

# Written vs. spoken eyewitness accounts: does modality of testing matter?

Citation for published version (APA):

Sauerland, M., & Sporer, S. L. (2011). Written vs. spoken eyewitness accounts: does modality of testing matter? *Behavioral Sciences & the Law*, 29(6), 846-857. <https://doi.org/10.1002/bsl.1013>

## Document status and date:

Published: 01/01/2011

## DOI:

[10.1002/bsl.1013](https://doi.org/10.1002/bsl.1013)

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

## Written vs. Spoken Eyewitness Accounts: Does Modality of Testing Matter?

Melanie Sauerland, Ph.D.\* and Siegfried L. Sporer, Ph.D.<sup>†,‡</sup>

---

**The aim of the current study was to test whether the modality of testing (written vs. spoken) matters when obtaining eyewitness statements. Writing puts higher demands on working memory than speaking because writing is slower, less practiced, and associated with the activation of graphemic representations for spelling words (Kellogg, 2007). Therefore, we hypothesized that witnesses' spoken reports should elicit more details than written ones. Participants ( $N=192$ ) watched a staged crime video and then gave a spoken or written description of the course of action and the perpetrator. As expected, spoken crime and perpetrator descriptions contained more details than written ones, although there was no difference in accuracy. However, the most critical (central) crime and perpetrator information was both more extensive and more accurate when witnesses gave spoken descriptions. In addition to cognitive factors, social factors are considered which may drive the effect. Copyright © 2011 John Wiley & Sons, Ltd.**

After witnessing a crime, eyewitnesses are often asked to give their account of the course of events as well as to describe the perpetrator(s). In actual cases, when a serious crime is investigated, these descriptions are usually obtained through a personal interview between the witness and a police officer. Thus, spoken reports are obtained.

In contrast, proceedings in civil procedures often require written accounts. Consequently, the modality of an eyewitness report depends on the seriousness of a crime and the centrality of a witness for a case. An archival analysis has shown that witnesses in real cases provide oral testimony which is subsequently summarized in a written report by the interviewing officer (Sporer, 1992a, 1996). In contrast, in many eyewitness studies, participants' reports are obtained in written (e.g., Heuer & Reisberg, 1990; Ibabe & Sporer, 2004; Migueles & García-Bajos, 1999; Sporer, 1992b; Wells, 1985) rather than spoken form (e.g., Pozzulo, Dempsey, & Crescini, 2009). It is the aim of the current study to investigate whether testing modality has an impact on the quality and quantity of eyewitness reports.

Only few studies have examined the issue of testing modality in the context of eyewitness testimony. Unfortunately, some of these have serious methodological problems. For example, an early study conducted by a legal scholar (Marshall, 1966) was criticized by Bekerian and Dennett (1990) to confound testing modality with retention interval (delay of testing). In the following, we will give an overview of the relevant studies, drawing from research carried out in various fields.

---

\*Correspondence to: Melanie Sauerland, who is now at the Section of Forensic Psychology, Department of Clinical Psychological Science, Maastricht University, P.O. Box 616, 6200 MD Maastricht, the Netherlands. E-mail: melanie.sauerland@maastrichtuniversity.nl

<sup>†</sup>Department of Psychology and Sports Science, University of Giessen, Germany.

<sup>‡</sup>This research was supported by a grant by the Deutsche Forschungsgemeinschaft (DFG Sp262/3-1 and DFG Sp262/3-2) to the second author.

An early study (Horowitz & Newman, 1964) analyzed speech production as a factor of modality. Specifically, participants had to discuss one of two topics (i.e., “What does a good teacher mean to you?”) either in writing or orally. Participants in Experiment 1 were given 2 min for this task. The results showed that spoken discussions elicited more major ideas (defined as meaningful and relevant) and subordinate ideas (elaboration of previously expressed ideas), but also more repetitions than written ones. However, this difference seemed to be due to the time limit. Therefore, participants of Experiment 2 were given 4, 8, or 12 min. Here, the difference in reported major ideas became non-significant after 10 min of writing vs. speaking. When looking at all expressed ideas, including subordinate and non-relevant ones, spoken reports were more productive than written ones at any time interval. Hence, although speaking was more productive, writing seemed more efficient. The greater ease of speaking, on the one hand, and the greater deliberateness and the more serious commitment that writing represents, on the other hand, could be possible explanations (Horowitz & Newman, 1964). Furthermore, speaking uses less muscular energy, is acquired much earlier in life (in early childhood, compared with age 6 years) and is practiced far more than writing.

