

Hvad måler Campbell Collaboration?

En kritisk kommentar til rapporten: 'School-Based Programs to Reduce Bullying and Victimization'

Af Inge Henningsen, Helle Rabøl Hansen og Dorte Marie Søndergaard

Campbell Collaboration har gennemført et omfattende arbejde med at indsamle data til en forskningsoversigt over antimobbeprogrammer. Der er meget information at hente i de 44 programmer, som rapporten inkluderer, og hvis hovedelementer er beskrevet i rapportens indledning. Campbells analyser har imidlertid så væsentlige begrænsninger, at centrale dele af rapportens konklusioner ikke kan siges at være holdbare. Vi har i det følgende eksemplificeret en række af vore indvendinger, og det er på den baggrund vores klare opfattelse, at rapporten ikke leverer den robuste og generaliserbare viden, der kræves, for at man kan tale om evidens i forhold til interventionsprogrammer.

I det lys er det næppe hensigtsmæssigt, når rapporten konkluderer, at tiden er moden til at udvikle et certificeringssystem (for antimobbeprogrammer) baseret på netop de fremgangsmåder, som rapporten har udpeget som virkningsfulde. Tilsvarende finder vi det betænkeligt, hvis rapportens konklusioner f.eks. i nationale publikationer om rapportens resultater fremstilles som fastslåede kendsgerninger – som et mindstemål må der refereres til de problematiseringer, som Campbell rapporten selv indeholder.

Det følgende er således ikke et direkte svar på de spørgsmål, som SFI Campbell har stillet, men en review kommentarer der knytter sig til rapportens grundlæggende metodiske og indholdsmæssige præmisser.

Udvælgelsesmetode

Campbell Collaboration har gennemgået 622 artikler/rapporter, der beskæftiger sig med antimobbeprogrammer. Af disse indgår 44 (7.2 %) i analysen, idet man kun har inkluderet undersøgelser, der indeholder sammenligning af en interventionsgruppe med en kontrolgruppe. De inkluderede programmer er således alene udvalgt på grundlag af deres interne evalueringskriterier. Campbell Collaborations koncept forpligter på den måde de forskere, der skriver forskningsoversigten, til at fravælge ganske omfattende mængder af forskningskundskab om det emne, de har i fokus – og fravælgelsen sker ikke på grundlag af kundskabens relevans endsige kvalitet, men alene på grundlag af de evalueringsmetoder, der har været knyttet til interventionsprogrammet.

Ambitionen med at udvikle metaanalyser hviler på en antagelse om, at en samlet analyse vil kunne bidrage til at skabe mere sikker viden end den, man kan få fra de enkelte undersøgelser. Metaanalyserne tænkes at kunne finde sammenhænge på tværs af undersøgelser og at bidrage til at gardere mod resultater, der kun beror på tilfældigheder. Campbell Collaboration har gjort et omfattende arbejde i forbindelse med udvælgelse og klassifikation af undersøgelser. Deres analyser har dog som nævnt

så væsentlige begrænsninger, at visse af rapportens konklusioner ikke kan siges at være holdbare.

Studiedesignet påvirker effektstørrelsen

Campbell rapporten baserer sig på 44 programmer med et meget varierende deltagerantal. Programmerne indeholder forskellige interventionselementer (Figur 2 og 3 i Farrington & Ttofi, 2009), de har forskelligt studiedesign, og der er brugt forskellige metoder til effektevaluering. Udvalgelseskriteriet har været, at programmerne (i generaliseret form) indeholder 'før og efter målinger' for interventionsgruppen og for en kontrol/sammenligningsgruppe, som ikke har modtaget nogen form for intervention. I analysen vurderes virkningerne af interventionen ved at sammenligne effekterne i interventions- og sammenligningsgruppen. Resultaterne fra de enkelte programmer er omregnet til et fælles effektmål (odds-ratio (OR), se Farrington & Ttofi p. 77¹) for reduktion af mobning (både blandt mobbere og mobbeofre). Programmerne er klassificeret i fire grupper afhængigt af det studiedesign, der danner baggrund for effektevalueringen (Se tabel 1):

