

Binghamton University

The Open Repository @ Binghamton (The ORB)

Bystander Reporting: Effects of Early Testing
and Accuracy Assessment for Critical Content:
Supplemental materials DOI: 10.22191/orb/
rrmiller/lab/1

Ralph Miller Laboratory

2019

EW Wasserman et al Supplementary Material Bystander Reporting (Supplementary material for Wasserman et al.)

Ralph Miller
rmiller@binghamton.edu

Follow this and additional works at: https://orb.binghamton.edu/bystander_reporting_suppmaterials

Recommended Citation

Miller, Ralph, "EW Wasserman et al Supplementary Material Bystander Reporting (Supplementary material for Wasserman et al.)" (2019). *Bystander Reporting: Effects of Early Testing and Accuracy Assessment for Critical Content: Supplemental materials DOI: 10.22191/orb/rrmiller/lab/1*. 5.
https://orb.binghamton.edu/bystander_reporting_suppmaterials/5

This Article is brought to you for free and open access by the Ralph Miller Laboratory at The Open Repository @ Binghamton (The ORB). It has been accepted for inclusion in Bystander Reporting: Effects of Early Testing and Accuracy Assessment for Critical Content: Supplemental materials DOI: 10.22191/orb/rrmiller/lab/1 by an authorized administrator of The Open Repository @ Binghamton (The ORB). For more information, please contact ORB@binghamton.edu.

1 Supplementary Material for:
2
3 Bystander Reporting: Effects of Early Testing and Accuracy Assessment for Critical Content
4

5
6 This file contains:

7
8 Supplementary Method
9

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

Supplementary Method

47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92

Participants

We recruited 114 SUNY-Binghamton undergraduates to participate in this experiment ($M = 19$ years of age, $SD = 1.01$). Sixty-nine participants were female and 45 participants were male. Participants were randomly assigned to either an ‘Immediate’ condition ($n = 54$), or a ‘Delay’ condition ($n = 60$). The difference in group sizes reflects the random procedure used in assigning participants to the two groups. We initially collected data from 189 participants, but the data of any participant who failed to fill out the Scantron correctly (i.e., failing to answer a question), or of any participant who failed to return after the 48-hour delay were eliminated from the analysis (75 participants eliminated). Therefore, the data for a final number of 114 participants were analyzed for the present experiment. The total number of participants in each group was based on sample sizes of 45-48 being appropriate for detecting differences between two groups based on an effect size of small to moderate, Cohen's $d = 0.30$ (Cohen, 1988). Participants were given partial credit of a course requirement for taking part in each half of the experiment. The protocol for this study was approved by the SUNY-Binghamton Institutional Review Board and all participants gave prior written informed consent in accordance with the Declaration of Helsinki.

Materials and Design

Participants initially signed up for both parts of the experiment, such that they completed the first part of the experiment on the first day, and returned for the second part 48 hours. All participants watched a video of a purse theft. The Immediate Group was tested immediately after watching the video, and returned 48 hours later for a second test. The Delay Group was tested only after a 48-hour delay. We use the term ‘Day One’ to refer to the first part of the experiment, during which all participants watched the video, but only the Immediate Group took a test. We use the term ‘Day Three’ to refer to the second part of the experiment that took place after a 48-hour delay, upon which the Immediate group took their second test, and the Delay Group took the test for the first time.

Stimulus video. The experiment started with on-screen instructions prompting participants to attend to a short video on the computer monitor. The video began with two patrons sitting and a cashier standing towards the back of a café. There was a white board towards the center right side of the screen that listed prices of goods. A few seconds into the video, a woman and male friend walk into the café and sit down at a table in the foreground of the scene. After sitting for a few moments, the couple walks towards the back of the café to place an order, leaving their belongings at the table. About 35 seconds into the video, a man who was sitting towards the front right corner of the café walks through the foreground of the scene, snatches the woman’s unattended purse, and leaves the café. A few moments later, the victim and friend notice the purse is gone and begin to search for it near their table. Some seconds later, they too exit the café, seemingly in search of the purse. The video was in color, was silent, lasted 1 minute, 6 seconds, and took up 85% of the height of the screen and 85% of the width of the screen. Participants sat with their eyes approximately 0.45 m from the center of the screen. The screen’s dimensions were 54 x 30 cm and the resolution was 2250 x 2450 pixels. Instructions appeared in black, Courier font style, text size 40, on a gray background. Next to the computer were a #2 pencil and a folder containing an answer sheet (Scantron) and a packet of instructions. The packet informed participants about how to anonymously fill out the Scantron in order for the experimenters to later match the Scantrons from Day One to the Scantrons from Day Three, so that the participants’ responses in each session could be linked. We used a forced-choice

93 procedure for the memory test, as it reflected the direct, short-answer often actually employed in
94 interviews (Clifford & George, 1996; Fisher et al., 1987; Ginet & Py, 2001). Additionally, using
95 a forced-choice procedure facilitated quantitative comparison of recall across different contextual
96 aspects of the target event. We assessed participants' recognition accuracy of information
97 concerning the perpetrator and aspects of contextual memory, including where and when the
98 event took place, and who else was present at the event.

