

1 Are we all on the same page? Subfield differences in open science practices in psychology

2

3

#### 4 Author Note

5 Availability Statement: All data, coding and analysis scripts are publicly available via  
6 the [Open Science Framework repository](#). The preregistration can be accessed [here](#). Conflict  
7 of Interest Disclosure: There were no conflicts of interest in relation to the authorship or  
8 publication of this article. Ethics Approval Statement: The study was approved by the  
9 UNSW Human Research Ethics Advisory Panel prior to data collection. Acknowledgements:  
10 The authors would like to acknowledge the tremendous efforts of the coding team. Each  
11 member of the coding team played an integral role in this investigation; their assistance is  
12 truly appreciated.

## Abstract

Although open science has become a popular tool to combat the replication crisis, it is unclear whether the uptake of open science practices has been consistent across the field of psychology. In this study, we tested whether data and material sharing, two prominent open science practices, differed as a function of psychological subfield at the distinguished journal, *Psychological Science*. The results showed that open data and open materials scores increased from 2014-2015 to 2019-2020. Of note, articles published in the field of developmental psychology generated considerably lower open data and open materials scores, compared to articles in other subfields. These findings are discussed in the context of why developmental psychologists may be slower to adopt open science practices, compared to other researchers, and how journals can more effectively encourage authors to practice open science across psychology. As part of our analyses, we also looked at how the awarding of Open Science Badges related to researchers' open science behaviours, showing that the shared data and materials were less usable expected, even for articles that earned badges. We propose an alternative open science initiative, that may be more effective in increasing the replicability of psychological research.

*Keywords:* open data; open materials; subfield differences; developmental psychology

Word count: 4748

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

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 (Open Science Collaboration, 2015). These findings illustrate the poor replicability of psychological research and the pressing need to rectify flawed research practices.

Open science practices, which increase the transparency of, and access to, scientific research, have been used to combat the replication crisis within psychology. (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 are designed to 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 and Open Materials Badges, for example, are awarded when the data and materials that are required to reproduce the methods and results of a study are shared publicly online. To date, over 75 journals (40 in Psychology) have adopted 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. found that data sharing increased dramatically from 2.5% of articles prior to badges to 39.4% of articles 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). Although their study simply described the proportion of articles that engaged in data and materials sharing before and after the policy change, the results led Kidwell et al. to conclude that Open Science Badges successfully incentivised the uptake of open science practices at *Psychological Science*.

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, the field of 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.' (Frank, 2020, March 6; Pfeifer, 2020, March 8). 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, 2021). A survey supporting these viewpoints showed that 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., (2020)). Developmental psychology researchers may be slower to adopt open science practices than those in other psychological disciplines, however, this possibility has yet to be empirically investigated.

Using meta-research, the study of research itself, we can empirically assess whether



Figure 1. Twitter thread about Society for Child Development open science policy.

developmental psychology is truly behind in the open science movement. Previous investigations, including Kidwell et al. (2016), have revealed that open science incentives can increase the use of open science practices. However, it is unclear whether Open Science Badges, have had the same impact across different psychological subfields and whether the effect is sustained over time. To address this research question, we used the open data from the Kidwell et al., (2016) study and designed a quantitative scoring system 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 applied the same coding system to articles published in the most recent 18 months (Jul 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 publishing in *Psychological Science* engaged with open

science practices at the same rate as researchers from other subdomains of psychology. Our methods and analysis plan were preregistered at the [Open Science Framework](#).

## Methods

### Design

This study had a quasi-experimental design; all articles were systematically assigned to one of seven subfields. For each article, we used coded variables to compute two scores that indexed the transparency of data and materials, respectively. Changes in data and material sharing were analysed over six-month intervals.

### Sample

The Kidwell et al., (2016) sample included all *Psychological Science* articles published between January 2014 and May 2015 ( $N = 367$ ), which were coded to evaluate the openness of their data and materials. To identify how data and material sharing may have changed since 2014-2015, our sample also included all *Psychological Science* articles that were published between July 2019 and December 2020 ( $N = 242$ ). Non-empirical articles that did not contain an experiment or analysis, including editorials, commentaries, replies, corrigenda, errata and retractions, were excluded from our analysis. After filtering out these non-empirical articles from the sample, 322 articles published between 2014-2015 and 193 articles published between 2019-2020, remained.

### Materials

To assess the transparency of data and materials for each article, Kidwell et al. (2016) employed a systematic coding system ([Kidwell system](#) and [variable definitions](#)). We downloaded the Kidwell et al. data from their [OSF repository](#) and filtered the dataset to only include *Psychological Science* articles published between January 2014 and May 2015.

In addition to the variables that Kidwell et al. had coded, we also coded for whether

the article specified their analysis software or not, and which type of analysis software had been specified (e.g., R, JASP, SPSS etc). These variables were important to include because when authors' identify analysis software, the analysis procedure can be easier to follow and the chance of successfully reproducing the analysis may increase (National Academies of Sciences, Medicine, & others, 2019). The same amended version of the Kidwell et al. coding system, including the two additional analysis software variables, was used to code the articles that were published between July 2019 and December 2020.

We designed an additional coding system ([subfield system](#) and [variable definitions](#)) to assign all the articles to one of seven psychological subfields: Developmental Psychology, Social Psychology, Cognition, Perception, Health Psychology, Cognitive Neuroscience and Behavioural Neuroscience. We identified these seven subfields as those that the majority of *Psychological Science* articles fall into, after thoroughly reviewing the journal website.

