Are we all on the same page? Subfield differences in open science practice

Christina Riochios[1] & Jenny L. Richmond[1]

[1] University of New South Wales

Author Note

School of Psychology, UNSW

The authors made the following contributions. Christina Riochios: Conceptualization, Methodology, Investigation, Data curation, Visualisation, Formal Analysis, Writing - Original Draft Preparation, Writing - Review & Editing; Jenny L. Richmond: Conceptualization, Methodology, Formal Analysis, Writing - Original Draft Preparation, Writing - Review & Editing, Supervision.

Abstract

Christina- can you insert abstract here please :)

One or two sentences providing a **basic introduction** to the field, comprehensible to a scientist in any discipline.

Two to three sentences of **more detailed background**, comprehensible to scientists in related disciplines.

One sentence clearly stating the **general problem** being addressed by this particular study.

One sentence summarizing the main result (with the words “**here we show**” or their equivalent).

Two or three sentences explaining what the **main result** reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge.

One or two sentences to put the results into a more **general context**.

Two or three sentences to provide a **broader perspective**, readily comprehensible to a scientist in any discipline.

Keywords: keywords

Word count: X

Are we all on the same page? Subfield differences in open science practice

INTRODUCTION

The field of psychology, like many other scientific disciplines, is currently facing a replication crisis, in which researchers are struggling to replicate existing findings. In 2015, a group of 270 psychological researchers attempted to replicate the findings of 100 psychology experiments. Whilst 97% of the original studies generated statistically significant findings, only 36% of the replication attempts were statistically significant. In addition, the replicated effects were, on average, half the size of the original effects (Collaboration, 2015). These findings illustrate the challenge of replicability in psychological research and the pressing need to rectify flawed research practices.

One strategy that has been used to combat the replication crisis within psychology is open science. Open science practices are those that increase the transparency of, and access to, scientific research (Klein et al., 2018). Open data and open materials practices, for example, involve researchers sharing their raw data and experimental materials on publicly accessible online repositories. These practices make it easier for others to replicate the methodology and reproduce the results from published work (Klein et al., 2018).

To encourage researchers to employ open science practices, many psychology journals have implemented incentives, like Open Science Badges. In 2013, the Center for Open Science established three Open Science Badges: Open Data, Open Materials and Preregistered, to acknowledge and reward researchers for their use of open science practices (Center for Open Science, 2021). The Open Data Badge and the Open Materials Badge are awarded when the data and materials that are required to reproduce the methods and results of a study are shared publicly online, whilst the Preregistered Badge is awarded when the design, hypotheses and/or analysis plan are publicly archived prior to data collection. To date, 75 journals (40 in Psychology) have adopted the COS Open Science badges (Center for Open Science, 2021).

At *Psychological Science*, the Association of Psychological Science's flagship journal, Open Science Badges appear to have been successful in encouraging researchers to adopt open science practices. In 2016, Kidwell et al. coded the frequency of data and material sharing in the 18 months before and after Open Science Badges were implemented at *Psychological Science*. (Kidwell et al., 2016). Kidwell et al. found that data sharing increased dramatically from 2.5% prior to badges to 39.4% following badges. Materials sharing also rose from 12.7% to 30.3%. Data and material sharing in control journals, such as the *Journal of Personality and Social Psychology*, which did not award badges, remained low over the same time period (Kidwell et al., 2016). These results led Kidwell et al. to conclude that Open Science Badges successfully incentivise the uptake of open science practices.

The support for open science continues to grow, however, it is not yet clear whether engagement with open science is consistent across different fields within psychology. Notably, developmental psychology has received significant criticism for its lack of receptivity towards open science. As Figure 1 illustrates, prominent developmental researchers, Prof Michael Frank and Dr. Jennifer Pfeifer, have labelled the Society for Research in Child Development's (SRCD) open science policy as 'weak' and as one that 'undervalues openness.' More recently, the Editor-in-Chief of Infant and Child Development, Prof Moin Syed, stated that the uptake of open science within the field of developmental psychology has been 'slow and uneven' (Syed, 2020). A survey supporting these viewpoints showed that, on average, 80% of researchers publishing in *Child Development* felt their institutions failed to provide adequate guidance or financial support for sharing data, (SRCD Task Force on Scientific Integrity and Openness Survey (2017), cited in Gennetian et al., (Gennetian, Tamis-LeMonda, & Frank, 2020). As such, developmental researchers may be slower to adopt open science practices than those in other psychological disciplines, however, this possibility has yet to be empirically investigated.

FIGURE 1

Meta-research, the study of research itself, can empirically assess whether developmental psychology is truly behind in the open science movement. Previous investigations, including Kidwell et al. (Kidwell et al., 2016), have revealed that open science incentives can increase the use of open science practices. However, what remains unclear is whether the uptake of open science has been consistent across psychological subfields and sustained over time. To address this research question, we used the open data from the Kidwell et al. study to examine whether rates of data and material sharing, following the implementation of Open Science Badges at Psychological Science, differed as a function of subfield. In addition, we conducted the same open data and materials coding for articles published in the most recent 18 months (July 2019-Dec 2020) to test whether the badges have continued to be impactful and whether the impact has been consistent across subfields. We were particularly interested in determining whether developmental psychology researchers papers published in developmental impact of badges has beenrovement sharing practices in the Whilst we expected to observe subfield differences in the use of open science practices, the nature and magnitude of these differences remained unclear. Study 1A was preregistered at the Open Science Framework: https://osf.io/3tsmy/.

