



UiO **• Institutt for spesialpedagogikk**
Det utdanningsvitenskapelige fakultet

Forskningsdesign: Ikke-eksperimentelle design og metaanalyse

Astrid Marie Jorde Sandsør



Mål

- Gjennomgå det dere har lest
- Supplere det dere har lest
- Gi perspektiver og innsikt
- Gi redskaper til å tenke

I dag

- Ikke-eksperimentelle studier
 - Tverrsnitt
 - Longitudinelle (prospektive)
- Indre validitet
 - Statistisk kontroll
- Systematic reviews

- Slides fra SPED 4010,
(Zachrisson, 2020)
- Lund et al (2002), Kapittel 8 & 9

Repetisjon

- Beskrivende studier
 - Ikke antagelser om kausalitet
 - Indre validitet ikke relevant
- Årsaksstudier
 - Trusler mot indre validitet

Tre klassiske kriterier

- Hume (1738-1740): «A Treatise of Human Nature» (vitenskapsteori)
 1. "The cause and effect must be contiguous in space and time.»
 2. "The cause must be prior to the effect.»
 3. "There must be a constant union betwixt the cause and effect. 'Tis chiefly this quality, that constitutes the relation.»

Indre validitet

- Viser studien et årsaksforhold?
 - Er sjokolade årsaken til Nobelpriser?
- Trusler: Er det alternative, plausible, forklaringer?
 - NB! Ikke alle trusler er plausible i alle studier
- Tenke gjennom ved design av en studie:
 - Hvordan kan vi gardere oss mot alternative forklaringer?

Trusler mot indre validitet

- Ikke-observert konfundering
- Retning
- Historie
- Modning
- Seleksjon
- Testing
- Frafall
- Instrumentering
- Regresjon mot gjennomsnittet
- Atypisk kontrollgruppe

Ikke-eksperimentelle studier: Definisjon

- Studier uten manipulering
 - Dvs. ikke eksperimentelle
- Studier/analyser hvor vi ikke sannsynliggjør forskjeller gjennom design
 - Dvs. ikke kvasieksperimentelle
- MEN: Noen studier kan ha elementer som gjør dem til kvasi-eksperimenter for noen problemstillinger, ikke andre
 - f.eks. analyser med/uten søsken-modeller

Fordeler

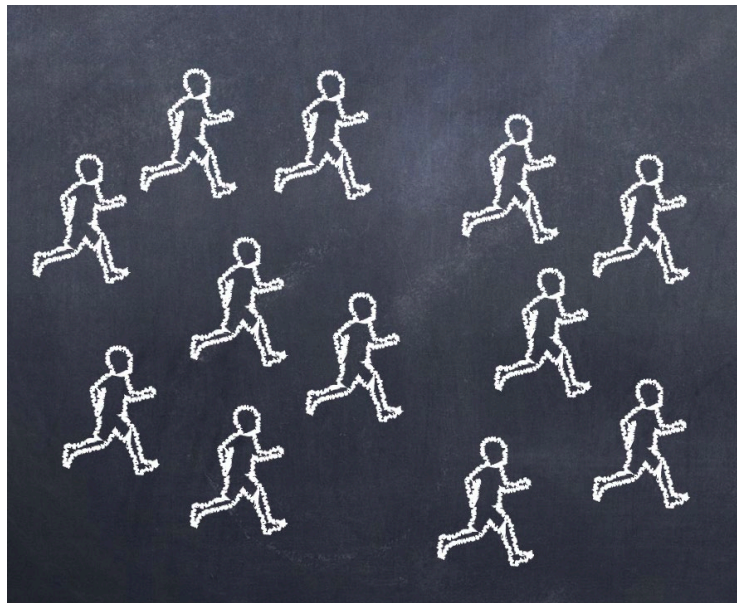
- Vi kan studere ting som vi ikke kan manipulere (praktisk eller etisk)
 - f.eks. kjønnsforskjeller
- Praktisk gjennomførbare
- Enklere logikk

To hovedtyper

- Kryss-seksjonell («tverrsnitt»)
 - Alle mål på ett tidspunkt
- Longitudinell («prospektiv», «lengdesnitt», «forløp», «panel»)
 - Mål på ulikt tidspunkt
- Kan være både beskrivende og ha ambisjoner om å si noe om årsakssammenhenger

Kryss-seksjonelt design

- Kun et måletidspunkt
- Avhengige og uavhengige variabler






- Vanlig fordi det er enkelt
- Sammenheng eller forskjeller mellom variabler på ett tidspunkt
- Svak indre validitet
- Potensielt god ytre validitet

Kryss-seksjonelt design

EUROPEAN EARLY CHILDHOOD EDUCATION RESEARCH JOURNAL, 2017
<https://doi.org/10.1080/1350293X.2018.1412050>




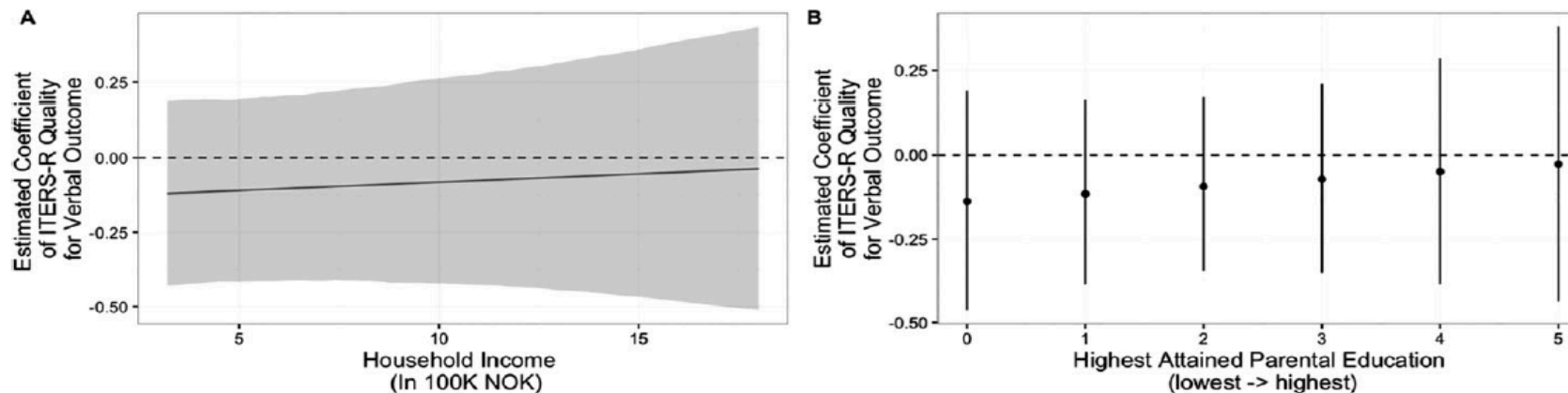
Is cognitive development at three years of age associated with ECEC quality in Norway?