Prior to data collection, each member of the coding team coded five trial articles, to confirm their understanding of the coding process. These trial articles were *Psychological Science* articles originally coded by Mallory Kidwell, the primary investigator in the Kidwell et al. (2016) study. Kidwell's coding acted as the standard to which coders' responses were compared. The senior coder in the current study generated the standard for the variables that weren't included in the Kidwell et al. coding system (i.e. those related to software and subfield). The trial articles varied in the transparency of their data and materials, and therefore, exposed coders to a representative range of coding outcomes.

The coding team coded both the trial and target articles via a Qualtrics survey, containing a series of multiple-choice questions. The questions were structured in an 'if-then' manner, with some questions only being asked if coders provided particular answers to the questions prior. For example, coders were only asked about the participants' age, if they had specified that the participants in the study were 'Humans' as opposed to 'Animals.'

## Procedure

After the investigation had been approved by the Human Research Ethics Advisory Panel, we assembled a team of volunteer coders, comprising of undergraduate psychology students. Once the coders completed the five trial articles and the senior coder was confident that each coder understood how to code all the variables, the coders were provided access to the target set of articles to begin coding using the Qualtric survey.

**Scoring procedure.** After all articles had been coded, we imported the data from Qualtrics into the software environment, R (R Core Team, 2020). For the articles that were published between 2014-2015, we combined the newly collected data related to software and subfield with the data from Kidwell et al. (2016). Each article, across both the 2014-2015 and 2019-2020 datasets, was assigned to one of the seven psychological subfields and received an open data and open materials score. The open data score indexed the extent to which the data were transparent, whilst the open materials score indexed the extent to which the materials were transparent. To calculate the scores, we weighted each coded variable according to how much it improved the transparency of the data and materials, respectively.

*Table 1: Open data scoring (left) and open materials scoring (right) criteria*

Variable	Score Assigned	Variable	Score Assigned
<b>Low-level transparency</b>		<b>Low-level transparency</b>	
Presence of data availability statement	1	Presence of materials availability statement	1
Data reported to be available	1	Materials reported to be available	1
Analysis software specified	1		
<b>Medium-level transparency</b>		<b>Medium-level transparency</b>	
Presence of data URL	2	Presence of materials URL	2
Data URL is functional	2	Materials URL is functional	2
Data located at URL	2	Materials located at URL	2
Data are downloadable	2	Materials are downloadable	2
Data correspond to article	2	Materials correspond to article	2
Data are complete	2	Materials are complete	2
<b>High-level transparency</b>		<b>High-level transparency</b>	
Codebook available with data	5	Explanation of materials/corresponding scripts	5
Analysis scripts available with data	5		

There were three levels of transparency: low-level transparency variables received a value of 1, moderate-level transparency variables received a value of 2 and high-level transparency variables received a value of 5 (see Table 1). We summed these scores so that each article received an open data score out of a possible 25 and an open materials score out



of a possible 19. Higher scores reflected a higher level of transparency.

**Reliability.** The senior coder randomly selected 25 empirical articles from the 2014-2015 dataset (8% of the empirical sample), ensuring that an equal number had been coded by each coder ( $n = 5$ ), and double-coded these software and subfield variable. Using the ‘kappa2’ function from the ‘irr’ package in R (Gamer, Lemon, & Singh, 2019), we ran a Cohen’s Kappa reliability analysis for subfield assignment, which revealed that the coding team had good reliability compared to the senior coder’s standard,  $k = .605$ , according to Fleiss’s (1981) guidelines. The percent agreement rating between the standard and the coding team was 72%. Upon examining cases where the standard and the coding team disagreed on an article’s subfield assignment, we found that the discrepancy could usually be attributed to the subject matter spanning across multiple subfields. Since our coding system did not account for the possibility of a study belonging to multiple subfields, the results from our reliability analysis may be conservative.

For the 2019-2020 sample of articles, the senior coder similarly selected 25 articles from the empirical sample (13%) and double-coded these articles. To assess reliability, each article received a total openness score, representing the sum of the open data and open materials score. We used the ‘icc’ function from the ‘irr’ package in R to generate an intraclass correlation coefficient (ICC) (Gamer, Lemon, & Singh, 2019). The ‘tolerance’ level was set at five Total Openness points; where scores fell within a five-point range of each other, they were considered to be equivalent.

The ICC analysis showed that the coding team had excellent reliability relative to the senior coder’s standard, according to Cicchetti’s (1994) guidelines,  $ICC = .905$ , 95% CI (.772, .962). As a secondary measure of inter-rater reliability, we also calculated the percent agreement between the standard and coders’ responses. The agreement rating between the coders and the standard was 73.7%, with a tolerance level of five Total Openness points.

**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)], *irr* [Version 0.84.1; Gamer, Lemon, and Singh (2019)], *janitor* [Version 2.1.0; Firke (2021)], *kableExtra* [Version 1.3.4; Zhu (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)], *scales* [Version 1.1.1; Wickham and Seidel (2020)], *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.

**Confirmatory Analyses.** Analysis of Variance (ANOVA) analyses were run to investigate differences in open data and open materials scores, separately, across the 2014-2015 and 2019-2020 datasets. In each analysis, we tested for a main effect of time, measured over three six-month intervals, and subfield. To ensure that there was a comparable number of articles in each subfield group, we combined Cognitive Neuroscience, Behavioural Neuroscience, Health Psychology and Perception into a single ‘Other’ category. As a result, a total of four subfield groups were included in our analysis: Developmental Psychology, Social Psychology, Cognition and Other. We report effect sizes in terms of generalised eta squared ( $\eta^2$ ).

**Exploratory Analyses.** After data collection, we were also interested in the

distribution of scores and how the spread of scores might differ by subfield. To illustrate this we generated two raincloud plots that illustrated the distribution of open data and open materials scores across 2019-2020. Raincloud plots visualise the distribution of scores in a dataset by showing the density of subjects at each level of the dependent measure (Allen, Poggiali, Whitaker, Marshall, & Kievit, 2019). In our case, where the violin plot was wider, the concentration of articles that received the corresponding open data or open materials score, was greater.