Methods

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

Participants

Material

Procedure

Data analysis

We used R [Version 4.0.3; R Core Team (2020)] and the R-packages *afex* [Version 1.0.1; Singmann, Bolker, Westfall, Aust, and Ben-Shachar (2021)], *apa* [Version 0.3.3; Gromer

(2020); Aust and Barth (2020)], *dplyr* [Version 1.0.7; Wickham, François, Henry, and Müller (2021)], *forcats* [Version 0.5.1; Wickham (2021a)], *ggeasy* [Version 0.1.3; Carroll, Schep, and Sidi (2021)], *gghalves* [Version 0.1.1; Tiedemann (2020)], *ggplot2* [Version 3.3.5; Wickham (2016)], *ggsignif* [Version 0.6.3; Constantin and Patil (2021)], *goodshirt* [Version 0.2.2; Gruer (2021)], *here* [Version 1.0.1; Müller (2020)], *janitor* [Version 2.1.0; Firke (2021)], *lme4* [Version 1.1.27.1; Bates, Mächler, Bolker, and Walker (2015)], *Matrix* [Version 1.3.4; Bates and Maechler (2021)], *papaja* [Version 0.1.0.9997; Aust and Barth (2020)], *patchwork* [Version 1.1.1; Pedersen (2020)], *purrr* [Version 0.3.4; Henry and Wickham (2020)], *readr* [Version 2.0.1; Wickham and Hester (2021)], *report* [Version 0.3.5; Makowski, Ben-Shachar, Patil, and Lüdecke (2021)], *stringr* [Version 1.4.0; Wickham (2019)], *tibble* [Version 3.1.4; Müller and Wickham (2021)], *tidyr* [Version 1.1.3; Wickham (2021b)], and *tidyverse* [Version 1.3.1; Wickham et al. (2019)] for all our analyses.

1A Results

1AData scores text

As illustrated in Figure 1, the two-way between-subjects ANOVA generated a significant main effect of time period, $F(2, 310) = 11.29$, $MSE = 41.51$, $p < .001$, $\hat{\eta}^2_G = .068$. However, the main effect of subfield, $F(3, 310) = 2.23$, $MSE = 41.51$, $p = .085$, $\hat{\eta}^2_G = .021$, and the interaction between time period and subfield, $F(6, 310) = 1.57$, $MSE = 41.51$, $p = .157$, $\hat{\eta}^2_G = .029$, were not statistically significant. Articles published in the latter part of 2014 had significantly higher open data scores than article published in the first half of the year, $t(99.74) = -2.87$, $p = .005$, however there was no further increase across the early part of 2015,$t(132.74) = -0.86$, $p = .391$. Critically, there was no significant difference in the open data scores as a function of subfield (see Figure 1a).