Erik Eliassen ^a, Henrik Daae Zachrisson ^b and Edward Melhuish ^{c,d}

- Er det sammenheng mellom barnehagekvalitet og kognitive ferdigheter?
 - Alle mål på ett tidspunkt
- Finner ingen sammenhenger
 - Trusler mot indre validitet styrker null-funn

Kryss-seksjonelt design

8  E. ELIASSEN ET AL.



- Tester om sammenhengen er lik for ulike nivåer av sosioøkonomisk bakgrunn (SES)
- Kan ikke utelukke 0-verdien fra noen av undergruppene

Kryss-seksjonelt design

- Hva kan de si noe om (og ikke)?
 - Sammenheng
 - Antakelser om kausalitet
 - Status quo – ved 3 års alder
 - Ikke endring
 - Ikke andre aldersgrupper

Kryss-seksjonelt design

- Andre eksempler
 - Er det kjønnsforskjell i ordflyt?
 - Er det sammenheng mellom tid brukt på lesetrening og ordflyt?
 - Cochleaimplantasjon og sosial deltakelse

Kryss-seksjonelt design

- Temporalitet (samtidighet)
 - Fører cochleaimplantat til sosial deltakelse eller fører sosial deltakelse til cochleaimplantat?
 - Eller er begge forklart av en ikke-observert variabel? (Familiebakgrunn? Utadvendthet?)
- Antatt temporalitet (litt mindre svak indre validitet)
 - Foreldres utdannelse kommer før nasjonale prøver selv om de måles samtidig
 - Men, er begge forklart av en ikke-observert variabel? (skolekvalitet der de bosetter seg?)

Kryss-seksjonelt design

- Alt er målt samtidig
- Kan si noe om sammenhenger... på et tidspunkt
- Svak indre validitet

Longitudinelle studier

- Trendstudier -> PISA
 - Sammenligning av ulike kohorter over tid
- Forløpsstudier (prospektive) -> NumLit
 - Følger personer over tid
- Kohortstudier -> MoBa
 - Ofte kombinasjon

Trendstudier

- F.eks. land følges over tid
 - De som måles er av lik alder
- Populasjon: Viser endring (f.eks. 16 åringer) over tid
- Individ: kryss-seksjonell studie
 - f.eks. lekser -> bedre resultat

Trendstudier - eksempel



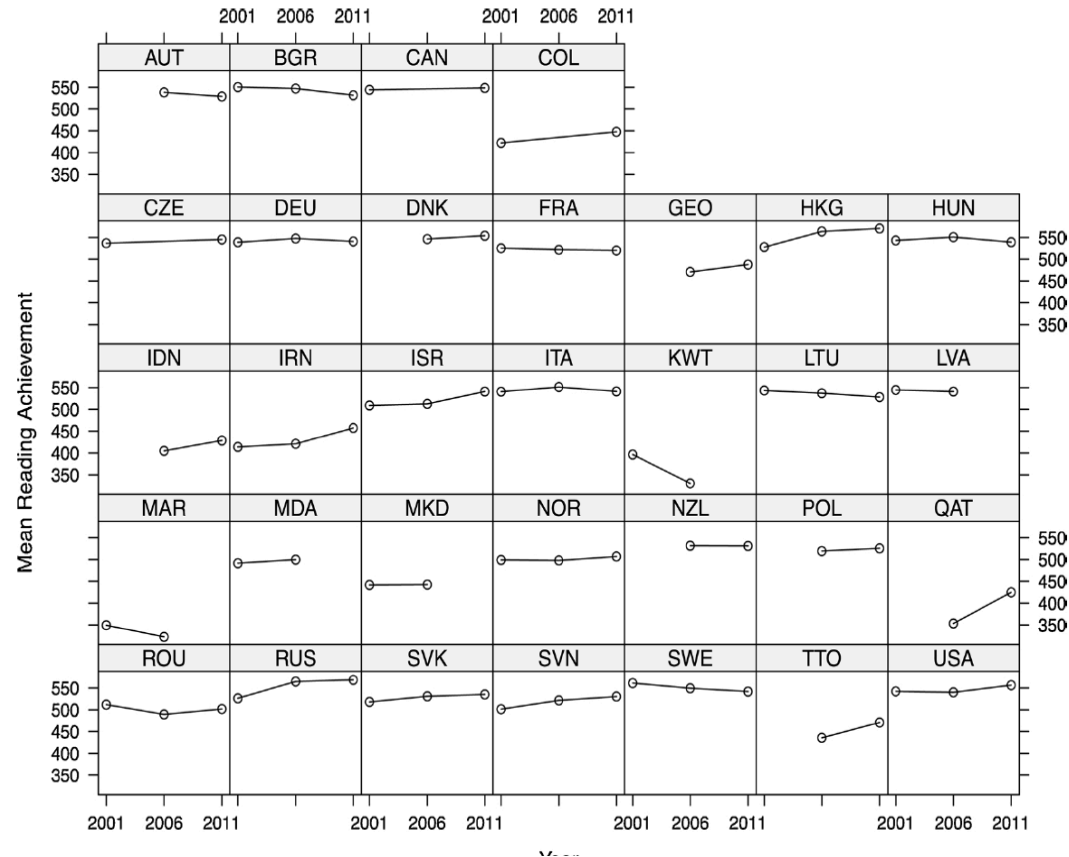
- Land er enhet over tid
- Endring i barnehagedekning -> endring i PISA/PIRLS-resultater
 - Elever målt på ett tidspunkt
 - Effekt på tvers av kohorter

