Errors and biases in Structural Equation Modeling
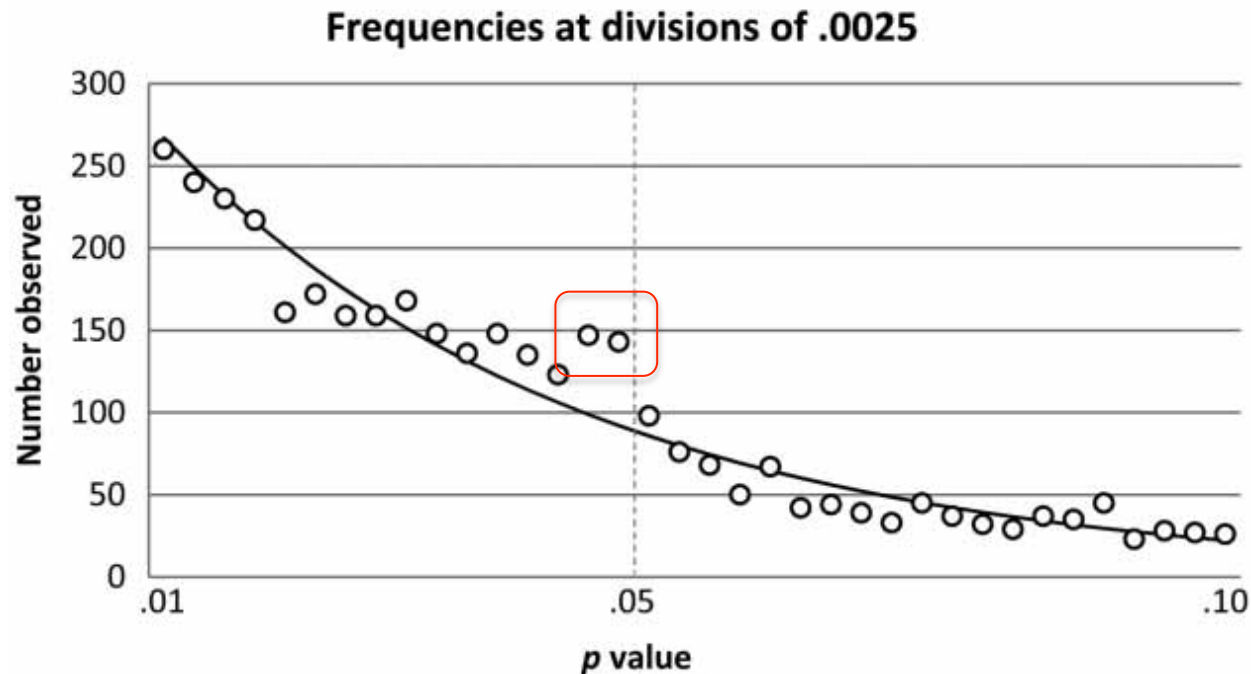
Jelte M. Wicherts

TILBURG ◆ UNIVERSITY

De empirical cycle

Grey zone map

Remove outliers (Z > |2|)

P>.05 → Call this a "failed study"

P>.05

Add 10 cases

P<.05 → Effect!

Perform new study

P<.05

Redo analysis with adapted dependent var.

P>.05

P<.05

Planned analysis

P>.05

P<.05

Write paper

Bakker, M., van Dijk, A, & Wicherts, J. M. (2012). The Rules of the Game Called Psychological Science. *Perspectives on Psychological Science, 7,* 543-554.
Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology. *Psychological Science, 22*, 1359 –1366.

3

The "bump" below p=.05

Frequencies at divisions of .0025

Based on 3,627 *p* values from 2008 issues of *Journal of Experimental Psychology: General*, *Journal of Personality and Social Psychology*, and *Psychological Science.*

Source: Masicampo, E. J. & Lalande, D. R. (2012). A peculiar prevalence of p values just below .05. *Quarterly Journal of Experimental Psychology, 65,* 2271-2279.