Tabel 1. *Gennemsnitlige estimerede effektstørrelser for mobning. Samlet og opdelt efter studiedesign*

Studiedesign	Effektstørrelse (OR)	
	Mobbere	Mobbeofre
Randomiserede, kontrollerede forsøg	1.12	1.14
Studier med før- og efter målinger	1.60 ^{**})	1.20 ^{*)}
Andre studier med interventions- og kontrolgruppe	1.20 ^{*)}	1.41 ^{**})
Alder-kohorte studier	1.51 ^{**})	1.44 ^{**})
Alle	1.37 ^{**})	1.29 ^{**})

Kilde: *Farrington & Ttofi (2009) tabel 8 og 9* ^{*)} signifikant $p < 0.05$ ^{**}) signifikant $p < 0.001$

Man kan i tabel 1 konstatere, at programmernes (estimerede) effekter ser ud til at afhænge af hvilket studiedesign, der er blevet anvendt. I de programmer, hvor man har foretaget en tilfældig udvælgelse af kontrolgrupper (randomisering), har programmerne ingen signifikant effekt hverken for mobbere eller for mobbeofre. I modsætning hertil viser programmer målt ved alder-kohorte studier i gennemsnit stærkt signifikante effekter i begge grupper. En sammenligning af de estimerede effektstørrelser for de fire typer af studiedesign viser da også, at de er signifikant forskellige ("we can conclude that effect sizes varied significantly across research designs" (ibid, p. 60)). Studiernes målte effekt har altså sammenhæng med, hvordan resultatevalueringen rent teknisk er tilrettelagt.

Der kan tænkes forskellige forklaringer på forskellene mellem de fire grupper: De kan skyldes statistiske tilfældigheder (usandsynligt i så stort materiale); de kan skyldes, at de fire evalueringmetoder måler forskelligt; at de "bedste" programelementer er knyttet til nogle bestemte evalueringdesign; eller at bestemte design fortrinsvis er

¹ OR=1 svarer til at der ikke har været nogen effekt, mens $OR > 1$ svarer til en positiv og $OR < 1$ svarer til en negativ effekt

brugt under særligt gunstige/ugunstige eksterne omstændigheder. Man kan også forestille sig, at der er tale om en kombination af disse forhold. Det er imidlertid afgørende for konklusionernes validitet, at programmernes effekt ikke bare er et artefakt fremkommet som konsekvens af mindre stringente studiedesign.

Campbell Collaborations rapport beskæftiger sig kort med fordele og ulemper ved de fire inkluderede studiedesign.

Overall, we conclude that these are the best four designs that have been used to evaluate the effects of anti-bullying programs, and we give credence to results obtained in all of them.

Farrington & Ttofi 2009:61

Er dette mon udtryk for, at man mener, at de fire studiedesign alle er lige troværdige, således at de fundne forskelle kan siges at repræsentere egentlige substantielle fund? Om det er rigtigt eller forkert er vanskeligt at afgøre, fordi metaanalysen indeholder en meget lang række af mulige kilder til inhomogenitet, der spiller sammen på komplicerede måder (studiedesign, geografisk område, programelementer, hvad er brugt som mobbedefinition, hvordan mobning er målt, etc.).

I en evidenssammenhæng er det første krav imidlertid, at man sammenligner "lige med lige", altså at de grupper, man betragter, er sammenlignelige i alle andre henseender end den, man er i gang med at undersøge. Vi skal i det følgende gå i detaljer med dele af metaanalysen og vise eksempler på, at dette ikke ser ud til at være tilfældet.

