UiO : Institutt for spesialpedagogikk
Det utdanningsvitenskapelige fakultet

Standarder for god forskning

Astrid Marie Jorde Sandsør

I dag

- Litt om valg av analyser
- Standarder for god forskning
- Forskningsprosessen
- God rapportering av forskning
- …med skrivetips

- Slides fra SPED 4010, (Zachrisson, 2020)

Om valg av design og analyser

- Dere har lært basics
  - Analyser ->mange muligheter
    - Forskjellige løsninger på samme problem
  - Dere kan mer om design enn om analyser…

Eksamen 2019

Ledelsen på Perslia ungdomsskole har nå besluttet prøve ut "LesBedre" på 8. trinn. Rektor ønsker imidlertid at utprøvingen av programmet følges opp med forskning. Hun har derfor spurt Universitetet i Oslo om å hjelpe skolen med å designe forskningsopplegg som kan gjennomføres parallelt med implementeringen for leseprogrammet.

Siden det er det stilt spørsmål om hvorvidt programmet også har effekt på lesere med lav lesekompetanse, så ønsker rektor at forskningen også skal kunne si noe om effekten av programmet hos svake versus flinke lesere.

a) Lag en konkret skisse til et forskningsopplegg som er egnet til å besvare rektors spørsmål. Skissen bør inneholde opplysninger om aktuelle forskningsspørsmål, utvalg, sentrale målinger og variabler, spesielle designelementer (tid, gruppering, randomisering) og mulige analyser. Begrunn dine valg.

Spørsmål fra 2020

Jeg lurte på, når man skal gjennomføre en pre-test post-test og analysere den, hva er den best egnede måten å gjøre det på?

Jeg har søkt som en gal på nettet, og jeg finner at mange skriver ANOVA (samt ulike varianter av ANOVA), t-test og regresjon. Men jeg forstår det ikke helt, er dette en smak og behag situasjon? Eller er det sånn at man bør gjennomføre alle sammen?

Blant annet står det i sensorveiledningen for HØST2019 at:
«Med hensyn til analyser kan ulike t-tester, variansanalyse, og regresjonsanalyse være aktuelle alternativer.
Kandidater som i denne sammenhenger greier å forklare analyseoppsettet og hvordan de ulike variablene behandles
(for eksempel oppsett med kovariater) i analysen skal belønnes for det. Hvis kandidaten greier å skissere et analyseoppsett
som er egnet til å avdekke interaksjonseffekter mellom opprinnelig lesenivå (sterk versus svak leser)
og effekten av leseprogrammet, så bør dette også belønnes i karaktersettingen.»
Og kan du eventuelt utdype hvorfor man bør velge den ene analysemåten fremfor den andre?

Forslag til løsning

- Utvalg, 8 kl = ca. 120
  - (357 elever jevnt fordelt 8, 9, 10 kl)
- Utfall: Nasjonale prøver, lesing
  - Effekt for elever med lave ferdigheter?
  - Finnes tidligere tester - andre elever, samme skole

Randomisering

- Randomisere klasse eller individer
  - Gjennomførbart? Smitteeffekt?
- Treatment/control pre/post-test
  - Interaksjon med leseferdigheter før
- Statistisk styrke er et problem

Randomisering: Analyse

- ANOVA to-gruppe, repetert
  - Interaksjon med leseferdigheter på pre-test
- Regresjon, f.eks.
  - Differanseskåre (T2 - T1)
  - Dummy (1,0) for treatment control
    - Interaksjon med pre-test
  - https://www.statsimprove.com/en/what-is-the-difference-between-anova-and-regression-and-which-one-to-choose/

Forskjell-i-forskjeller (Difference-in-differences, DD)

- Nasjonale prøver over noen år for 8. og 9. klasse.
  - for 8. klasse (2020 treatment) og 9. klasse kl (2020 control)
    - Er forskjellen mellom 8. og 9. klasse prøver i 2020 mindre enn tidligere år?
    - Er forskjellen mellom de som scorer på de laveste nivåene i 8. og 9. klasse prøver i 2020 mindre enn tidligere år?
- Andre ting som kan påvirke 8. klasse men ikke 9. klasse som bryter med antakelsene?

Forskjell-i-forskjeller: Analyse

- Plotte data= tidstrender før 2020 for 8. klasse og 9. klasse. Er det felles trend? (avgjørende antakelse)
- Hvis felles trend: Regresjon der man kjører en forskjell-i-forskjeller analyse

$$y_{st} = \gamma Treat_s + \lambda Post_t + \delta(Treat_s \times Post_t) + \epsilon_{st};$$

- δ: Effekten av leseprogrammet

Husk: Ingen perfekte løsninger

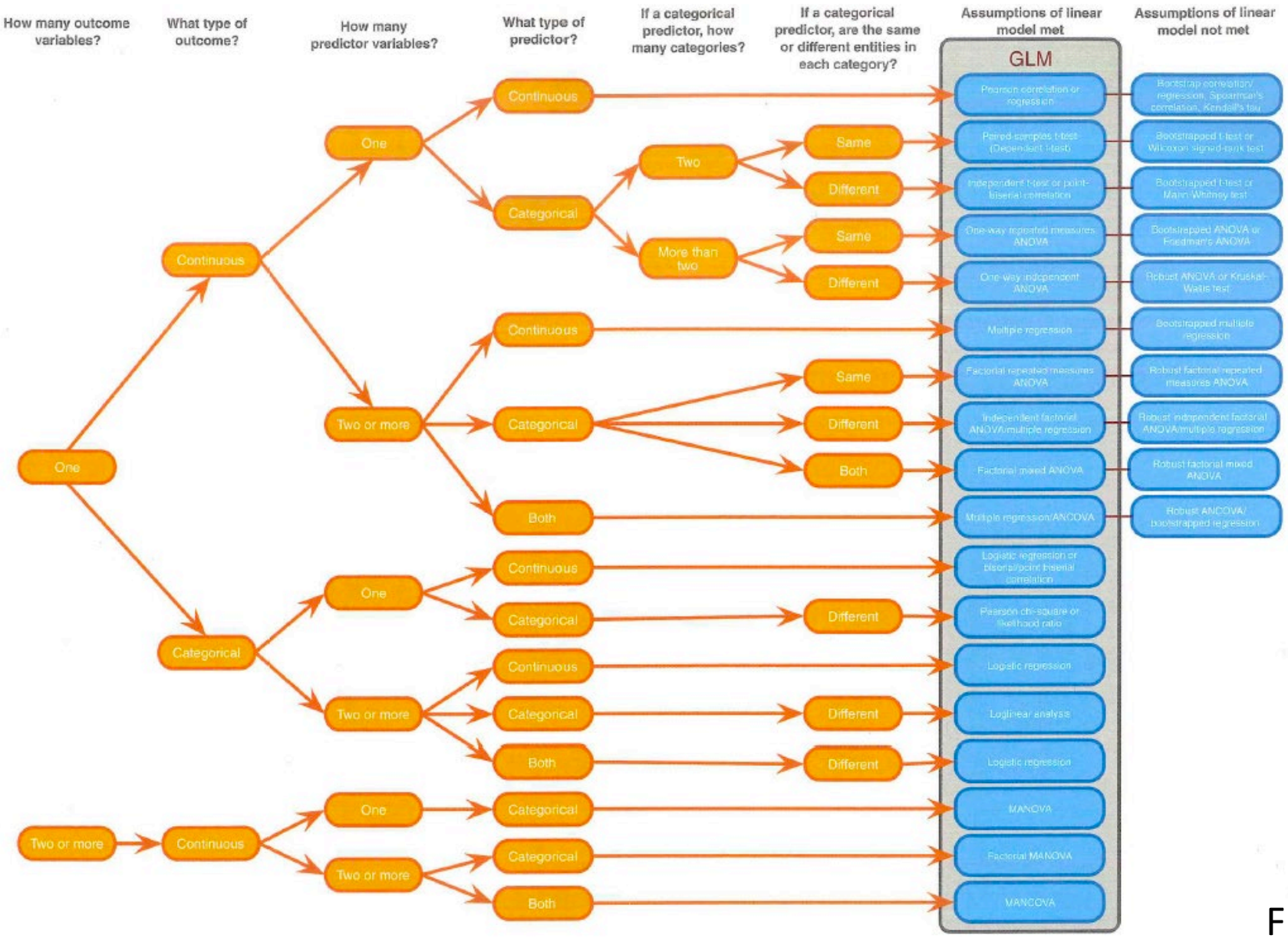
- Begrunne valg
  - diskutere styrker/svakheter
  - gjøre praktiske (etiske) virderinger
    - Hva fungerer i den virkelige verden