More recently, Kellogg (2007) argued that writing puts higher demands on working memory than speaking because writing is slower, less practiced, and requires the activation of graphemic representations for spelling words. Accordingly, less attention might be devoted to retrieval for written recall than for spoken recall. To test these assumptions, participants were asked to recall the content of Bartlett’s (1932) famous Native Indian story on the *War of the Ghosts*. Supporting the hypothesis, spoken reports were more complete and more accurate (proportion correct), but also elicited more distortions, than written reports. Apparently, less attention was paid to reconstructing the propositions of the story in written than in spoken recall. This notion was further supported by the finding that the deficit for writing emerged primarily for the middle and final parts of the story. The author concluded that spoken recall seemed to have the advantage of greater accessibility to memory traces.

More evidence for the superiority of spoken recall comes from the medical field (Bergmann, Jacobs, Hoffmann, & Boeing, 2004). In a survey of medical history, good agreement between a personal interview and a written questionnaire administered about 2 years later was obtained for serious diseases requiring hospitalization (e.g., malignant tumors) or continuing medical care (e.g., diabetes mellitus). For less well-defined diseases (e.g., benign tumors) or diseases with intermittent appearance (e.g., ulcer of the stomach) agreement was low. This was mainly due to participants not reporting diseases on the questionnaire they had previously reported in the interview. The authors suggest that the absence of an interviewer may have led participants to judge less serious diseases as not being important enough to report, although over-reporting of diseases in the interviews was also possible.

Altogether, these studies suggest that modality does matter for the recall of textual narratives, with spoken reports eliciting more correct but also more incorrect information. For the eyewitness field, the effect of testing modality for *visual* information may be more relevant than for *textual* narratives. Two studies addressed this issue. An early experiment (Lipton, 1977) used a film that showed a staged shooting and robbery of a pedestrian in a park. No differences regarding completeness or proportion correct were found when comparing written and spoken reports. Unfortunately, neither effect sizes nor means and standard deviations to calculate them were reported to assess the statistical power of the tests conducted. Also, no tests of interactions with the other

independent variables investigated (sex, retention interval between encoding and recall, structure of recall) were reported. In an attempt to clarify the effect of recall modality on memory for visual stimuli, Bekerian and Dennett (1990) presented participants with a visual narrative consisting of 16 color slides, the last of which showed a car accident. In this study, spoken reports elicited more correct descriptors and higher accuracy (proportion correct) than written ones. The authors argued that participants in the writing condition may have filled gaps in their storyline with schema-consistent information, thus leading to less accurate descriptions.

With the possible exception of Lipton's (1977) study, the available data overall suggest that the modality of an eyewitness report could be an important factor that determines how informative and accurate an eyewitness report will be. We tested this hypothesis by presenting participants with a color video showing a staged theft. Participants described the thief and the course of events in writing or orally. We hypothesized that witnesses' spoken reports should contain more details than written ones. For description accuracy, only a very small number of studies exist, yielding inconsistent results. Therefore, we refrained from formulating a directed hypothesis about description accuracy in spoken vs. written descriptions. Furthermore, we investigated whether differences between written and spoken recall emerge only for central as opposed to peripheral details (see Ibabe & Sporer, 2004).

## METHOD

### Participants

In all, 192 individuals (96 males, 96 females; age 19–52 years, median = 24 years) completed this experiment. Most participants were psychology majors (62%) who received course credit for their participation. Other participants were students with other areas of study and persons of various occupations. They were randomly assigned to the experimental conditions and tested individually at two different German universities.

### Stimulus Film

The stimulus film was taken from an earlier study by Franzen and Sporer (1994). The film showed the theft of an expensive pair of sunglasses in an optometrist's store. Altogether five amateur actors (one woman and four men) participated in the film, which lasted 6:30 min. The target person could be seen for 18 s. A close-up showing the target's head and shoulders in half-profile lasted for about 2 s. For the remaining time the target was filmed from a distance of several meters, where his head and whole torso could be seen. The actual theft took about 30 s (looking at the sunglasses on the rack, taking a pair and putting it in the pocket).

### Procedure

All instructions and the line-up presentation were programmed with SuperLab 1.75 ([www.cedrus.com](http://www.cedrus.com)). Participants were asked to watch the film closely and pay attention to every detail. After viewing the film, participants completed a 30-min filler task consisting of 40 general knowledge questions. In the following, participants were

instructed to either give a detailed spoken or written description of the theft they had witnessed earlier, imagining they were making a witness statement for a real police investigation. In the written condition, participants wrote down their accounts on paper. In the spoken condition, participants told their accounts to a co-investigator and were videotaped while doing so. Participants were first instructed to give a free report of the crime and were subsequently asked to answer eight open-ended questions concerning the course of events on another sheet (see Appendix A). The same procedure was followed for the description of the thief. First, participants were asked to describe the target with their own words so that “another person should be able to recognize him in a crowd.” Then, 12 open-ended questions concerning the target’s appearance followed (see Appendix B). The description instructions were in line with the standard description instructions as used in previous studies (e.g., Finger & Pezdek, 1999; Meissner & Brigham, 2001).