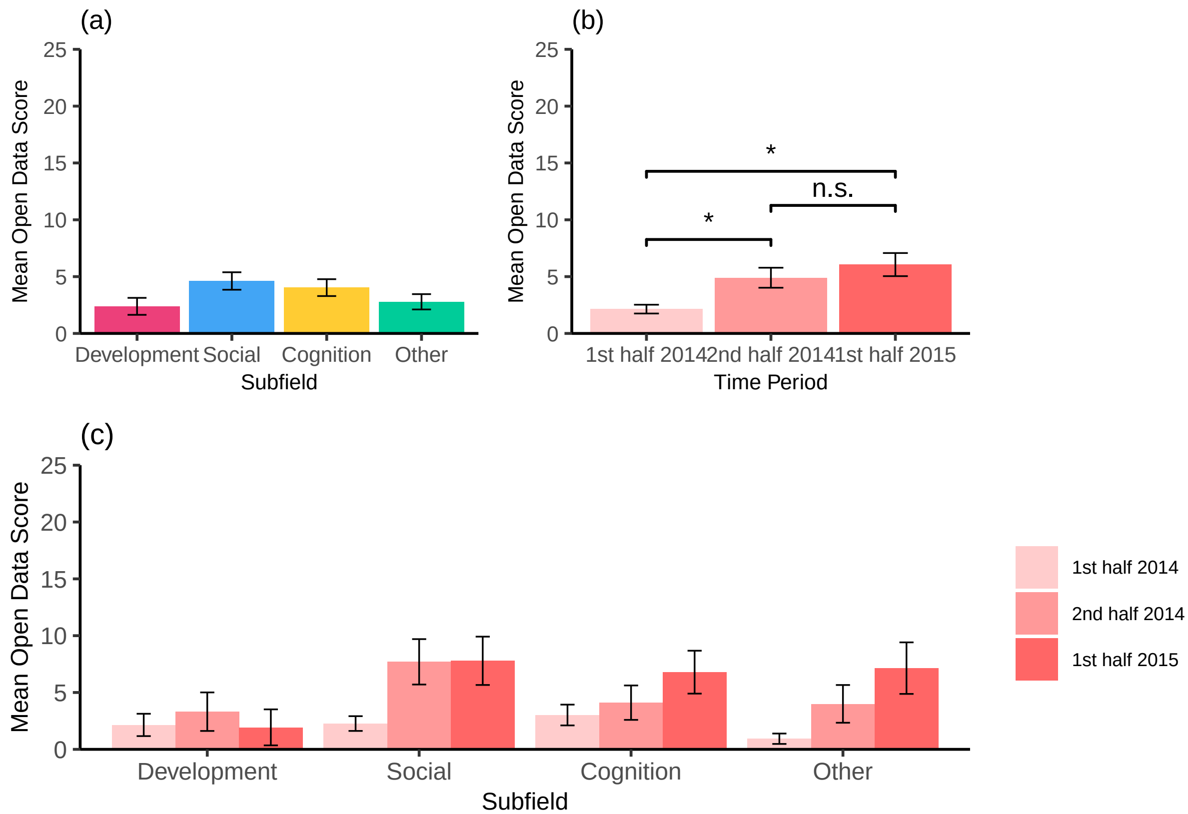


Figure 1: Mean open data scores for article published in 2014 and 2015, as a function of subfield and time period.

1AMaterial scores text

In contrast, for open materials scores, two-way between-subjects ANOVA generated a significant main effect of subfield, $F(3, 310) = 4.03$, $MSE = 32.16$, $p = .008$, $\hat{\eta}^2_G = .038$, and a significant main effect of time period, $F(2, 310) = 4.74$, $MSE = 32.16$, $p = .009$, $\hat{\eta}^2_G = .030$. However the interaction between time period and subfield, $F(6, 310) = 0.85$, $MSE = 32.16$, $p = .530$, $\hat{\eta}^2_G = .016$, was not statistically significant. Papers in developmental psychology had lower open materials scores than those in both social, $t(153.62) = -2.75$, $p = .007$, and cognition, $t(153.74) = -3.20$, $p = .002$, but developmental open materials scores did not differ from papers allocated to the other subfield category, $t(137.95) = -1.19$, $p = .236$. Open materials scores increased incrementally across 2014 and 2015; papers published in the first half of 2015 had significantly higher open materials scores than those published in the first half of 2014, $t(95.44) = -2.55$, $p = .012$.

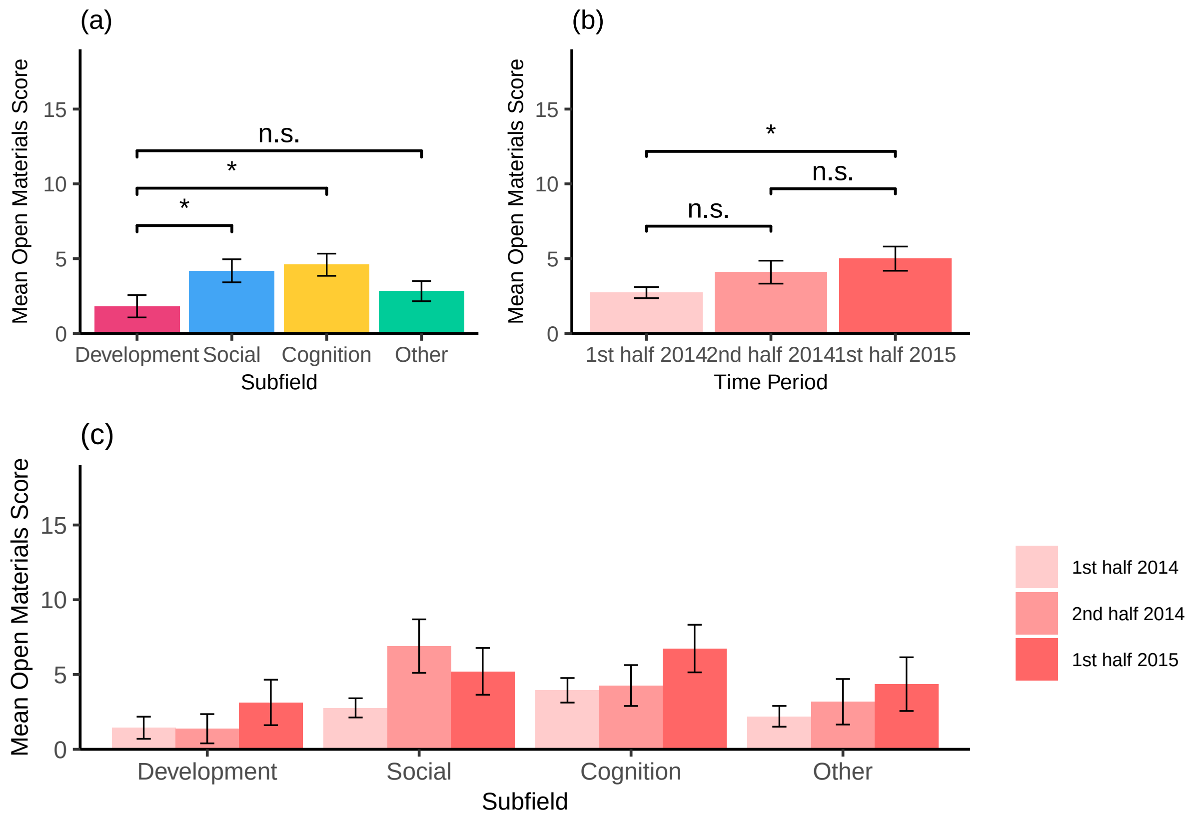


Figure 2: Mean open materials scores for articles published in 2014 and 2015, as a function of subfield and time period.