For eksamen

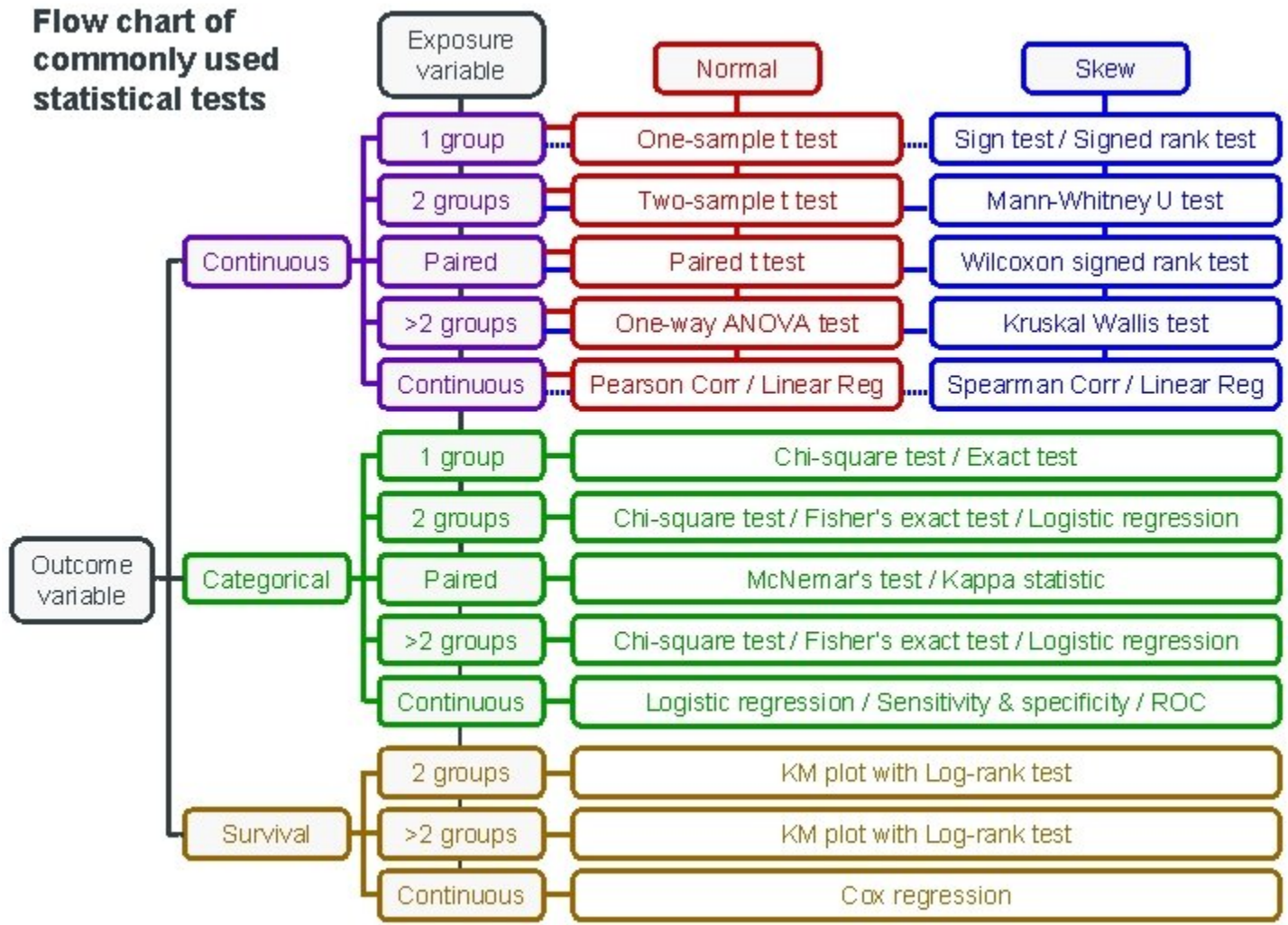
- Velg design og analyser dere er trygge på
  - La dere gjerne inspirere av gode eksempler i artikler dere leser
  - begrunn valgene
    - fordeler/ulemper (f.eks. styrke)
  - Sensorene vet hva dere har lært

For masteroppgaven

- Diskuter med veileder
- Forvent at dere må lære noe mer


How many outcome variables?
What type of outcome?
How many predictor variables?
What type of predictor?
If a categorical predictor, how many categories?
If a categorical predictor, are the same or different entities in each category?
Assumptions of linear model met
Assumptions of linear model not met
GLM
One
Two or more
Continuous
Categorical
Continuous
Categorical
One
Two or more
Two
More than two
Same
Different
Same
Different
Continuous
Categorical
Both
Same
Different
Both
Continuous
Categorical
One
Two or more
Continuous
Categorical
Both
Different
Different
Different
One
Two or more
Continuous
Categorical
Categorical
Both
Pearson correlation or regression
Paired-samples t-test (Dependent t-test)
Independent t-test or point-biserial correlation
One-way repeated measures ANOVA
One-way independent ANOVA
Multiple regression
Factorial repeated measures ANOVA
Independent factorial ANOVA/multiple regression
Factorial mixed ANOVA
Multiple regression/ANCOVA
Logistic regression or biserial/point biserial correlation
Pearson chi-square or likelihood ratio
Logistic regression
Loglinear analysis
Logistic regression
MANOVA
Factorial MANOVA
MANCOVA
Bootstrap correlation/ regression, Spearman's correlation, Kendall's tau
Bootstrapped t-test or Wilcoxon signed-rank test
Bootstrapped t-test or Mann-Whitney test
Bootstrapped ANOVA or Friedman's ANOVA
Robust ANOVA or Kruskal-Wallis test
Bootstrapped multiple regression
Robust factorial repeated measures ANOVA
Robust independent factorial ANOVA/multiple regression
Robust factorial mixed ANOVA
Robust ANCOVA/ bootstrapped regression