Willingness to share research data is related to the strength of the evidence

Data shared? ■ No ■ Yes
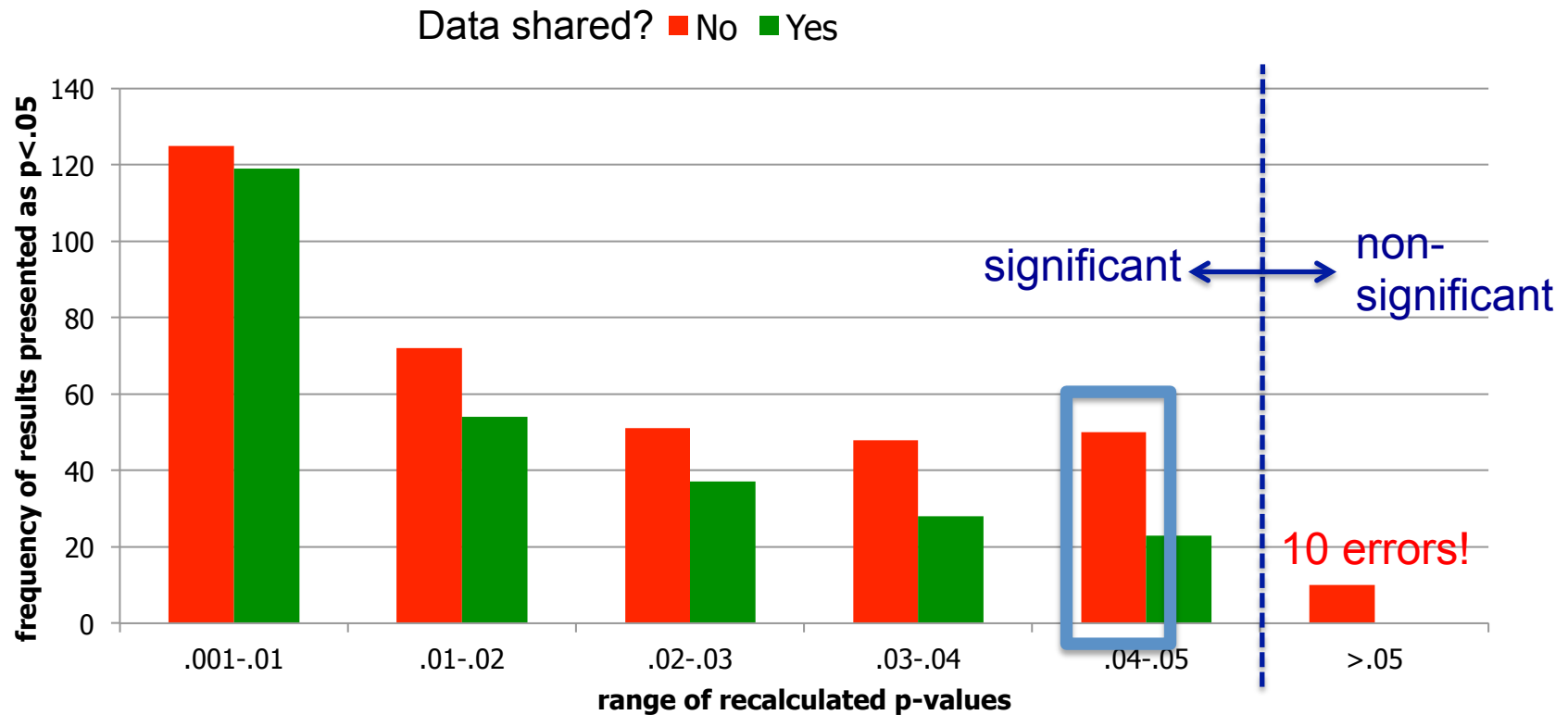
Source: Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. PLoS ONE, 6, e 26828.
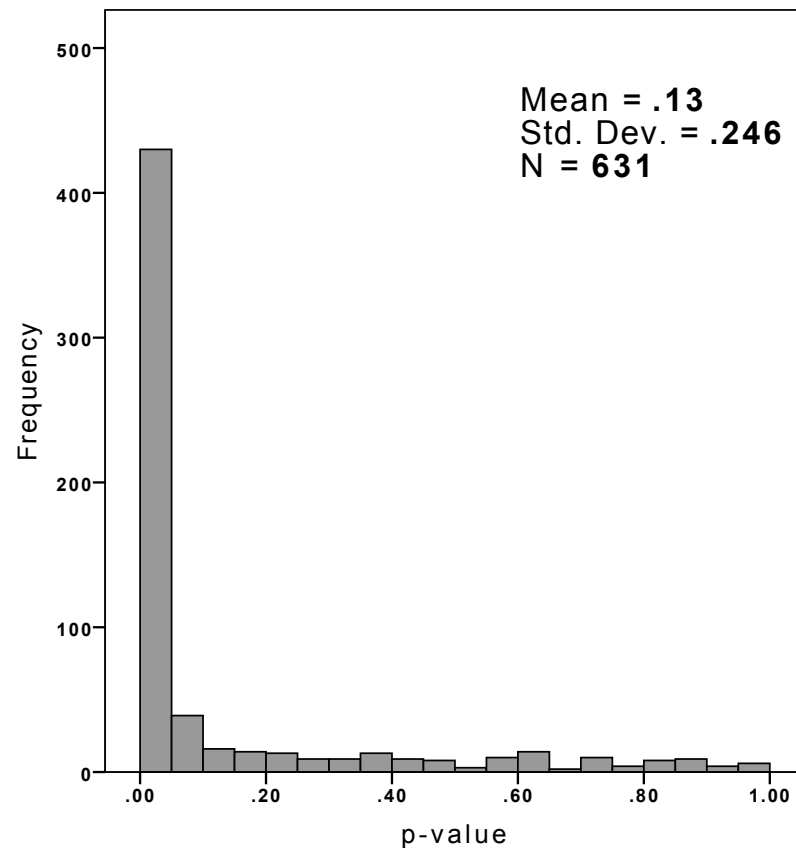
5

SEM: fitting models

Now the goal is to *fit* a model.

Via the fit function, the hypothesized model
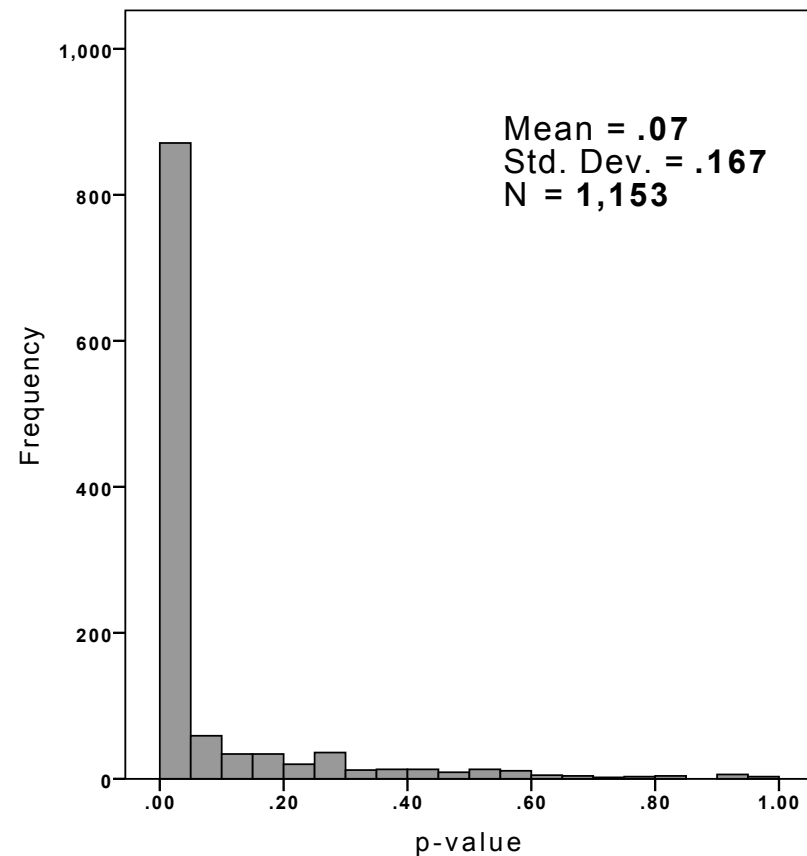(incl. distributional assumptions)
provides a chi-square test.
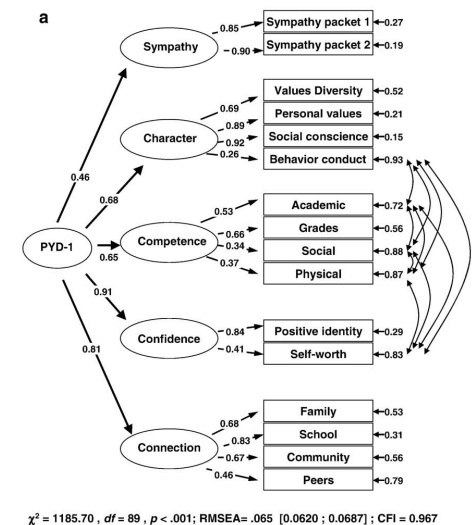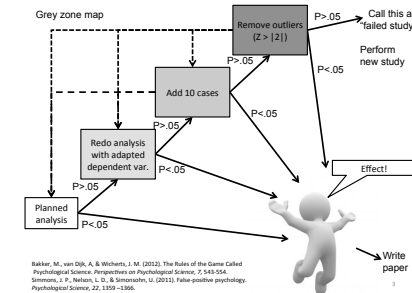
Results from analyses aimed at either p<.05 (NHST) or p>.05 (SEmodel fit)
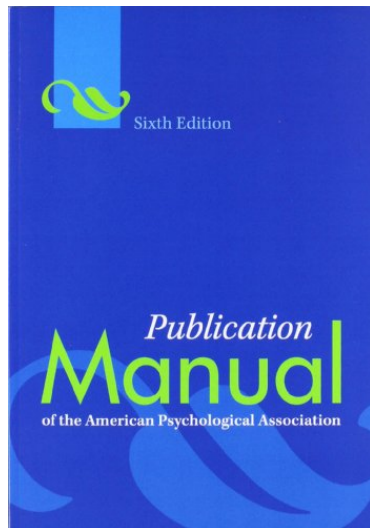
Reporting SEMs & model fit