Trendstudier - eksempel

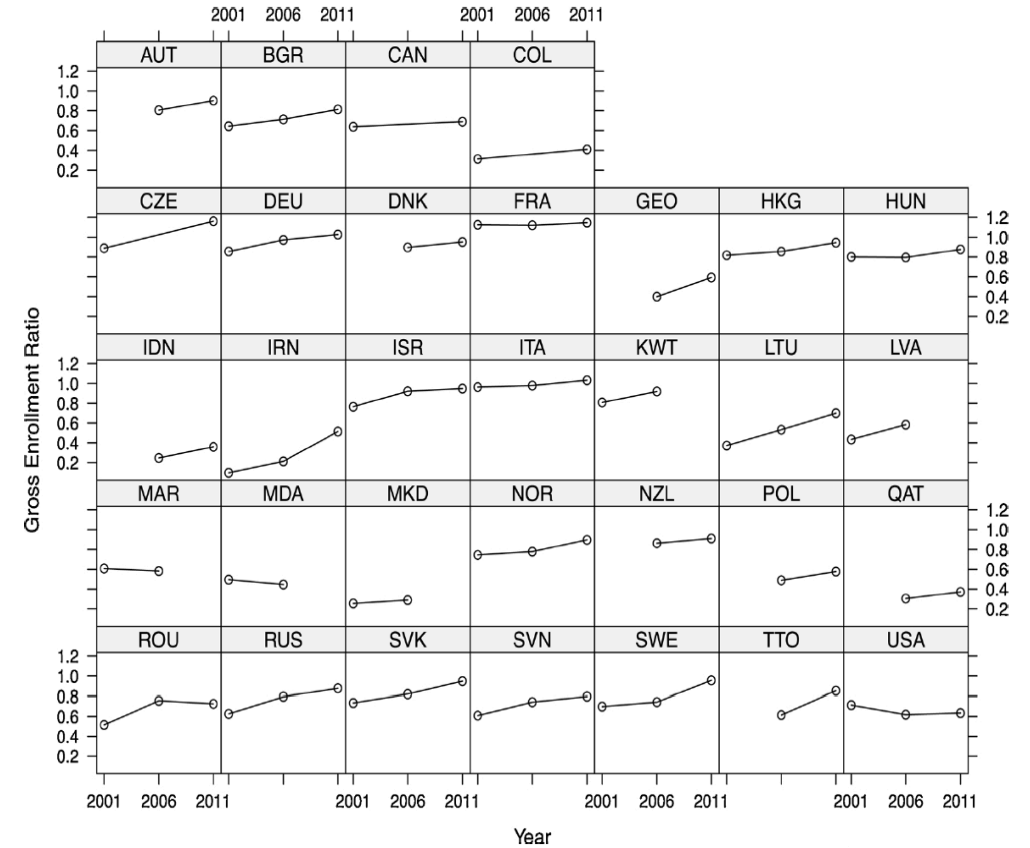
R. Strietholt et al.

International Journal of Educational Development 79 (2020) 10228

International Journal of Educational Development 79 (2020) 10228



Endring i testskår

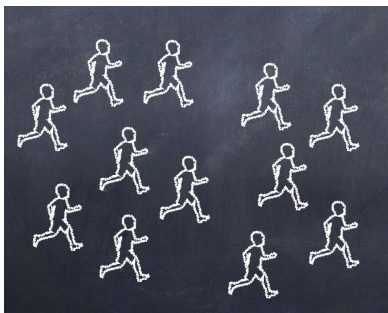


Endring i barnehagedekning

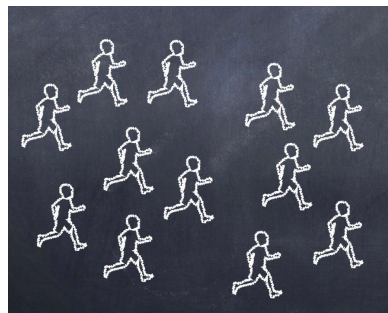
Longitudinelle studier på individnivå

- Flere måletidspunkt

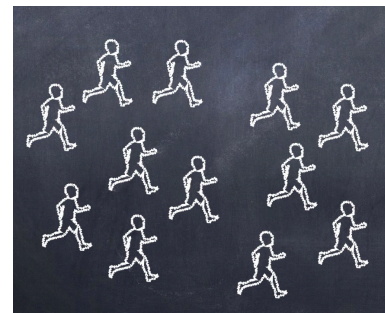
• T1



T2



T3



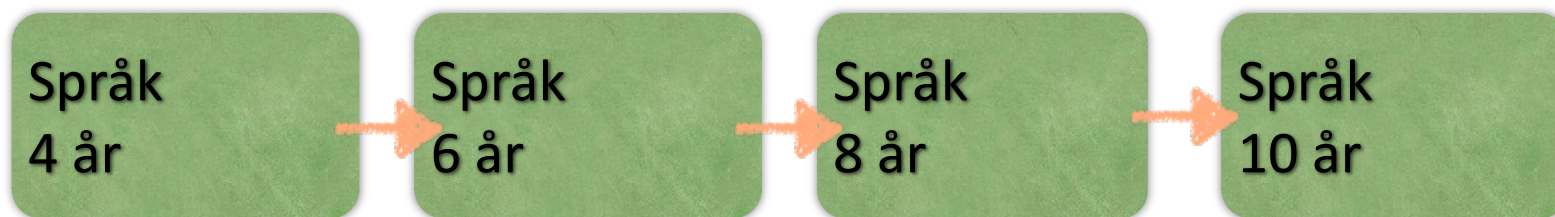
- Avhengige og/eller uavhengige variable + tid
- Dager-måneder-år

Longitudinelle studier på individnivå

- Mer krevende -> frafall
 - Sammenheng eller forskjeller mellom variabler over tid
 - Potensielt svak indre validitet
 - Potensielt god ytre validitet