Field, 2018


Flow chart of commonly used statistical tests
Outcome variable
Exposure variable
Normal
Skew
Continuous
1 group
One-sample t test
Sign test / Signed rank test
2 groups
Two-sample t test
Mann-Whitney U test
Paired
Paired t test
Wilcoxon signed rank test
>2 groups
One-way ANOVA test
Kruskal Wallis test
Continuous
Pearson Corr / Linear Reg
Spearman Corr / Linear Reg
Categorical
1 group
Chi-square test / Exact test
2 groups
Chi-square test / Fisher's exact test / Logistic regression
Paired
McNemar's test / Kappa statistic
>2 groups
Chi-square test / Fisher's exact test / Logistic regression
Continuous
Logistic regression / Sensitivity & specificity / ROC
Survival
2 groups
KM plot with Log-rank test
>2 groups
KM plot with Log-rank test
Continuous
Cox regression

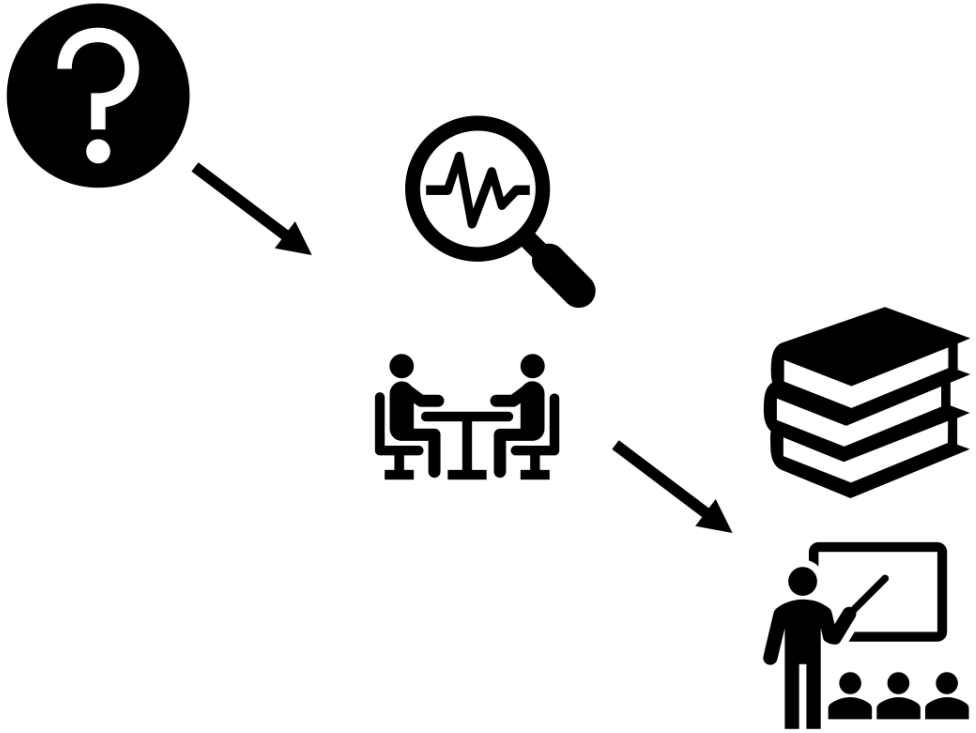
Standards for High-Quality Research and Analysis (RAND)

- The problem should be well formulated, and the purpose of the study should be clear.
- The study approach should be well designed and executed.
- The study should demonstrate understanding of related studies.
- The data and information should be the best available.
- Assumptions should be explicit and justified.
- The findings should advance knowledge and bear on important policy issues.
- The implications and recommendations should be logical, warranted by the findings, and explained thoroughly, with appropriate caveats.
- The documentation should be accurate, understandable, clearly structured, and temperate in tone.
- The study should be compelling, useful, and relevant to stakeholders and decisionmakers.
- The study should be objective, independent, and balanced.

God forskning er

- Etterrettelig
- Systematisk
- Gjennomsiktig
- Etterprøvbar

En prosess-
anvendelsen av vitenskapelig metode



God og dårlig forskning

Vi må spørre:

- Hvordan er begreper operasjonalisert?
- Hvilke alternative forklaringer er det?
- Hvilken kontekst er resultatene gyldige i?

Kleven & Hjardemaal, 2018

Egenskaper ved en god forsker

1. Entusiasme
2. Åpenhet
3. Sunn fornuft
4. Rolle-taking
5. Oppfinnsomhet
6. Tillit til egne vurderinger
7. Påpasselig og opptatt av detaljer
8. Evne til å kommunisere
9. Ærlighet

(Hall, 1984)

Egenskaper ved en god forsker

«En god forsker forsøker å vise at hun/han tar feil—en dårlig forsker forsøker å vise at hun/han har rett»

Hilsen Henrik Zachrisson

Forskningsprosessen

Sette sammen det dere allerede har lært

- Hypotese/forskningsspørsmål
- Design
- Sampling
- Validitet og reliabilitet
- Analyser

Forskningsprosessen

"The garden of forking paths" «Gelman & Loken, 2014»

- Strategiske, legitime valg i design/analyse, f.eks.
  - Utvalgstørrelse
  - Konstruksjon av variabler
  - Analyse
    - Valg av kovariater
    - Uteliggere
- Har betydning for resultatene
  - Alle valg må begrunnes
  - Rapporteres slik at de (i prinsippet) kan etterprøves

Robusthetssjekker

- Hva skjer hvis dere gjør et annet legitimt valg?
  - Konstruerer variabler annerledes?
  - Gjør en annen analyse?
  - Er resultatet robust på tvers av subgrupper?

- Også en måte å teste ut alternative hypoteser
  - Put your theory to the test

Faremomenter – «The troubling trio»

- Vær skeptisk til studier med
  - Lav statistisk styrke
  - Overraskende resultater
  - P-verdier nærme .05
- Risiko for at det er vanskelig å replikere resultatene

Editorial

Replication in Psychological Science

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Psychological Science
1–6

DOI: 10.1177/0956797615616374
pss.sagepub.com
SAGE

Dette skal dere IKKE gjøre – «P-hacking»

- «trimme» analysen -> p<.05
  - legge til eller trekke fra variabler
  - velge «beste sub-sample» post hoc
  - kun rapportere noen resultater

