EasyChair Preprint

№ 12975

# Effects of a Source Evaluation Intervention on Sourcing Skills: Replication and Extension

Magalí A. Martínez, Franco Londra, Gastón Saux and Debora I. Burin



April 9, 2024

# Effects of a source evaluation intervention on sourcing skills: Replication and extension

Sourcing is a process that involves attending to, retrieving, representing, and evaluating information about the sources of textual contents, in order to establish relations between them (Bråten & Braasch, 2018; Saux et al., 2017). Some relevant source features are the authors' occupation, credentials, motivation, as well as the documents' media quality (i.e., presence of an information validation process prior to publication) (Pérez et al., 2018). This is an advanced skill required to be a proficient reader in a multi-document environment (Salmerón et al., 2018).

The present study tested the effects of a college classroom intervention on source features evaluation. It is a pre-registered (https://osf.io/8ga4s/?view_only=38d7630a8c6e4eb6aee200540eec175d) replication and extension (to college students, in a Latin American context) of Pérez et al. (2018). In an experimental design (intervention vs control), we analysed sourcing skills before, after, and 6 to 8 weeks post- intervention.

This report concerns the main pre-registered hypotheses:

1. Trained students' rating scores for worse quality items would decrease after the intervention, compared to the control group.
2. To control that the predicted result was not caused by an overall effect of global decrease in the ratings, better quality items rating scores would not drop in post-test evaluations.

## Method

*Participants*

Participants were 266 (218F, 45M, 3 other gender; mean age 24.18, SD = 7.78) students from an introductory psychology course at a Latin American public university.

After considering pre-registered inclusion criteria, 229 participants were included in the analyses (intervention group: n=94; control group: n=135).

*Materials*

*Workshops content*. Workshop materials were adapted to the local sample from the original study (Pérez et al., 2018). Two workshops were implemented: the first dealt with author source dimensions (position, motivation), and the second introduced media quality (pre-publishing validation). They included group discussion of examples, explicit explanations about each source dimension, and individual practice with contrasting cases in which students' answers were discussed in groups with instructor guidance, until reaching consensus. The second workshop ended summarizing both workshops, and with an exercise where students had to rate the three source dimensions on a certain author giving information.

*Sourcing skills: Knowledge Application (KA) task*. Students were instructed to rate from 0 (absolutely not) to 4 (absolutely yes) whether they would visit the links presented when looking for trustworthy information for a time-constrained college task. Descriptions of nine websites with different combinations of the three source dimensions were presented: good or acceptable quality (at least two reliable features out of three total features), and poor or bad quality (only one or no reliable features). There were two items of each quality, and a filler (non-relevant) item.

We also measured participants' working memory capacity, reading comprehension, and vocabulary —an index of verbal ability—; not analysed here.

*Procedure*

Data were collected from March 2021 to November 2022. All participants were enrolled in the same semestral course. Each semester two classes were randomly assigned to the intervention or control group. Due to COVID-19 pandemic restrictions, during 2021 all

activities were implemented in online classes. In the other two semesters classes, workshops and sourcing skills assessments were implemented in-person.

Both groups received five sessions of no more than 90 minutes. Sessions 2 and 3 were one week apart and involved the intervention workshops (trained) or regular classroom activities (control); assessments were performed in Sessions 1 (pre-test), 4 (post-test at one week) and 5 (post-test at 6-8 weeks).

*Data analyses*

All analyses were conducted in R version 4.2.1 (R Core Team, 2022). The KA task items were categorised as Item Type bad (poor and bad links) and good (acceptable and good links). A general linear mixed effects model with Participants and Items as random factors, Group (control, trained), Phase (0, 1, 2), and Item Type (good, bad) as fixed factors, and KA task ratings for each item as dependent variable, was followed by two linear mixed effects (LME) models, one for each link type (bad, good). The analyses script (https://osf.io/2k6bz/?view_only=cea79297a0aa4f33aa35bcf8e13333cf) and database (https://osf.io/t9zs2/?view_only=73c23ecc235d4208ba59c39c81def8ef) are available in OSF.