Virkningsfulde programmer

I rapporten udpeges 19 programmer, der ser ud til at være ("appeared to be") effektive (Farrington & Ttofi, 2009: p. 58), når det gælder om at reducere mobning. Grundlaget her er en statistisk vurdering af effektstørrelserne i de enkelte studier og ikke metaanalysen. Disse programmer har ifølge målingerne meget varierende effekter, og de enkelte programmer kan indenfor målingerne vise sig at have meget forskellig effekt, når det gælder at reducere hhv. antal mobbere og antal mobbeofre. Her giver metaanalysens systematiseringer anledning til nogle interessante spørgsmål.

Der er f.eks. kun et ud af ni randomiserede studier, som viser en signifikant effekt på mobning, mens 7 ud af 9 interventionsprogrammer med alder-kohorte design gav signifikant positive resultater. Er dette udtryk for faktiske effekter, eller er det sådan, at nogle af disse programmer blot fremstår som effektive, fordi det benyttede studiedesign (alders-kohorte designet) har lav validitet? Om programmerne reelt reducerer mobning, eller om det bare ser sådan ud, fordi der ikke er kontrolleret effektivt for konfunderende effekter, er afgørende for vurderingen og brugen af Campbells analyse. Man kunne have forventet en nærmere undersøgelse af dette forhold. Nedenfor er samspillet mellem effekt og studiedesign elaboreret.

Samspil mellem lokalitet og studiedesign

I det følgende betragtes de tre lokaliteter Norge (7), Europa (uden Norge) (17), US/Canada/Australien/New Zealand (16) med antal programmer i parentes². (Se

² En lille undersøgelse fra Sydafrika er udeladt af analysen, fordi den ikke rigtigt hører til nogen steder.

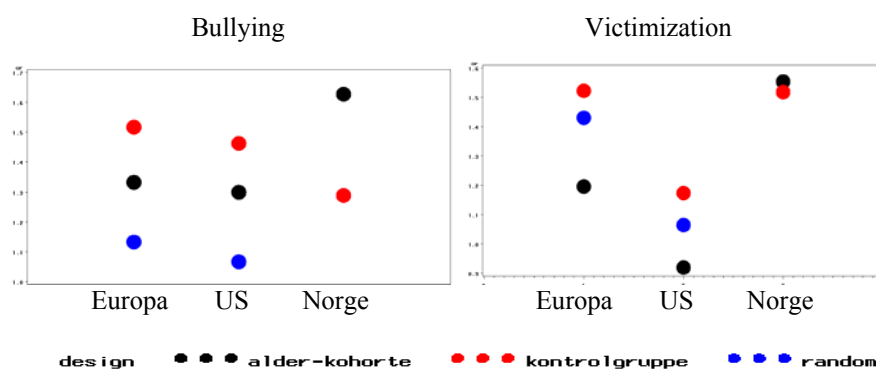
Farrington & Ttofi, 2009, figur 3). I tabel 2 er effektstørrelser for mobbere og mobbede for de tre lokaliteter beregnet.

Tabel 2 Effektstørrelser for mobbere og mobbede opdelt efter lokalitet.

Lokalitet	OR (Bullying)	OR (Victimization)
Europa (uden Norge)	1.32	1.30
US/Canada etc.	1.33	1.13
Norge	1.58	1.55

Der er signifikant forskel på effektstørrelser mellem lokaliteter. Dette er ikke overraskende, da mobbeintervention må formodes at være kontekstafhængig, således at det ikke er de samme programmer, der virker alle steder. I de inkluderede programmer er der således en kompliceret sammenhæng mellem studiedesign og lokalitet, idet alle studiedesign ikke er brugt lige meget alle steder (US/Canada har mange randomiserede undersøgelser, Norge har fortrinsvis alder-kohorte design.) Nedenstående figurer viser, hvad der sker med OR for hhv. mobbere og mobbede, når man kontrollerer for både land og studiedesign.