Rapportering av forskning

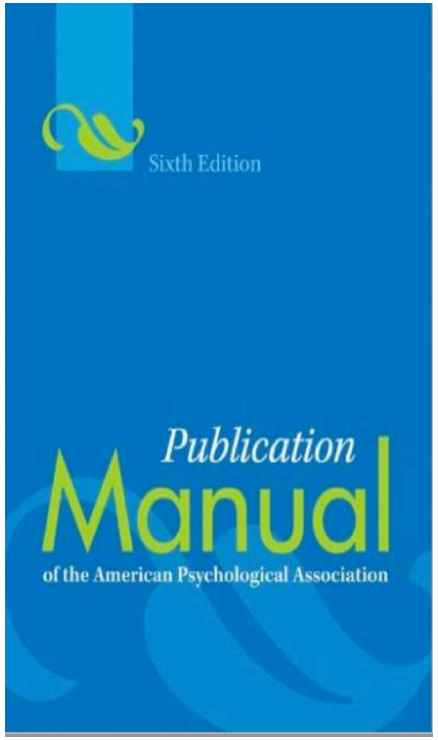
- En gjennomgang av alle elementene i en forskningsrapport
- Adaptert fra kurs på CEMO om akademisk skriving (2016)
- På engelsk…

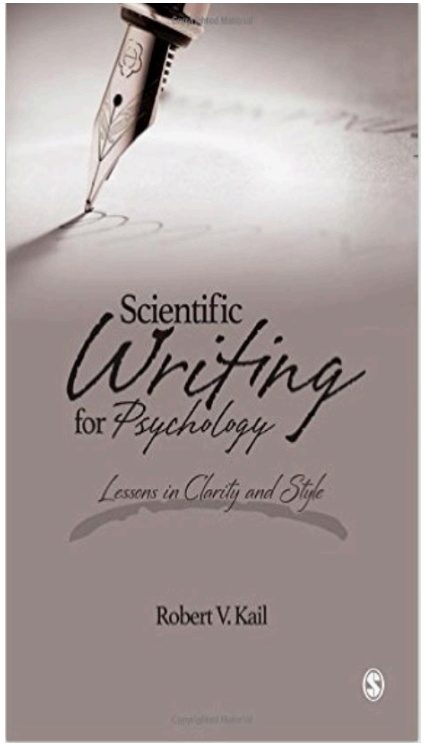
Style and Stuff

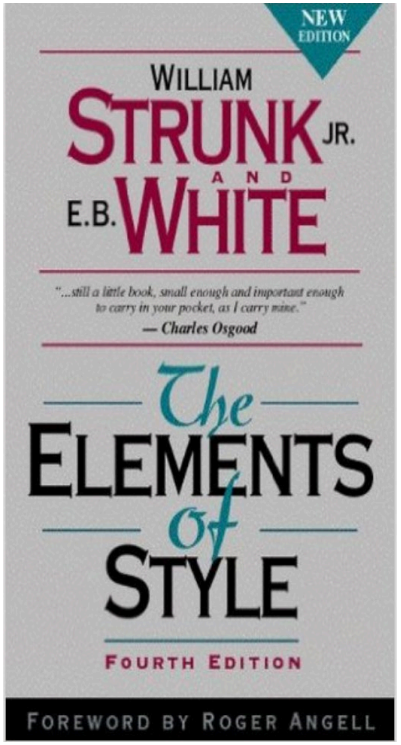
- Online resources
  - http://www.apastyle.org
  - https://laeringsbloggen.com/hvordan-skrive-akademisk-tekst-3-boker-med-gode-rad/
  - sokogskriv.no
  - http://www.nature.com/scitable/topicpage/effective-writing-13815989

Style and Stuff

Writings about writing

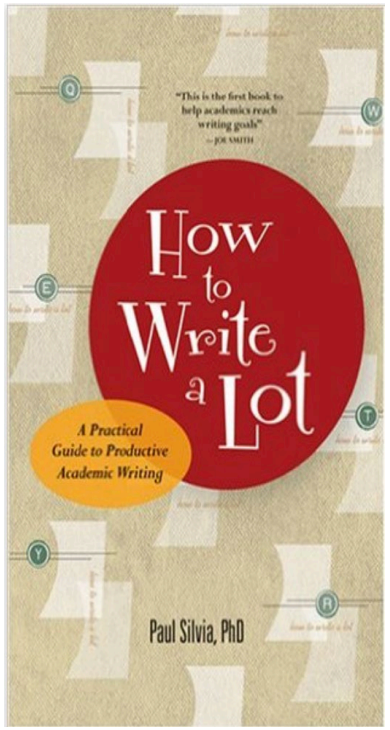


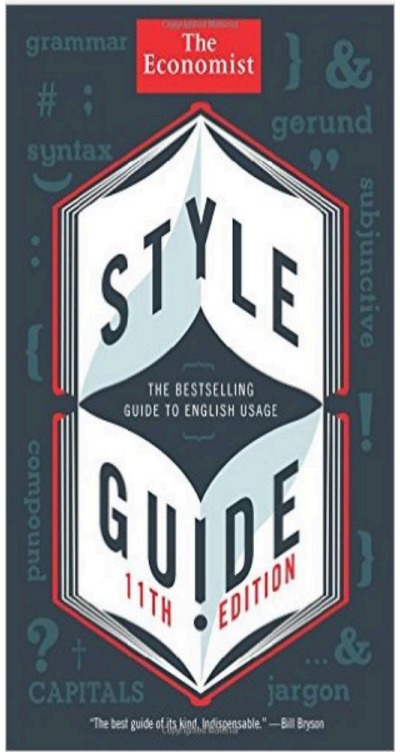




Style and Stuff

More writings about writing





Style and Stuff

George Orwell's Writing Rules

1. Never use a metaphor, simile, or other figure of speech which you are used to seeing in print.
2. Never use a long word where a short one will do.
3. If it is possible to cut a word out, always cut it out.
4. Never use the passive where you can use the active.
5. Never use a foreign phrase, a scientific word, or a jargon word if you can think of an everyday English equivalent.
6. Break any of these rules sooner than say anything outright barbarous.

From Orwell's essay "Politics and the English Language."

Artikkelbasert master

- Velg et tidsskrift du eller veilederen din kjenner godt
- Les om tidsskriftets mål og omfang
- Hvilke artikler planlegger du å sitere – kanskje du kan bidra til å «fortsette» den akademiske samtalen i samme tidsskrift
- Er artikkelen din på et passende nivå for tidsskriftet?
  - Design, data, implikasjoner, omfang
- Er det relevante forskere i redaksjonen til tidsskriftet?
- Finnes det et plan B tidsskrift med samme format?

Psychological Assessment
1995, Vol. 7, No. 3, 228-237

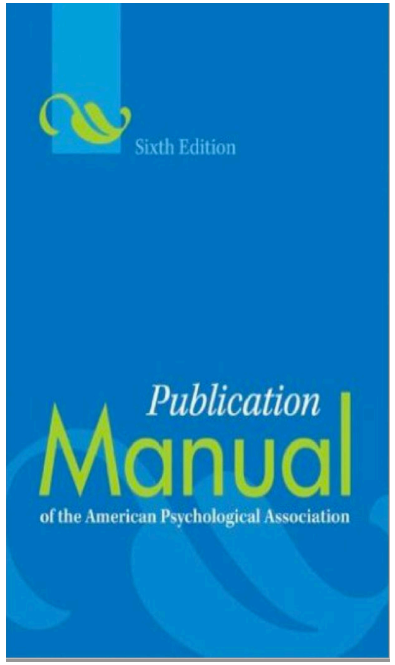
1040-3590/95/$3.00

Preparing and Evaluating Research Reports

Alan E. Kazdin
Yale University

The Title

- Summarize the main idea
  - If possible, with style
- Explanatory when standing alone
- Used for indexing
- Be concise, avoid unnecessary words



The Title: The authoritative

Journal of Child Psychology and Psychiatry 47:12 (2006), pp 1263–1271 doi:10.1111/j.1469-7610.2006.01682.x

Is there an epidemic of child or adolescent depression?

