

Nonrespons Bij Schriftelijke Vragenlijsten, Telefonische Interviews en Face-to-face Interviews; Een Vergelijkende Meta-analyse

Joop J. Hox¹ en Edith D. de Leeuw²

Summary

Nonresponse in mail, telephone, and face-to-face surveys; A comparative meta-analysis.

This article reports a meta-analysis of 45 studies that explicitly compare the response rates obtained using a mail, telephone or face-to-face survey. The data analysis uses a generalized hierarchical linear model. Sampling procedure (local, select versus random, general), saliency of topic, and researcher's background (university, government vs. marketing research) had a positive effect on the response. The face-to-face condition achieved the highest response rate (80.5%), the telephone survey the next highest (78.2%), and the mail survey the lowest (73.5%). There is a significant interaction with the year of publication: the response to face-to-face and telephone surveys is going down in the time period covered by the meta-analysis (1947 to 1992), but the response to mail surveys is going up slightly. We attribute this to the large amount of research on response problems with mail surveys, and recommend more research and development in this direction for face-to-face and telephone methods.

¹Vakgroep Pedagogische Wetenschappen, Faculteit POW, Universiteit van Amsterdam

²Vakgroep Methoden en Technieken, Faculteit SCW, Vrije Universiteit

Correspondentie: J.J. Hox, Pedagogische Wetenschappen, UvA, IJsbaanpad 9, 1076 CV Amsterdam. Dit artikel is gebaseerd op een paper gepresenteerd op de 3e International Workshop on Household Survey Nonresponse, Voorburg, CBS, 14-16 sept. 1992.

1. Inleiding

Eén van de belangrijkste problemen die onderzoekers tegenkomen bij het uitvoeren van een surveyonderzoek is het probleem van de nonrespons; het niet kunnen verkrijgen van metingen van alle beoogde respondenten. Nonrespons dient te worden onderscheiden van een andere, vergelijkbare foutenbron, namelijk het ontbreken van sommige elementen in de steekproef omdat het betreffende populatiesegment ontbreekt in het operationele steekproefkader (de zogenaamde coverage error, men denke aan 'onbekend adres' bij het werken met adreslijsten, of 'geen telefoon' bij telefonisch onderzoek). Bij nonrespons zijn de betreffende elementen wel degelijk in de operationele steekproef opgenomen, maar om een of andere reden (onbereikbaar, weigering) worden van hen geen gegevens verkregen (Groves & Lyberg, 1988).

De respons bij een surveyonderzoek wordt beïnvloed door een groot aantal factoren, zoals het onderwerp, het soort steekproef, of de onderzoeksorganisatie van de onderzoekers (cf. Goyder, 1987). Ook de methode van dataverzameling is van belang. In het algemeen resulteren face-to-face interviews in een hogere respons dan vergelijkbare telefonische of postenquêtes. Zo rapporteren Groves en Kahn (1979, p. 76) een respons van 74% voor een face-to-face survey en 70% voor een gelijksoortige telefonische enquête. Steeh (1981) rapporteert een daling van de respons van 73% naar 66% bij een overgang van face-to-face naar telefonisch onderzoek. Goyder, tenslotte, vergelijkt in zijn boek (Goyder, 1987) de responsgegevens van 385 postenquêtes, 112 face-to-face surveys, en 53 telefonische surveys, uitgevoerd in de Verenigde Staten en Canada. Hij vindt een gemiddelde respons van 67.3% voor face-to-face interviews, 60.2% voor telefonische surveys, en 58.4% voor postenquêtes.

Er zijn aanwijzingen dat de respons bij met name face-to-face surveys in de meeste westerse landen in de loop van de tijd lager is geworden. Steeh (1981) rapporteert een daling bij twee grote face-to-face surveys in de Verenigde Staten, Goyder (1987) rapporteert een dalende respons voor de Verenigde Staten en Canada, en Sugiyama maakt melding van een dalende respons bij Japans surveyonderzoek (Sugiyama, 1991). In Nederland rapporteren Bethlehem en Kersten (1981, 1986) nonrespons ratio's voor CBS surveys die oplopen van 13% à 18% in de zeventiger jaren tot 18% à 52% in 1983. Het CBS vindt daarbij geen opvallende verschillen tussen face-to-face en telefonische surveys (Kerssemakers, 1985).