1B Results

1BData scores text

Consistent with the results from Study 1A, open data scores also increased significantly across 2019 and 2020, $F(2, 181) = 3.68$, $MSE = 70.92$, $p = .027$, $\hat{\eta}^2_G = .039$, however across this time period, open data scores differed significantly by subfield, $F(3, 181) = 3.31$, $MSE = 70.92$, $p = .021$, $\hat{\eta}^2_G = .052$. The interaction between time period and subfield, $F(6, 181) = 1.11$, $MSE = 70.92$, $p = .358$, $\hat{\eta}^2_G = .035$, was not statistically significant. When we compared the open data scores from papers published in developmental psychology to each of the other subfield categories (Figure 2a), we found that papers in developmental psychology had significantly lower open data scores than papers in cognition, $t(58.75) = -2.65$, $p = .010$. The magnitude of open data scores did not differ from papers published in social psychology, $t(71.68) = -0.42$, $p = .675$ or those that fell into the other category, $t(71.12) = -0.31$, $p = .756$. In terms of changes over time, open data scores increased significantly from late 2019 into 2020, $t(126.73) = -2.70$, $p = .008$, but were stable across 2020, $t(62.65) = 0.98$, $p = .331$.

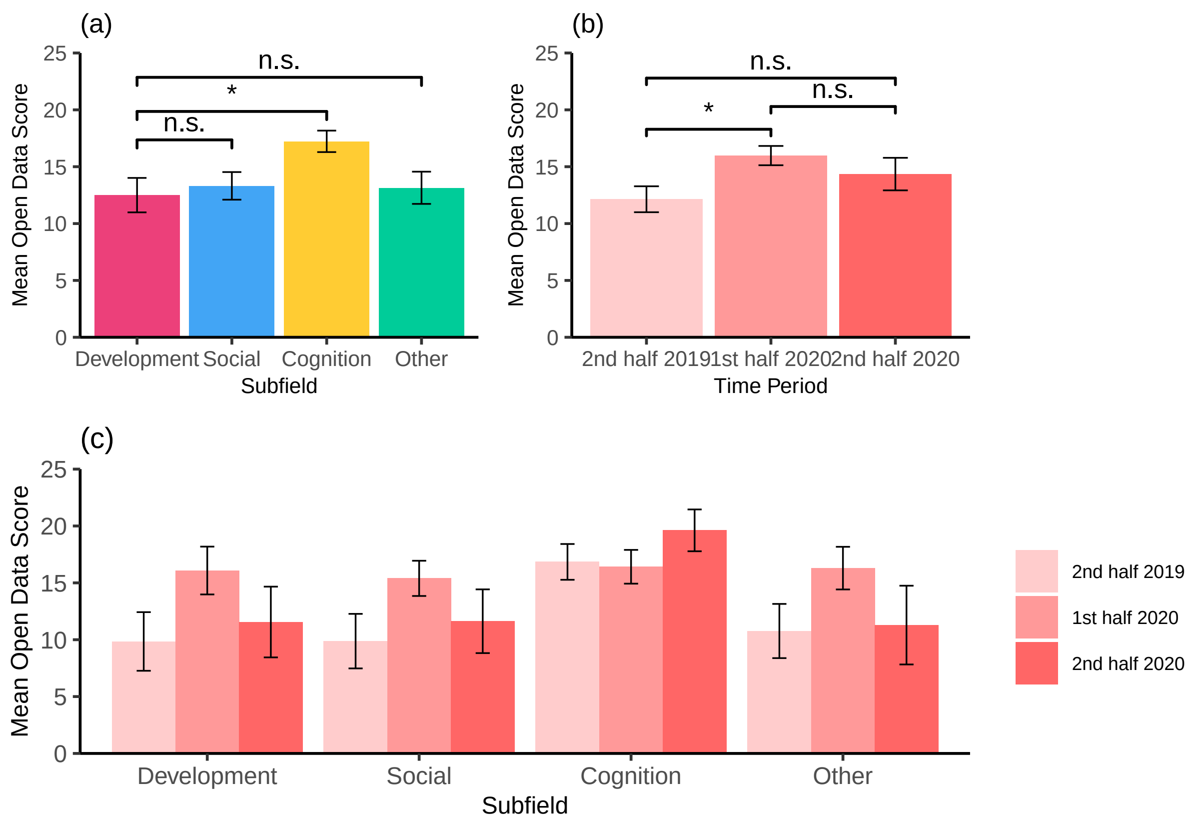


Figure 3: Mean open data scores for articles published in 2019 and 2020, as a function of subfield and time period.