## Results

A significant interaction between Group, Phase, and Item Type, showing better performance of the trained group at phase 1, estimate = 1.05, 95%CI = [0.72, 1.37] and phase 2, estimate = 0.81, 95%CI = [0.49, 1.13], *p*s < 0.001, was followed by analyses on good and bad items. Fig. 1 shows mean predicted ratings for bad and good test items, by Group and Phase.

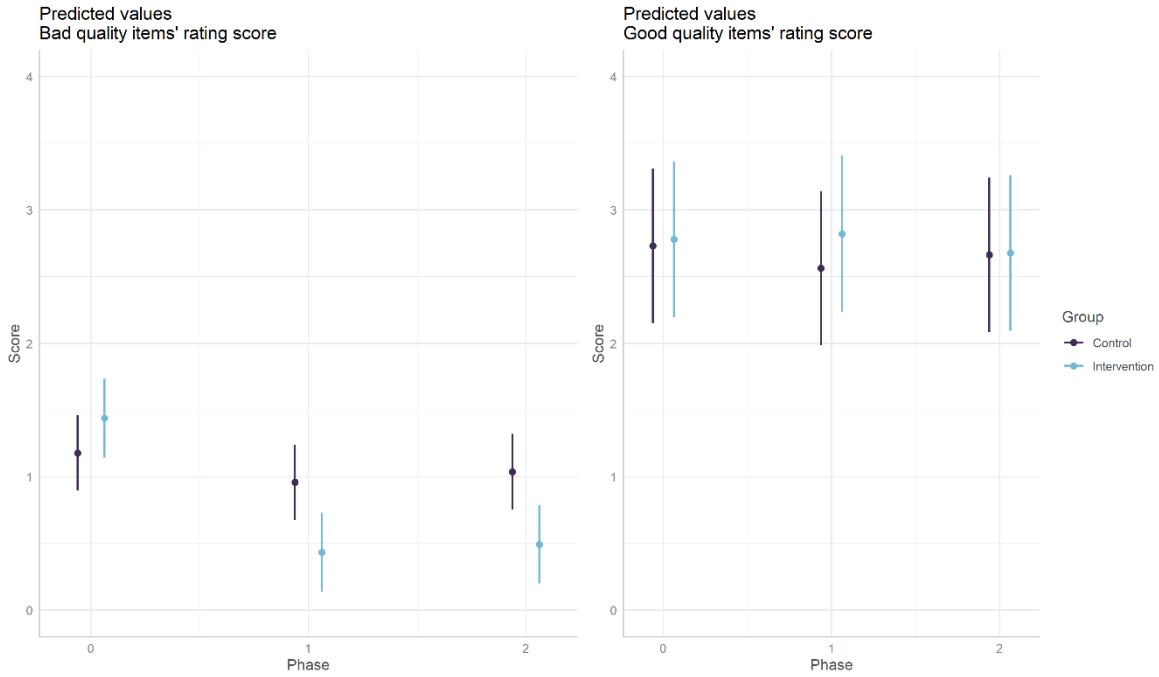


**Fig. 1.** Mean predicted ratings for bad (left panel) and good (right panel) items, by Group and Phase.

Table 1 shows the effects of Group and Phase on KA ratings for each bad quality item. In phase 0, control group rated bad items significantly lower than intervention group, estimate = -0.261, 95%CI = [-0.456, -0.0653], $p = 0.009$. Nevertheless, in phase 1 control group rated bad items significantly higher than intervention group, estimate = 0.525, 95%CI = [0.328, 0.722], $p < 0.0001$, and phase 2, estimate = 0.543, 95%CI = [0.349, 0.7362], $p < 0.0001$.