We also wanted to learn how Open Science Badges related to researchers' data and materials sharing practices. To generate two corresponding figures, we filtered the 2019-2020 dataset to only include the articles that had received an Open Data Badge and an Open Materials Badge, respectively. We then plotted the percentage of these articles that met a series of data and materials sharing criteria, described in the Results section below.

**Preregistration.** We preregistered our aims, hypotheses, design, and planned analysis procedure for the study at the [OSF](#). Whilst we attempted to follow each of the proposed procedures as closely as possible, we made one notable modification, choosing not to normalise the Open Data and Open Materials Scores (so that they were both out of 100). Since our study was focussed on measuring subfield differences and changes over time, within each type of score, rather than comparing the differences between the two scores, we ultimately realised that normalising the scores was not necessary. All the materials, data and analysis scripts from the study can be accessed via the [OSF](#).

## Results

We first used the open data from Kidwell et al., (2016) and analysed whether open data and open materials scores improved across the 2014-2015 period and differed by subfield. As illustrated in Figure 2A, during the period immediately following the badge policy change, open data scores were uniformly low across subfields,  $F(3, 310) = 2.23$ ,  $MSE = 41.51$ ,  $p = .085$ ,  $\hat{\eta}_G^2 = .021$ . Whilst open data scores increased over

time,  $F(2, 310) = 11.29$ ,  $MSE = 41.51$ ,  $p < .001$ ,  $\hat{\eta}_G^2 = .068$ , this improvement did not differ  
 as a function of subfield  $F(6, 310) = 1.57$ ,  $MSE = 41.51$ ,  $p = .157$ ,  $\hat{\eta}_G^2 = .029$ .

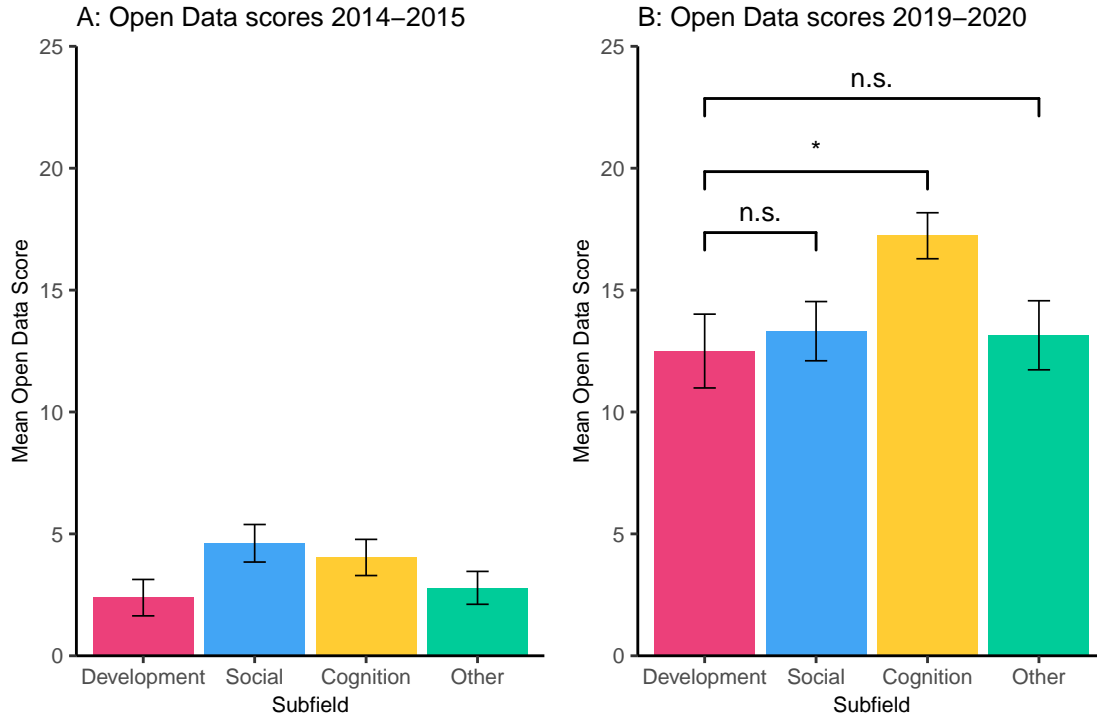


Figure 2. Mean open data scores for articles published in *Psychological Science* between 2014-2015 and 2019-2020 as a function of subfield.

Across the 2019-2020 period, open data scores also increased significantly,  
 $F(2, 181) = 3.68$ ,  $MSE = 70.92$ ,  $p = .027$ ,  $\hat{\eta}_G^2 = .039$ , and differed by subfield,  
 $F(3, 181) = 3.31$ ,  $MSE = 70.92$ ,  $p = .021$ ,  $\hat{\eta}_G^2 = .052$ . When we compared the open data  
 scores from papers published in developmental psychology to each of the other subfield  
 categories (Figure 2B), we found that papers in developmental psychology had significantly  
 lower open data scores ( $M = 12.50$ ,  $SD = 8.83$ ) than papers in cognition ( $M = 17.23$ ,  $SD =$   
 $7.37$ ),  $t(58.75) = -2.65$ ,  $p = .010$ , but did not differ from papers published in social  
 psychology ( $M = 13.32$ ,  $SD = 9.17$ ),  $t(71.68) = -0.42$ ,  $p = .675$  or those that fell into the  
 other category ( $M = 13.15$ ,  $SD = 9.07$ ),  $t(71.12) = -0.31$ ,  $p = .756$ . There was no  
 significant time by subfield interaction,  $F(6, 181) = 1.11$ ,  $MSE = 70.92$ ,  $p = .358$ ,  $\hat{\eta}_G^2 = .035$ .