N=206
X^2 (DF=125) = 417.4, p<.001
RMSEA = .103 (90%CI: [.092,.115])
SRMR = .066
CFI = .948, NNFI = .928, GFI = .830

Table 2
Fit Indices for Nested Sequence of Cross-Sectional Models

Model	χ^2	NFI	PFI	χ^2_{diff}	ΔNFI
1. Mobley's (1977) measurement model	443.18*	.92	.67		
2. Quit & search intentions	529.80*	.89	.69		
Difference between Model 2 and Model 1				86.61*	.03
3. Search intentions & thoughts of quitting	519.75*	.90	.69		
Difference between Model 3 and Model 1				76.57*	.02
4. Intentions to quit & thoughts of quitting	546.97*	.89	.69		
Difference between Model 4 and Model 1				103.78*	.03
5. One withdrawal cognition	616.97*	.87	.70		
Difference between Model 5 and Model 1				173.79*	.05
6. Hom, Griffeth, & Sellaro's (1984) structural model	754.37*	.84	.71		
Difference between Model 6 and Model 5				137.39*	.03
7. Structural null model	2,741.49*	.23	.27		
Difference between Model 7 and Model 6				1,987.13*	.61
8. Null model	3,849.07*				

Note. NFI = normed fit index; PFI = parsimonious fit index.
* $p < .05$.

Source: Hom & Griffeth, 1991, Journal of Applied Psychology

Reporting

We drew a random sample of 242 articles that referred to AMOS, LISREL, or M-PLUS manuals and selected only those that used one of these packages to fit SEM/CFA models on data.

Of the 1286 models....
- 322 (25%) were reported with correlation matrix
- 322 (25%) included a full path model.
- 1159 (90.1%) reported the DF
- 1165 (90.6%) reported the Chi-square
- 1024 (79.6%) reported RMSEA
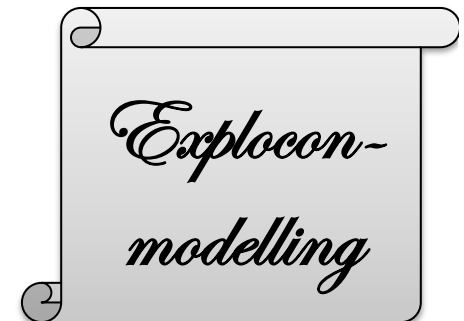- 936 (72.8%) reported CFI

Example (N=16,030)
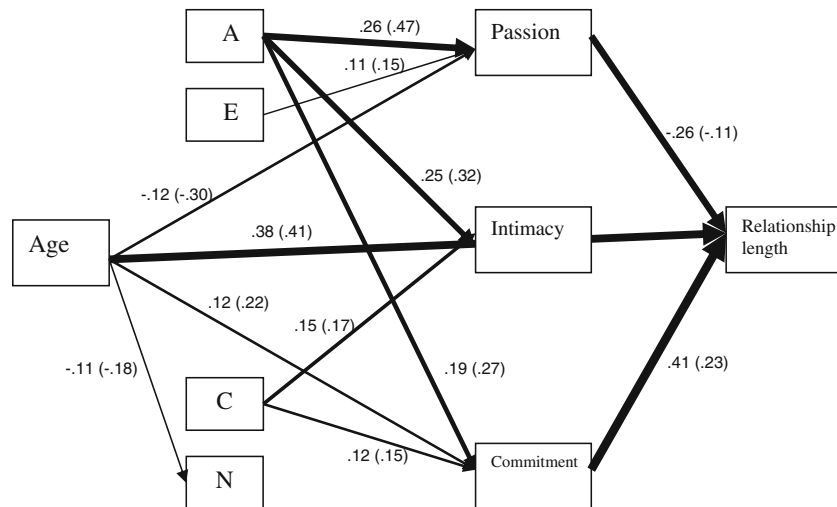
The saturated model, where only a variable directly to the left of another was allowed to influence it, did not fit the data well: $\chi^2 = (10\,df, p < .01)\,4230.2$, GFI $= .95$, AGFI $= .77$, PGFI $= .21$, RMSEA $= .16$, AIC $= 4300.2$.