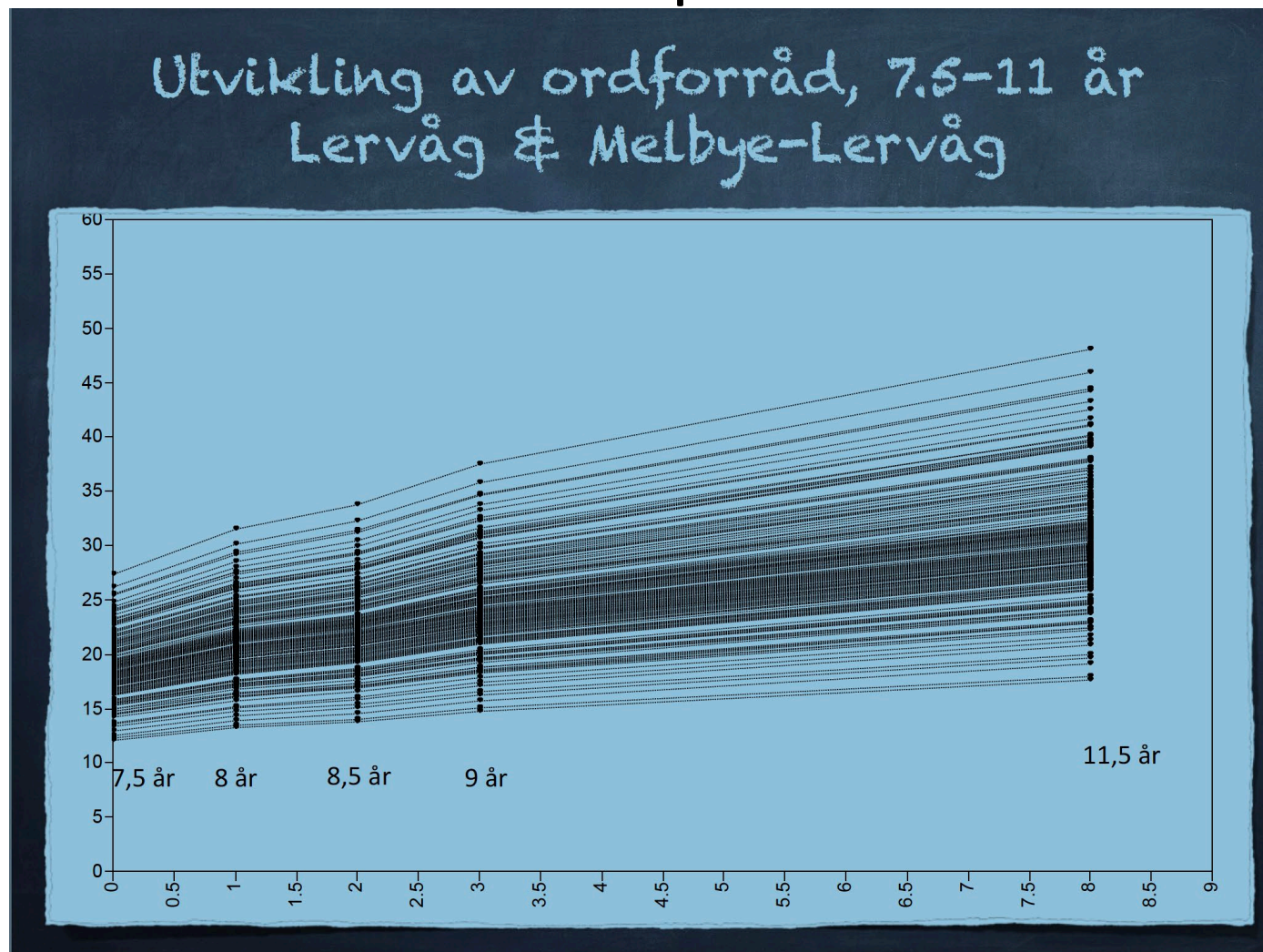
Longitudinelle studier på individnivå

- Endring over tid



- Måler vi det samme over tid?
- «Intercept» (nivå f.eks. ved T1)
- «slope» (endring over tid)

Longitudinelle studier på individnivå - Eksempel



Longitudinelle studier på individnivå - Eksempel


Research Report

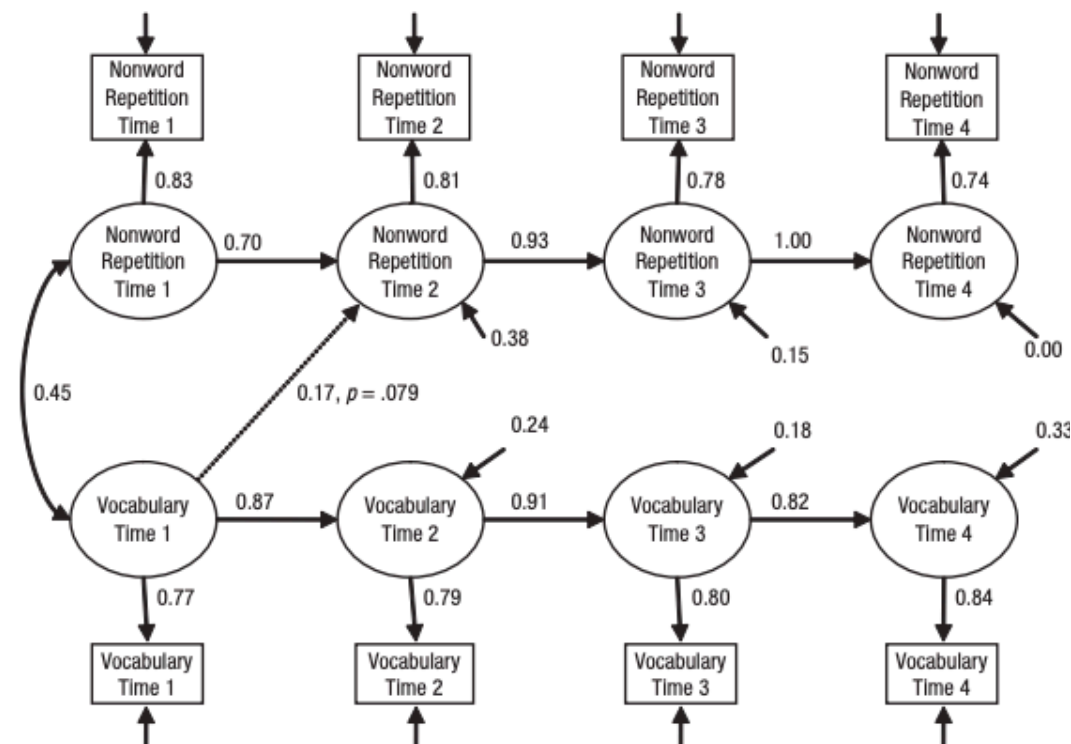
Nonword-Repetition Ability Does Not Appear to Be a Causal Influence on Children's Vocabulary Development

Monica Melby-Lervåg¹, Arne Lervåg², Solveig-Alma Halaas Lyster¹,
Marianne Klem¹, Bente Hagtvet¹, and Charles Hulme^{1,3}

¹Department of Special Needs Education, University of Oslo; ²Department of Educational Research, University of Oslo; and
³Division of Psychology and Language Sciences, University College London

PSYCHOLOGICAL SCIENCE

Psychological Science
23(10) 1092–1098
© The Author(s) 2012
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797612443833
<http://jps.sagepub.com>




Longitudinelle studier på individnivå - Eksempel

- Er det endring i kjønnsforskjell i ordflyt fra 2.-4. klasse?
- Er det sammenheng mellom tid brukt på lesetrening i 2. klasse og endring i ordflyt gjennom året?
- Er det endring i sosial deltagelse for barn med cochleaimplantat gjennom barneårene?