### Coding of Description Details

Two coders received training in which they individually coded all details named in 10 cases and discussed the discrepancies with each other and the trainer. Subsequently, all descriptions were coded for correct, incorrect or subjective details (cf. Finger & Pezdek, 1999). Subjective details (e.g. “looked tired”, “unfriendly face”) were excluded from further analyses. The few descriptors, on which coders disagreed even after conferring, were removed from the analyses.

#### *Perpetrator description*

Intercoder correlations were computed for clothing, body, facial and other details (e.g., posture, nationality) and separately for free reports and the open-ended questions. Due to the non-ambiguity of the coding system the mean intercoder correlations after Fisher’s *Z*-transformation of free descriptions and open-ended questions were equally high for body, facial, and other details ( $r$  values = 0.99). The mean intercoder correlation after Fisher’s *Z*-transformation for clothing details was  $r = 0.86$  for free reports and  $r = 0.99$  for open-ended questions. Estimates of age, height and weight of the target were excluded from analysis due to a lack of standard for coding “accuracy” (e.g., is a response of “20–25 years” to be considered accurate when a person is 25 years old?).

#### *Crime description*

Crime descriptions were coded as central action, central descriptive, peripheral action or peripheral descriptive details. Descriptive details included all details regarding objects, persons or the location. Any details offered with regard to the description of the perpetrator were accounted for in the perpetrator description and not coded here. Action details referred to any observable behaviors. Central details were defined as all details that could be pertinent for clarifying the incident and identifying the thief. Intercoder correlations after Fisher’s *Z*-transformation were high for central action (free report:  $r = 0.96$ , open-ended questions:  $r = 0.95$ ), central descriptive ( $r$  values = 0.90 vs. 0.93), and peripheral descriptive details ( $r$  values = 0.99 vs. 0.99). For peripheral action details the values were  $r = 0.80$  (free report) and  $r = 0.95$  (open-ended questions).

For data analysis, description scores combining free descriptions and open-ended questions were formed in addition to the categories listed above. When a participant named the same detail in the free report and the open-ended questions, it was only counted once. The subsequent analyses were carried out with the total number of correct and incorrect details. Additionally, description accuracy scores<sup>1</sup> were computed [description accuracy = correct details/(correct + incorrect details)].

## RESULTS

An alpha level of  $\alpha = 0.05$  was used for all inferential analyses. Cohen's (1988)  $d$  and partial  $\eta^2$  are reported as measures of effect size for ANOVAs (see Sporer & Cohn, 2011).

The number of descriptors that participants reported in free reports and open-ended questions as well as the respective proportions correct are presented in Tables 1 (crime description) and 2 (description of perpetrator). Note that the total number of details is not the sum of free report and open-ended questions, as descriptors named in both phases of the descriptions were only counted once.

### Crime Descriptions

To test the hypothesis that spoken descriptions contain more information, we computed two one-way ANOVAs with modality of description (written vs. spoken) as the independent variable and the total number of descriptors (description quantity) as the dependent variable. Spoken crime ( $M = 29.34$ ;  $SD = 5.71$ ) descriptions contained more details than written ones ( $M = 18.80$ ;  $SD = 4.00$ ) [ $F(1, 190) = 219.20$ ,  $p < 0.001$ ,  $d = 2.13$ ]. There was no difference in the accuracy (in %) of spoken ( $M = 82.29\%$ ;  $SD = 9.63$ ) vs. written descriptions ( $M = 81.11\%$ ;  $SD = 9.85$ ) [ $F(1, 190) = 0.70$ ,  $p = 0.402$ ,  $d = 0.12$ ].