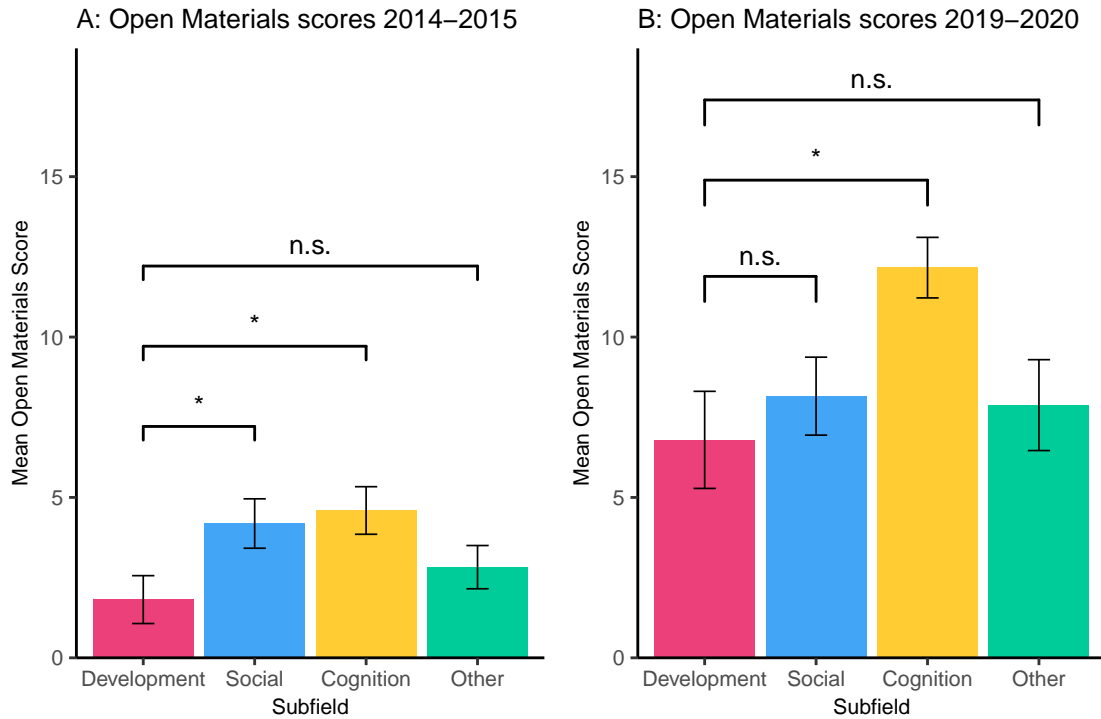


Figure 3. Mean open materials scores for articles published in *Psychological Science* between 2014-2015 and 2019-2020 as a function of subfield.

For open materials scores across 2014-2015, there were significant main effects of both subfield,  $F(3, 310) = 4.03$ ,  $MSE = 32.16$ ,  $p = .008$ ,  $\hat{\eta}_G^2 = .038$ , and time period,  $F(2, 310) = 4.74$ ,  $MSE = 32.16$ ,  $p = .009$ ,  $\hat{\eta}_G^2 = .030$  (see Figure 3A). Papers in developmental psychology had lower open materials scores ( $M = 1.82$ ,  $SD = 6.02$ ) than those in both social ( $M = 4.19$ ,  $SD = 7.34$ ),  $t(153.62) = -2.75$ ,  $p = .007$ , and cognition ( $M = 4.59$ ,  $SD = 7.07$ ),  $t(153.74) = -3.20$ ,  $p = .002$ , but developmental open materials scores did not differ from papers allocated to the other subfield category ( $M = 2.83$ ,  $SD = 5.84$ ),  $t(137.95) = -1.19$ ,  $p = .236$ . The interaction between time period and subfield,  $F(6, 310) = 0.85$ ,  $MSE = 32.16$ ,  $p = .530$ ,  $\hat{\eta}_G^2 = .016$ , was not statistically significant.

As illustrated in Figure 3B, there were also subfield differences in open materials scores during the 2019-2020 period,  $F(3, 181) = 5.24$ ,  $MSE = 53.87$ ,  $p = .002$ ,  $\hat{\eta}_G^2 = .080$ .

Consistent with open data scores, papers published in developmental psychology had significantly lower open materials scores ( $M = 6.79$ ,  $SD = 8.83$ ) than papers published in cognition, ( $M = 6.79$ ,  $SD = 8.83$ ),  $t(61.36) = -3.45$ ,  $p = .001$ , however, open materials scores did not differ between developmental and social psychology ( $M = 6.79$ ,  $SD = 8.83$ ),  $t(68.36) = -0.84$ ,  $p = .406$ , nor between developmental psychology and the other subfield category ( $M = 6.79$ ,  $SD = 8.83$ ),  $t(70.45) = -0.62$ ,  $p = .539$ . There were no changes in open materials scores across the time period between mid-2019 and the end of 2020,  $F(2, 181) = 0.37$ ,  $MSE = 53.87$ ,  $p = .694$ ,  $\hat{\eta}_G^2 = .004$ , and differences in subfield did not vary over time,  $F(6, 181) = 0.48$ ,  $MSE = 53.87$ ,  $p = .822$ ,  $\hat{\eta}_G^2 = .016$ .