1BMaterials scores text

As in Study 1A, for open materials scores in 2019 and 2020, there was a significant main effect of subfield, $F(3, 181) = 5.24$, $MSE = 53.87$, $p = .002$, $\hat{\eta}^2_G = .080$, however the main effect of time period, $F(2, 181) = 0.37$, $MSE = 53.87$, $p = .694$, $\hat{\eta}^2_G = .004$, and the interaction between time period and subfield, $F(6, 181) = 0.48$, $MSE = 53.87$, $p = .822$, $\hat{\eta}^2_G = .016$, were not statistically significant (see Figure 2). Consistent with open data scores, papers published in developmental psychology had significantly lower open materials scores than paper published in cognition, $t(61.36) = -3.45$, $p = .001$, however, open materials scores did not differ between developmental and social psychology, $t(68.36) = -0.84$, $p = .406$, or between developmental psychology and the other subfield category, $t(70.45) = -0.62$, $p = .539$. There were no additional changes in open materials scores across this period between mid-2019 and the end of 2020.

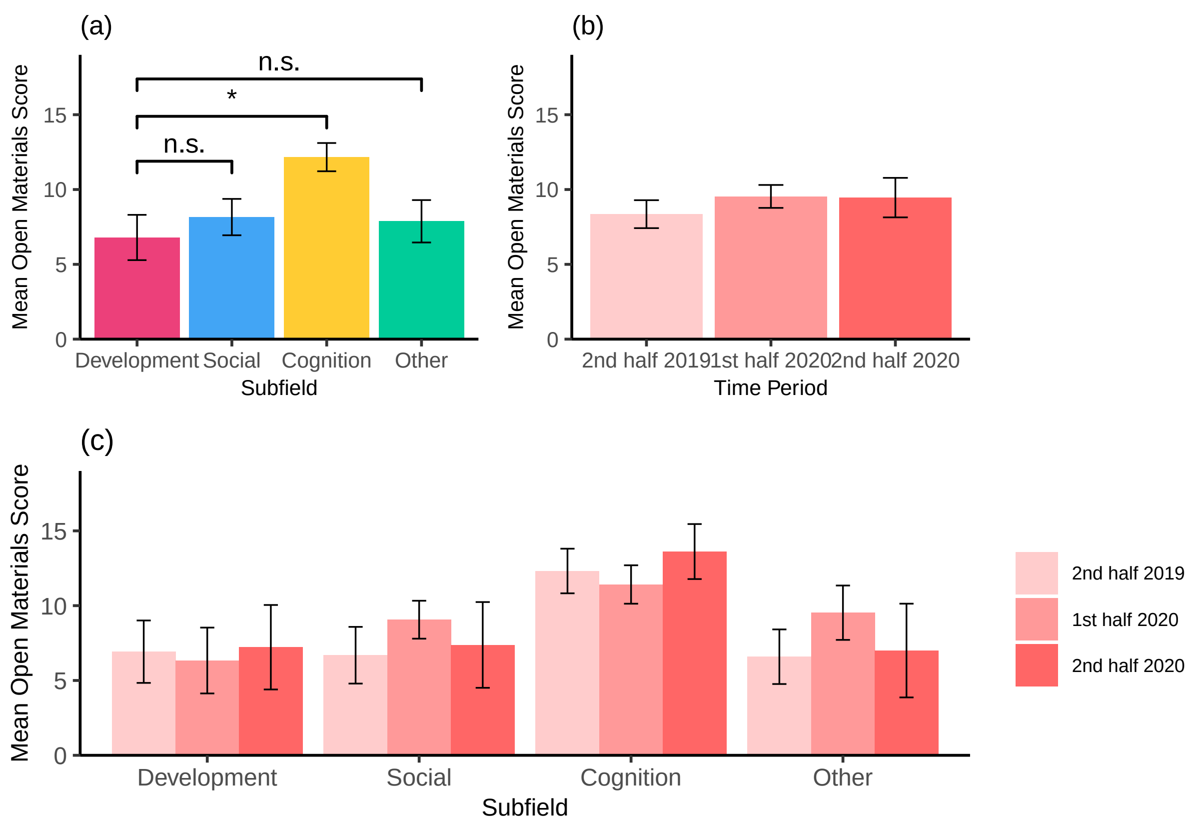


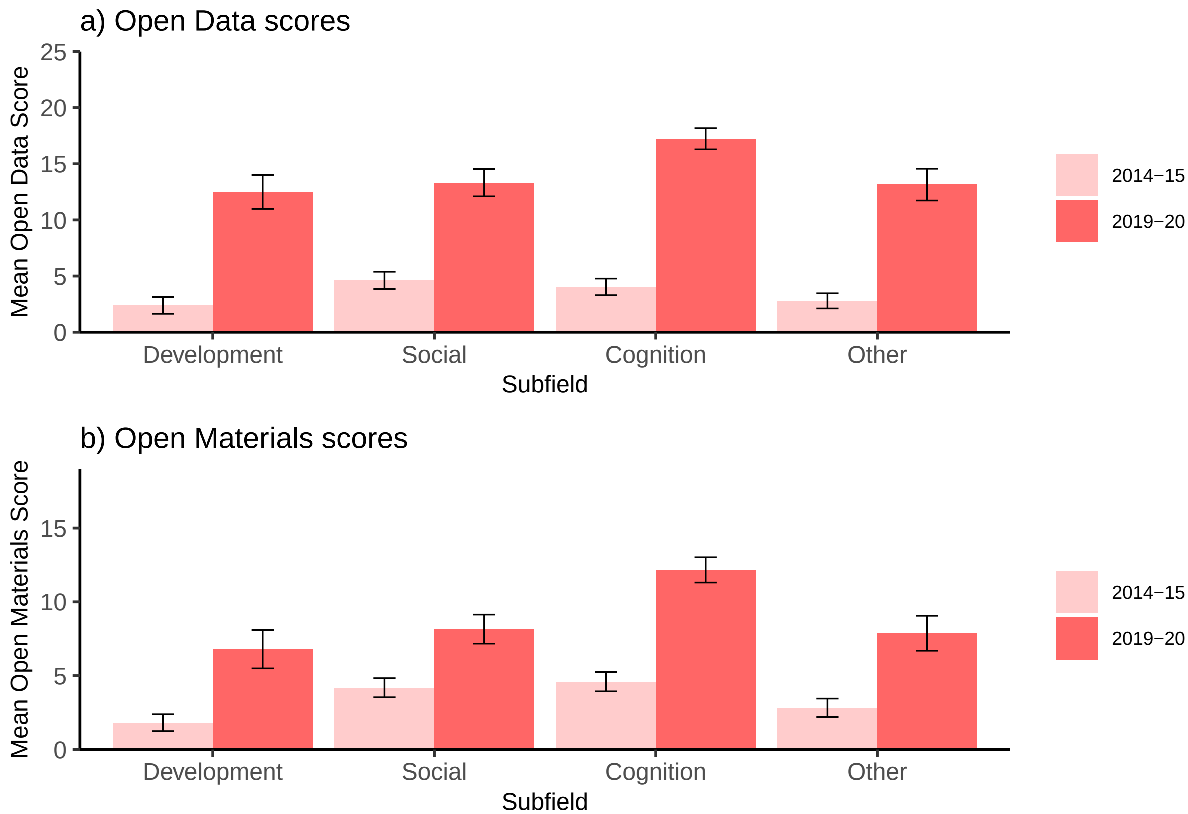
Figure 4: Mean open materials scores for articles published in 2019 and 2020, as a function of subfield and time period.