In the next step we wanted to test whether this result pattern holds across all parts of the reports, or only for the most central parts of the descriptions. Therefore, we split the descriptors into central and peripheral ones and ran separate analyses for these descriptor categories. We found that spoken reports ( $M = 20.59$ ;  $SD = 3.37$ ) elicited more central information than written ones ( $M = 14.66$ ;  $SD = 2.56$ ) [ $F(1, 190) = 187.88$ ,  $p < 0.001$ ,  $d = 1.97$ ]. Furthermore, this information was also more accurate ( $M = 81.84\%$ ;  $SD = 8.49$  vs.  $78.92\%$ ;  $SD = 10.85$ ) [ $F(1, 190) = 4.30$ ,  $p = 0.039$ ,  $d = 0.30$ ]. We also found that spoken reports elicited more peripheral information ( $M = 8.75$ ;  $SD = 3.33$  vs.  $M = 4.14$ ;  $SD = 2.41$ ) [ $F(1, 190) = 120.66$ ,  $p < 0.001$ ,  $d = 1.58$ ], but there was no significant difference in description quality ( $M = 90.17\%$ ;  $SD = 10.20$  vs.  $87.41\%$ ;  $SD = 20.52$ ) [ $F(1, 190) = 1.39$ ,  $p = 0.239$ ,  $d = 0.17$ ].

Finally, we compared written and spoken accounts that were made in free reports vs. open-ended questions. Two mixed-model ANOVAs were computed with modality of description (written vs. spoken) as the fixed factor and quantity (quality) of the

<sup>1</sup> Using description accuracy as a predictor of identification accuracy has been criticized in previous research (Sporer, 1996; Wells, 1985) because the accuracy coefficient does not differentiate between more and less detailed descriptions. A description consisting of only one correct detail will obviously have a higher accuracy score (100%) than a description consisting of 10 details with two of them wrong (80%). However, the minimum of total person descriptors provided by participants in our experiment was eight ( $M = 17.28$ ;  $SD = 4.2$ ), while the minimum of total crime descriptors provided was nine ( $M = 27.29$ ;  $SD = 10.5$ ). This made analyses with description accuracy appropriate and justifiable.

Table 1. Means, 95% confidence intervals (CIs), and standard deviations of recall of total number of central and peripheral crime details and resulting accuracy for free report, open-ended questions, and their combination

	Free report			Open-ended questions			Free report and open-ended questions		
	Central	Peripheral	Total	Central	Peripheral	Total	Central	Peripheral	Total
	<b>Number of descriptors</b>								
Written ( <i>n</i> = 96)	<i>M</i>	7.91*	10.32*	9.24*	2.05*	11.29*	14.66*	4.14*	18.80*
	95% CI	7.34–8.47	9.49–11.15	8.94–9.54	1.85–2.25	10.86–11.71	14.14–15.18	3.65–4.63	17.99–19.61
	<i>SD</i>	2.78	4.10	1.49	0.98	2.08	2.56	2.41	4.00
Spoken ( <i>n</i> = 96)	<i>M</i>	12.43	17.51	14.65	6.00	20.65	20.59	8.75	29.34
	95% CI	11.56–13.31	16.08–18.93	14.21–15.08	5.54–6.46	19.92–21.37	19.91–21.27	19.91–21.27	8.07–9.43
	<i>SD</i>	4.33	7.05	2.15	2.25	3.59	3.37	3.33	5.71
Accuracy (proportion correct)	<i>d</i>	1.24	1.25	2.91	2.27	3.17	1.97	1.58	2.13
	<i>M</i>	0.91	0.98*	0.92	0.71*	0.82†	0.79†	0.87	0.81
	95% CI	0.89–0.93	0.96–0.98	0.91–0.94	0.68–0.75	0.77–0.88	0.77–0.81	0.83–0.92	0.79–0.82
Spoken	<i>SD</i>	0.09	0.07	0.08	0.17	0.28	0.11	0.21	0.10
	<i>n</i>	96	87	96	96	96	96	96	96
	<i>M</i>	0.90	0.94	0.91	0.79	0.89	0.82	0.90	0.82
Accuracy	95% CI	0.88–0.92	0.92–0.97	0.89–0.93	0.77–0.82	0.86–0.92	0.80–0.84	0.88–0.92	0.80–0.84
	<i>SD</i>	0.10	0.11	0.09	0.11	0.14	0.09	0.10	0.10
	<i>n</i>	96	93	96	96	96	96	96	96
Accuracy	<i>d</i>	-0.17	-0.38	-0.18	0.57	0.30	0.30	0.17	0.12

Note: Means in columns with a † differ significantly at  $p < 0.05$  ( $*p < 0.01$ ). Effect size *d* refers to the comparison between written and spoken reports in each column. Free report and open-ended questions do not add up exactly, as details named in both interview phases were counted as only one detail when summed up.