**Table 1.** LME model summary for the bad quality items rating scores, by Group and Phase

| | | value | |
|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* |
| (Intercept) | 1.18 | 0.89 – 1.46 | **<0.001** |
| Intervention group (baseline: control) | 0.26 | 0.07 – 0.46 | **0.009** |
| Phase 1 (baseline: 0) | -0.22 | -0.36 – -0.08 | **0.002** |
| Phase 2 (baseline: 0) | -0.14 | -0.28 – -0.01 | **0.042** |
| Intervention group by phase 1 | -0.79 | -1.00 – -0.57 | **<0.001** |
| Intervention group by phase 2 | -0.80 | -1.01 – -0.59 | **<0.001** |
| **Random Effects** | | | |
| $\sigma^2$ | 1.05 | | |
| $\tau_{00\ id}$ | 0.22 | | |
| $\tau_{00\ item}$ | 0.07 | | |
| ICC | 0.22 | | |
| $N_{\ id}$ | 229 | | |
| $N_{\ item}$ | 4 | | |
| Observations | 2346 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.076 / 0.277 | | |

Table 2 shows the effects of Group and Phase on KA ratings for each good quality item. There were no significant differences between groups for the good items in phase 0, nor in phase 2, *p*s > 0.05; but in phase 1 control group rated good items significantly lower than intervention group, estimate = -0.2589, 95%CI = [-0.448, -0.07], $p = 0.0073$. There were no significant differences between the ratings in each phase nor in the control group, nor in the intervention group, ps > 0.05.

**Table 2.** LME model summary for the good quality items rating scores, by Group and Phase

| | | value | |
|---|---|---|---|
| *Predictors* | *Estimates* | *CI* | *p* |
| (Intercept) | 2.73 | 2.15 – 3.31 | **<0.001** |
| Intervention group (baseline: control) | 0.05 | -0.14 – 0.24 | 0.608 |
| Phase 1 (baseline: 0) | -0.17 | -0.32 – -0.02 | **0.030** |
| Phase 2 (baseline: 0) | -0.07 | -0.22 – 0.08 | 0.379 |
| Intervention group by phase 1 | 0.21 | -0.03 – 0.45 | 0.084 |
| Intervention group by phase 2 | -0.03 | -0.27 – 0.20 | 0.773 |
| **Random Effects** | | | |
| $\sigma^2$ | 1.35 | | |
| $\tau_{00\ id}$ | 0.10 | | |
| $\tau_{00\ item}$ | 0.33 | | |
| ICC | 0.24 | | |
| $N_{\ id}$ | 229 | | |
| $N_{\ item}$ | 4 | | |
| Observations | 2357 | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.004 / 0.245 | | |

## Discussion

These results align with our hypotheses: despite being initially outperformed by control students, trained students became more critical to bad quality websites after the intervention. These results were not caused by a global decrease in ratings, since good quality links showed no significant differences in their rating scores across all testing phases. In sum, there is evidence that trained students improved their source evaluation skills. Future analysis of this study will report efficacy in terms of knowledge transfer, including a high-stakes academic essay.

## References

Bråten, I., & Braasch, J. (2018). The role of conflict in multiple source use. In J. Braasch, I. Bråten, & M. McCrudden (Eds.), *Handbook of multiple source use* (pp. 184-201). Routledge.

Pérez, A., Potocki, A., Stadtler, M., Macedo-Rouet, M., Paul, J., Salmerón, L., & Rouet, J.-F. (2018). Fostering teenagers' assessment of information reliability: Effects of a classroom intervention focused on critical source dimensions. *Learning and Instruction*, *58*, 53-64. https://doi.org/10.1016/j.learninstruc.2018.04.006

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org

Salmerón, L., Strømsø, H., Kammerer, Y., Stadtler, M., & van den Broek, P. (2018). Comprehension processes in digital reading. In M. Barzillai, J. Thomson, S. Schroeder, & P. van den Broek (Eds.), *Studies in Written Language and Literacy* (Vol. 17, pp. 91-120). John Benjamins Publishing Company. https://doi.org/10.1075/swll.17.04sal

Saux, G., Britt, A., Le Bigot, L., Vibert, N., Burin, D., & Rouet, J.-F. (2017). Conflicting but close: Reader's integration of information sources as a function of their disagreement. *Memory & Cognition*, *45*(1), 151-167. https://doi.org/10.3758/s13421-016-0644-5