AB ACROSS time

AB data/materials text

When looking at open data scores across both time periods, we found that scores increased over time, $F(1, 507) = 248.98$, $MSE = 55.28$, $p < .001$, $\hat{\eta}^2_G = .329$, and differed as a function of subfield, $F(3, 507) = 3.74$, $MSE = 55.28$, $p = .011$, $\hat{\eta}^2_G = .022$. However, the interaction between time period and subfield, $F(3, 507) = 2.28$, $MSE = 55.28$, $p = .078$, $\hat{\eta}^2_G = .013$, was not statistically significant, suggesting that there were no subfield differences in the magnitude of open data scores over time (see Figure Xa). A similar pattern was found for open material scores; scores increased over time $F(1, 507) = 90.18$, $MSE = 40.21$, $p < .001$, $\hat{\eta}^2_G = .151$, and differed by subfield, $F(3, 507) = 8.17$, $MSE = 40.21$, $p < .001$, $\hat{\eta}^2_G = .046$, but there was no evidence that the magnitude of improvement over time differed by subfield (see Figure Xb), $F(3, 507) = 2.07$, $MSE = 40.21$, $p = .103$, $\hat{\eta}^2_G = .012$.

t-tests probably not necessary, b/c point was really to see if there was an interaction, also maybe the most appropriate plot is just the interaction one, rather than subfield + time