Figur 1. OR for bullying and victimization opdelt efter lokalitet og studiedesign



Det fremgår klart af disse figurer, at studiedesignet uanset lokaliteten påvirker den estimerede effektstørrelse, og at interventionseffekten – særligt for mobbede – ser ud til at afhænge af lokaliteten. Men man ser også, at de norske undersøgelser er specielle (ikke bare er effekten højere, men virkning af studiedesign er byttet om i forhold til undersøgelser på andre lokaliteter). Det er vores opfattelse, at man må have afklaret det inhomogene forhold mellem lokalitet og studiedesign, før resultaterne fra metaanalysen generaliseres.

Virkningsfulde design- og programelementer

Campbell Collaboration har suppleret den samlede analyse med analyser af en række design- og programelementer, hvis effektivitet man ønsker at identificere særskilt (Farrington & Ttofi 2009, fig. 2 og 3). Dette er gjort gennem en række separate test (marginale analyser), hvor man for hvert element har sammenlignet de gennemsnitlige effektstørrelser i programmer med og uden det pågældende element (Farrington & Ttofi 2009, tabel 12 og 13). På den måde identificeres visse design- og programelementer som virkningsfulde på grundlag af statistisk signifikans.

Denne undersøgelse af enkeltstående design- og programelementer har imidlertid nogle betydelige svagheder. Da elementerne optræder sammen i programmerne (i vekslende kombinationer), har man i de udførte statistiske analyser ikke kontrolleret for indflydelse fra andre programelementer.

Variables that might help to explain differential treatment effects in meta-analysis (e.g. elements of the intervention) cannot be assumed to be statistically independent. Researchers should try to disentangle the relationships among them and identify those that truly have significant independent relationships with effect sizes³. Multivariate techniques can be used to solve this problem in meta-analysis

Farrington & Ttofi 2009:67

Og videre

It should be borne in mind, however, that we have discovered the program elements that are most highly correlated with effectiveness. This does not prove that they cause effectiveness [...]

Farrington & Ttofi 2009:69

Campbell Collaboration er således ganske klar på, at man ikke kan vide, om man har fundet de programelementer, der ”virker”. Imidlertid spiller disse begrænsninger/ advarsler en ringe rolle andre steder i Campbell-rapporten, hvor resultaterne fra de enkeltstående analyser tværtimod behandles, som om deres holdbarhed er dokumenteret. Det er således også disse resultater, der uden nogen form for forbehold refereres som ’fund’ i udkastet til den danske artikel fra SFI Campbell. Vil man påstå robust viden om enkeltelementers virkning, er det imidlertid afgørende, at man rent faktisk finder de elementer, der virker, ikke bare dem der i et udvalg af interventionsprogrammer har optrådt sammen med dem, der virkede. Vi skal i det følgende afsnit argumentere for, at der i Campbell Collaborations rapport i høj grad må antages at være tale om sådan ”spurious correlation”.

Ovenfor viste vi, at effekten af interventionsprogrammerne varierede med lokalitet, selv når man kontrollerede for effekten af studiedesign, og forfatterne konstaterer da også, at Norge spiller en helt speciel rolle.

Programs implemented in Norway seem to work best and this could be related to the long tradition of bullying research in Scandinavian countries. Other factors are that Scandinavian schools are of high quality, with small classes and well-trained teachers, and there is a Scandinavian tradition of state intervention in matters of social welfare.

Farrington & Ttofi 2009:73

Hvis det imidlertid er sådan, at der er i Norge er særligt gunstige forhold for mobbeintervention, så vil programelementer, der fortrinsvis er med i de norske undersøgelser komme til at fremtræde som virkningsfulde – uanset om de er det eller ej – fordi de afprøves under mere gunstige omstændigheder end programelementer, der anvendes i andre lande⁴.

³ Det anføres, at multivariate analyser kan afhjælpe disse problemer, og der udføres faktisk multivariate analyser, hvor kun et begrænset antal af de udpegede elementer viser sig at have en virkning, men disse fund spiller ingen væsentlig rolle i de senere konklusioner.

⁴ Når Norge kan få mærkbar indflydelse skyldes det, at metaanalyser indeholder en del meget store norske undersøgelser.