Table 2 Bivariate correlations between the Big Five factors, love dimensions, age, and relationship length

Variable	1	2	3	4	5	6	7	8	9	10
1. Age	–	−.18**	.46**	−.09**	−.12**	.13**	−.05**	−.02**	.08**	−.11**
2. Sex		–	−.07**	.05**	−.03**	−.02*	.10**	.22**	.06**	.21**
3. RL			–	.08**	−.13**	.38**	−.05**	.01	.09**	−.01
4. Intimacy				–	.54**	.56**	.14**	.30**	.20**	.02**
5. Passion					–	.40**	.18**	.29**	.11**	−.05**
6. Commitment						–	.01	.21**	.17**	−.03*
7. E							–	.20**	.02**	−.18**
8. A								–	.20**	.01
9. C									–	.00
10. N										–

N neuroticism, *C* conscientiousness, *A* agreeableness, *E* extraversion, *RL* relationship length

* $p < .05$; ** $p < .01$

Explocon-modelling

Source: Ahmetoglu et al. (2010). The Relationship Between Dimensions of Love, Personality, and Relationship Length. Archives of Sexual Behavior
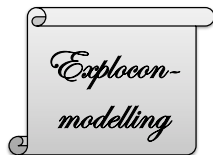
10

Table 4
Model fit statistics for cross-sectional models.

	Grade 5		Grade 6		Grade 7	
	M1	M2	M1	M2	M1	M2
χ^2	2611.22	1185.7	5330.11	1867.34	2879.09	2234.86
df	99	89	147	136	114	107
Change in χ^2		1425.52		3462.77		644.23
RMSEA	0.094	0.065	0.111	0.066	0.098	0.083
Confidence interval	(.091, .097)	(.062, .069)	(.108, .113)	(.064, .069)	(.095, .101)	(.080, .086)
CFI	0.930	0.967	0.907	0.971	0.932	0.950

Note. M1 (model 1): no correlated residuals. M2 (model 2): correlated residuals for Harter subscales.

Explocon-modelling

153 input – 49 pars = DF 104
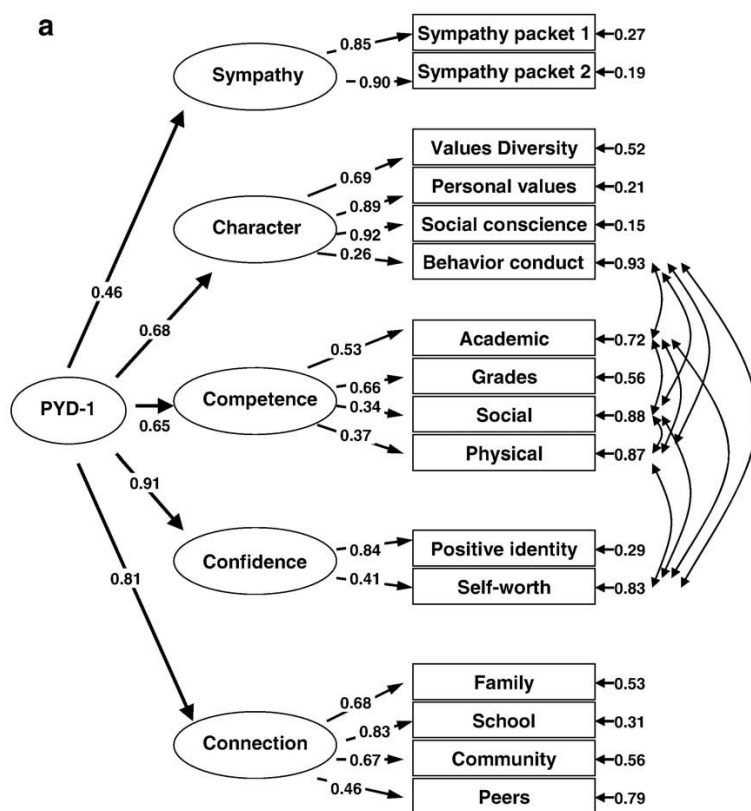
$\chi^2 = 1185.70$, *df* = 89 , *p* < .001; RMSEA= .065 [0.0620 ; 0.0687] ; CFI = 0.967
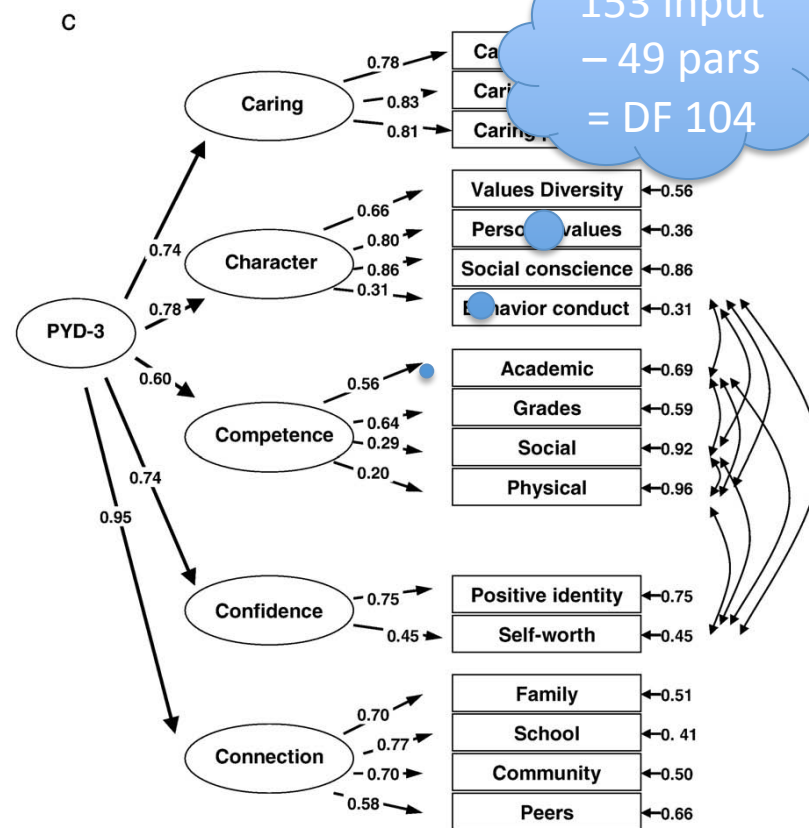
$\chi^2 = 2234.862$, *df* =107 , *p* < .001; RMSEA= .083 [0.080 ; 0.086]; CFI = 0.950