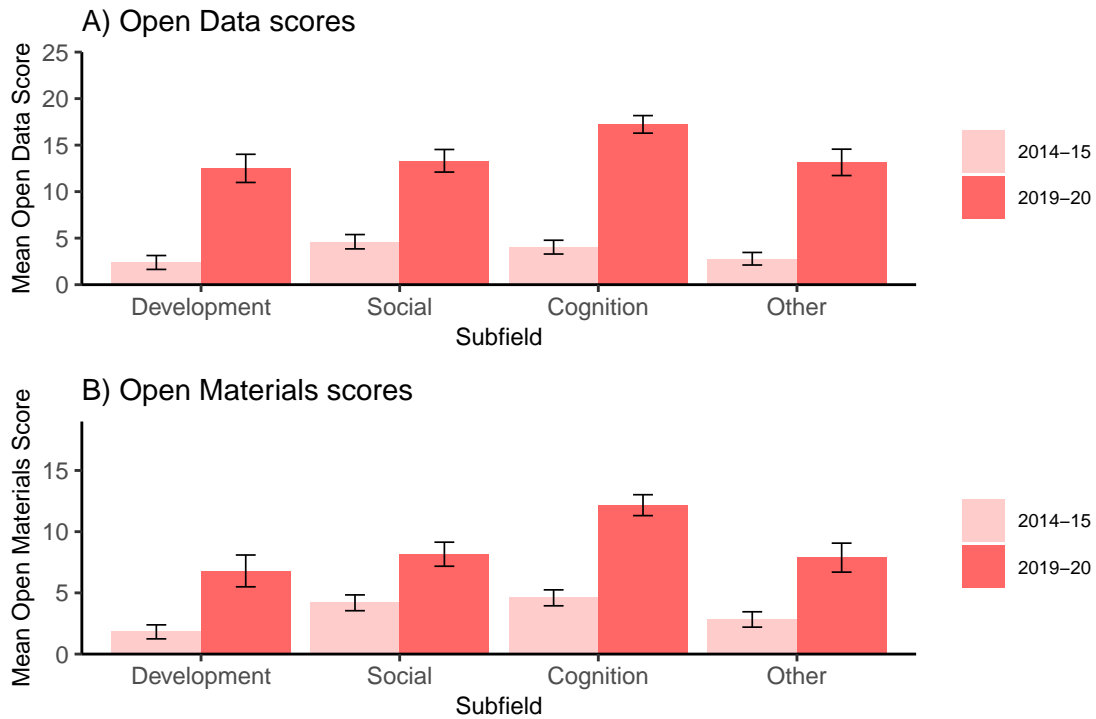


Figure 4. Mean open data and open materials scores for articles published in *Psychological Science* as a function of subfield and time period.

It is clear that since the introduction of Open Science Badges in 2014, papers published in *Psychological Science* have become more open over time and that most recently, developmental psychology has lagged behind some, but not all, subfields. To determine

whether the rate of improvement from 2014-2015 through 2019-2020 differed significantly by subfield, we combined the data across the two coded time periods and looked for subfield by time interactions in both open data and open materials scores. As illustrated in Figure 4, there was no evidence that the magnitude of improvement over time differed by subfield for either both open data scores  $F(3, 507) = 2.28$ ,  $MSE = 55.28$ ,  $p = .078$ ,  $\hat{\eta}_G^2 = .013$  or open materials scores  $F(3, 507) = 2.07$ ,  $MSE = 40.21$ ,  $p = .103$ ,  $\hat{\eta}_G^2 = .012$ .

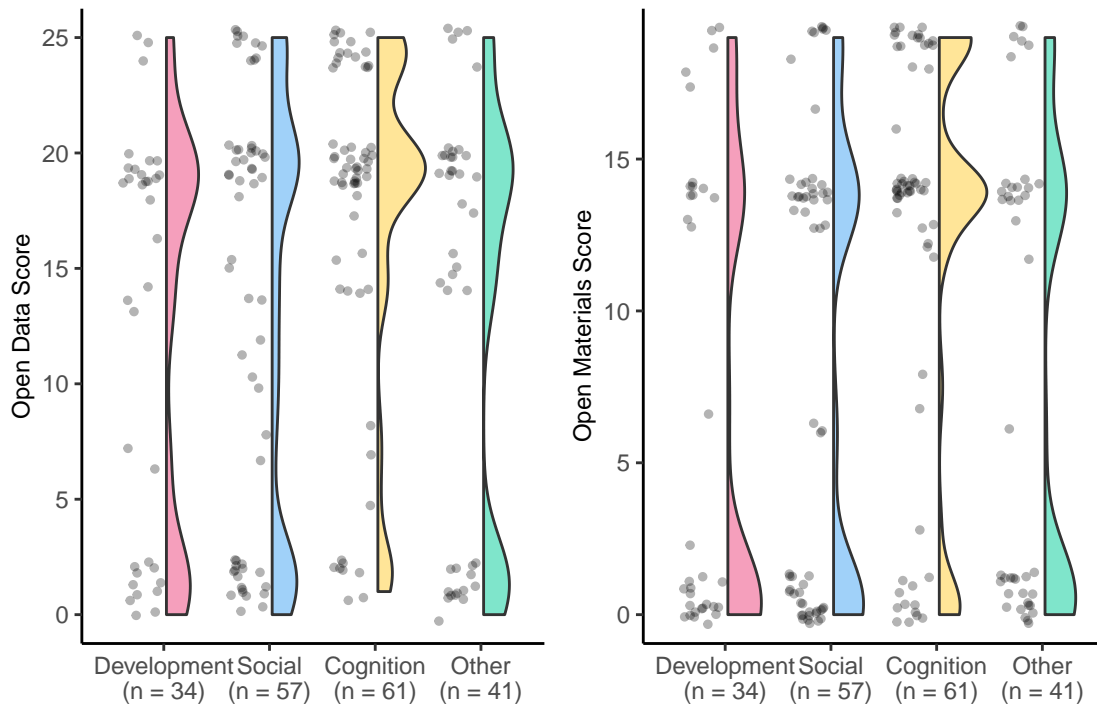


Figure 5. Distribution of open data and open materials scores earned by articles published in *Psychological Science* between 2019 and 2020 as a function of subfield

Our confirmatory analyses showed that on average, open data and materials scores for papers published in *Psychological Science* have increased markedly across all subfields, however, scores within each subfield varied widely. To explore changes in variability over time, we used raincloud plots (Allen, Poggiali, Whitaker, Marshall, & Kievit, 2019) to represent the distribution of open data and materials scores across subfields. Figure 5 illustrates that the majority of papers score on the upper half of the scale, however, there are

287 still one third of papers published that receive scores less than 5.