Table 2. Means, 95% confidence intervals, and standard deviations of total number of perpetrator descriptors and accuracy of free report, open-ended questions, and their combination

Number of descriptors	Free report				Open-ended questions				Free report and open-ended questions			
	Body	Clothes	Face	Total	Body	Clothes	Face	Total	Body	Clothes	Face	Total
	Written ( <i>n</i> = 96)	<i>M</i> 0.59 95% CI 0.47–0.72 <i>SD</i> 0.61	5.28* 4.68–5.88 2.97	4.52 4.11–4.94 2.05	10.49† 9.73–11.25 3.76	1.43 1.27–1.59 0.79	7.14* 6.61–7.67 2.61	5.49* 5.18–5.80 1.54	14.86 14.14–15.59 3.57	1.54 1.36–1.73 0.91	8.23* 7.63–8.83 2.95	7.01* 6.66–7.41 2.01
Spoken ( <i>n</i> = 96)	<i>M</i> 0.45 95% CI 0.33–0.56 <i>SD</i> 0.57	4.15 3.61–4.68 2.63	4.47 4.03–4.91 2.17	9.17 8.38–9.96 3.91	1.26 1.15–1.36 0.50	5.31 4.81–5.82 2.49	7.40 7.10–7.69 1.45	14.92 14.26–15.58 3.27	1.39 1.27–1.51 0.59	6.25 5.64–6.86 3.02	8.25 7.86–8.67 2.11	16.93 16.07–17.79 4.24
Accuracy (proportion correct)	<i>d</i> -0.25	-0.40	-0.02	-0.35	-0.26	-0.71	1.28	0.02	-0.20	-0.66	0.60	-0.17
Written	<i>M</i> 0.35 95% CI 0.22–0.48 <i>SD</i> 0.47	0.74 0.69–0.79 0.24	0.72 0.68–0.77 0.23	0.70 0.67–0.74 0.17	0.38† 0.30–0.47 0.42	0.80 0.77–0.83 0.15	0.64* 0.61–0.68 0.18	0.71† 0.69–0.73 0.11	0.39 0.30–0.47 0.40	0.78 0.75–0.81 0.15	0.67* 0.63–0.70 0.17	0.71 0.69–0.73 0.11
Spoken	<i>M</i> 0.34 95% CI 0.19–0.49 <i>SD</i> 0.45	0.73 0.67–0.79 0.30	0.68 0.64–0.73 0.24	0.69 0.65–0.73 0.19	0.51 0.42–0.59 0.41	0.75 0.70–0.80 0.23	0.75 0.73–0.77 0.12	0.74 0.72–0.77 0.11	0.48 0.40–0.56 0.39	0.74 0.69–0.78 0.23	0.73 0.71–0.76 0.12	0.73 0.71–0.75 0.10
	<i>n</i> 52	90	96	96	94	95	96	96	96	96	96	96
	<i>d</i> -0.02	-0.04	-0.17	-0.06	0.31	-0.26	0.72	0.31	0.23	-0.21	0.42	0.15

Note: Means in columns with a † differ significantly at  $p < 0.05$  ( $*p < 0.01$ ). Effect size *d* refers to the comparison between written and spoken reports in each column. Free report and open-ended questions do not add up exactly, as details named in both interview phases were counted as only one detail when summed up. Furthermore, body, clothes and face details do not add up to the numbers in the columns superscripted total due to a very small number of details that could not be assigned to one of the three categories.



descriptions in free reports vs. open-ended questions as the repeated measures factor. The results revealed an interaction between medium and phase for both description quantity [ $F(1, 190) = 7.28, p = 0.008, \text{partial } \eta^2 = 0.037$ ] and description accuracy [ $F(1, 190) = 29.04, p < 0.001, \text{partial } \eta^2 = 0.133$ ].

The analysis of simple main effects showed that spoken free reports ( $M = 17.51$ ;  $SD = 7.05$ ) elicited more information than written free reports ( $M = 10.32$ ;  $SD = 4.10$ ) [ $F(1, 190) = 74.44, p < 0.001, d = 1.25$ ]. However, there was no significant difference regarding the accuracy of spoken ( $M = 90.96\%$ ;  $SD = 8.75$ ) vs. written ( $M = 92.42\%$ ;  $SD = 7.68$ ) free reports [ $F(1, 190) = 1.52, p = 0.218, d = -0.18$ ]. For open-ended questions, spoken answers ( $M = 20.65$ ;  $SD = 3.59$ ) again led to higher description quantity than written answers ( $M = 11.29$ ;  $SD = 2.08$ ) [ $F(1, 190) = 487.48, p < 0.001, d = 3.17$ ]. Furthermore, spoken answers ( $M = 82.29\%$ ;  $SD = 9.63$ ) to open-ended questions were significantly more accurate than written ones ( $M = 73.23\%$ ;  $SD = 15.60$ ) [ $F(1, 190) = 23.46, p < 0.001, d = 0.70$ ].