Een hoge respons bij een survey onderzoek is met name belangrijk, omdat anders het gevaar bestaat van vertekening van de onderzoeksresultaten als gevolg van selectieve nonrespons (nonresponse bias). Alle onderzoeksorganisaties die gebruik maken van survey onderzoek worden geconfronteerd met het respons probleem, en er is dan ook een groot aantal onderzoeken uitgevoerd naar de effectiviteit van verschillende maatregelen om de respons te

verhogen (voor het Nederlandse taalgebied zie daarvoor o.a. Nederhof, 1982; Hox, De Leeuw & Duijx, 1984). De hoeveelheid onderzoek op dit terrein is dermate uitgebreid dat er inmiddels meerdere overzichtsartikelen zijn verschenen. Verscheidene van deze overzichtsartikelen (o.a. Heberlein & Baumgartner, 1978; Goyder, 1982; Yu & Cooper, 1983; Fox, Crask & Kim, 1988) gebruiken meta-analyse (Glass, McGaw & Smith, 1981) om de empirische bevindingen systematisch te beschrijven.

De belangrijkste resultaten in deze meta-analyses zijn de volgende. Herhaalde contactpogingen hebben, net als het geven van een kleine beloning of attentie, een positief effect op de respons. Het belang van het onderwerp (saliency), heeft eveneens een positief effect. Ook de onderzochte populatie en de onderzoeksorganisatie is van belang: in het algemeen leveren steekproeven uit een geografisch dichtbijgelegen en selecte populatie een hogere respons op dan een aselechte steekproef uit de gehele bevolking, en overheidsinstanties of universiteiten behalen hogere responspercentages dan het marktonderzoek of het bedrijfsleven.

In dit artikel worden de resultaten gepresenteerd van een meta-analyse van empirisch vergelijkend onderzoek naar verschillen in respons tussen face-to-face, telefonische, en per post uitgevoerde surveys. Hierbij staan twee vragen centraal: 1) wat is het effect van verschillende methoden van dataverzameling op de response, en 2) welke evidentie is er voor het optreden van interactie-effecten, dat wil zeggen differentiële effecten van achtergrondvariabelen (bijv. jaartal) op de grootte van het verschil tussen verschillende dataverzamelmethode. De gebruikte analysemethode is een zgn. multiniveau model, een type model dat zeer geschikt is om variërende effectgroottes zoals responsgegevens te modelleren (Bosker & Snijders, 1990; Hox, 1991).

2. Opzet van het onderzoek

Meta-analyse is een verzameling methoden voor het statistisch analyseren en samenvatten van de resultaten van een groot aantal empirische onderzoeken op een specifiek terrein (Glass, McGaw & Smith, 1981; Light & Pillemer, 1984). De opzet van een meta-analytisch literatuuronderzoek is in hoge mate te vergelijken met de opzet van een gewoon empirisch onderzoek, met onderzoeksdelen probleemstelling, dataverzameling (verzamen van de relevante artikelen en papers), coderen, en statistische analyse. Dikwijls is er sprake van één centrale afhankelijke variabele.

In de vroegste meta-analyses (cf. Rosenthal, 1984) wordt doorgaans voor ieder onderzoek afzonderlijk en voor alle onderzoeken tezamen de significantie van het gerapporteerde effect bepaald. Recentelijk is de aandacht meer komen te liggen op het analyseren en verklaren van verschillen in de effectgroottes van uiteenlopende onderzoeken (Hunter, Schmidt & Jackson, 1982; Fricke & Treinies, 1985; Wolf, 1986; Hunter & Schmidt, 1990). Als verklarende variabelen worden achtergrondgegevens van de verzamelde onderzoeken gebruikt, zoals jaar van publikatie, soort publikatie, soort steekproef, en respondent-kenmerken.

Het gebruik van statistische analysemethoden om het effect van achtergrondvariabelen op de gevonden resultaten na te gaan is het belangrijkste verschil tussen meta-analyse en het traditionele literatuuroverzicht (Bangert-Drowns, 1986). De specifieke aard van meta-analytische gegevens heeft geleid tot de ontwikkeling van hierop toegesneden statistische methoden (zie o.a. Hedges & Olkin, 1986). In de volgende twee paragrafen wordt de opzet beschreven van de hier gerapporteerde meta-analyse en wordt het gehanteerde analysemodel nader toegelicht.