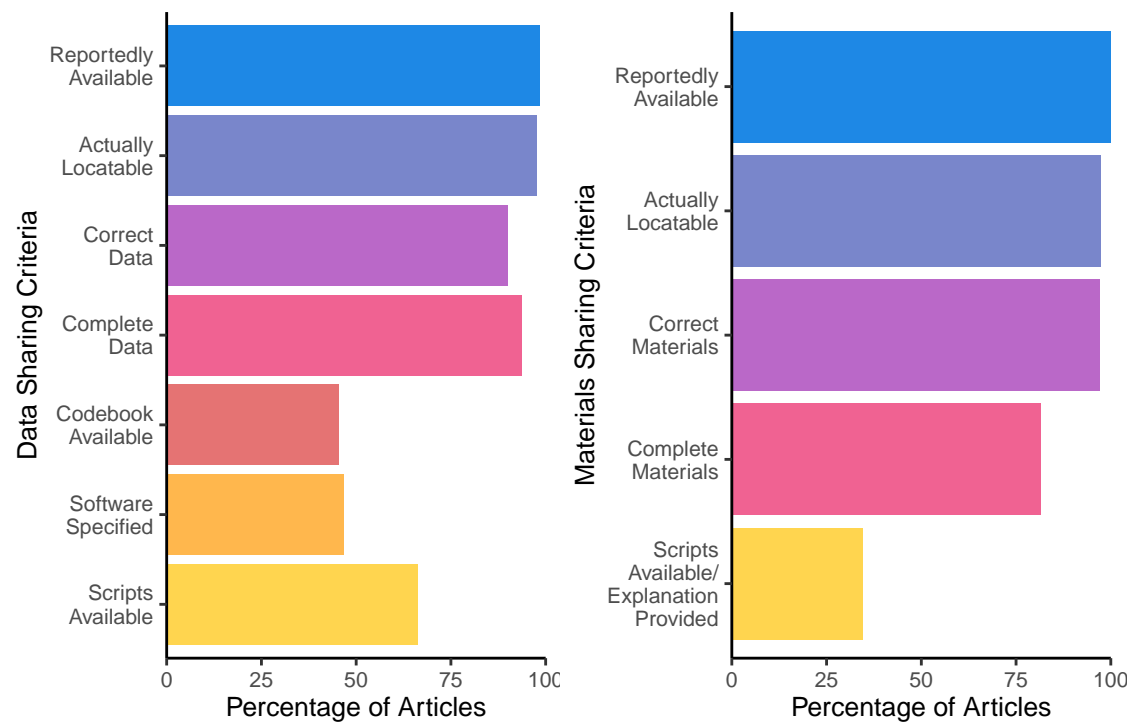


Figure 6. Proportion of articles published in *Psychological Science* in 2019-2020 that earned an Open Data Badge (left) or Open Materials Badge (right) and engaged with sharing criteria behaviours

288 We were surprised how few articles received very high open data and materials scores  
289 even in 2019-2020. In order to receive very high scores, article authors needed to engage in  
290 behaviours that make shared resources most useful (i.e. sharing data with a accompanying  
291 codebook and analysis script). We were particularly interested in how common this kind of  
292 metadata sharing was among papers that had earned an Open Data or Open Materials  
293 Badge. To produce Figure 6, we filtered articles published within the 2019-2020 window for  
294 those that were awarded open data and materials badges and then plotted the proportion of  
295 those articles that shared codebooks and scripts along with complete data.

296 Figure 6 shows that the vast majority of papers earning an open data badge had



complete data available, however, less than half shared a codebook and only 66% included an analysis script. Similarly for open materials, most articles earning a badge shared raw materials on an open repository, but a relatively small percentage of articles also shared an script and/or detailed explanation of how to use the materials in a replication study.

## Discussion

In the past few years, there has been concern from some academics that developmental psychology was lagging behind in its use of open science practices, compared to other psychological subfields. The results of the current study provide empirical support for this idea. Our analysis showed that since the introduction of Open Science Badges at *Psychological Science* in 2014, open science practices have generally improved, however, developmental psychology articles coded most recently had significantly lower open data and open materials scores relative to cognition articles.

There are several factors that may be contributing to lower open data and open materials scores in developmental psychology. Notably, practicing open science may pose a greater reputational risk to developmental scientists compared to researchers from other subdisciplines (Gilmore, Cole, Verma, Van Aken, & Worthman, 2020). Participants in developmental research are temperamental and unpredictable, which makes it difficult for researchers to stick to strict experimental protocols (Peterson, 2016). For example, if a child is getting fussy, the experimenter may deviate from the experimental protocol and allow the parent to complete the paradigm with them (Slaughter & Suddendorf, 2007). These “off-protocol” decisions make protocols difficult to replicate and add noise to experimental data (Peterson, 2016). Researchers may be reluctant to share data and materials online, out of fear that they will be criticised for a lack of scientific rigor, and that their reputation may be harmed (Gilmore, Cole, Verma, Van Aken, & Worthman, 2020). It is possible that the perceived risks of data and material sharing in developmental psychology may impact openness and transparency

The scarcity of data in developmental psychology may further impede data sharing. Developmental scientists usually recruit their participants from off-campus locations (Peterson, 2016) making recruitment a time consuming and expensive process and sample sizes generally small (Davis-Kean & Ellis, 2019). In contrast, cognition researchers are typically able to recruit large samples of participants on campus or from online platforms (Benjamin, 2019). According to the law of supply and demand, rare commodities are more highly valued (Steuart, 1767). Given that willingness to share decreases as the value of an item increases (Hellwig, Morhart, Girardin, & Hauser, 2015) it is possible that developmental researchers are less likely to share data simply because it is more highly valued than data in other subfields.