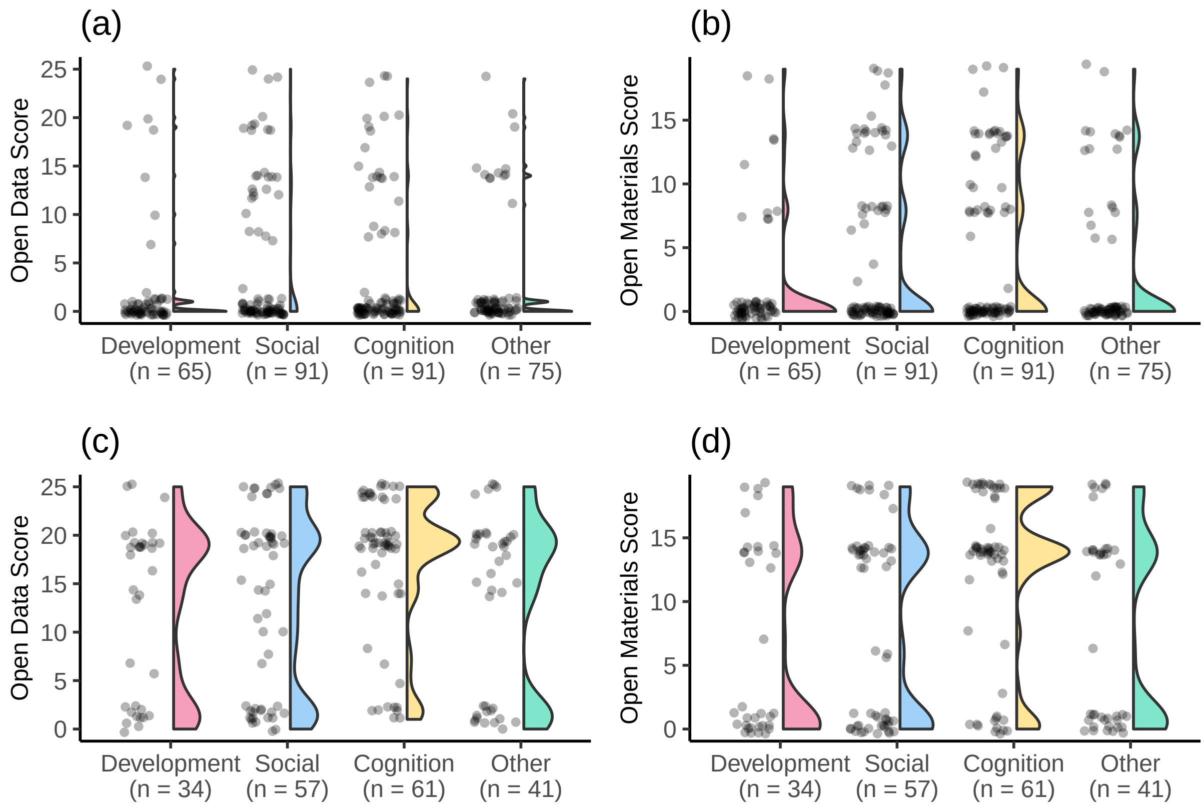


including just the combo of interaction plots here. Christina- can you remove the code that makes the subfield and time plots from the chunks above and work out how to make this patchwork plot share y axis labels and legends

JENNY UP TO HERE

STUDY 1B EXPLORATORY

RAINCLOUD PLOTS



OPEN SCIENCE BADGES

Bind all statistics together.

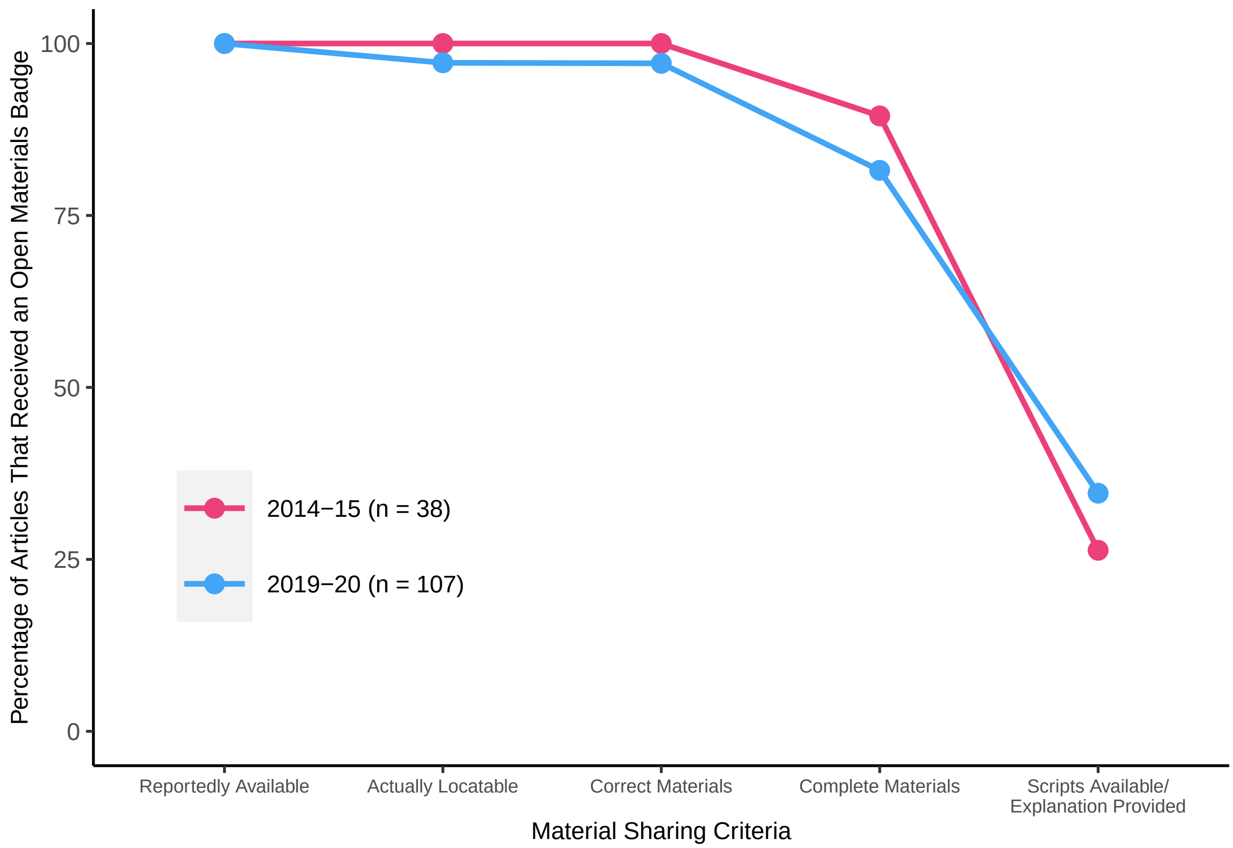
Select only relevant 1B variables.