2.1 Opzet van de meta-analyse

Voor het opsporen van de relevante onderzoeksliteratuur is on-line gezocht in de titelbestanden van Psychological Abstracts, Sociological Abstracts, Dissertation Abstracts, Dialog/SSCI en de SRM catalogus. Daarnaast is in drie Europese nieuwsbrieven een oproep geplaatst. Tenslotte is op basis van de literatuurlijst van de gevonden publikaties een sneeuwbal zoekprocedure uitgevoerd. Gezocht is naar publikaties die resultaten bevatten van vergelijkend onderzoek naar de effecten van de gehanteerde dataverzamelingsmethode op de kwaliteit van de verzamelde gegevens. In die gevallen dat meerdere aparte publikaties (deel)resultaten rapporteren van één enkel onderzoek, zijn deze publikaties gecodeerd als één enkele publikatie (zie de Leeuw, 1992).

Uit de beschikbare verklarende variabelen zijn in eerste instantie alleen die variabelen gekozen waarvan op grond van de eerdere meta-analyses kan worden aangenomen dat zij mogelijk belangrijk zijn. De drie dataverzamelingscondities zijn op conditieniveau gecodeerd door middel van twee (0/1) dummy-variabelen die coderen voor telefonische enquête en postenquête, de face-to-face methode is daarmee de vergelijkingsconditie. Op studieniveau zijn de volgende variabelen gecodeerd: soort steekproef (1=beschikbare groep, 2=bestaand panel, 3=lokaal random, 4=landelijk random); soort publikatie (1=intern, 2=extern); vorm publikatie

(1=tijdschrift, 2=boek, rapport); jaartal (van publikatie, 1947=0, 1948=1 etc); indringendheid vragenlijst (0=niet,..., 2=sterk bedreigend); saliency (0=niet,..., 2=sterk); kwaliteit onderzoek (z-score, gebaseerd op een homogeniteitsanalyse van acht indicatoren van onderzoekskenmerken zoals weergave methodologische details in verslag, interne en externe validiteit van het gebruikte design); onderzoeksorganisatie (1=overheid, 2=universiteit, 3=marktonderzoek, 4=bedrijfsleven).

In totaal zijn er 60 publikaties gecodeerd. Als controle op de betrouwbaarheid van de codering is deze uitgevoerd door twee onafhankelijke codeurs aan de hand van een gedetailleerd codeboek. De intercodeur-correlatie is 0.93.

Bij een analyse van de respons is het uiteraard belangrijk om te weten of de gepubliceerde respons gebaseerd is op de 'completion rate' of op de 'response rate'. De completion rate is de respons als percentage van de totale uitgezette steekproef, de response rate is de respons als percentage van de effectieve uitgezette steekproef. (Bij de response rate wordt de nonrespons gecorrigeerd voor fouten in het steekproefkader, zoals niet bestaande adressen, overleden respondenten; voor een bespreking van verschillende definities van 'respons' zie o.a. Kviz, 1977; De Bie, 1987). De response rate is vanzelfsprekend vrijwel altijd hoger dan de completion rate. Bij 12 van de 60 publikaties blijkt het onmogelijk om uit de gepresenteerde gegevens het type respons eenduidig te achterhalen (in vier van deze gevallen is het niet eens mogelijk de omvang van de steekproef te achterhalen). Als gevolg van ontbrekende waarnemingen op de verklarende variabelen vervallen nog drie studies. De meta-analyse is dientengevolge uiteindelijk uitgevoerd op de resultaten van 45 onderzoeken.

De oudste publikatie is van 1947, de meest recente van 1992. Over zes van deze onderzoeken is alleen intern gepubliceerd, over 39 is extern gepubliceerd: vijf in boekvorm en 34 in de vorm van een tijdschriftartikel. De tijdschriftartikelen zijn verspreid over 20 verschillende tijdschriften; het tijdschrift met het grootste aantal van deze publikaties is *Public Opinion Quarterly*. Van deze 45 onderzoeken was het merendeel Amerikaans (32) of Nederlands (8). De meeste van deze onderzoeken vergeleken slechts twee van de drie dataverzamelmethode met elkaar; het totaal aantal geïmplementeerde dataverzamelingscondities is 99 (38 face-to-face, 35 telefonisch, 26 post).