Source: Phelps et al. (2009). The structure and developmental course of Positive Youth Development (PYD) in early adolescence. Journal of Applied Developmental Psychology
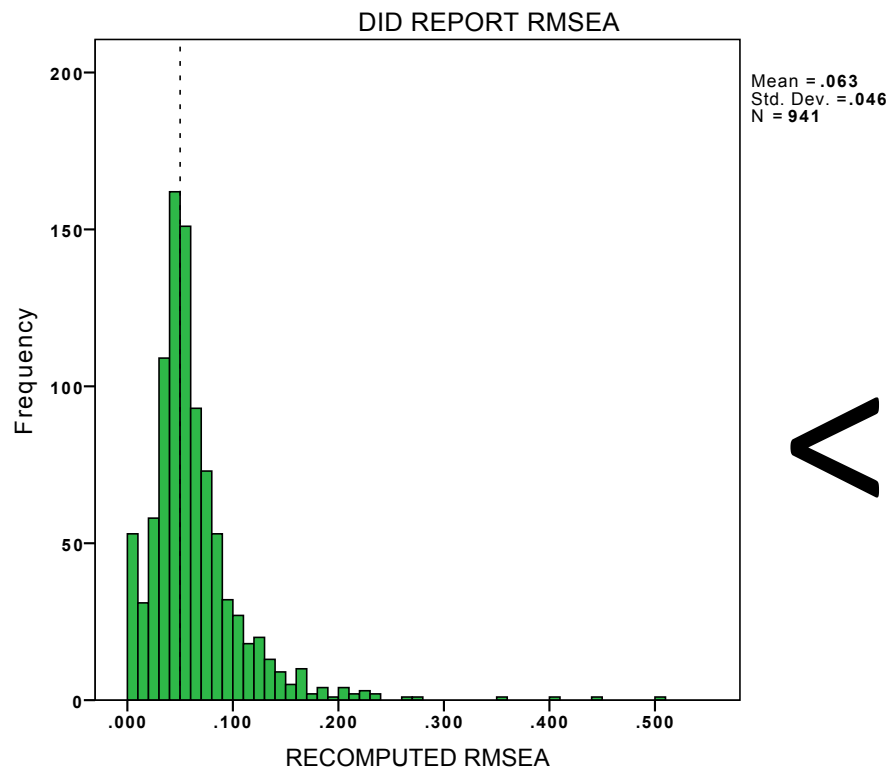
11

Are statistical results checked by (co-)authors and reviewers?

Simple effects analyses within each of the two levels of valence were conducted, revealing a significant main effect of subtype upon the proportion of positive words falsely recalled, $F(2, 65) = 3.02$, $p = .05$, $\eta_p^2 = .09$, and the proportion of negative words falsely recalled, $F(2, 64) = 4.45$, $p < .05$, $\eta_p^2 = .12$.
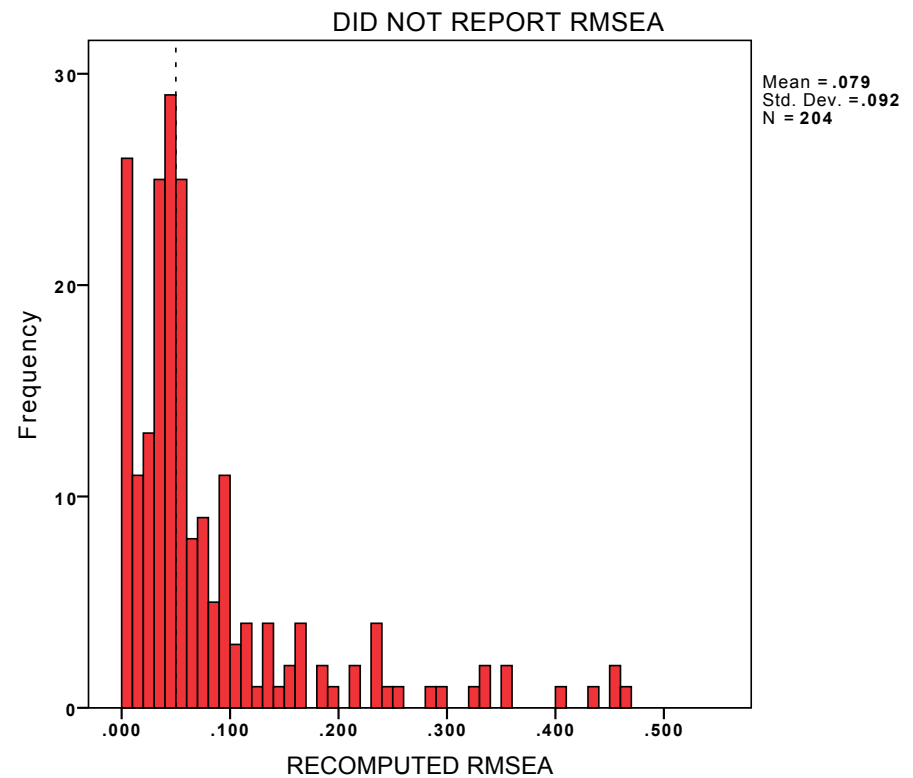
$p = .06$