For at undersøge dette, har vi gentaget Campbell Collaborations analyse af program- og designelementer uden de 7 norske programmer. Hvis de signifikante effekter faktisk var udtryk for generaliserbare sammenhænge, skulle man forvente at finde samme signifikante effekter i den nye analyse. Tabel 3, der bygger på Farrington & Ttofi 2009, tabel 12, 13 og 14 og egne beregninger, viser hvilke program- og designelementer, der er fundet signifikante i Campbell rapportens marginale analyser. Tabellen viser yderligere, hvilke elementer der ville være signifikante, hvis de norske interventionsundersøgelser ikke var inkluderet; og hvilke der i rapporten er fundet signifikante ved den multivariate analyse. Som man ser, finder Campbells marginale analyser langt flere signifikante programelementer, end man finder, når Norge er udeladt, eller når der er kontrolleret for samvariation gennem en multivariat analyse. Dette kan kun tages som et tegn på, at en lang række af de signifikante program- og designelementer, som Campbell Collaboration har fundet, ikke er virkningsfulde i sig selv, men kun gennem deres association med andre virkningsfulde elementer (spurious correlation). Det er endvidere bemærkelsesværdigt, at visse programelementer (*Disciplinary methods* og *Based on Olweus*) ligefrem skifter ”fortegn” og går fra at have ingen eller signifikant positiv effekt til at have en negativ effekt, når de norske undersøgelser udelades.

Tabel 4 Signifikante program- og designelementer i tre analyser

	Alle programmer (Marginal)		Uden Norge (Marginal)		Alle programmer (Regression)	
	Mobbere.	Mobbefre.	Mobbere.	Mobbefre.	Mobbere.	Mobbefre.
Program Elements						
Parent training/meetings	▲	▲	▲		▲	
Playground supervision	▲					
Intensity for children	▲	▲			▲	
Intensity for teachers	▲	▲	▲	▲		
Duration for children	▲	▲	▲			▲
Disciplinary methods	▲	▲		●		
Duration for teachers	▲	▲		+		
Classroom management	▲					
Teacher training	▲					
Classroom rules	▲					
Whole-school policy	▲					
School conferences	▲					
Information for parents	▲					
Cooperative group work	▲					
Work with peers		●				●
Videos		▲				▲
Design Features						
Age of children	▲	▲	▲			
Outcome measure	▲	▲		+		
Publication date	▲	▲		+		
Based on Olweus	▲			●		
Total elements	▲					
In Norway	▲	▲				
Not in US and Canada		▲		▲		
Sample size	+					
Design		▲				

Kilde: Farrington & Ttofi 2009, tabel 12, 13 og 14 og egne beregninger. ▲ $p < 0.05$, + $p < 0.10$, ● $p < 0.05$ (negativt fortegn)

Ovenstående overvejelser peger på, at resultaterne af metaanalysen meget langt fra er entydige, og at de ikke giver baggrund for håndfaste konklusioner. Vi vil ikke påstå, at de resultater, vi har fundet, når vi f.eks. udelader de norske resultater, er ”sandere” end resultaterne fra den fulde analyse, men de er anderledes, og det burde de ikke være, hvis resultaterne fra Farrington & Ttofi var robuste og generaliserbare.

Disciplinary methods

Disciplinary methods var et af de programelementer, der skiftede fortegn alt efter hvilke andre programelementer, de blev associeret med. I den samlede analyse havde de programmer, der indeholder elementet *Disciplinary methods*, signifikant bedre resultater end de programmer, der ikke indeholdt elementet. (Den gennemsnitlige effekt estimeredes til OR=1.44 mod OR=1.21 for dem, der ikke indeholdt elementet). Når de 7 norske programmer fjernes, giver de programmer, der indeholder elementet *Disciplinary methods*, signifikant dårligere resultater end de programmer, der ikke indeholder elementet. (Her estimeres den gennemsnitlige effekt til OR=1.00 mod OR=1.28 for dem, der ikke indeholder elementet). I tabel 5 er programmer, der indeholder programelementet *Disciplinary methods* opdelt efter geografisk område.