Het totaal aantal potentiële respondenten dat in de 45 onderzoeken is benaderd is te schatten op 106310 (in die gevallen waarin alleen een response rate vermeld wordt is een precieze bepaling van de uitgezette steekproef niet altijd mogelijk), waarvan er 79696 respondeerden. De totale respons in deze 45 onderzoeken kan dus geschat worden op 75%.

2.2 Analysemethode

Het achterliggende idee bij een meta-analyse is dat een statistische analyse wordt uitgevoerd met als data de uitkomsten van eerder onderzoek (Walberg & Haertel, 1980). In feite wordt daarbij het onderzoeksmateriaal behandeld alsof het afkomstig is uit één enkel omvangrijk onderzoek. Daarbij doen zich verschillende onderscheidbare methodologische problemen voor (Bangert-Drowns, 1986).

Een gecompliceerd probleem is doorgaans het eenduidig bepalen van de afhankelijke variabele omdat deze in verschillende onderzoeken op verschillende wijze geoperationaliseerd kan zijn. In ons geval doet dit probleem zich voor doordat de respons in sommige gevallen geoperationaliseerd is als de completion rate en in andere gevallen als de response rate, waarbij enkele studies beide operationaliseringën hanteren. Als afhankelijke variabele is in onze analyse gekozen voor de completion rate (het aantal respondenten gedeeld door de gehele uitgezette steekproef), omdat deze het meest zuivere beeld geeft van het aantal respondenten dat bij een bepaalde uitgezette steekproef te verwachten is. In die gevallen waar deze niet bepaald kan worden, is de response rate genomen (completion rate gecorrigeerd voor kaderfouten). In de lijst van verklarende variabelen is vervolgens een dummy variabele opgenomen die aangeeft welk type respons (completion rate of respons rate) beschreven wordt. Mogelijke verschillen in het respons proces tussen completion rate en response rate zijn geanalyseerd door het opnemen van interactietermen tussen deze dummy variabele en de andere verklarende variabelen.

Een tweede probleem betreft het statistische model dat ten grondslag ligt aan het samenvoegen van de verschillende gegevens. Allereerst zijn de beschikbare gegevens niet altijd onafhankelijk. In ons geval treedt onderlinge afhankelijkheid op omdat de steekproef van 99 dataverzamelingscondities die vergeleken worden afkomstig is uit slechts 45 verschillende onderzoeken. Aangenomen mag worden dat de condities die in hetzelfde onderzoek zijn opgenomen meer op elkaar lijken (b.v. zelfde onderwerp, zelfde populatie, zelfde veldwerkorganisatie) dan condities uit verschillende onderzoeken. Dit resulteert in een correlatie tussen waarnemingen in de verschillende condities binnen één onderzoek; deze kunnen in het algemeen niet verondersteld worden onafhankelijk te zijn.

Een ander statistisch probleem is dat de beschikbare gegevens niet noodzakelijk homogeen zijn. Wanneer de gepubliceerde onderzoeken homogeen zijn, dan verschillen de gerapporteerde resultaten niet méér van elkaar dan op grond van de steekproefgroottes van de respectievelijke onderzoeken verwacht mag worden. In dit geval is het zinvol om de onderzoeken samen te nemen en de gemeenschappelijke effectgrootte te bepalen. Wanneer

de uitkomsten van de onderzoeken méér van elkaar verschillen dan op grond van pure steekproefvariatie verwacht mag worden, dan zijn de gegevens heterogeen, en is de voor de hand liggende onderzoeksvraag welke achtergrondgegevens deze systematische variatie in onderzoeksuitkomsten kunnen verklaren.

