1007/s10979-008-9146-8
- Ginsburg, N., & Karpiuk, P. (1994). Random number generation: Analysis of the responses. *Perceptual and Motor Skills*, *79*, 1059–1067. doi:10.2466/pms.1994.79.3.1059
- Ginsburg, N., & Karpiuk, P. (1995). Simulation of human performance on a random generation task. *Perceptual and Motor Skills*, *81*, 1183–1186. doi:10.2466/pms.1995.81.3f.1183
- Goldsmith, M., Koriat, A., & Weinberg-Eliezer, A. (2002). Strategic regulation of grain size in memory reporting. *Journal of Experimental Psychology: General*, *131*, 73–95. doi:10.1037/0096-3445.131.1.73
- Gould, J. D. (1978). How experts dictate. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 648–661. doi:10.1037/0096-1523.4.4.648
- Gould, J. D., & Boies, S. J. (1978). How authors think about their writing, dictating, and speaking. *Human Factors*, *20*, 495–505. doi:10.1177/001872087802000415
- Grabowski, J. (2007). The writing superiority effect in the verbal recall of knowledge: Sources and determinants. In Rijlaarsdam, G. (Series Ed.) and M. Torrance, L. van Waes & D. Galbraith (Volume Eds.), *Writing and cognition: Research and applications* (Studies in Writing, Vol. 20, pp. 165–179). Amsterdam: Elsevier.
- Horowitz, M. W., & Newman, J. B. (1964). Spoken and written expression: An experimental analysis. *Journal of Abnormal and Social Psychology*, *68*, 640–647. doi:10.1037/h0048589
- Kane, M., & Engle, R. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *9*, 637–671. doi:10.3758/bf03196323
- Kellogg, R. T. (2007). Are written and spoken recall of text equivalent? *American Journal of Psychology*, *120*, 415–428.
- 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
- Krix, A. C., Sauerland, M., Merkelbach, H., Gabbert, F., & Hope, L., (2013). *Effective retrieval support for eyewitnesses requires good source memory*. Manuscript submitted for publication.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174. doi:10.2307/2529310
- Lehrer, J. (2010, Dec 13). The truth wears off. *The New Yorker*. Retrieved from http://www.newyorker.com/reporting/2010/12/13/101213fa_fact_lehrer

- Lipton, J. P. (1977). On the psychology of eyewitness testimony. *Journal of Applied Psychology*, *62*, 90–95. doi:10.1037/0021-9010.62.1.90
- Mitchell, K. J., Johnson, M. K., Raye, C. L., & Greene, E. J. (2004). Prefrontal cortex activity associated with source monitoring in a working memory task. *Journal of Cognitive Neuroscience*, *16*, 921–934. doi:10.1162/0898929041502724
- Peters, M., Giesbrecht, T., Jelicic, M., & Merckelbach, H. (2007). The random number generation task: Psychometric properties and normative data of an executive function task in a mixed sample. *Journal of the International Neuropsychological Society*, *13*, 626–634. doi:10.1017/s1355617707070786
- Peters, M., Jelicic, M., Haas, N., & Merckelbach, H. (2006). Mild executive dysfunctions in undergraduates are related to recollecting words never presented. *International Journal of Neuroscience*, *116*, 1065–1077. doi:10.1080/00207450600808768
- Powell, M. B., Fisher, R. P., & Wright, R. (2005). Investigative interviewing. In N. Brewer & K. D. Williams (Eds.), *Psychology and law: An empirical perspective* (pp. 11–42). New York: Guilford.
- Rosenthal, R. (2002). Covert communication in classrooms, clinics, courtrooms, and cubicles. *American Psychologist*, *57*, 839–849. doi:10.1037/0003-066X.57.11.839
- Sauerland, M., & Sporer, S. L. (2011). Written vs. spoken eyewitness accounts: Does mode of testing matter? *Behavioral Sciences and the Law*, *29*, 846–857. doi:10.1002/bsl.1013
- Schooler, J. W. (2011). Unpublished results hide the decline effect. *Nature*, *470*, 437. doi:10.1038/470437a
- Shaffer, J. P. (1986). Modified sequentially rejective multiple test procedures. *Journal of the American Statistical Association*, *81*, 826–831. doi:10.1080/01621459.1986.10478341
- Sporer, S. L., & Cohn, L. D. (2011). Meta-analysis. In B. D. Rosenfeld & S. D. Penrod (Eds.), *Research methods in forensic psychology* (pp. 43–62). New York: Wiley.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. New York: Pearson.
- Unsworth, N., & Brewer, G. A. (2010a). Individual differences in false recall: A latent variable analysis. *Journal of Memory and Language*, *62*, 19–34. doi:10.1016/j.jml.2009.08.002
- Unsworth, N., & Brewer, G. A. (2010b). Variation in working memory capacity and intrusions: Differences in generation or editing? *European Journal of Cognitive Psychology*, *22*, 990–1000. doi:10.1080/09541440903175086
- Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*, 104–132. doi:10.1037/0033-295x.114.1.104
- Unsworth, N., Heitz, R. R., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505. doi:10.3758/bf03192720
- Vallano, J. P., & Schreiber Compo, N. (2011). A comfortable witness is a good witness: Rapport-building and susceptibility to misinformation in an investigative mock-crime interview. *Applied Cognitive Psychology*, *25*, 960–970. doi:10.1002/acp.1789
- Vredevelde, A., Hitch, G. J., & Baddeley, A. D. (2011). Eyeclosure helps memory by reducing cognitive load and enhancing visualisation. *Memory & Cognition*, *39*, 1253–1263. doi:10.3758/s13421-011-0098-8
- Wagstaff, G. F., Wheatcroft, J., Cole, J. C., Brunas-Wagstaff, J., Blackmore, V., & Pilkington, A. (2008). Some cognitive and neuropsychological aspects of social inhibition and facilitation. *European Journal of Cognitive Psychology*, *20*, 828–846. doi:10.1080/09541440701469749
- Wechsler, D. (1997). *Administration and scoring manual for the Wechsler Adult Intelligence Scale* (3rd ed.). San Antonio, TX: Psychological Corporation.
- Zajonc, R. B. (1965). Social facilitation. *Science*, *149*, 269–274. doi:10.1126/science.149.3681.269