Summary tables.

Combine 1A and 1B stats together.

Plot combined data.

Plot combined data.


Percentage of Articles That Received an Open Materials Badge
100
75
50
25
0
2014−15 (n = 38)
2019−20 (n = 107)
Reportedly Available
Actually Locatable
Correct Materials
Complete Materials
Scripts Available/
Explanation Provided
Material Sharing Criteria

Discussion

References

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown.* Retrieved from https://github.com/crsh/papaja

Bates, D., & Maechler, M. (2021). *Matrix: Sparse and dense matrix classes and methods*. Retrieved from https://CRAN.R-project.org/package=Matrix

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Carroll, J., Schep, A., & Sidi, J. (2021). *Ggeasy: Easy access to 'ggplot2' commands.* Retrieved from https://CRAN.R-project.org/package=ggeasy

Collaboration, O. S. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251).

Constantin, A.-E., & Patil, I. (2021). ggsignif: R package for displaying significance brackets for 'ggplot2'. *PsyArxiv*. https://doi.org/10.31234/osf.io/7awm6

Firke, S. (2021). *Janitor: Simple tools for examining and cleaning dirty data.* Retrieved from https://CRAN.R-project.org/package=janitor

Gennetian, L. A., Tamis-LeMonda, C. S., & Frank, M. C. (2020). Advancing transparency and openness in child development research: opportunities. *Child Development Perspectives*, *14*(1), 3–8.

Gromer, D. (2020). *Apa: Format outputs of statistical tests according to APA guidelines*. Retrieved from https://CRAN.R-project.org/package=apa

Gruer, A. (2021). *Goodshirt: R client for the good place quotes API.*

Henry, L., & Wickham, H. (2020). *Purrr: Functional programming tools.* Retrieved from https://CRAN.R-project.org/package=purrr

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., … others. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS Biology*, *14*(5), e1002456.

Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Mohr, A. H., … others. (2018). A practical guide for transparency in psychological science. *Collabra: Psychology*, *4*(1).

Makowski, D., Ben-Shachar, M. S., Patil, I., & Lüdecke, D. (2021). Automated results reporting as a practical tool to improve reproducibility and methodological best practices adoption. *CRAN*. Retrieved from https://github.com/easystats/report

Müller, K. (2020). *Here: A simpler way to find your files.* Retrieved from https://CRAN.R-project.org/package=here

Müller, K., & Wickham, H. (2021). *Tibble: Simple data frames.* Retrieved from https://CRAN.R-project.org/package=tibble

Pedersen, T. L. (2020). *Patchwork: The composer of plots.* Retrieved from https://CRAN.R-project.org/package=patchwork

R Core Team. (2020). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2021). *Afex: Analysis of factorial experiments.* Retrieved from https://CRAN.R-project.org/package=afex

Syed, M. (2020). Infant and child development: A journal for open, transparent, and inclusive science from prenatal through emerging adulthood.

Syed, M. (2020). Infant and child development: A journal for open, transparent, and inclusive science from prenatal through emerging adulthood.

Tiedemann, F. (2020). *Gghalves: Compose half-half plots using your favourite geoms.* Retrieved from https://CRAN.R-project.org/package=gghalves

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis.* Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

Wickham, H. (2019). *Stringr: Simple, consistent wrappers for common string operations.* Retrieved from https://CRAN.R-project.org/package=stringr

Wickham, H. (2021a). *Forcats: Tools for working with categorical variables (factors).* Retrieved from https://CRAN.R-project.org/package=forcats

Wickham, H. (2021b). *Tidyr: Tidy messy data.* Retrieved from https://CRAN.R-project.org/package=tidyr

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., & Müller, K. (2021). *Dplyr: A grammar of data manipulation.* Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham, H., & Hester, J. (2021). *Readr: Read rectangular text data.* Retrieved from https://CRAN.R-project.org/package=readr

Table 1

Between-subjects ANOVA for Open Materials Scores

Effect	F	df_1	df_2	MSE	p	$\hat{\eta}^2_G$
Time period	4.74	2	310	32.16	.009	.030
Subfield groups	4.03	3	310	32.16	.008	.038
Time period × Subfield groups	0.85	6	310	32.16	.530	.016

Table 2

Between-subjects ANOVA for Open Materials Scores

Effect	F	df_1	df_2	MSE	p	$\hat{\eta}^2_G$
Time period	0.37	2	181	53.87	.694	.004
Subfield groups	5.24	3	181	53.87	.002	.080
Time period × Subfield groups	0.48	6	181	53.87	.822	.016