### Perpetrator Descriptions

The analyses of the perpetrator descriptions were carried out analogously to the analyses regarding the descriptions of the crime. Unexpectedly, written and spoken perpetrator descriptions did not differ in overall quantity [ $F(1, 190) = 2.58, p = 0.248, d = -0.17$ ] or accuracy [ $F(1, 190) = 1.02, p = 0.315, d = 0.15$ ].

The perpetrator descriptions consisted primarily of information regarding the facial appearance (43%) as well as the clothing (44%). Few details referred to the body of the perpetrator (5%). For police investigations, facial descriptors are most critical for tracing a perpetrator, whereas clothing can easily be changed. Therefore, we tested whether the quantity and quality of face descriptions varied as a function of description modality. The results showed that spoken perpetrator descriptions contained more facial descriptors ( $M = 8.25$ ;  $SD = 2.11$ ) than written ones ( $M = 7.01$ ;  $SD = 2.01$ ) [ $F(1, 190) = 17.48, p < 0.001, d = 0.60$ ]. Furthermore, spoken face descriptions were also more accurate ( $M = 73.11\%$ ;  $SD = 12.05$ ) than written ones ( $M = 66.93\%$ ;  $SD = 17.00$ ) [ $F(1, 190) = 8.45, p = 0.004, d = 0.42$ ]. In contrast, written clothing descriptions ( $M = 8.22$ ;  $SD = 2.95$ ) contained more information than spoken ones ( $M = 6.25$ ;  $SD = 3.02$ ) [ $F(1, 190) = 21.12, p < 0.001, d = -0.66$ ], although there was no significant difference in description accuracy ( $M = 77.67\%$ ;  $SD = 15.16$  vs.  $73.64\%$ ;  $SD = 22.79$ , respectively) [ $F(1, 190) = 2.09, p = 0.150, d = -0.21$ ].

Thus, when looking at the most crucial information, spoken reports were qualitatively and quantitatively superior to written eyewitness accounts. When looking at less critical pieces of information, written accounts elicited more descriptors, while there was no significant difference in the level of accuracy between the two description modalities.

Again, we compared written and spoken accounts made in free reports vs. open-ended questions as reported above. The results revealed an interaction between testing modality and phase for both description quantity [ $F(1, 190) = 9.16, p = 0.003, \text{partial } \eta^2 = 0.046$ ] and description accuracy [ $F(1, 190) = 5.05, p = 0.026, \text{partial } \eta^2 = 0.026$ ]. Analyses of simple main effects showed that written free reports ( $M = 10.49$ ;  $SD = 3.76$ ) elicited more information than spoken free reports ( $M = 9.17$ ;  $SD = 3.91$ ) [ $F(1, 190) = 5.72, p = 0.018, d = -0.35$ ]. However, there was no significant difference in the quality of written ( $M = 70.28\%$ ;  $SD = 17.39$ ) vs. spoken ( $M = 69.17\%$ ;  $SD = 18.61$ )

free reports [ $F(1, 190) = 0.18, p = 0.669, d = -0.06$ ]. In answers given to open-ended questions, there was no difference in the quantity of written ( $M = 14.86; SD = 3.57$ ) vs. spoken answers ( $M = 14.92; SD = 3.27$ ) [ $F(1, 190) = 0.01, p = 0.908, d = 0.02$ ]. Spoken answers ( $M = 74.34\%; SD = 10.83$ ) to open-ended questions were, however, significantly more accurate than written ones ( $M = 70.96\%; SD = 11.19$ ) [ $F(1, 190) = 4.53, p = 0.035, d = 0.31$ ].

## DISCUSSION

The aim of the current study was to test whether the modality of testing (written vs. spoken) has an effect on the quantity and quality of eyewitness statements. While witnesses in real criminal cases mostly give an oral account of the event and perpetrator, a look at the available literature shows that the vast majority of studies use written reports (e.g., Heuer & Reisberg, 1990; Migueles & García-Bajos, 1999; Sporer, 1992b; Wells, 1985).

We hypothesized that witnesses' spoken reports should elicit more details than written ones. The hypothesis was supported for crime descriptions. Specifically, participants reported many more crime details when speaking than when they were writing ( $d = 2.13$ ). Also, when looking at central vs. peripheral details, spoken accounts were more detailed than written ones ( $d = 1.97$  vs.  $1.58$ ). To be truly beneficial, however, the modality of recall should lead to a gain not only in description *quantity*, but also in description *quality*. Spoken central crime descriptors were significantly more accurate than written ones ( $d = 0.30$ ), while there was no difference for peripheral crime details. Thus, as expected, spoken reports led to significantly more crime details and, for the most critical information, also to a significant increase in accuracy.