Tabel 5 *Programmer, der indeholder programelementet Disciplinary methods opdelt efter geografisk område. Effekt for mobbeofre*

	Program	OR for mobbeofre
Europa (uden Norge)	Sprober et al. (2006)	1.15
	Fekkes et al. (2006)	1.25
	Rican et al. (1996)	2.46
	Alle	1.26
US/Canada	Kaiser-Ulrey (2003)	0.65
	Rosenbluth et al. (2004)	0.70
	Pagliocca et al. (2007)	0.92
	Pepler et al. (2004)	0.94
	Melton et al. (1998)	1.06
	Alle	0.95
Norge	Olweus/Bergen 2	1.43
	Olweus/Oslo 2	1.48
	Olweus/New National	1.59
	Olweus/Oslo 1	1.81
	Olweus/Bergen	2.89
	Alle	1.62

De europæiske programmer har en effekt, der ligger helt på linje med de programmer, der ikke indeholder elementet (OR=1.26 mod OR=1.21). Derimod ligger programmerne fra US/Canada markant under og programmerne fra Norge markant over de øvrige programmer. Disse resultater kalder på en nøjere analyse af, hvordan og hvorfor resultaterne på denne måde hænger sammen med geografisk område, og det burde være oplagt, at metaanalysen i sin nuværende form ikke giver baggrund for

at udtale sig om virkningen af programelementet *Disciplinary methods* i forhold til at reducere antal mobbeofre.

En anden vinkel

I de foregående analyser har vi holdt os inden for metaanalysens forståelsesramme, hvor man nok taler om samvariation, men ikke om synergi. Man kunne imidlertid anlægge en helt anden vinkel på problematikken.

Måske er det netop ikke særlig hensigtsmæssigt at adskille og måle effekt af disse elementer enkeltvis. Måske er det den samtidige tilstedeværelse af mange elementer – det at problematikken mødes på mange niveauer, fra mange vinkler, gennem mange forskellige kræfter og kanaler – der afstedkommer effekt? Frem for at søge enkeltelementers effekt og måle den som isoleret fænomen, burde man måske tænke i evalueringsmetoder, som netop havde fokus på sammenhænge mellem forskellige spektre af elementer og effekter skabt gennem netop sådanne strategisk tilrettede flervinklede indsatser.

Rapportens mobbebegreb

Campbell rapportens definition og afgrænsning af mobbebegrebet er hentet fra det, man kan kalde for mobbeforskningens første paradigme, hvor mobning forstås som gentagende aggressive handlinger fra en eller flere personer over tid i et asymmetrisk magtforhold, og hvor fokus primært rettes mod individuelle egenskaber hos hhv. mobber og mobbet. Grundlæggeren og den førende teoretiker indenfor dette paradigme er svensk/norske professor Dan Olweus. I rapportens afsnit om mobbedefinition citeres dog flere mobbeforskere. En del af disse forskere bl.a. Roland (1989) og Samivalli (2002) har delasppekter med i deres mobbeafgrænsning, som indikerer, at mobning her forstås som et fænomen, der udover individuelle egenskaber også indeholder visse sociale dynamikker. Samtlige nævnte referencer i rapporten er, trods henvisningen til forskere der også inkluderer disse delasppekter i deres definitionspraksis, alle hentet fra mobbeforskningens første paradigme, hvis fællesnævner og udgangspunkt er en forståelse af mobning som et møde mellem personer med svage og stærke personsegenskaber.