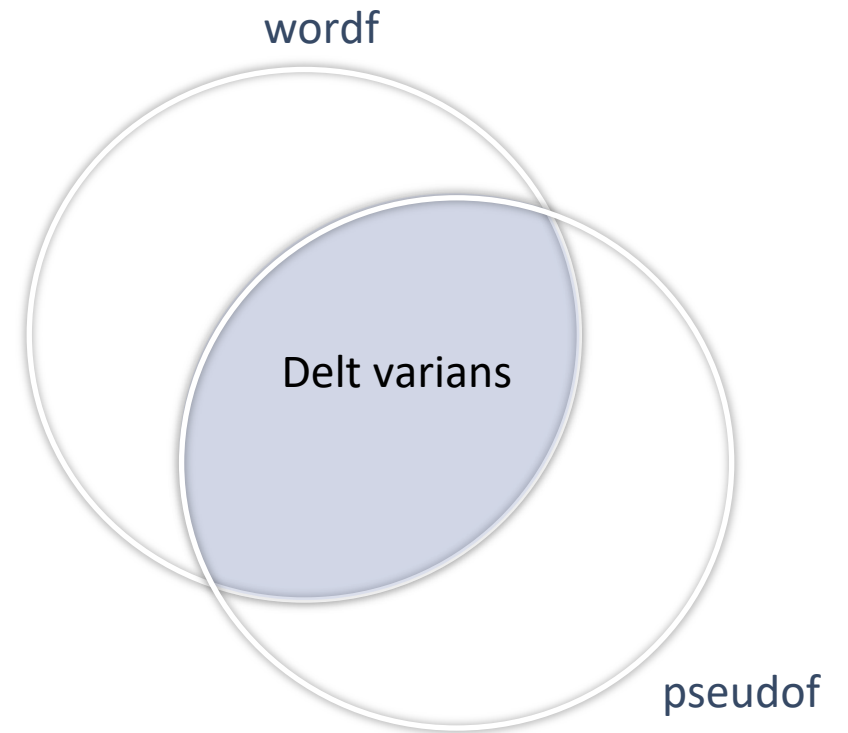
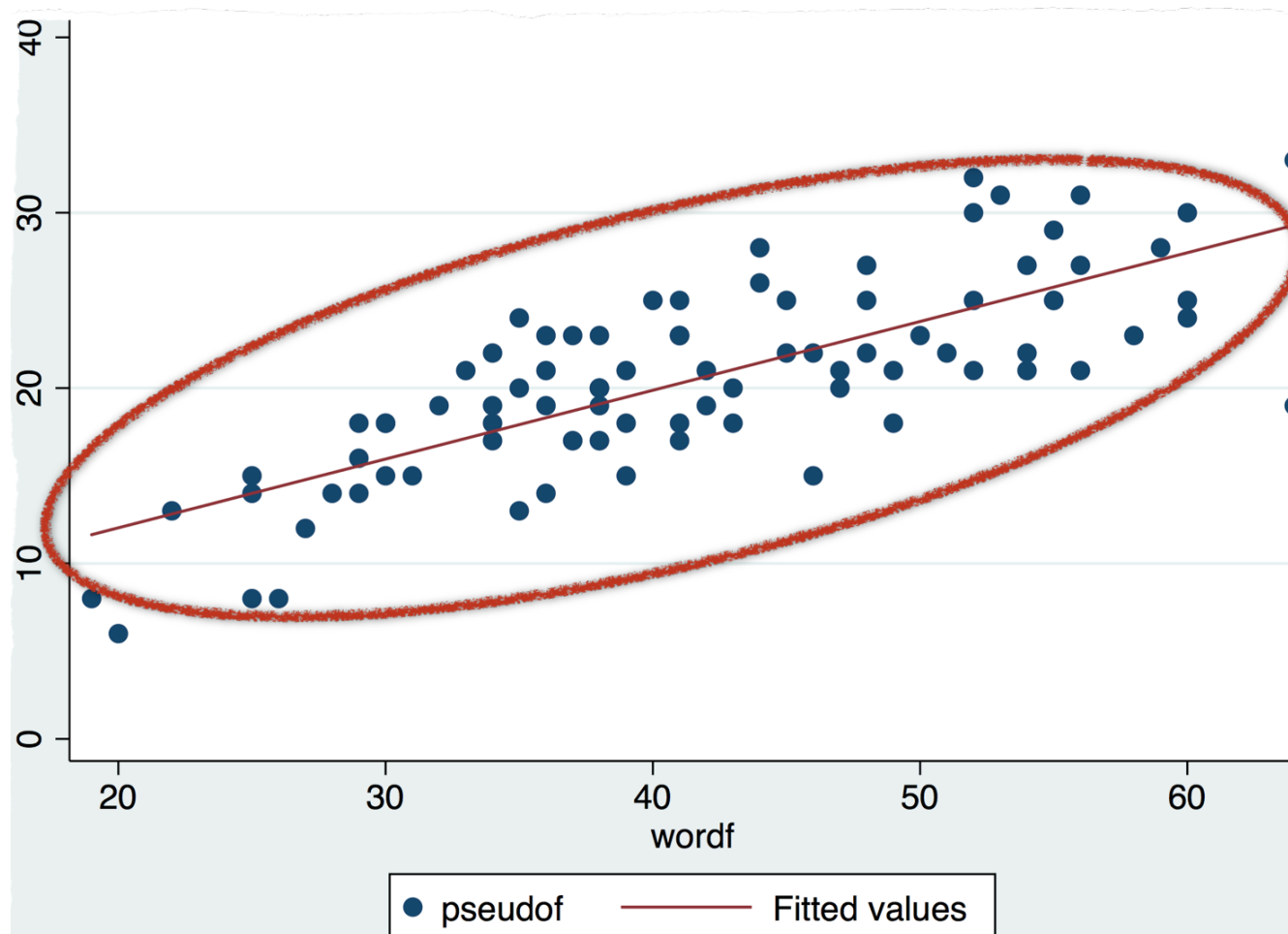
Hvordan styrke intern validitet?

- Kontrollere statistisk for alternative hypoteser
 - Skyldes forskjellen en tredjevariabel?
- Multivariat regresjon/variensanalyse
 - «kontrollere» for andre variabler
 - «ta bort» alternative forklaringer
- Effekt isolert for disse alternativene

Hvordan styrke intern validitet?