Er zijn verschillende methoden beschikbaar om te onderzoeken of gegevens homogeen zijn (Hedges & Olkin, 1985) of om onderlinge afhankelijkheden te verdisconteren (Rosenthal & Rubin, 1986). Raudenbush en Bryk (1985, 1987) wijzen er op dat meta-analyse kan worden opgevat als een specifiek multiniveau model. Er is sprake van hiërarchisch geordende gegevens: er zijn een aantal studies, en binnen elke studie zijn een (in ons geval wisselend) aantal condities opgenomen. De afhankelijke variabele (in ons geval de respons) is gedefinieerd op het laagste niveau, en verklarende variabelen kunnen zijn gedefinieerd op zowel het niveau van de geïmplementeerde condities als op het hogere studie-niveau. De afhankelijkheid tussen de condities binnen dezelfde studies wordt in een multiniveau analyse in het statistisch model verdisconteerd (vgl. Bosker & Snijders, 1990; Hox, 1991). De vraag of de effecten van de geïmplementeerde condities in de verschillende studies homogeen zijn, vertaalt zich binnen een multiniveau analyse in de vraag of de regressiecoëfficiënten (intercept en regressiehellings) van de variabelen die de condities beschrijven als gefixeerd mogen worden beschouwd. Wanneer deze regressiecoëfficiënten significante random componenten hebben, is de implicatie dat de geïmplementeerde condities in de verschillende studies niet steeds een gelijk effect hebben, met andere woorden, de studies zijn dan heterogeen. De variatie van de random regressiecoëfficiënten kan vervolgens geanalyseerd worden met behulp van verklarende variabelen op het studieniveau.

De afhankelijke variabele 'respons' is een proportie. De analyse van proporties kan worden uitgevoerd door middel van een gegeneraliseerd lineair model (GLM, zie McCullagh & Nelder, 1989). Gegeneraliseerde lineaire modellen zijn opgebouwd uit drie componenten: 1) een lineaire regressievergelijking, 2) een specifieke foutenverdeling, en 3) een link functie die de voorspelde waarden transformeert naar de schaal van de geobserveerde waarden. Voor proporties wordt doorgaans een logit of probit link functie gebruikt met een binomiale foutenverdeling. Gegeneraliseerde lineaire multiniveau modellen zijn beschreven door Wong en Mason (1985), Longford (1988, 1990), Mislevy en Bock (1989) en Goldstein (1991). Onze analyse is gebaseerd op een logit link en uitgevoerd met het door Longford ontwikkelde programma VARCL (Longford, 1990). De uiteindelijke modeltoetsing ging via een stapsgewijze procedure, beginnende bij het meest simpele nulmodel.

We schrijven P_{ij} voor de geobserveerde proportie respondenten in conditie i van studie j , en π_j voor de corresponderende proportie in de populatie. Dan is

$$P_{ij} = \pi_{ij} \quad (1)$$

waarbij P_{ij} gegeven π_{ij} een binomiale verdeling heeft. In het door Longford in VARCL geïmplementeerde binomiale model wordt de variantie van de verdeling van P_{ij} gegeven door $\sigma^2(\pi_{ij}/(1-\pi_{ij}))$. De parameter σ^2 is een schaalfactor die in het binomiale geval op 1.00 wordt gesteld en niet geïnterpreteerd. Wanneer er sprake is van overdispersie ten opzichte van de binomiale verdeling kan deze gemodelleerd worden door de schaalfactor σ^2 een waarde te geven groter dan 1.00.

De logit link functie wordt gegeven door $\text{logit}(x) = \text{LN}(x/(1-x))$. Het gebruik van een logistisch regressie model voor de proportie P geeft een lineair regressie model voor de logits van de proporties.

Het simpelste multiniveau model voor $\text{logit}(P_{ij})$ luidt:

$$\text{logit}(P_{ij}) = \beta_{0j} \quad (2),$$

waarbij

$$\beta_{0j} = \gamma_{00} + \delta_{0j} \quad (3),$$

zodat

$$\text{logit}(P_{ij}) = \gamma_{00} + \delta_{0j} \quad (4).$$

In dit model, het zogenaamde nulmodel (in het engels doorgaans aangeduid als het 'intercept only' model), wordt de algemene proportie aangegeven door de intercept γ_{00} , de proporties verschillen voor de verschillende studies j , en de error-termen δ_{0j} zijn normaal verdeeld op de logit-schaal.