E. Jane Costello,[1] Alaattin Erkanli,[2] and Adrian Angold[1]
[1]Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, USA; [2]Department of Biostatistics and Bioinformatics, Duke University School of Medicine, USA

Comorbidity

Adrian Angold, E. Jane Costello, and Alaattin Erkanli
Duke University Medical Center, Durham, U.S.A.

J. Child Psychol. Psychiat. Vol. 40, No. 1, pp. 57–87, 1999
Cambridge University Press.

0021–9630/99 $15.00 + 0.00

The Title: The playful

American Economic Journal: Economic Policy 3 (May 2011): 97–129
http://www.aeaweb.org/articles.php?doi=10.1257/pol.3.2.97

No Child Left Behind: Subsidized Child Care and Children's Long-Run Outcomes†

By Tarjei Havnes and Magne Mogstad*

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

Money for nothing? Universal child care and maternal employment

Tarjei Havnes [a,*], Magne Mogstad [b]

[a] Dept. of Economics, University of Oslo, Postboks 1095 Blindern, 0317 Oslo, Norway
[b] Research Dept. of Statistics Norway, Pb 8131 Dep, 0033 Oslo, Norway

The Title: The descriptive

AERA Open
July-September 2016, Vol. 2, No. 3, pp. 1–18
DOI: 10.1177/2332858416657343

Recent Trends in Income, Racial, and Ethnic School Readiness Gaps at Kindergarten Entry

Sean F. Reardon
Stanford University
Ximena A. Portilla
MDRC

Journal of Educational Psychology

0022-0663/16/$12.00 http://dx.doi.org/10.1037/edu0000147

Process Mediates Structure: The Relation Between Preschool Teacher Education and Preschool Teachers' Knowledge

Sigrid Blömeke
Centre for Educational Measurement at the University of Oslo

Lars Jenßen and Marianne Grassmann
Humboldt University of Berlin

Simone Dunekacke
Leibniz Institute for Science and Mathematics Education

Hartmut Wedekind
Alice Salomon University of Applied Science

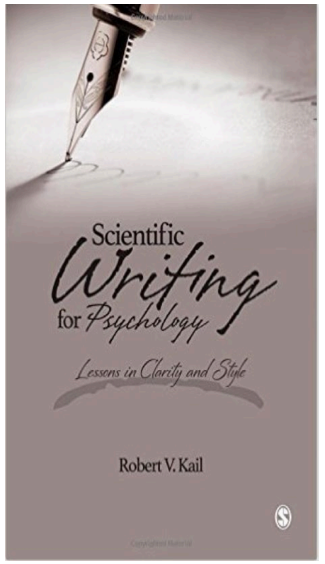
The Abstract

- Brief (100-200 words) comprehensive summary
- The single most important paragraph
  - Read most often
  - Basis for deciding to read the whole article
  - Used for systematic reviews
- Do not include information not in the main article
  - Include sample/methods/results

Table 1
Major Questions to Guide Journal Article Preparation

Abstract
What were the main purposes of the study?
Who was studied (sample, sample size, special characteristics)?
How were participants selected?
To what conditions, if any, were participants exposed?
What type of design was used?
What were the main findings and conclusions?

The introduction (the most difficult)

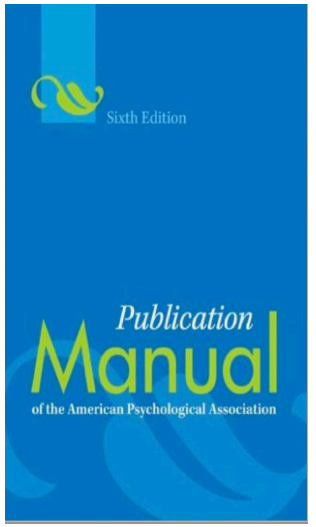


What is the function of the introduction?

- Convince the readers that this study is important
  - Timely and critical to the field?
  - Does this study offer new insights?
  - Are the methods well suited to answer the questions?
- A bad intro screws a manuscript
  - The “Ah-effect”
- An intro sets the scene – as in a first date

Questions to be answered:

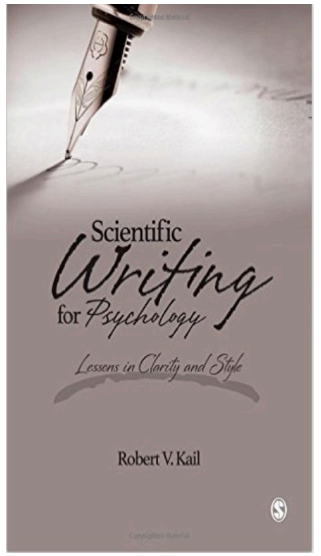
- Why is this problem important?
- How does the study relate to previous work?
  - What is similar, what is new?
- What are the hypotheses?
  - Links to theory?
- How is the design relevant for the hypotheses?
- What are the theoretical or practical implications of the study?



Introduction

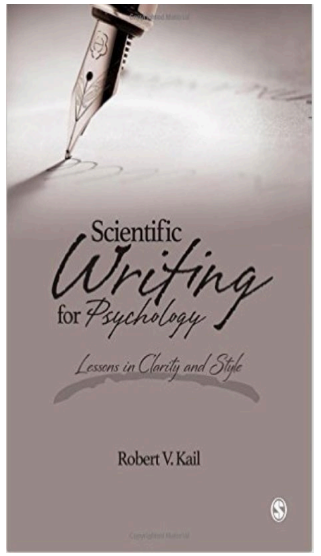
What is the background and context for the study?
What in current theory, research, or clinical work makes this study useful, important, or of interest?
What is different or special about the study in focus, methods, or design to address a need in the area?
Is the rationale clear regarding the constructs to be assessed?
What specifically were the purposes, predictions, or hypotheses?

Some tweaks to the opening paragraph(s)



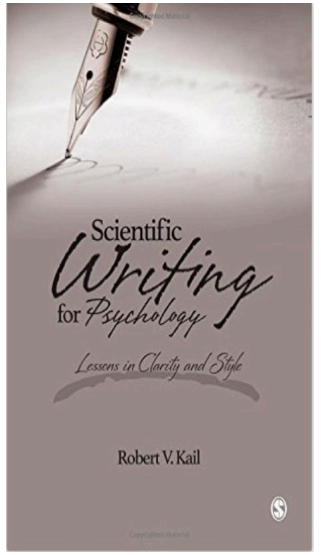
- Hook the reader
- Make the research problem real life
  - Not: Social psychologists have studied many types of cooperation but have ignored interactions that require physical cooperation.
  - Rather: When people enter a building, others often hold the door for them. Although such interactions are commonplace, we know little about forms of social interaction like these that involve physical cooperation.