For perpetrator descriptions, our hypothesis found only partial support. Whereas no difference in description quantity occurred between written vs. spoken total descriptions, participants in the spoken condition reported significantly more central perpetrator (i.e., face) details than participants in the written condition ( $d = 0.60$ ). In contrast, peripheral perpetrator (i.e., clothing) details were reported more frequently in the written condition than in the spoken condition ( $d = -0.66$ ). Paralleling the results for crime descriptions, central perpetrator details were more accurate in spoken descriptions than in written ones ( $d = 0.42$ ), while there was no difference for peripheral perpetrator details ( $d = -0.21$ ).

All in all, the results suggest that obtaining spoken rather than written eyewitness accounts can be beneficial. The effect sizes obtained are generally substantial and may direct us to an important new avenue of research and application. The most critical information about the crime and the perpetrator was both more extensive and more accurate when witnesses gave spoken descriptions. This could have important implications for the investigation of crimes and obtaining information from eyewitnesses. It is possible, of course, that modality of testing interacts with other factors in such a way that some persons may profit more from spoken recall, and others may profit more from written recall. Such factors could include age or educational background. Future studies should address these issues. Another limitation concerning the applicability of the current results is that participants in the written and spoken conditions were asked the same questions. In real cases, however, police officers may ask additional questions which may lead to an increase in both correct and incorrect details.

For the remainder of this discussion we speculate about possible mechanisms that could underlie the spoken report advantage that we observed in the current study

and, furthermore, cause its particularly beneficial effects for central details. It is conceivable that both cognitive (Kellogg, 2007) and social (Bergmann et al., 2004) factors are responsible for the effect. Specifically, writing places higher demands on working memory than speaking because writing is slower, less practiced, and entails the activation of graphemic representations for spelling words (Kellogg, 2007). Indeed, cognitive factors may account for the differential results regarding clothing vs. facial perpetrator descriptions. It is known that person descriptions are generally meager (Sporer, 1996) and that we suffer from lack of vocabulary when trying to describe even a familiar person or someone standing right in front of us. Clothing, in contrast, is much easier to describe.

When cognitive load is high, performance for effortful retrieval from long-term memory is reduced (Moscovitch, 1994). Consequently, memory retrieval should benefit from a reduction in cognitive load (cf. Vredeveldt, Hitch, & Baddeley, 2011). When engaged in a cognitively more demanding task (written description), witnesses may thus choose to describe primarily those details that require less effort (clothing) rather than those that require more effort (face). If, however, cognitive load is lower (spoken descriptions), witnesses may have more working memory capacity available to describe those details that require more effort.<sup>2</sup>

On the other hand, social influences introduced by the presence of an interviewer cannot be ruled out (Bergmann et al., 2004; Rosenthal, 2002). Specifically, the differential effect of recall modality on central vs. peripheral perpetrator details could also suggest that social factors play an important role in the effect. Specifically, interviewers in the spoken conditions may have subtly signaled to the participants (although instructed not to) to speak more about central details than about peripheral ones. It is known that such influences occur (Rosenthal, 2002) and that people are not aware of exerting them (Greathouse & Kovera, 2009).

The degree to which the effect is due to social factors could be tested by interviewing some participants in person (as done in the present study) and by having another group of participants make their spoken statements with the aid of a voice recorder, that is, with no interviewer present. If the spoken advantage is primarily due to social factors, their influence should be diminished in the voice-recorder-only condition. The role of cognitive factors, on the other hand, could be tested by increasing the cognitive load during spoken reports. For example, participants could be instructed to memorize a word list before making their statement which they would have to recall afterwards. If cognitive factors are the predominant determinant producing the spoken advantage, this effect should decrease in spoken conditions that involve additional cognitive demands. Our guess is that the effect is due to a mixture of both, social and cognitive factors. This is for future research to determine.

## REFERENCES

- Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, England: Cambridge University Press.
- Bekerian, D. A., & Dennett, J. L. (1990). Spoken and written recall of visual narratives. *Applied Cognitive Psychology*, 4, 175–187.

---

<sup>2</sup>We would like to thank Alana Krix for offering this explanation.

