Results

Exclusion criteria

Prior to analysing the data, we have excluded participants based on a priori set criteria. Participants who have spent less than or equal to 90 seconds on the practice text were excluded (1 exclusion). Further, we wanted to exclude participants who have had no correct answers on the final test (0 exclusions). Finally, we have excluded participants who have stated that they have reading deficits (3 exclusions). This left us with a total sample of 203 participants. The descriptives for the sample are shown in Table 1. There is another set of exclusion criteria based on the number of times the participants have read each of the three texts. These are used in robustness check analyses (see suplementary materials).

Interpolated activity effect

Our first two hypotheses are concerned with the effects of different interpolated activities on the total number of correct answers and total number of intrusive distractors chosen. To test these hypotheses, we have focused only on the groups which have not received feedback, since there was no feedback option for the rereading group ($n = 122$). We conducted a one-way MANOVA with interpolated activity as the independent variable and the total number of correct and intrusive options chosen as dependent variables. The correlation between our DVs calculated on the whole sample is -0.71 (95% CI: [-0.77, -0.63], $p = 4.793 \times 10^{-32}$). Boxplots for the groups in this analysis are shown in Figure 1.

Pillai's V for the analysis is 0.126, $p = 0.004$ (Wilks' $\Lambda = 0.875$, $p = 0.003$; Hotelling-Lawley's trace = 0.142, $p = 0.003$; Roy's largest root = 0.137, $p = 4.912 \times 10^{-4}$. The effect size, calculated as $\omega^2_{mult} = 0.109$ (bootstrap$_{\mathrm{R}\,=\,10000}$ median = 0.132, BC_α 95% CI = [0.012, 0.201]). To further inspect the relationship of the interpolated activities with our dependent variables, we have conducted a Roy-Bargmann stepdown analysis, as suggested by Tabachnick and Fidell (2012; a linear discriminant analysis with the same aim is available in the supplementary materials). The total number of correct answers was a priori chosen to be the higher priority variable. Therefore, we first conducted an ANOVA with interpolated activity type as the indepedent variable and the total number of correct answers as the dependent variable.

As could be expected, the ANOVA points to an interpolated activity effect, with $F(2, 119) = 7.541$, $p = 8.254 \times 10^{-4}$. Following the ANOVA, we conducted an ANCOVA, with the total number of correct answers as the covariate, and the total number of intrusors as the dependent variable. As could be expected, the results imply a main effect of the total number of correct answers ($F(1, 118) = 79.674$, $p = 6.873 \times 10^{-15}$). After correcting for the number of correct answers, there is no evidence for an effect of interpolated activity on the total number of chosen intrusors ($F(2, 118) = 0.844$, $p = 0.433$. For now, we may claim that we do not have any evidence to support our second hypothesis that the type of interpolated activity will have an effect on the number of intrusors.

In order to test our first hypothesis, we have contrasted (i) the rereading group with the two test groups, and (ii) the two test groups with each other, taking only the total number of correct answers as the DV. The first contrast finds no evidence of a difference between the rereading group and the two test groups ($t = 1.355$, $p = 0.178$, $g_s = 0.19$, 95% CI = [-0.19, 0.57], Cohens's $U_{3,g_s} = 57.6\%$, probability of superiority = 55.39%). However, there is a difference between the two test groups ($t = 3.62$, $p = 4.34 \times 10^{-4}$, $g_s = 0.66$, 95% CI = [0.21, 1.1], Cohens's $U_{3,g_s} = 74.43\%$, probability of superiority = 67.88%). These two findings are not in line with our predictions.

Table 1: Descriptive statistics for the number of correct answers and chosen intrusors broken down by experimental condition.

Measure	Condition	n	M	SE	SD	min	max	skew	kurtosis
Total correct	Content, feedback	41	13.22	0.508	3.25	2	19	-0.800	1.503
	Content, no feedback	42	12.79	0.465	3.02	7	19	0.039	-0.775
	General, feedback	40	10.97	0.533	3.37	1	17	-0.481	0.462
	General, no feedback	40	10.47	0.449	2.84	5	16	-0.053	-0.986
	Rereading	40	10.88	0.443	2.80	4	17	-0.141	-0.253
Total intrusors	Content, feedback	41	3.15	0.258	1.65	0	7	0.292	-0.351
	Content, no feedback	42	3.38	0.257	1.67	0	7	0.203	-0.385
	General, feedback	40	4.17	0.318	2.01	0	8	0.024	-1.124
	General, no feedback	40	4.58	0.288	1.82	1	9	0.328	-0.484
	Rereading	40	4.62	0.350	2.21	1	10	0.272	-0.537