Keep the topic sentence in mind



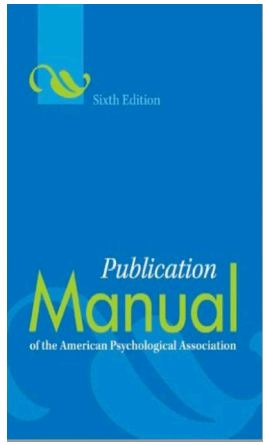
- Make it alive
  - Not: Facial expressions are cues to people's behaviors
  - Rather: People often use other's facial expressions to understand their behavior

The fine art of paragraphs (Ch. 4)



- One main idea
- Expressed in the first sentence (Topic sentence)
- Followed by supporting/elaborating sentences
- Closing/concluding sentence

Methods section



Main goals:

- Give the reader all information needed replicate your study, in detail (Section 2.06)
  - Descriptions of how the study was conducted, including the conceptual and operational definitions of variables
- Give the reader all information needed to interpret your findings
  - Evaluate the appropriateness of your methods
  - Evaluate the validity, reliability, and robustness of your results

Method

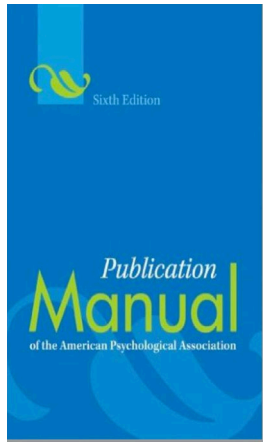
Participants
- Who were the participants and how many of them were there in this study?
- Why was this sample selected in light of the research goals?
- How was this sample obtained, recruited, and selected?
- What are the participant and demographic characteristics of the sample (e.g., gender, age, ethnicity, race, socioeconomic status)?
- What if any inclusion and exclusion criteria were invoked (i.e., selection rules to obtain participants)?
- How many of those participants eligible or recruited actually were selected and participated in the study?
- Was informed consent solicited? How and from whom, if special populations were used?

Design
- What is the design (e.g., longitudinal, cross-sectional) and how does the design relate to the goals of the study?
- How were participants assigned to groups or conditions?
- How many groups were included in the design?
- How were the groups similar and different in how they were treated in the study?
- Why were these groups critical to address the questions of interest?

Assessment
- What were the constructs of interest and how were they measured?
- What are the relevant reliability and validity data from previous research (and from the present study) that support the use of these measures for the present purposes?
- Were multiple measures and methods used to assess the constructs?
- Are response sets or styles relevant to the use and interpretation of the measures?
- How was the assessment conducted? By whom (as assessors/observers)? In what order were the measures administered?
- If judges (raters) were used in any facet of assessment, what is the reliability (inter- or intrajudge consistency) in rendering their judgments/ratings?

Procedures
- Where was the study conducted (setting)?
- What materials, equipment, or apparatuses were used in the study?
- What was the chronological sequence of events to which participants were exposed?
- What intervals elapsed between different aspects of the study (e.g., assessment occasions)?
- What procedural checks were completed to avert potential sources of bias in implementation of the manipulation and assessments?
- What checks were made to ensure that the conditions were carried out as intended?
- What other information does the reader need to know to understand how participants were treated and what conditions were provided?

The results section



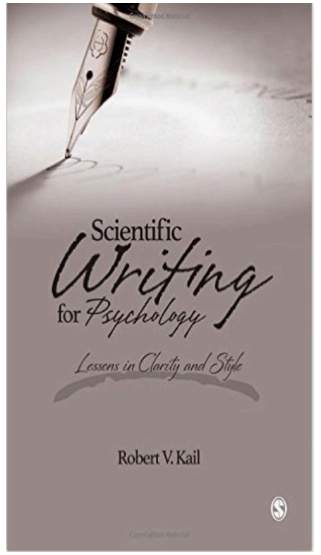
The most important part of your paper

- Present all results informing your conclusion
- Be transparent & unbiased
  - Report necessary and sufficient details
  - Never omit results contrary to your expectations

Results

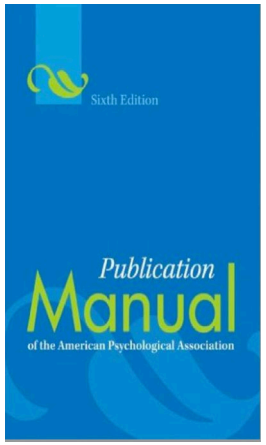
What were the primary measures and data on which the predictions depend?
What are the scores on the measures of interest for the different groups and sample as a whole (e.g., measures of central tendency and variability)?
How do the scores compare with those of other study, normative, or standardization samples?
Are groups of interest within the study similar on measures and variables that could interfere with interpretation of the hypotheses?
What analyses were used and how specifically did these address the original hypotheses and purposes?
Were the assumptions of the data analyses met?
If multiple tests were used, what means were provided to control error rates?
If more than one group was delineated, were they similar on variables that might otherwise explain the results (e.g., diagnosis, age)?
Were data missing due to incomplete measures (not filled out completely by the participants) or due to loss of participants? If so, how were these handled in the data analyses?
Are there ancillary analyses that might further inform the primary analyses or exploratory analyses that might stimulate further work?

Write the results with style



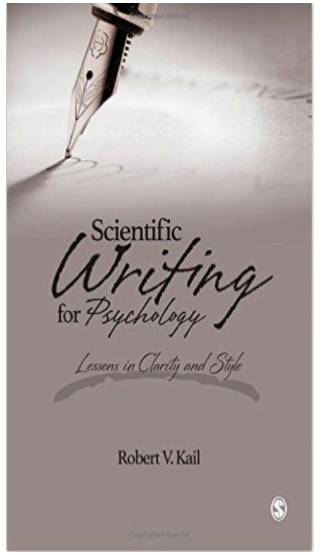
- Tell a story
  - Not only report numbers (numbers are boring!)
- Results in the foreground – analyses in the back
  - What happened? – Justify this by analyses
- Stats is to the results what citations are to the intro
  - Give credibility to the story
- Think of the results section as a photo essay
  - Pictures (figures and tables) pull the story

Reporting stats (NHST)



- Choose meaningful unit of reporting
  - Unstandardized
  - Standardized
  - Mix
- Effect size measure
- Report *SE* or confidence interval
- Omnibus tests/fit statistics