- Delt varians
 - Kontrollere / betinge
 - Tar bort delt varians mellom X, Y og kontrollvariabler
- Varians = Mål på spredning $\sigma = \text{SD}(X) = \sqrt{\text{VAR}(x)}$
- Eksempel: Er sammenhengen mellom ordflyt (wordf) og pseudoordflyt (pseudof) forklart av lesehastighet?

Plot og venn-diagram



Bivariat regresjon

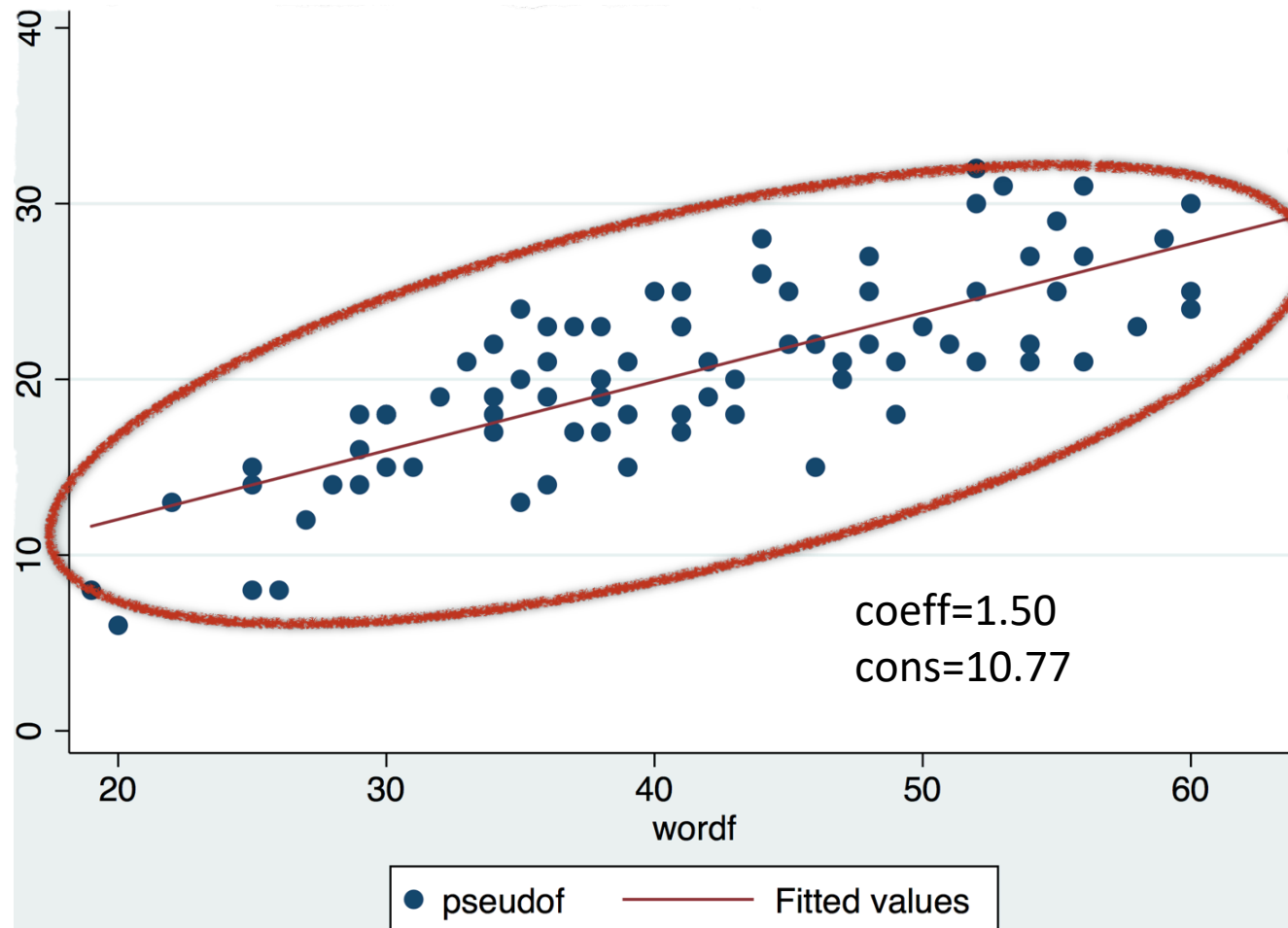
sum wordf pseudof

Variable	Obs	Mean	Std. Dev.	Min	Max
wordf	90	41.58889	10.64852	19	64
pseudof	90	20.5	5.438626	6	33

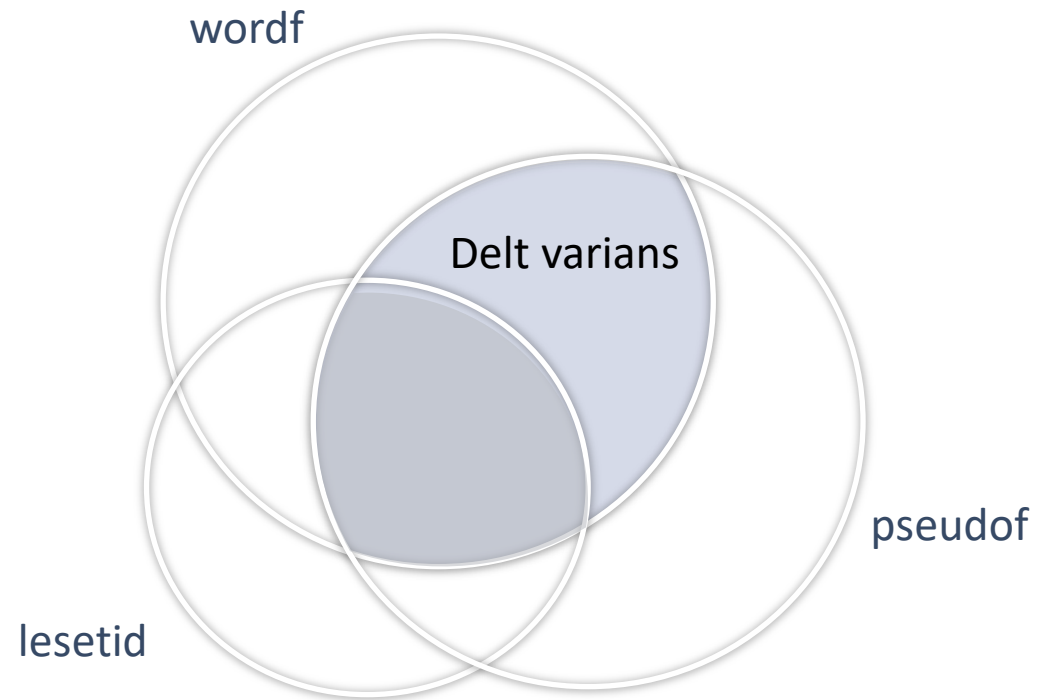
Source	SS	df	MS	Number of obs	=	90
Model	5949.40408	1	5949.40408	F(1, 88)	=	126.39
Residual	4142.38481	88	47.0725546	Prob > F	=	0.0000
				R-squared	=	0.5895
				Adj R-squared	=	0.5849
Total	10091.7889	89	113.390886	Root MSE	=	6.8609

wordf	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
pseudof	1.503324	.133721	11.24	0.000	1.237581 1.769066
_cons	10.77075	2.835075	3.80	0.000	5.136635 16.40487