Finally, the methods that developmental psychologists use may make it particularly difficult to share materials openly. As Peterson (2016) reports, in developmental studies, experimental stimuli are typically constructed by hand and are set up manually by research assistants. The physical nature of these experimental paradigms may make them more difficult, and sometimes impossible, to share online. In contrast, computer-based experimental paradigms are becoming increasingly popular in cognition and social psychology. These paradigms, which can be automated and run online, make it relatively easy to upload materials to online repositories (Paxton & Tullett, 2019). Subfield differences in the types of materials researchers employ may explain why developmental psychologists are less likely to share materials than researchers in other subfields.

Although developmental psychology appears to be lagging behind other subfields, there is cause for optimism. Open data and materials scores for developmental psychology articles published in *Psychological Science* improved from 2014 to 2020 at the same rate as articles in other subfields. It seems that developmental psychology researchers, at least those who are looking to publish in *Psychological Science*, are keeping up with their colleagues and becoming more and more likely to adopt open data and open materials into their research

workflow.

It is clear that open data and materials practices are becoming more common, however, the current findings highlight the significant progress that has yet to be made in the open science movement across the field of psychology. We were surprised to see that in 2019-2020 a large proportion of articles received extremely low scores open data and open materials scores. In addition, very few articles were awarded the highest possible open data and open materials score, indicating that even when data and materials were shared, they were often not accompanied by a codebook, analysis script and/or explanation of the materials. Roche et al. (2015) suggest that without these metadata, open data and open materials may be not be usable, both for the purpose of reproducing the findings of a particular study and conducting novel research. Like all open science incentives, Open Science Badges are not an end to themselves; although the aim is to increase the transparency of research methods, the ultimate goal is to improve the replicability. Whilst Open Science Badges appear to incentivise researchers to share their data and materials, if they do not increase the availability of metadata, then their value in overcoming the replication crisis, remains debatable.

Our results also raise concerns about how well Open Science Badges criteria are adhered to, in practice. According to the COS, Open Data Badges can only be awarded if a ‘data dictionary’ such as a codebook, or other related metadata is made available (Center for Open Science, 2013a). Similarly, for articles to be awarded an Open Materials Badge, the authors must provide a sufficiently detailed explanation of how the materials were used in the study, and how they can be reproduced, if they can’t be shared digitally (Center for Open Science, 2013b). We found that only 45% of the articles that were awarded an Open Data Badge in 2019-2020, shared a codebook, and only 35% of those awarded an Open Materials Badge provided an explanation of their materials. These results not only suggest that a very small proportion of the articles that received an Open Data and/or Open

Materials Badge were truly deserving of one, but they also show that the criteria for Open Science Badges may be applied inconsistently. Further research is required to identify whether this issue is specific to *Psychological Science*, or if it is a broader issue observed across all journals that award Open Science Badges. In any case, the potentially inconsistent application of the criteria for Open Science Badges questions how valid and reliable they are as indicators of transparency and usability.

Where to from here? Open Science Badges have clearly had an impact on author behaviour at *Psychological Science* but that there is still work to do in making psychology a truly open science. *Psychological Science* was ideally suited for our subfield comparison due to its broad publishing scope. However, because *Psychological Science* implemented Open Science Badges, and is one of psychology's top tier journals, publishing only a very small subset of high quality and novel research articles, it is unclear whether the results from the current study reflect the field of psychology as whole. Future meta-research should focus on open science practices across a broader range of psychology journals to assess whether it is the badges per se, or a broader shift in research workflow that has resulted in improved transparency at *Psychological Science*.

Although Open Science Badges may encourage authors to be more transparent in their research, it is possible that they are rewarding researchers for doing the bare minimum, and not actually pushing the field toward a more replicable science. Perhaps journals should consider employing an open science scoring system, instead. Such a system (see (Hartshorne & Schachner, 2012; Yang, Youyou, & Uzzi, 2020) for related examples) would involve psychology journals awarding each article they publish a "Reproducibility Score" that indexes the likelihood of the findings being successfully reproduced based on the transparency of the data and materials. To maximise objectivity and to minimise time costs, an automated algorithm would generate the Reproducibility Score (Altmejd et al., 2019; Yang, Youyou, & Uzzi, 2020). Future research should test whether, compared to Open Science Badges, scores

may be a more precise and meaningful indicator of transparency and potential replicability.

To conclude, the present study provides support for the existence of subfield differences in the uptake of open science practices, across the field of psychology. Whilst the findings indicated that researchers' use of open science practices have increased since *Psychological Science* introduced Open Science Badges in 2014, there appears to be considerable progress yet to be made. Although Open Science Badges do not appear to be as valuable in overcoming the replication crisis as they seem, an open science scoring system may provide a promising alternative. Overall, we hope that the results of the study enhance the way open science is endorsed and applied across psychological subfields.