The interaction between feedback and interpolated activity type

The remaining hypotheses deal with the effect of feedback on the total number of correct answers and the total number of intrusors. Therefore, these analyses are carried out only on the data from participants in the general and content related test conditions (n = 163). Boxplots for these groups are shown in Figure 2. To test these hypotheses, we first conducted a two-way MANOVA with interpolated activity and feedback as independent variables, and total number of correct answers and total number of intrusors as the dependent variables.

Pillai's V for the interpolated activity effect (calculated with type III sums of squares) is 0.071, p = 0.003 (Wilks' Λ = 0.929, p = 0.003; Hotelling-Lawley's trace = 0.08, p = 0.003; Roy's largest root = 0.08, p = 0.003) confirming the main effect of interpolated activity type. The effect size ω^2_{mult} = 0.065 (bootstrap$_{\text{R = 10000}}$ median = 0.072, BC_α 95% CI = [0.007, 0.139]).

On the other hand, we find no evidence for an effect of giving feedback on the linear combination of our two dependent variables — Pillai's V = 0.003, p = 0.8 (Wilks' Λ = 0.997, p = 0.8; Hotelling-Lawley's trace = 0, p = 0.8; Roy's largest root = 0, p = 0.8). The effect size is ω^2_{mult} = -0.003 (bootstrap$_{\text{R = 10000}}$ median = 0.003[1]).

Furthermore, we find no evidence for an interaction effect between activity type and feedback — Pillai's V = 0.001, p = 0.941 (Wilks' Λ = 0.999, p = 0.941; Hotelling-Lawley's trace = 0, p = 0.941; Roy's largest root = 0, p = 0.941). The effect size ω^2_{mult} = -0.005 (bootstrap$_{\text{R = 10000}}$ median = 0.003[2]). Both the feedback and the interaction estimates of ω^2_{mult} are to be considered to be zero, given their negative values.

Again, we have conducted a follow-up Roy-Bargmann stepdown analysis. In the ANOVA model with the total number of correct answers as the dependent variable and the type of interpolated activity, feedback and their interaction as predictors, only the type of activity seems to be relevant ($F(1, 159) = 11.2, p = 0.001$). In the second step, we fit an ANCOVA model with the total number of correct answers as the covariate. In this model, the type of interpolated activity ceases to be a relevant predictor ($F(1, 155) = 0.175, p = 0.676$).

[1] The BC_α 95% CI for this estimate is $[-0.006, 0.004]$. Our guess is that this odd result is due to the fact that most of the density is concentrated around 0, causing an unreliable estimate.

[2] Same as footnote 1. BC_α 95% CI = $[-0.006, -0.005]$.