99 **Test questions.** The test consisted of 34 forced-choice questions pertaining to the video,
100 with an additional two questions at the end asking the participants whether they had watched the
101 video today and whether they had previously answered questions about the video. The 34
102 questions focused on four main categories: 'Perpetrator,' 'When' (temporal information),
103 'Where' (spatial information), and 'Who' (who else was at the crime scene). 'Perpetrator'
104 questions emphasized the perpetrator's physical appearance and clothing. 'Who' questions asked
105 about features of the other people at the event, such as the victim's hair and shirt color, and what
106 the other patrons were doing at the event (i.e., drinking coffee or texting). 'When' questions
107 asked about the temporal information, such as the duration of actions, order of actions, and the
108 time of year in which the event occurred, based on the date appearing on a whiteboard menu,
109 trees outside a window, and the clothing of the people in the video. 'Where' questions alluded to
110 the relative spatial locations of objects and people, such as where the cash register was located,
111 the seating orientation of the victim relative to the victim's friend, and features of objects, such
112 as their color. For all but the last two questions, the answer choices consisted of one correct
113 answer, two plausible foils (incorrect answer choices), a "None of the above" choice, and an "I
114 do not know" choice. Both foil options, "I do not know" and "None of the above" were all coded
115 as incorrect answers. The last two questions could only be answered with "yes" or "no", and
116 their purpose was simply to provide evidence that we had correctly matched the Scantrons from
117 each participant. See Appendix for representative test questions.

118 **Pilot study.** Prior to the present experiment, we had run a pilot study designed to identify
119 a set of questions from each content area ('Perpetrator', 'When' [temporal information], 'Where'
120 [spatial information] and 'Who' [who else was at the event]) that on average would be equally
121 apt to be answered correctly and that demonstrated a relatively high degree of inter-item
122 reliability within each of the designated content areas. We tested 72 participants, half of whom
123 were assigned to an Immediate condition (tested immediately after watching the video of a purse
124 being stolen and the same participants tested again 48 hours later) and half of whom were
125 assigned to a Delay condition tested for the first time 48 hours after watching the video. The set
126 of questions consisted of 49 target questions from four main content areas ('Perpetrator',
127 'When', 'Where' and 'Who'). There was only one order of questions, and one question from
128 each of the four content areas appeared on a page at least until questions concerning a given
129 content were exhausted. Informed by analysis of the data from only participants who were tested
130 immediately after watching the purse-snatching video, we eliminated potential questions for the
131 experiment reported here in which 10% or fewer participants answered the question correctly, or
132 90% or more of participants answered the question correctly. We additionally removed questions
133 that had very low point bi-serial correlations (i.e., less than -0.10) with other questions within a
134 content area, such that performance on one question in a content area did not accurately predict
135 performance on other questions in that content area. We then matched the three content areas for
136 the quantity of items by removing questions with the lowest point bi-serial correlations, which
137 we recalculated having removed items using the previous criteria. In total, we removed three

138 items from the ‘Perpetrator’ category and four items from each of the ‘Who’, ‘What’ and ‘When’
139 categories.

140 **Order of test questions and participant instructions.** The selected 34 questions for the
141 present experiment included nine questions in each context category, which allowed us to
142 maintain comparable sensitivity. The ‘Perpetrator’ category contained only seven questions due
143 to the lack of further testable content concerned solely with the perpetrator. The first page of the
144 question packet provided instructions on how to fill out the Scantron in order for us to pair
145 Scantrons from the two sessions representing the same participant. For the present experiment,
146 on Day One, the following pages of the packet contained questions regarding the video for only
147 the Immediate Group. Specifically, each page of the question packet contained one question
148 from each content area, except for two pages that did not include a ‘Perpetrator’ question due to
149 there being fewer perpetrator questions. The order of the four types of questions from each
150 content area on each page was randomized. No questions from the same content area appeared in
151 immediate succession (e.g., if the last question on a page was a ‘Perpetrator’ question, then the
152 first question on the next page was from any content area except ‘Perpetrator’). On Day One,
153 after the first page, the Delay Group was presented with a page informing them that this part of
154 the experiment was complete. We had six different versions of the test that contained different
155 pseudorandomized orders of questions. The Immediate group was given the same questions, in a
156 different order, on their ‘Day One’ and ‘Day Three’ tests. On Day Three, the first page of the
157 packet was identical to the one that both groups had seen 48 hours earlier. The subsequent pages
158 contained the 36 questions regarding the video. The last page of the packet thanked participants
159 for taking part in the experiment, informed them that we were studying memory of different
160 types of information, and asked them not to discuss the experiment with anyone else. Participants
161 were randomly assigned to cubicles constrained by counterbalancing between groups, and were
162 seated in the same cubicle for both parts of the experiment.