Writing – results in the front

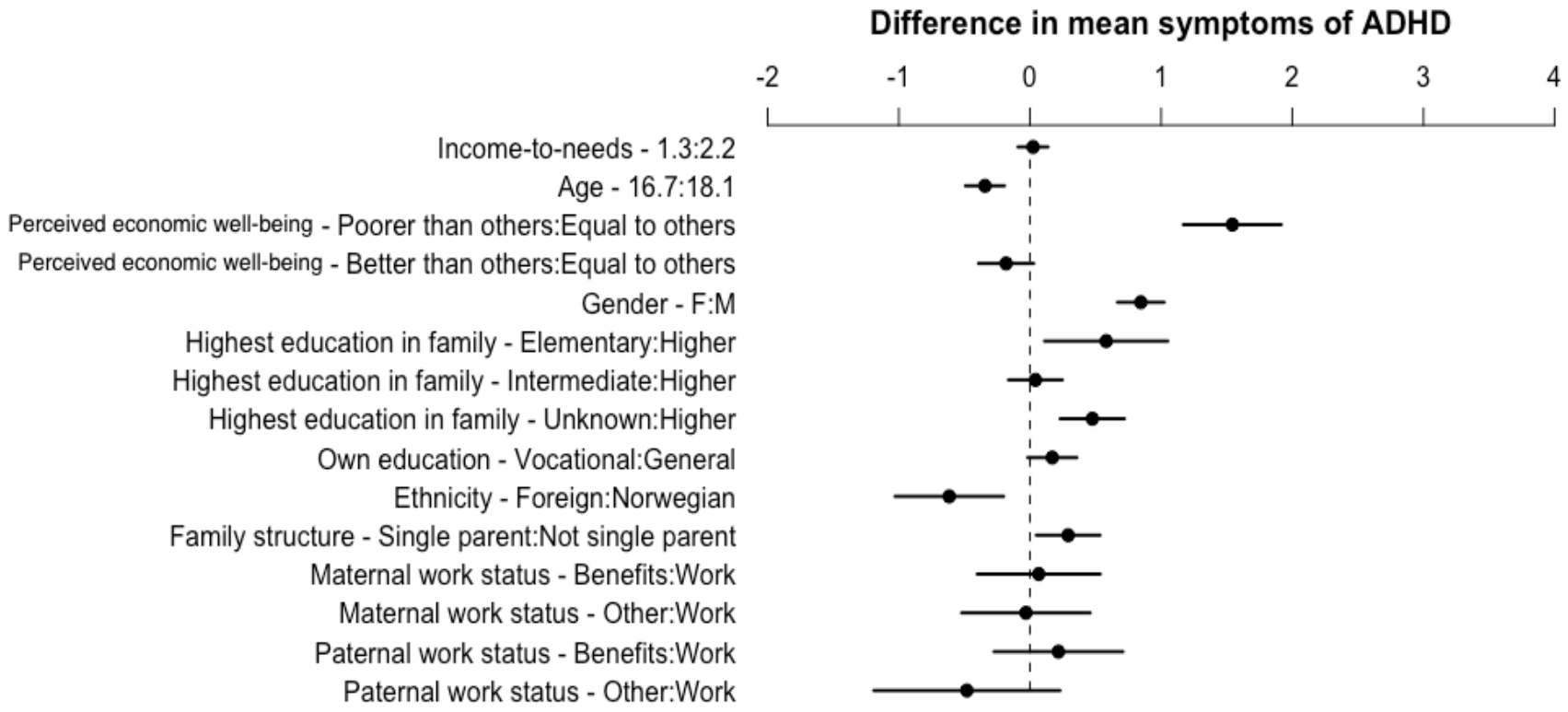


- There was a significant main effect for type of video game, $F(1,65) = 6.24$, $\eta 2 = .08$
- Cheating was greater in violent video game condition than in the nonviolent game condition, $F(1,65) = 6.24$, $\eta 2 = .08$
- Participants cheated more when they played violent videogame than when they played a nonviolent video game, $F(1,65) = 6.24$, $\eta 2 = .08$

Figures

- A picture says more than a thousand words…
- Figures make people remember your story
  - Easier when you talk about your results
- Get inspired
  - keep notes of cool graphs in papers
  - search the web

Graphical inferential stats



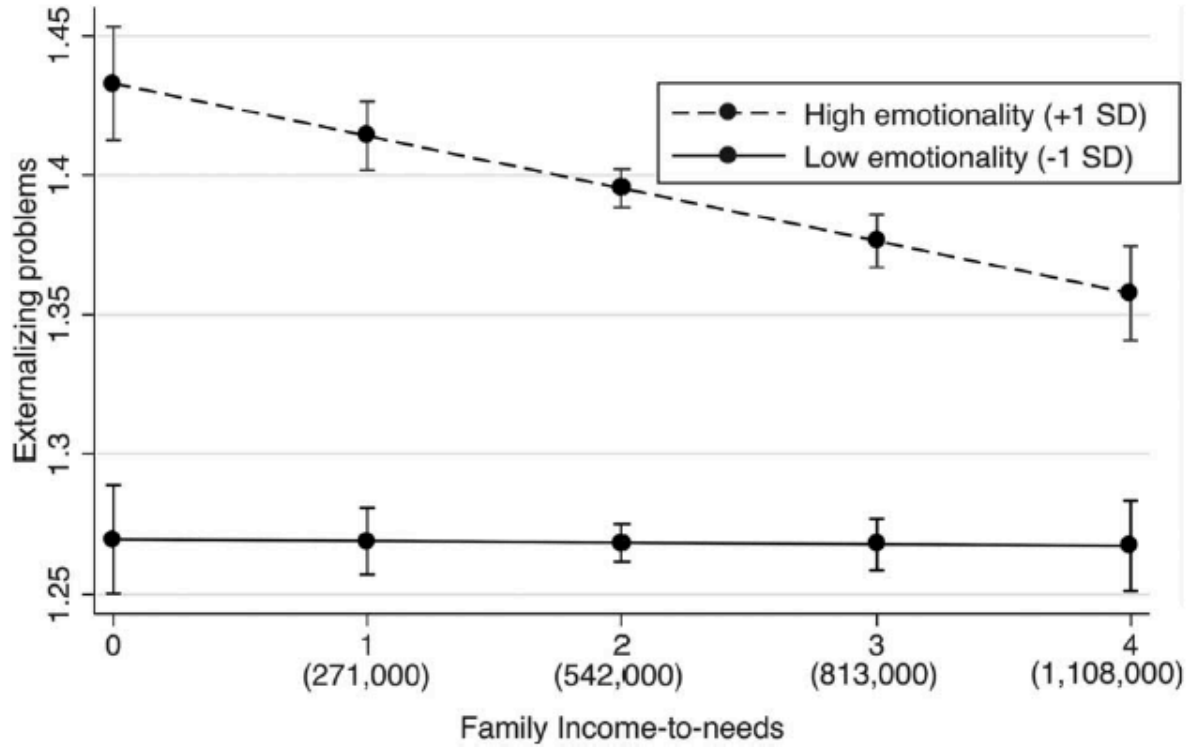
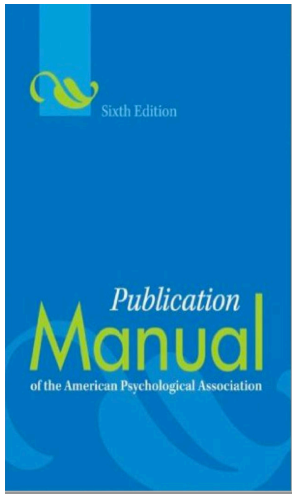
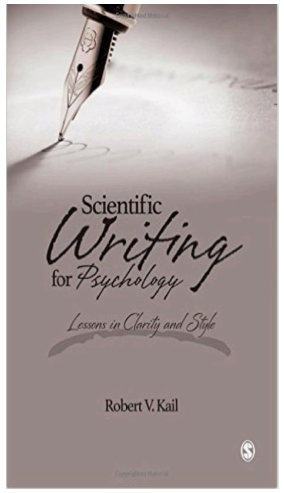


Figure 1. Externalizing problems in 5-year-old children as a function of the interaction between income-to-needs and average emotionality measured when children were 1.5 and 3. Values on the y-axis are 1/6 of an SD apart. Error bars represent 95% CIs. Values on the x-axis are income-to-needs ratios, with examples (in brackets) of income in 2006 Norwegian kroner for a family of 2 adults and 2 children.