## References

- Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2019). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, 4. <https://doi.org/10.12688/wellcomeopenres.15191.1>
- Altmejd, A., Dreber, A., Forsell, E., Huber, J., Imai, T., Johannesson, M., . . . Camerer, C. (2019). Predicting the replicability of social science lab experiments. *PloS One*, 14(12), e0225826. <https://doi.org/10.1371/journal.pone.0225826>
- 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>
- Benjamin, A. S. (2019). Editorial. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(2). <https://doi.org/10.1037/xlm0000688>
- Carroll, J., Schep, A., & Sidi, J. (2021). *Ggeasy: Easy access to 'ggplot2' commands*. Retrieved from <https://CRAN.R-project.org/package=ggeasy>
- Center for Open Science. (2013a). Open data badge criteria. Retrieved from <https://osf.io/g6u5k/>
- Center for Open Science. (2013b). Open materials badge criteria. Retrieved from <https://osf.io/gc2g8/>

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284. <https://doi.org/10.1037/1040-3590.6.4.284>

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284. <https://doi.org/10.1037/1040-3590.6.4.284>

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

Davis-Kean, P. E., & Ellis, A. (2019). An overview of issues in infant and developmental research for the creation of robust and replicable science. *Infant Behavior and Development*, 57, 101339. <https://doi.org/10.1016/j.infbeh.2019.101339>

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

Fleiss, J. L. (1981). Balanced incomplete block designs for inter-rater reliability studies. *Applied Psychological Measurement*, 5(1), 105–112. <https://doi.org/10.1177/014662168100500115>

Fleiss, J. L. (1981). Balanced incomplete block designs for inter-rater reliability studies. *Applied Psychological Measurement*, 5(1), 105–112. <https://doi.org/10.1177/014662168100500115>

Frank, M. [@mcxfrank]. (2020, March 6). At the same time, this policy statement is weaker than it should be! Openness does not just cause harm. It also reduces harm - often dramatically [tweet]. Retrieved from <https://twitter.com/mcxfrank/status/1103068416791855104>

Gamer, M., Lemon, J., & Singh, I. F. P. (2019). *Irr: Various coefficients of interrater reliability and agreement*. Retrieved from <https://CRAN.R-project.org/package=irr>

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. <https://doi.org/10.1111/cdep.12356>

Gilmore, R. O., Cole, P. M., Verma, S., Van Aken, M. A., & Worthman, C. M. (2020). Advancing scientific integrity, transparency, and openness in child development research: Challenges and possible solutions. *Child Development Perspectives*, 14(1), 9–14. <https://doi.org/10.1111/cdep.12360>

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*.

Hartshorne, J., & Schachner, A. (2012). Tracking replicability as a method of post-publication open evaluation. *Frontiers in Computational Neuroscience*, 6, 8. <https://doi.org/10.3389/fncom.2012.00008>

Hellwig, K., Morhart, F., Girardin, F., & Hauser, M. (2015). Exploring different types of sharing: A proposed segmentation of the market for “sharing” businesses. *Psychology & Marketing*, 32(9), 891–906. <https://doi.org/10.1002/mar.20825>

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. <https://doi.org/10.1371/journal.pbio.1002456>

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). <https://doi.org/10.1525/collabra.158>

Makowski, D., Ben-Shachar, M. S., Patil, I., & Lüdtke, 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>

National Academies of Sciences, Engineering, Medicine, & others. (2019).  
*Reproducibility and replicability in science*. National Academies Press.  
<https://doi.org/10.17226/25303>

Open Science Collaboration. (2015). Estimating the reproducibility of psychological  
science. *Science*, 349(6251). <https://doi.org/10.1126/science.aac4716>

Paxton, A., & Tullett, A. (2019). Open science in data-intensive psychology and  
cognitive science. *Policy Insights from the Behavioral and Brain Sciences*, 6(1),  
47–55. <https://doi.org/10.1177/2372732218790283>

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

Peterson, D. (2016). The baby factory: Difficult research objects, disciplinary

standards, and the production of statistical significance. *Socius*, 2,  
2378023115625071. <https://doi.org/10.1177/2378023115625071>

Pfeifer, J. [@jennDSN]. (2020, March 8). Reflecting on my lukewarm reaction – agree  
it seemed to undervalue openness, as nice but not full optional, bc it’s risky and  
hard [tweet]. Retrieved from  
<https://twitter.com/jennDSN/status/1103891773909168128>

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/>

Roche, D. G., Kruuk, L. E., Lanfear, R., & Binning, S. A. (2015). Public data  
archiving in ecology and evolution: How well are we doing? *PLoS Biology*, 13(11),  
e1002295. <https://doi.org/10.1371/journal.pbio.1002295>

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>

Slaughter, V., & Suddendorf, T. (2007). Participant loss due to “fussiness” in infant  
visual paradigms: A review of the last 20 years. *Infant Behavior and Development*,  
30(3), 505–514. <https://doi.org/10.1016/j.infbeh.2006.12.006>

Steuart, J. (1767). *An inquiry into the principles of political economy* (Vol. 2). Oliver  
& Boyd.

Syed, M. (2021). Infant and child development: A journal for open, transparent, and  
inclusive science from prenatal through emerging adulthood. *Infant and Child*  
*Development*, 30(1). <https://doi.org/10.1002/icd.2215>

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>

Wickham, H., & Seidel, D. (2020). *Scales: Scale functions for visualization*. Retrieved

from <https://CRAN.R-project.org/package=scales>

Yang, Y., Youyou, W., & Uzzi, B. (2020). Estimating the deep replicability of

scientific findings using human and artificial intelligence. *Proceedings of the  
National Academy of Sciences*, 117(20), 10762–10768.

<https://doi.org/10.1073/pnas.1909046117>

- 548       Zhu, H. (2021). *kableExtra: Construct complex table with 'kable' and pipe syntax*.  
549       Retrieved from <https://CRAN.R-project.org/package=kableExtra>