Fra årtusindeskiftet suppleredes imidlertid det første paradigme med andre grundlæggende forståelser af mobning. Eriksson et al. (2002) påpeger, at mobning er et kontekstsensitivt fænomen. I Danmark har de børneretlige miljøer (Børnerådet, Børns Vilkår, Red Barnet) igennem de sidste 10 år kritiseret den individorienterede tilgang til mobbeproblematikken og fremhævet eksempler på, at en sådan forståelse kan føre til eksklusion af enkeltelever fra lokale skolemiljøer. I eXbus arbejder vi med en tilgang, der sætter fokus på mobning som resultatet af et samspil mellem mange (sociale, kulturelle, historiske, individuelle og materielle) kræfter i og omkring skoleklasser. Den socialdynamiske mobbeforståelse i NGO-miljøerne og blandt et stigende antal forskere anser mobning som et gruppefænomen, der opstår og bliver til i sociale sammenhænge som f.eks. en skoleklasse, hvor børn er henvist til at fungere sammen qua formaliserede rammer.

I Campbell-rapporten er det socialdynamiske paradigme (andet paradigme) kun repræsenteret i minimal grad, og når det sker, da i forbindelse med mindre eller

sekundære dele af de evaluerede programmer. Campbell-rapporten kan i den forstand ikke anses for at levere en bred og gennemgribende undersøgelse af antimobbeprogrammer i almindelighed, men i højere grad ses som en undersøgelse af første-paradigme programmer. Disse programmer er de ældste i feltet og har derfor en længere implementeringshistorie bag sig. Samtidig er der qua definitionstypen og altså det første paradigmes mere statiske problem- og personforståelse, tale om en type programmer, der har langt nemmere ved at korrespondere med en så enkel evidensforståelse som den, der ligger til grund for Campbell-rapportens arbejde. En fænomenforståelse, der fordrer større sensitivitet overfor et fænomens sociale kompleksitet, sådan som tilfældet er i det andet paradigme, ville selvsagt kræve evidensbaseret foretaget gennem måleredskaber, som var i stand til at adressere og fange kompleksitet – og det er der ikke tale om med de metoder, som Campbell rapporten gør brug af. Det kan være en af årsagerne til, at det netop er disse 44 programmer, der er udtaget til evidensanalyser.

Konklusioner

Campbell Collaboration har udarbejdet en omfattende forskningsoversigt over antimobbeprogrammer. Der er, som vi nævnte indledningsvis, meget information at hente i de 44 programmer, som er inkluderet i metaanalysen, og hvis hovedelementer er beskrevet kort i rapportens indledning. Så meget desto mere kan det undre, at rapporten grundet sin bundethed til et ganske bestemt udvælgelseskoncept med baggrund i en særlig form for evalueringsstandarder, men uden begrundelse i programmernes indhold, kommer til at negligere så stor en del af de indsatser og de kundskaber om mobning og antimobbepraktisser, som det her bliver tilfældet. I alt 578 af de identificerede programmer er således frasorteret i rapportens oversigt. Udvalgsstrategien kommer på den måde til at fungere direkte kontraindikativt i forhold til en ambition om at generere stadig mere omfattende kundskab om mobning som fænomen.

Givet det metaanalytiske arbejdes begrænsninger kunne man have ønsket sig en mere forsigtig formulering af rapportens konklusioner.

Efterspørgslen efter ny viden og nye handlemuligheder er meget stærk på dette felt. I det lys er det uheldigt, hvis rapporter af denne art snarere kommer til at bremse end at bidrage støttende til kundskabsudviklingen. I Campbell-rapporten bliver det første tilfældet, når det konkluderes, at der nu er genereret tilstrækkelig meget og tilstrækkelig robust kundskab til, at der kan udvikles et system for certificering baseret på netop de fremgangsmåder, som rapporten har udpeget som virkningsfulde. Som indledningsvis nævnt mener vi netop ikke, at rapporten giver den robuste og generaliserbare viden, der kræves for at kunne tale om evidens i forhold til interventionsprogrammer, langt mindre giver baggrund for at iværksætte et akkrediteringsprogram som foreslået af Campbell Collaboration.

Reference:

Farrington, David P. & Maria M. Ttofi (2009) School-Based Programs to Reduce Bullying and Victimization. Campbell Systematic Reviews 2009:6. Open access article: www.campbellcollaboration.org.