The discussion section





Purpose:

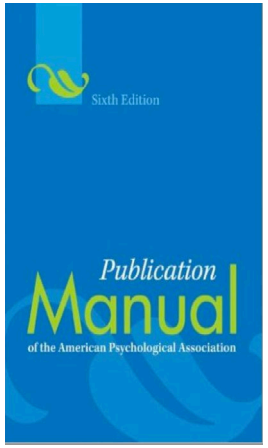
- Evaluate and interpret the implications of your results
  - With respect to your hypotheses (why supported/rejected?)

..or

- To consider the issues raised in the introduction in light of your results

Structure

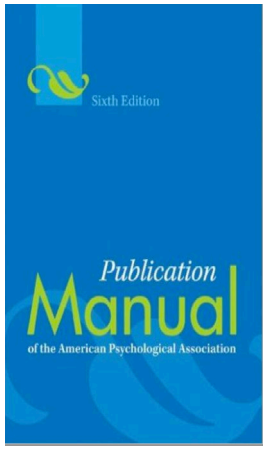
Recap

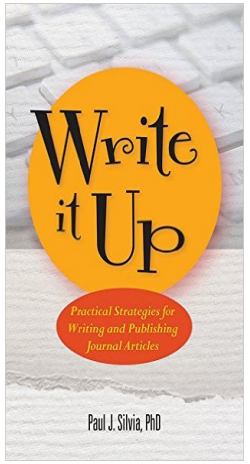


1. A clear statement of support or not for hypotheses (summary of results in light of hypotheses)
2. Similarities and differences with previous research
   - Post hoc explanation in case of nonsupport for hypotheses/differences from previous research
   - Don't repeat previous points, but contextualize (sample, design, country)
   - Substantive discussions of effect sizes
3. Potential mechanisms/explanations (beyond data)

Connect

Resolve





Structure (*cont'd*)

Keep it short, relevant, and positive

4. Limitations

- Sources of bias (e.g., threats to internal validity)
- Quality of measures
- Other limitations (e.g., sample size, sampling) or alternative explanations

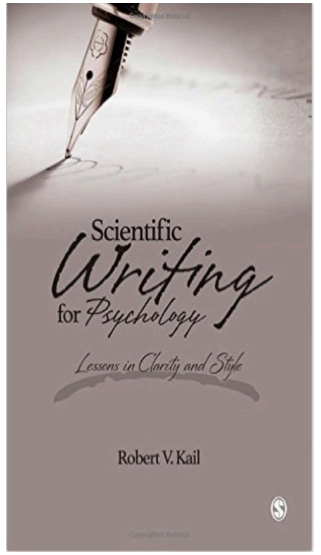
5. Implications

- What is the practical implication of your results?

6. Conclusion

- Sum up the discussion – Why should the field attend to this study?

Alternative placement of the limitations section



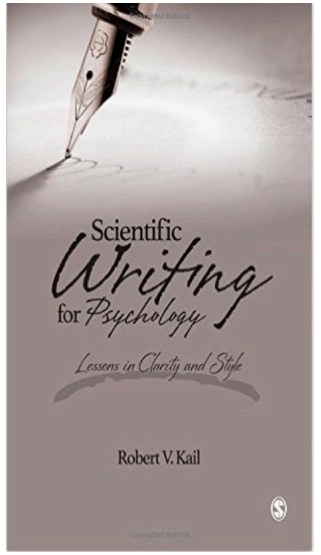
- Move it up front (after the summary)
  - Your paper ends with a more positive focus
  - Don’t kill your reader’s enthusiasm at the end
  - You can discuss the findings in light of the limitations

- Don’t kill your findings
  - Be transparent about weaknesses
  - If possible, suggest implications for your findings

Discussion

What were the major findings of the study?
How do these findings add to research and how do they support, refute, or inform current theory?
What alternative interpretations can be placed on the data?
What limitations or qualifiers must be placed on the study given methodology and design issues?
What research follows from the study to move the field forward?

Avoid common mistakes…



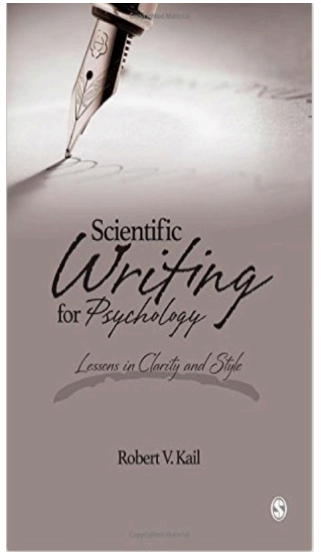
1. Too long: don't talk about every tiny detail
   - Stick with the big story
2. Poor organization: use APA or other structure
3. Doesn't discuss: avoid repeating a list of results
   - Put results in perspective
4. Off topic: stick with your story
5. Inconsistent tone: don't be too bold, don't be to cautious
6. Too much self-praise: drop "we are the first to…" or "this is important because…" show it, rather than say it.

End your paper with a bang, not a whimper

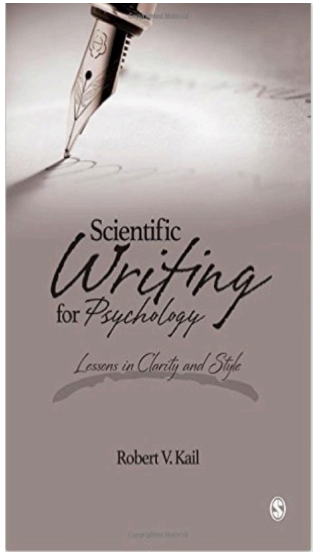
- Avoid banal endings
  - "Answering these lingering questions is a topic for further research."

(Section 3.08)

- Decide on your take-home message
  - Find a cool way of stating it in a sentence



Update your first-sentence hook



- When people enter a building, others often hold the door for them. Although such interactions are commonplace, we know little about forms of social interactions like these that involve physical cooperation.
- Thus, the present findings highlight the role of shared effort in driving physical cooperation, and they lead to some practical advice: if you want the door held for you as you enter the building, arrive with a friend, not alone

Ta-men-hjem: God forskning er

- Etterrettelig
- Systematisk
- Gjennomsiktig
- Etterprøvbar