- 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.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Finger, K., & Pezdek, K. (1999). The effect of the cognitive interview on face recognition accuracy: Release from verbal overshadowing. *Journal of Applied Psychology*, *84*, 340–348.
- Franzen, S., & Sporer, S. L. (1994). Personenverwechslungen durch irreführende Rekonstruktionsbilder: Zum Einfluss nachträglicher Informationen und der Wiederherstellung des Wahrnehmungskontextes. [Person mixups as a function of misleading composites: On the influence of postevent information and context reinstatement]. In Sporer S. L., & Meurer D. (Eds.), *Die Beeinflussbarkeit von Zeugenaussagen* (pp. 207–236). Marburg: N.G. Elwert.
- Greathouse, S. M., & Kovera, M. B. (2009). Instruction bias and lineup presentation moderate the effects of administrator knowledge on eyewitness identification. *Law and Human Behavior*, *33*, 70–82.
- Heuer, F., & Reisberg, D. (1990). Vivid memories of emotional events: The accuracy of remembered minutiae. *Memory & Cognition*, *18*, 496–506.
- Horowitz, M. W., & Newnam, J. B. (1964). Spoken and written expression: An experimental analysis. *Journal of Abnormal and Social Psychology*, *6*, 640–647.
- Ibabe, I., & Sporer, S. L. (2004). What you ask is what you get: On the influence of question form on accuracy and confidence. *Applied Cognitive Psychology*, *18*, 711–726.
- Kellogg, R. T. (2007). Are written and spoken recall of text equivalent? *The American Journal of Psychology*, *120*, 415–428.
- Lipton, J. P. (1977). On the psychology of eyewitness testimony. *Journal of Applied Psychology*, *62*, 90–95.
- Marshall, J. (1966). *Law and psychology in conflict*. Indianapolis: Bobbs-Merrill.
- Meissner, C. A., & Brigham, J. C. (2001). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology*, *15*, 603–616.
- Miguelles, M., & Garcia-Bajos, E. (1999). Recall, recognition, and confidence patterns in eyewitness testimony. *Applied Cognitive Psychology*, *13*, 257–268.
- Moscovitch, M. (1994). Cognitive resources and dual-task interference effects at retrieval in normal people: The role of the frontal lobes and medial temporal cortex. *Neuropsychology*, *8*, 524–534.
- Pozzulo, J. D., Dempsey, J., & Crescini, C. (2009). Preschoolers' person description and identification accuracy: A comparison of the simultaneous and elimination lineup procedures. *Journal of Applied Developmental Psychology*, *30*, 667–676.
- Rosenthal, R. (2002). Covert communication in classrooms, clinics, courtrooms, and cubicles. *American Psychologist*, *57*, 839–849.
- Sporer, S. L. (1992a, March). Person descriptions in an archival analysis of criminal cases. Paper presented at the Biennial Meeting of the American Psychology-Law Society in San Diego, California.
- Sporer, S. L. (1992b). Postdicting eyewitness identification accuracy: Confidence, decision-times and person descriptions among choosers and non-choosers. *European Journal of Social Psychology*, *22*, 157–180.
- Sporer, S. L. (1996). Psychological aspects of person descriptions. In Sporer S. L., Malpass R. S., & Koehnken G. (Eds.), *Psychological issues in eyewitness identification* (pp. 53–86). New Jersey: Lawrence Erlbaum Associates.
- Sporer, S. L., & Cohn, L. D. (2011). Meta-analysis. In Rosenfeld B. D., & Penrod S. D. (Eds.), *Research methods in forensic psychology* (pp. 43–62). New York: Wiley.
- Vredevelde, A., Hitch, G. J., & Baddeley, A. D. (2011). Eye-closure helps memory by reducing cognitive load and enhancing visualization. *Memory and Cognition*. doi: 10.3758/s13421-011-0098-8
- Wells, G. (1985). Verbal descriptions of faces from memory: Are they diagnostic of identification accuracy? *Journal of Applied Psychology*, *70*, 619–626.

## APPENDIX A: OPEN-ENDED QUESTIONS REGARDING THE THEFT

1. What did the salesperson do, so that the culprit could take the eyeglasses without being noticed?
2. Where did the culprit put the eyeglasses?
3. How many persons were in the optometrist's store during the whole film?
4. Who was in the store at the same time as the culprit?
5. What did the other customers do while the culprit took the eyeglasses?

6. What did the culprit say to the salesperson?
7. For how long did the culprit stay in the store?
8. How long did the film take altogether ( \_\_min \_\_s)?

## **APPENDIX B: OPEN-ENDED QUESTIONS REGARDING THE THIEF**

1. How old do you think the culprit was?
2. How tall was the culprit (in cm)?
3. Describe the culprit's build!
4. Describe the culprit's clothing!
5. Describe the culprit's hair color!
6. Describe the culprit's hairdo!
7. Describe the culprit's face shape!
8. Which of the culprit's special features caught your eye?
9. Did the culprit wear headgear? If yes, what kind?
10. Did the culprit wear glasses? If yes, what did they look like?
11. Did the culprit have a beard? If yes, what did it look like?
12. Did the culprit speak in a dialect or did he have an accent? If yes, which?