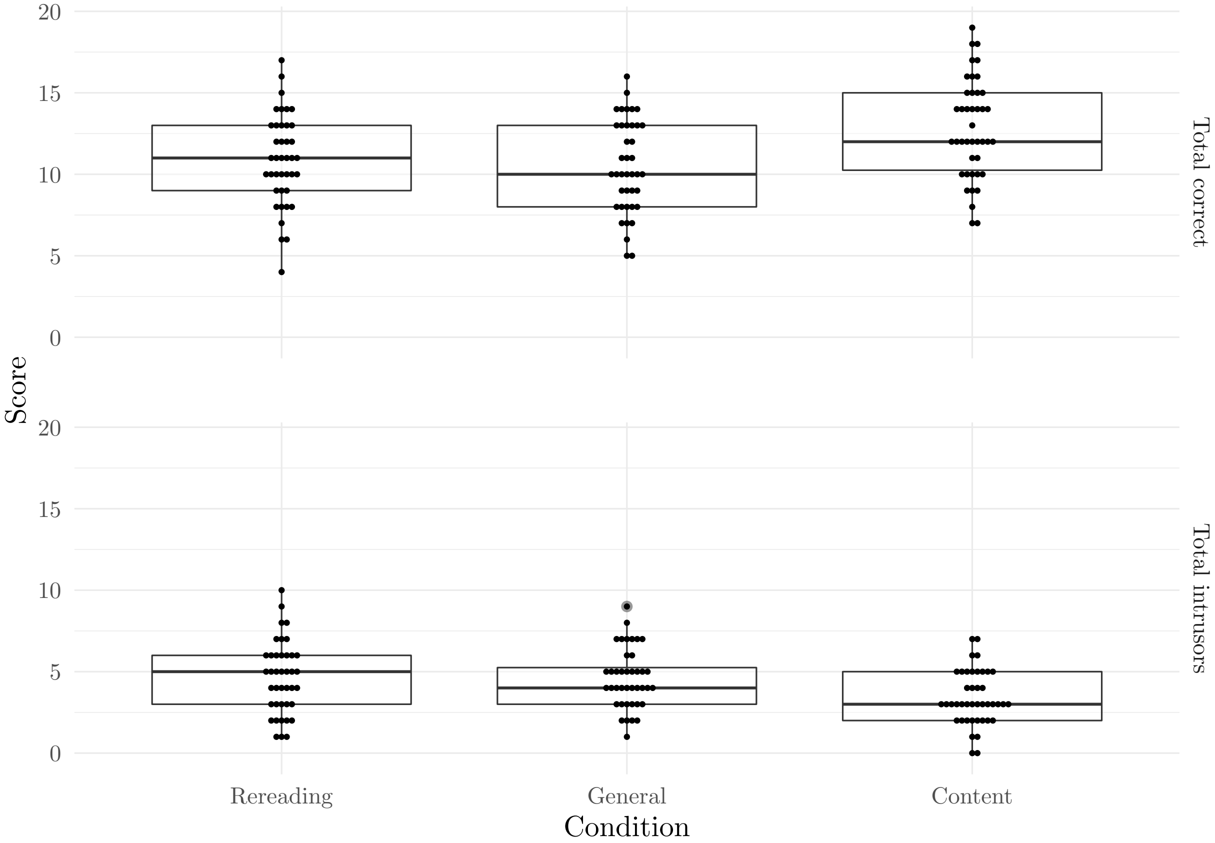

Figure 1: Boxplots broken down by experimental conditions included in the first MANOVA, and dependent variable, with overlayed raw scores.

Notes

Plots created using `ggplot2` (Wickham, 2016). Bootstrap conducted using the `boot` package (Canty & Ripley, 2017). Methods and analyses written using `rmarkdown` (Allaire et al., 2019) and `knitr` (Xie, 2019). The package `car` (Fox & Weisberg, 2011) was used to obtain type III sums of squares. `compute.es` (Re, 2013) was used to obtain effect sizes for contrasts. `kableExtra` was used to help generate tables (Zhu, 2019). Other utilities used are `tidyverse` (Wickham, 2017), `magrittr` (Bache & Wickham, 2014), `here` (Müller, 2017), `conflicted` (Wickham, 2018), `psych` (Revelle, 2018).