Bivariat regresjon



Venn-diagram – legge til lesetid



Multippel regresjon

Sammenheng
wordf pseudof
«kontrollert»
for rtime

Variable	Obs	Mean	Std. Dev.	Min	Max
wordf	90	41.58889	10.64852	19	64
pseudof	90	20.5	5.438626	6	33
rtime_a	90	21.14849	10.71277	1.674	50.6385

Source	SS	df	MS	Number of obs	=	90
Model	5967.47659	2	2983.7383	F(2, 87)	=	62.94
Residual	4124.3123	87	47.4058885	Prob > F	=	0.0000
				R-squared	=	0.5913
				Adj R-squared	=	0.5819
Total	10091.7889	89	113.390886	Root MSE	=	6.8852

wordf	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
pseudof	1.862225	.5965649	3.12	0.002	.6764882	3.047963
rtime_a	-.1869985	.3028622	-0.62	0.539	-.7889699	.4149729
_cons	7.368004	6.20214	1.19	0.238	-4.95942	19.69543

Multippel regresjon

10 Ribeiro and Zachrisson

Table 2

Fixed Effects Models Predicting Teacher-Rated Child PA From Challenging Peers in the Group (Model 2, n = 956) Across Ages 2

	Model 1 (peer average PA)	
	Coefficient (SE)	ES
Peer average PA	.403 (.086)**	.22
One challenging peer	—	—
Two or more	—	—
Group size	−.008 (.009)	.004
How long known the child	.006 (.003)	.003
Hours in care	.004 (.005)	.002
Single mother	−.113 (.214)	.06
Maternal employment	−.010 (.074)	.005
Maternal depression	.048 (.12)	.03
Peer Average PA × Gender	—	—
One Challenging Peer × Gender	—	—
Two or More × Gender	—	—

* $p < .05$, ** $p < .01$ ES = effect size; PA = physical aggression

Statistisk kontroll – trusler mot intern validitet

- Har vi kontrollert for ALLE alternative hypoteser?
- Er alle målene våre gode nok?
 - f.eks. tid brukt på lesing; foreldres jobb
- Vanligvis kan vi ikke utelukke alle alternative hypoteser (hvordan kan vi vite?)
- Men: det er mulig
 - Hvis teorien er god nok... og målene er gode nok ... valide og reliable
 - ... Så kan en eksperimentell effekt gjenskapes

Multivariate Behavioral Research, 44:828–847, 2009
Copyright © Taylor & Francis Group, LLC
ISSN: 0027-3171 print/1532-7906 online
DOI: 10.1080/00273170903333673

 Psychology Press
Taylor & Francis Group

2008 Saul B. Sells Award Address Paper

How Bias Reduction Is Affected by
Covariate Choice, Unreliability, and
Mode of Data Analysis: Results From
Two Types of Within-Study
Comparisons

Thomas D. Cook and Peter M. Steiner
*Institute for Policy Research
Northwestern University*

Steffi Pohl
Friedrich-Schiller-Universität, Jena, Germany

Ikke-eksperimentell design

- Kryss-seksjonelle - ett tidspunkt
- Longitudinelle - flere tidspunkt
- Statistisk kontroll - «kontrollerer» for alternative hypoteser

Tenkeoppgave

- Sverige og Norge valgte i fjor vår ulike strategier for å håndtere pandemien
 - Sykdom, død, og økonomi har vært ulik
- Er dette en test på om den ene vs. den andre modellen var best?
- Hvordan kan vi lage et design som gir oss gode svar?

Systematic reviews

- Samle studiene som prøver å besvare den samme problemstillingen
- Syntesedannende undersøkelser, systematic review, metaanalyse
 - Sammenligne effektstørrelser, hva er gjennomsnittlig resultat og hvor mye varierer resultatene?
 - Hvorfor er resultatene forskjellig?
 - Hva lærer vi om forskningsspørsmålet når vi vi samler enkeltstudiene?
- Metoden kan også brukes til å få oversikt over et forskningsfelt eller forskningsspørsmål – systematisk litteraturgjennomgang
 - Oppsummere relevante studier
 - Målet kan være å få kunnskap om et felt eller finne trekk ved hva som fungerer, selv når studier ikke er helt sammenlignbare
 - Omfattende systematic review (f.eks. Cochrane) eller Rapid review (forenklet)

Systematic reviews

- Handler om ytre validitet
 - Kan funnene fra enkeltstudier generaliseres?
 - Hvilken effektstørrelse tror vi på?
- Kvalitative vurderinger i hva som skal med, hvilke begrensninger vi legger i utvalget av studier
- Kvalitative vurderinger i kvaliteten på studiene
- Kvantitative vurderinger når utvalgte studier sine effektstørrelser sammenlignes

Systematic reviews

- Idealet, «Best case»: samle sammen rådata fra masse ulike forsøk til et datasett og teste hypotese og underhypoteser
- Realiteten: samle sammen resultatet av studier
 - Telle +, - , 0-funn? Hva med størrelsen på effekten?
 - Glass (1976): Metaanalyse som statistisk analyse av effektstørrelser
 - (Effektstørrelser ble gjennomgått i forelesningen 14. oktober)
- Sammenligner vi epler og pærer
 - Er det samme tiltak?
 - Måler de samme utfall?
 - Bør et studie med lav styrke få like stor vekt som en med høy styrke?
 - Bør et studie med dårlig design få like stor vekt som et studie med godt design?