163 **Procedure**

164 All participants completed the task in individual cubicles devoted to computer-based
165 psychology experiments. Upon arriving, the experimenter reminded participants that this was the
166 first part of the study, and that they were to return in two days. On Day One, participants were
167 asked to read and sign the Informed Consent form. Then, they were told to follow the
168 instructions on the computer screen, and when prompted to do so, should follow the directions in
169 the packet in a folder next to the computer. In addition, participants were asked not to use cell
170 phones during the experiment, nor to discuss the experiment with anyone else during or after the
171 experiment.

172 All participants viewed the following instructions upon sitting down at their computers:
173 “Thank you for participating in our study. The experiment depends on your participation both
174 today and two days from now. You will be shown a video shortly. Pay close attention to the
175 video. Press [SPACEBAR] to start the video.” After watching the video, participants were asked
176 to turn to the folder next to their computers. Participants in the Immediate Group received
177 printed instructions that they would be taking a test, and would need to answer all questions. At
178 the end of the test booklet, these participants were informed that this part of the experiment was
179 over, and that they should return in exactly 48 hours. The Delay Group participants were
180 informed that this part of the experiment was over, and that they should return in exactly 48
181 hours. Upon returning 48 hours later, all participants saw the following instructions: “Thank you
182 for returning today. Now please open the folder next to the computer and carefully follow the
183 directions provided.” All participants were then informed that they would be taking a test on the

184 video they had previously viewed, and that the experiment was complete when they finished the
185 test.

186 **Statistical Analysis**

187 Participant accuracy was determined by calculating the mean number of questions correct
188 for each content category. First, a 2 x 4 mixed-design analysis of variances (ANOVA) was
189 performed to assess forgetting over a 48-hour delay with Immediate Day One and Delay Day
190 Three as a between-subjects factor, and content area as within-subjects factors. This was
191 followed by planned contrasts to examine the change in accuracy across the delay for each
192 content area. Second, a 2 x 4 mixed-design ANOVA with Immediate Day Three and Delay Day
193 Three as a between-subjects factor, and content area as within-subjects factors was used to
194 examine the effect of an initial immediate test on test performance 48 hours later. Subsequent
195 planned contrasts were conducted to examine the effect of immediate testing on later accuracy
196 across the different content areas. Third, a 2 x 4 fully within-subject ANOVA was conducted to
197 compare the first test performance of Group Immediate with the second test performance of
198 Group Immediate to assess differences in the effects of early testing on later testing. Fourth,
199 exploratory Pearson correlations were calculated to determine whether for Group Immediate
200 there was a relationship between performances on Day One and Day Three as a function of the
201 category information. Fifth, a composite score combining contextual 'Who', 'Where' and
202 'When' information was created to assess the overall relationship of all contextual information
203 relative to 'Perpetrator' information by performing Pearson correlations. Additionally, Pearson
204 correlations were used to examine whether memory of one context area was correlated with
205 memory of the other context areas. To see how many Group Immediate participants changed
206 their responses from Day One to Day Three, and in which direction the changes in responses
207 occurred, we calculated Group Immediate's Day One percentages correct for the 'Perpetrator',
208 'Where', 'When' and 'Who' questions. We then calculated the shares of the above four
209 percentages that changed to any Incorrect answer on Day Three for each category. Additionally,
210 we calculated, for each category, the percentage of answers that changed from a Correct answer
211 to each Incorrect response. We also calculated, for each category, what percentage of answers
212 changed from any Incorrect answer on Day One to a Correct answer on Day Three. We
213 additionally examined how many participants in each group (Group Delay and Group Immediate
214 on both Day One and Day Three) responded "I do not know". We calculated the number of
215 Incorrect answers each group provided, and then calculated the percentage of participants who
216 responded "I do not know" out of the total number of Incorrect answers for each group. Lastly,
217 we calculated the number of Incorrect answers per group omitting the "I do not know" response
218 as an Incorrect answer. Results were considered significant when $p < .05$.