Method: a representative sample of 257 papers Recomputed 4720 p-values from NHST and checked for consistency
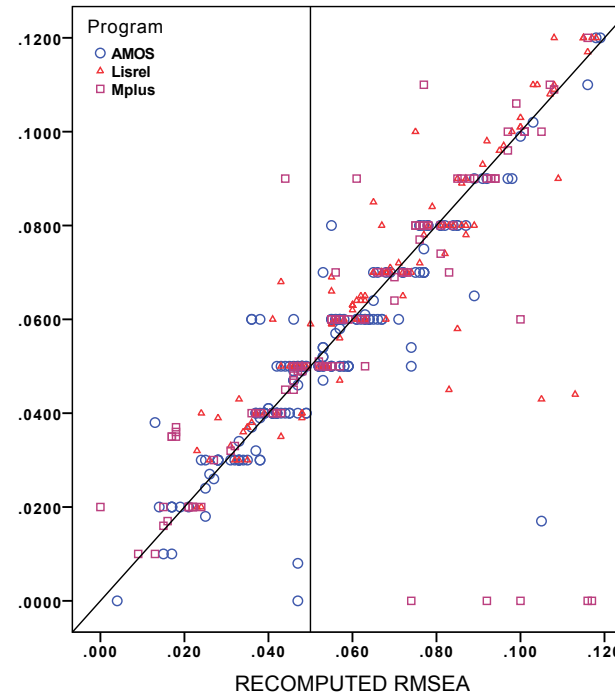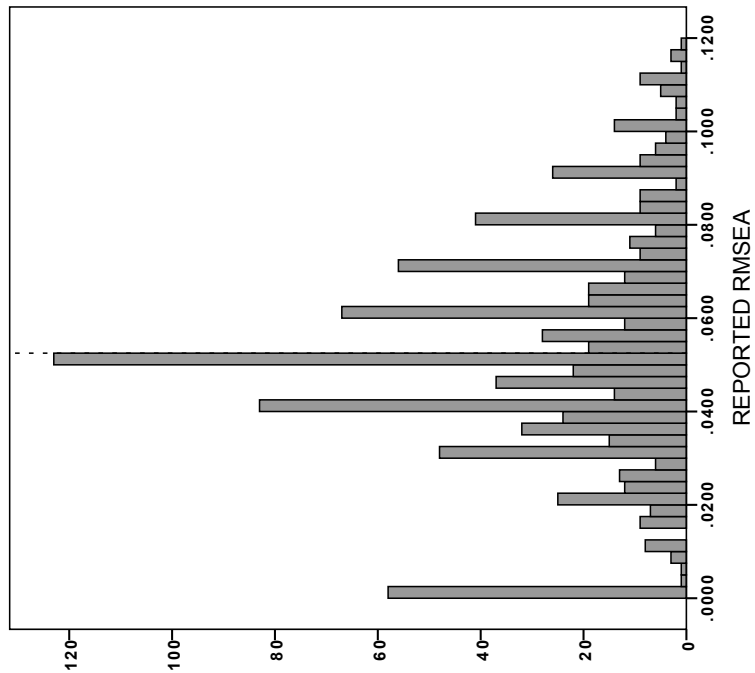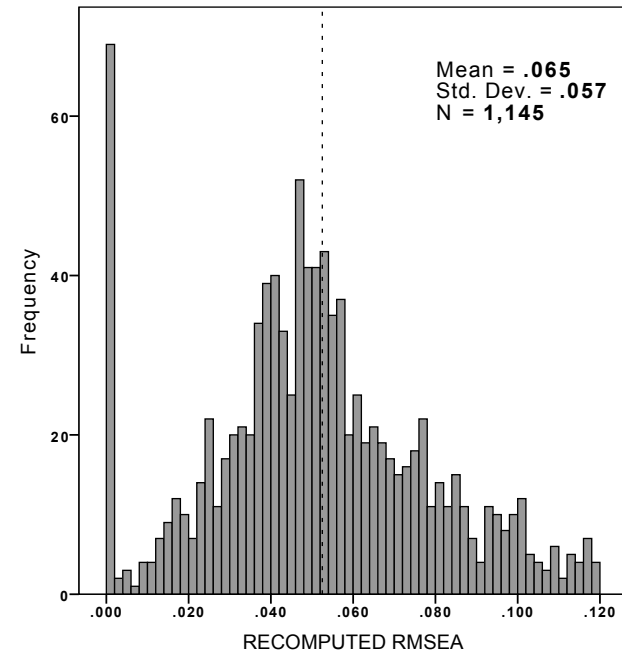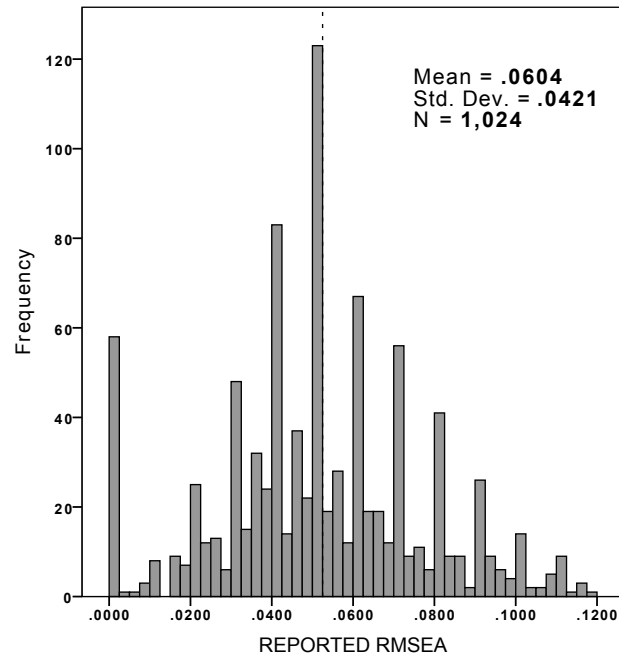
Results: 128 papers (50%) contained at least one error
39 papers (15%) contained at least one error related to $p = .05$
Conclusion: Errors predominantly led to "better" results

Source: Bakker, M. & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. Behavior Research Methods, 43, 666-678.

RMSEA: to report or not to report?

27% of 450 discrepancies were >.005

14

EXPLOCON MODELLING

Check Modification indices and adapt model

run EFA

CFI<.95
RMSEA >.05
X^2 : p <.05

CFI >.95

CFI>.95
RMSEA <.05
X^2 : p >.05

Check CFI (or any of its alternatives)

CFI>.95

RMSEA>.05

Check RMSEA and dismiss X^2 test (N = large)

"Excellent fit"!