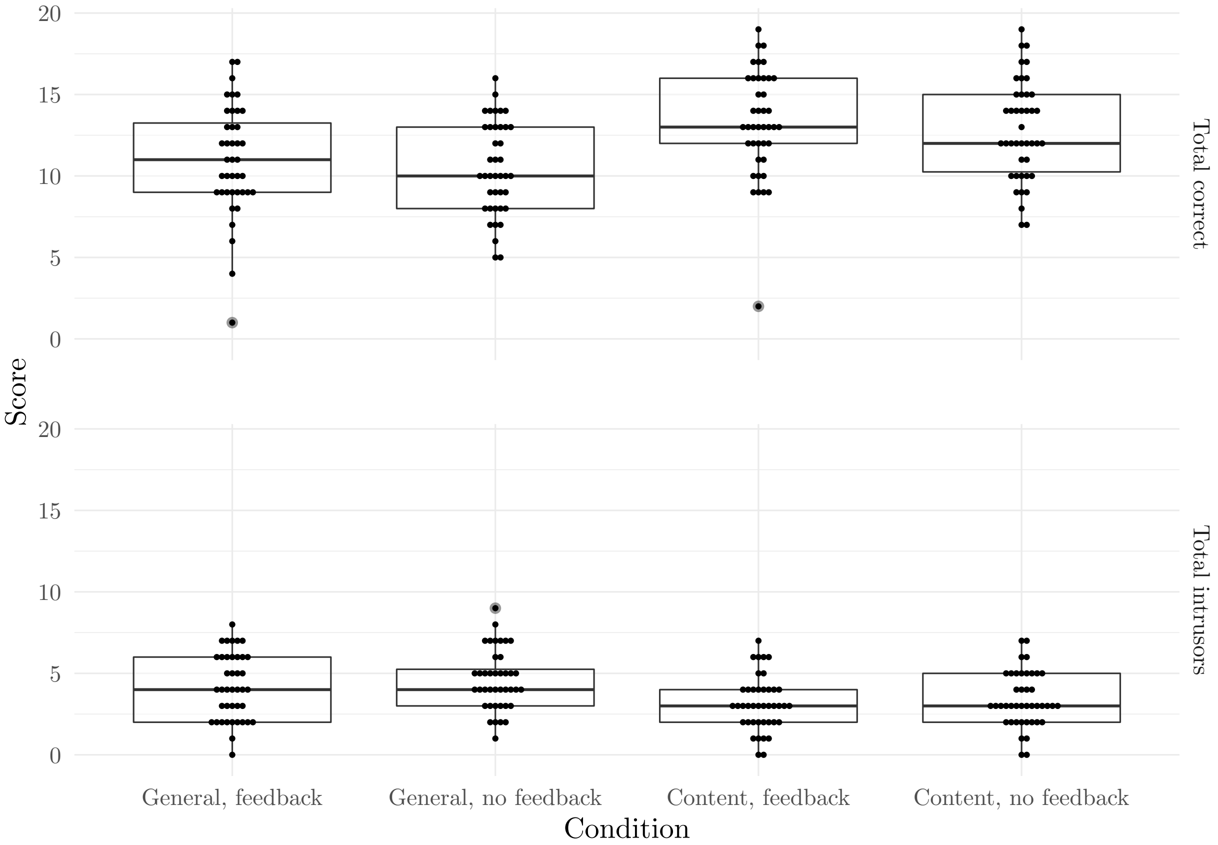

Figure 2: Boxplots broken down by experimental conditions included in the second MANOVA, and dependent variable, with overlayed raw scores.

Table 2: Full ANOVA and ANCOVA models for the second Roy-Bargmann stepdown analysis.

Term	*SS*	*df*	*F*	*p*
ANOVA				
Activity	109.393	1	11.200	0.001
Feedback	3.904	1	0.400	0.528
Activity x Feedback	0.045	1	0.005	0.946
Residuals	1553.046	159		
ANCOVA				
Activity	0.301	1	0.175	0.676
Feedback	0.173	1	0.100	0.752
Total correct	63.216	1	36.760	0.000
Activity x Feedback	0.813	1	0.473	0.493
Activity x Total correct	0.862	1	0.501	0.480
Feedback x Total correct	0.130	1	0.075	0.784
Activity x Feedback x Total correct	1.229	1	0.715	0.399
Residuals	266.551	155		

References

Allaire, J. J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., … Iannone, R. (2019). *rmarkdown: Dynamic Documents for R*. Retrieved from https://rmarkdown.rstudio.com

Bache, S. M., & Wickham, H. (2014). *magrittr: A Forward-Pipe Operator for R*. Retrieved from https://CRAN.R-project.org/package=magrittr

Canty, A., & Ripley, B. D. (2017). *boot: Bootstrap R (S-Plus) Functions*.

Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression* (Second ed.). Thousand Oaks CA: Sage. Retrieved from http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Müller, K. (2017). *here: A Simpler Way to Find Your Files*. Retrieved from https://CRAN.R-project.org/package=here

Re, A. C. D. (2013). *compute.es: Compute Effect Sizes*. Retrieved from http://cran.r-project.org/web/packages/compute.es

Revelle, W. (2018). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, Illinois: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Tabachnick, B. G., & Fidell, L. S. (2012). *Using Multivariate Statistics*. Pearson.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Retrieved from http://ggplot2.org

Wickham, H. (2017). *tidyverse: Easily Install and Load the 'Tidyverse'*. Retrieved from https://CRAN.R-project.org/package=tidyverse

Wickham, H. (2018). *conflicted: An Alternative Conflict Resolution Strategy*. Retrieved from https://CRAN.R-project.org/package=conflicted

Xie, Y. (2019). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. Retrieved from https://yihui.name/knitr/

Zhu, H. (2019). *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax*. Retrieved from https://CRAN.R-project.org/package=kableExtra