Systematic reviews

1. Problemstilling (forskningsspørsmålet)
2. Kriterier for inkludering (geografi, utvalg, publiseringstyper, tidsrom..) og utvikling av søkestrenger
3. Søk (google scholar, databaser for artikler, «snøballmetode», e-poster til forskere...)
4. Rensing av søk (f.eks. vurdere basert på abstract)
5. Kvalitetsvurdering av inkluderte studier (skjematisk analyse)
6. Sammenstilling av resultater + dokumentasjon på fremgangsmåte (effektstørrelser, eventuelt mer generelle funn)

Systematic reviews

- Forskeren blir en integrert del av arbeidet – umulig å unngå skjønnsmessige vurderinger
 - Bør ha god kjennskap til forskningsfeltet
 - Bør ha god kjennskap til forskningsdesign, også fra andre fagfelt
 - Ofte velger forskere bare å inkludere eksperimenter...
 - ...men kvasi-eksperimenter kan si noe om effekt dersom de har et godt design!
 - (hva er best, et eksperiment med 10 klasserom, eller et kvasi-eksperiment med alle landets klasserom?)
 - Gode kvasi-eksperimenter har som regel mye mindre effektstørrelser enn ikke-eksperimentelle metoder. Hva tror vi mest på?
 - Krever at forskeren kjenner til metodene

Systematic reviews

- Hvordan sammenligne effektstørrelser?
 - Standardisere forskjellene mellom tiltaksgruppen og kontrollgruppen
 - Glass $g = \frac{\bar{Y}_T - \bar{Y}_K}{s_K}$, Cohen $d = \frac{\bar{Y}_T - \bar{Y}_K}{s_p}$, s_p er veid gjennomsnitt av s_K og s_T
 - Pearsons r , korrelasjon mellom variabler (krever ikke mer standardisering)
 - Mulig å transformere mellom de ulike, og regne ut basert på ulike rapporterte test-statistikker (se Lund et al boka)
 - Cohen: liten/medium/stor effekt basert på r , MEN ikke å anbefale uten å ha tenkt seg nøye om!
- Så kan man lage et gjennomsnitt av alle effektstørrelsene, eventuelt med vektorer for antall observasjoner, standardavvik, osv.
- Og teste om effektstørrelsene er konsistente

Systematic reviews

- Analyse av tredjevariabler
 - Kan vi undersøke mulige årsaker til ulike effektstørrelser?

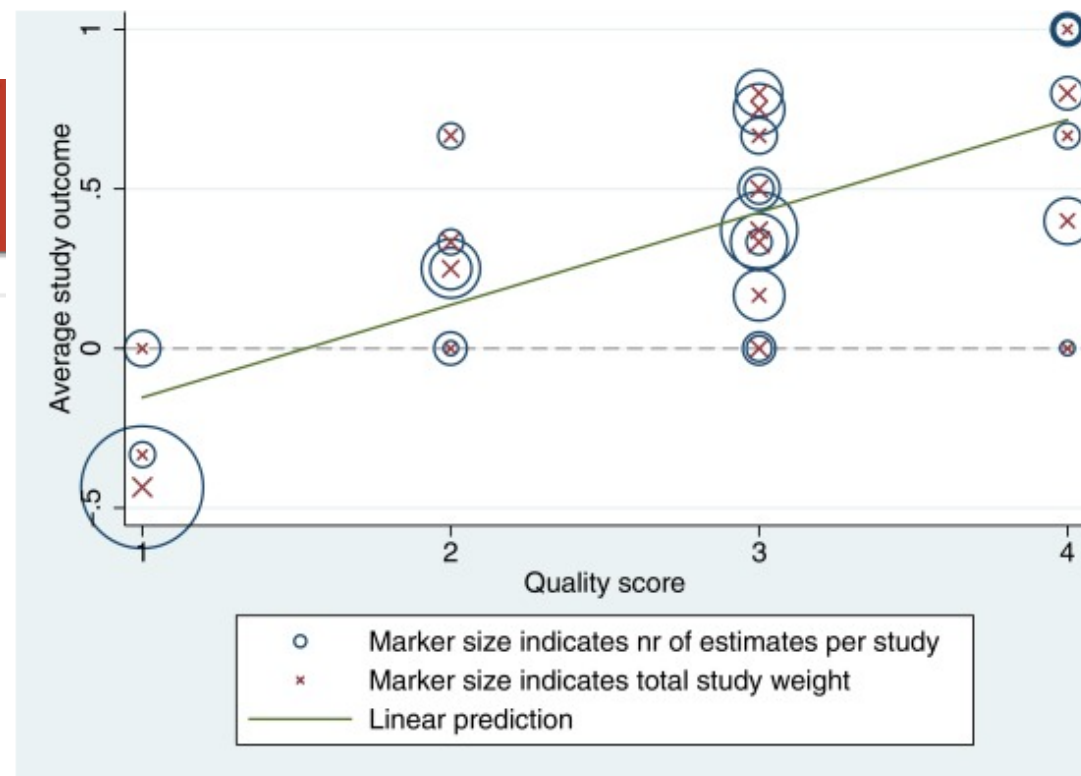


Economics of Education Review

Volume 66, October 2018, Pages 206-222

Do children benefit from universal early childhood education and care? A meta-analysis of evidence from natural experiments

Thomas van Huizen  , Janneke Plantenga



Systematic reviews

Working Memory Training Does Not Improve Performance on Measures of Intelligence or Other Measures of “Far Transfer”: Evidence From a Meta-Analytic Review

Monica Melby-Lervåg¹, Thomas S. Redick²,
and Charles Hulme³

¹Department of Special Needs Education, University of Oslo; ²Department of Psychological Sciences, Purdue University; and ³Division of Psychology and Language Sciences, University College London, and Department of Special Needs Education, University of Oslo

Perspectives on Psychological Science
2016, Vol. 11(4) 512–534
© The Author(s) 2016
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/17456916166635612
pps.sagepub.com



Phonological Skills and Their Role in Learning to Read: A Meta-Analytic Review

Monica Melby-Lervåg and
Solveig-Alma Halaas Lyster
University of Oslo

Psychological Bulletin
2012, Vol. 138, No. 2, 322–352

Charles Hulme
¹University College London and University of Oslo

The effect of linguistic comprehension instruction on generalized language and reading comprehension skills: A systematic review

Campbell Systematic Reviews. 2019;15:e1059.
<https://doi.org/10.1002/cl2.1059>

Kristin Rogde^{1,3} | Åste M. Hagen² | Monica Melby-Lervåg² | Arne Lervåg³

Preschool predictors of later reading comprehension ability: a systematic review

A Campbell Systematic Review
2017:14

Hanne Næss Hjetland, Ellen Irén Brinchmann, Ronny Scherer, Monica Melby-Lervåg