RMSEA <.05

X^2: P<.05

Planned model

X^2: P>.05

Write paper

So many SEMers…

De empirical cycle
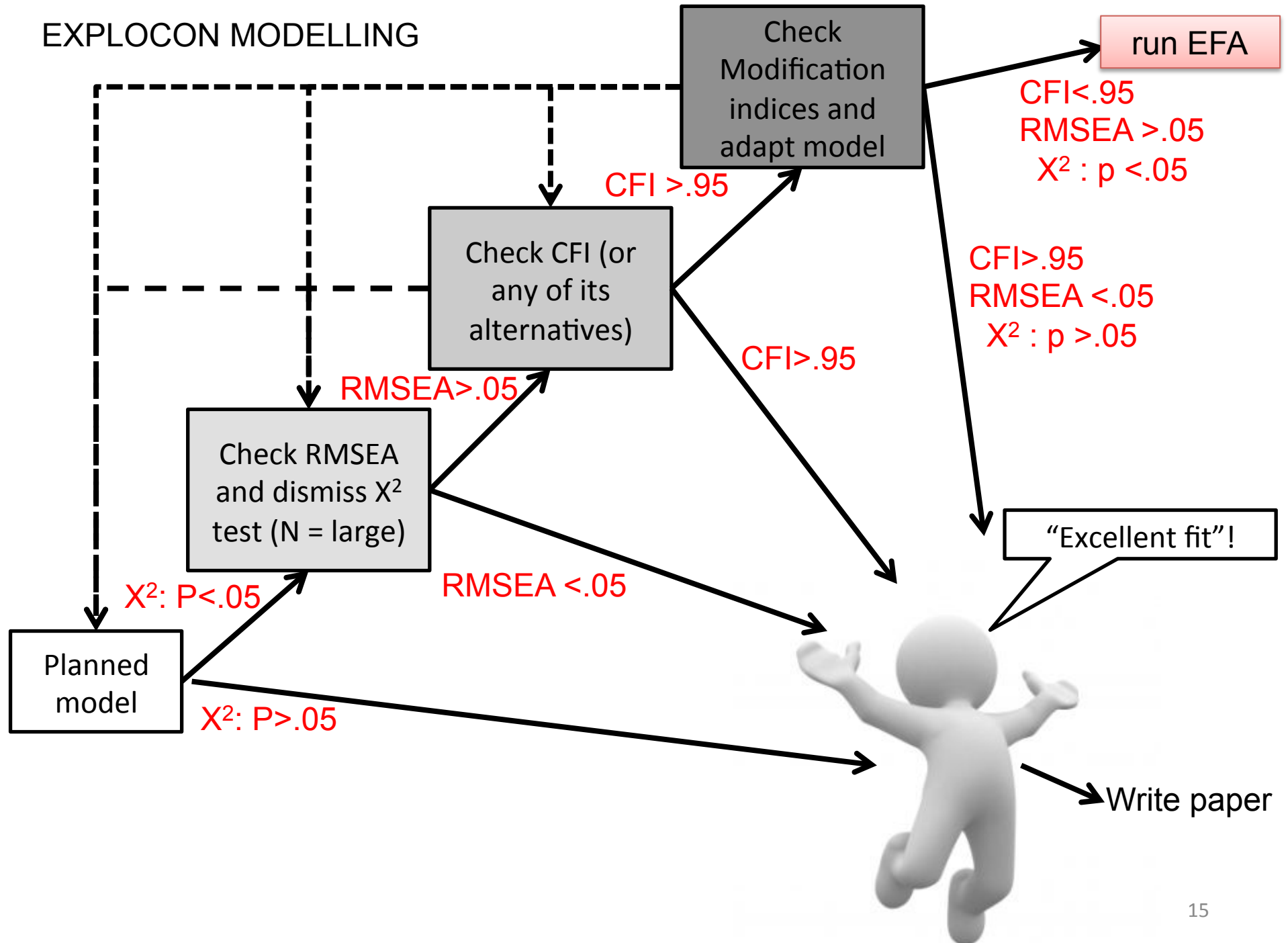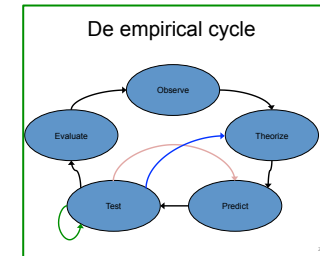
- report models that do not fit

- employ explocon modelling (adapt models, select "best" fit measures, etc.)

- do not conduct proper cross-validation

- do not report SEM results in a replicable manner

- make reporting errors (to their benefit?)

- misreport RMSEAs to reach rule-of-thumb

Researchers (and SEMers) are only human!

