

Novel tools for open science

Sheila K. Keener
Strome College of Business
Old Dominion University

November 14, 2025

Tools to Promote Accuracy, Transparency, and Automation

statcheck

✓ “Spellchecker for stats”

- Introduced in 2015 by Nuijten and colleagues.
- Determines whether reported p -values match the accompanying test statistics and dfs.
- Requires results to be in APA format (t -tests, F -tests, correlations, z -tests, Q -tests, and Chi-Square tests).
- Web-based app: <https://statcheck.io>; R package: [statcheck](#)

Upload files (pdf, html, or docx):

Browse... Sample_Paper.pdf

Upload complete

Settings:

☒ Try to identify and correct for one-tailed tests

Generating the report may take a few moments. [Download report](#)

Copy Print Download table ▾

Search:

File	Statistical reference	Computed p-value	Consistency
1 Sample_Paper.pdf	F(2, 96) = 5.84, p = .005	0.00404	Inconsistency
2 Sample_Paper.pdf	F(2, 96) = 4.95, p = .009	0.00900	Consistent
3 Sample_Paper.pdf	F(2, 96) = 2.91, p = .003	0.05930	Decision Inconsistency

```
#statcheck
install.packages("statcheck")
install.packages("readtext")
library(statcheck)
library(readtext)
setwd ("/Users/sheilakeener/Dropbox/CARMA/")
```

```
#Option 1: directly enter text
text<- "(F(2,96) = 5.84, p = .005)"
statcheck(text)
```

```
#Option 2: full text document
doc<-readtext("Sample_Paper.docx")
res<-statcheck(doc$text)
res
```

```
#Option 3: pdf document
checkPDF("Sample_Paper.pdf")
```

metacheck (formerly papercheck)

✓ Manuscript screener

- Developed by the [ScienceVerse / MetaCheck](#) team
- [Modules](#) that look for:
 - Missing effect sizes
 - Imprecise p -values (e.g., "<.05" or "ns")
 - References to papers in the Retraction Watch database
 - p -values described as "marginally significant"
 - Inconsistent p -values (using statcheck)
- Other modules are in development

```
install.packages("devtools")
devtools::install_github("scienceverse/metacheck")
library(metacheck)
install.packages(c("shinydashboard", "shinyjs", "shiny.i18n", "DT", "waiter"))
```

```
pdf_file<-file.choose()
xml_file<-pdf2grobid(pdf_file)
paper<-read(xml_file)
file.copy(xml_file, "/Users/sheilakeener/Dropbox/CARMA/Sample_Paper.xml")
```

```
#opens shiny app
metacheck_app()
```

🔴 Missing Effect Sizes

No effect sizes were detected for any t-tests or F-tests. The Journal Article Reporting Standards state effect sizes should be reported.

text	section	div	p	s	test_text	test
We found that $F(2, 96) = 5.84, p = .005$.	results	10	1	1	$F(2, 96) = 5.84$	F-test

🔴 Exact P-Values

You may have reported some imprecise p-values

text	p_comp	p_value	section	div	p	s
$p < .1$	<	0.10	intro	2	13	3

🟢 RetractionWatch

You cited no papers in the Retraction Watch database (as of 2025-05-20)

🟢 Marginal Significance

No effects were described as marginally/borderline/close to significant.

🔴 StatCheck

We detected possible errors in test statistics

LLMs (e.g., ChatGPT, Gemini)



Assist with Coding, Annotating, & Troubleshooting

- Explain what code does, step-by-step, to help you:
 - More clearly annotate code to share.
 - Adapt it correctly.
- Troubleshoot errors in code.
- Spot obvious errors in code.
- Translate Code (R <-> Python code)
- **Still requires human verification - LLMs make mistakes.**

Sample Prompt: Assess the clarity and reproducibility of this code. Identify anything confusing, ambiguous, or not well-documented. Suggest improvements to comments, structure, naming, and workflow explanation to make the code easier for someone else to rerun.



Support Transparency & Reporting Standards

- Evaluate manuscripts against the Journal Article Reporting Standards (JARS), Meta-analytic Reporting Standards (MARS), APA format, etc.
- Compare manuscript to pre-registration.
- Identify missing reporting elements or unclear descriptions in documentation/methods.
- **Adjust privacy settings before uploading manuscripts; enterprise/business versions provide more protection.**

Sample Prompt: Compare this manuscript to the APA Journal Article Reporting Standards (JARS-Qual). Identify which required reporting elements are present, which are incomplete, and which are missing. Provide the results in a clear table with three columns: Required Item, Status (Present / Incomplete / Missing), and Suggested Fix. Then give me a concise narrative summary of the biggest reporting gaps.

Other useful tools

✓ Granularity-related inconsistency means (GRIM) Test

- Developed by [Brown & Heathers \(2017\)](#).
- Examines whether reported means are mathematically possible given the number of Likert-type items and sample size.

N	0.64	rounded	mean	items:	1
1	0.64	1.00	1.00		
2	1.28	1.00	0.50	Instructions:	
3	1.92	2.00	0.67	1. Enter the number of Likert-type items in the scale (typically 1) in cell F1	
4	2.56	3.00	0.75	2. Enter the fractional part of the mean to test in cell B1	
5	3.20	3.00	0.60	3. Look for your sample size in Column A	
				4. The corresponding number in Column D is the correctly reported mean.	
6	3.84	4.00	0.67	Green=consistent, white=inconsistent	
7	4.48	4.00	0.57		
8	5.12	5.00	0.63		
9	5.76	6.00	0.67		
10	6.40	6.00	0.60		
11	7.04	7.00	0.64		
12	7.68	8.00	0.67		

✓ Degrees of freedom calculator

- Developed by Green et al., and cited in [Cortina et al. \(2017\)](#).
- [Shiny app](#) that lets you calculate the correct dfs for SEM/CFA and path models.

"We tested a three-factor CFA (12 items). The model had acceptable fit to the data, $\chi^2(48)=78.32, p=.011$, CFI = .99, TLI = .98, RMSEA = .045 (90% CI [.022, .065]), SRMR = .032"

Output

Knowns = 78

Unknowns = 27

Degrees of Freedom = 51

Feel free to contact me with questions later: skeener@odu.edu