Testing vs. Fitting

Finding a well fitting SE model is not the same as testing the model. Yet, many SEMers typically approach it as though they are doing the latter.
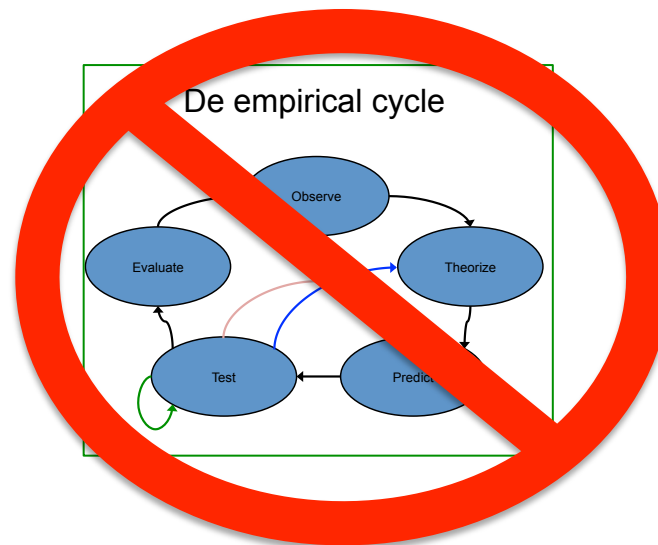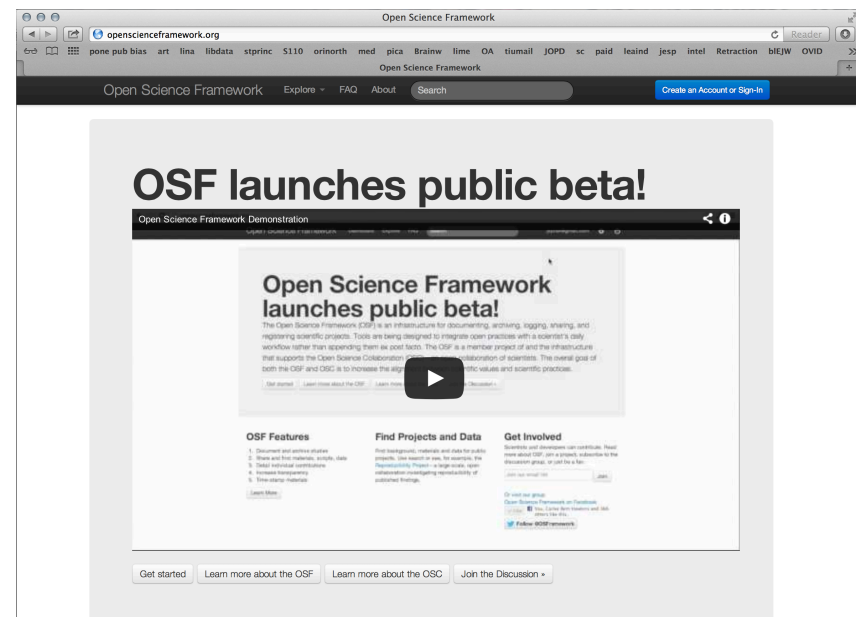
Avoiding explocon modelling

For *confirmatory* factor analyses and SEM applications that aim to "test" a certain structural model (i.e., *confirmatory* studies): pre-register the modelling approach via OSF, and/or use cross validation sample

Sources: Nosek, B. A., Spies, J., & Motyl, M. (2012). Scientific Utopia: II - Restructuring Incentives and Practices to Promote Truth Over Publishability. *Perspectives on Psychological Science, 7*, 615-631.
Wagenmakers et al. (2012). An Agenda for Purely Confirmatory Research. *Perspectives on Psychological Science, 7,* 632-638.

Avoiding explocon modelling

In *exploratory* analyses make sure that the phrasing is correct: you aim to find a model that gets you RMSEA<.05, CFI>.95, etc. And consider it a model-comparison enterprise.

Be careful of any test that is in the model (including those related to important parameters). Or go Bayesian

Avoiding errors: the copilot model

- Let your co-authors (or colleagues) replicate your analyses

- Exercise openness concerning analytic choices

- Share data & scripts with collaborators

Source: Wicherts, J. M. (2011). Psychology must learn a lesson from fraud case. *Nature, 480*, 7.

Replicability

Always enable replicability of results by "peers" by publishing covariance matrices and scripts and/or by publishing the data (e.g., via the Journl of Open Psychology Data)

http://openpsychologydata.metajnl.com

Table 2 Bivariate correlations between the Big Five factors, love dimensions, age, and relationship length

Variable	1	2	3	4	5	6	7	8	9	10
1. Age	–	−.18**	.46**	−.09**	−.12**	.13**	−.05**	−.02**	.08**	−.11**
2. Sex		–	−.07**	.05**	−.03**	−.02*	.10**	.22**	.06**	.21**
3. RL			–	.08**	−.13**	.38**	−.05**	.01	.09**	−.01
4. Intimacy				–	.54**	.56**	.14**	.30**	.20**	.02**
5. Passion					–	.40**	.18**	.29**	.11**	−.05**
6. Commitment						–	.01	.21**	.17**	−.03*
7. E							–	.20**	.02**	−.18**
8. A								–	.20**	.01
9. C									–	.00
10. N										–

N neuroticism, *C* conscientiousness, *A* agreeableness, *E* extraversion, *RL* relationship length

$* p < .05; ** p < .01$

Sources: Wicherts, J. M. (2013). Science revolves around the data [Editorial]. *Journal of Open Psychology Data* 1(2).
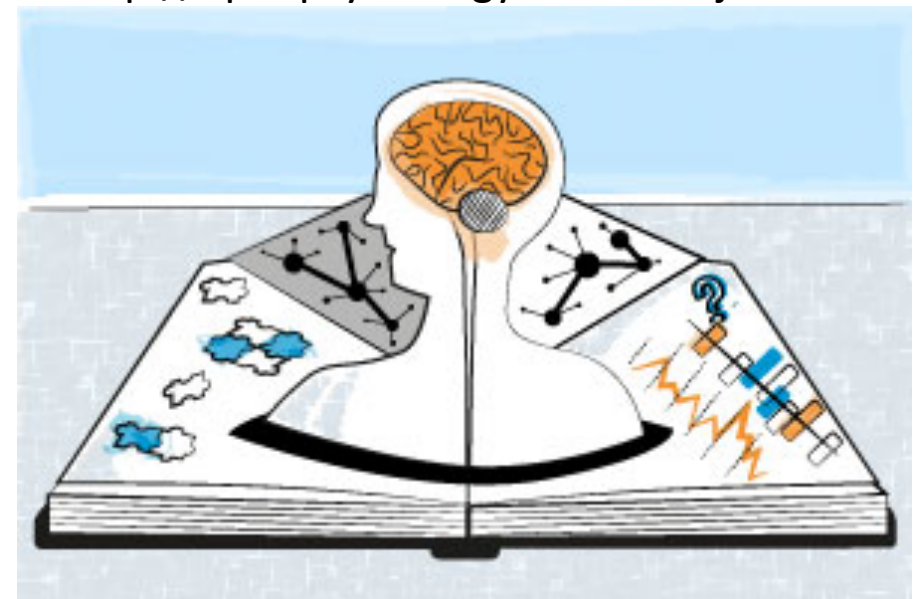
Thank you!

Netherlands Organisation for Scientific Research