Het volledige multiniveau model voor één verklarende variabele X_{ij} op het conditieniveau en één verklarende variabele Z_j op het studieniveau luidt:

$$\text{logit}(P_{ij}) = \beta_{0j} + \beta_{1j}X_{ij} \quad (5),$$

waarbij

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + \delta_{0j} \quad (6),$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + \delta_{1j} \quad (7),$$

zodat

$$\text{logit}(\mathbf{P}_{ij}) = \gamma_{00} + \gamma_{10}\mathbf{X}_{ij} + \gamma_{01}Z_j + \gamma_{11}\mathbf{X}_{ij}Z_j + \delta_{0j} + \delta_{1j}\mathbf{X}_{ij} \quad (8).$$

Model (8) geeft aan dat de proportie respondenten in conditie i van studie j verklaard wordt uit een lineair regressiemodel, waarbij zowel de intercepten als de regressiehellingen van de verklarende variabele $\mathbf{X}$ voor de verschillende studies kunnen variëren. De residuele errortermen δ_{0j} en δ_{1j} hebben variantie σ_0^2 en σ_1^2 .

Longford (1990) past een iteratief algoritme toe om de parameters van dit model te schatten, inclusief de bijbehorende standaardfouten. De gebruikte procedure is een quasi-likelihood procedure (cf. Longford, 1988; McCullagh & Nelder, 1989). Bij deze procedure zijn de parameterschattingen asymptotisch correct, maar de correctheid van de standaardfouten en van de deviantie als maat voor de passing van het model staat enigszins ter discussie (McCullagh & Nelder, 1989; zie ook Snijders, 1991).

De interpretatie van de regressiecoëfficiënten gamma is in termen van de normaal verdeelde variabele gedefinieerd door de (nonlineaire) logit transformatie. Voor de passing van het model en de standaardfouten van de regressieparameters kunnen de schattingen van het gegeneraliseerde lineaire multiniveau model gebruikt worden, voor de interpretatie van de resultaten in termen van de voorspelde proporties respondenten verdient het de voorkeur om de voorspellingen terug te rekenen naar de proporties.

3. Resultaten

3.1. Respons per conditie univariaat bekeken

Als alle publikaties een vergelijking zouden geven van alle drie de dataverzamelmethode, en zowel een completion rate als een respons rate zouden rapporteren, dan konden de resultaten simpel samengevat worden met de gemiddelde respons per conditie, zoals in onderstaande Tabel 1:

Tabel 1. Completion rate (CR) en respons rate (RR) per conditie (percentages).

	F-t-F	Tel.	Post	Totaal
CR	70.9	63.1	60.9	65.4
RR	73.5	70.3	68.2	71.0

Echter, de meeste onderzoeken vergelijken slechts twee condities, en hoewel sommige onderzoeken zowel de completion rate (CR) als de respons rate (RR) rapporteren, geven de meeste publikaties slechts een van deze twee. De cellen in Tabel 1 zijn dus enerzijds niet onafhankelijk, omdat sommige studies zowel onder 'CR' als onder 'RR' vermeld worden, en anderzijds voor een deel gebaseerd op niet vergelijkbare studies. Het zou immers mogelijk kunnen zijn, bijvoorbeeld, dat de lagere respons van de postenquête deels gebaseerd is op een verschil in de studies waarin postenquêtes wel of niet zijn opgenomen. In het algemeen worden postenquêtes en telefonische interviews immers vaker ingezet bij grote landelijk gespreide bevolkingssteekproeven, die doorgaans een wat lagere respons kennen dan lokale steekproeven. De resultaten van een directe vergelijking zoals in Tabel 1 zijn daarom potentieel misleidend. In plaats daarvan dient een statistisch model toegepast te worden dat rekening houdt met zowel de afhankelijkheden binnen de studies als de verschillen tussen de studies. Het hierboven beschreven multiniveau model voorziet hierin.

3.2. Resultaten modelpassing

De verklarende variabelen zijn dataverzamelmethode, soort steekproef, soort publikatie, vorm publikatie, jaartal, indringendheid vragenlijst, saliency, kwaliteit onderzoek, en onderzoeksorganisatie (zie ook 2.1). De afhankelijke variabele is de completion rate (CR). In die gevallen waar deze niet bepaald kan worden, is de response rate (RR) gebruikt. Aan de lijst van verklarende variabelen is daarvoor een dummy variabele toegevoegd die aangeeft welk type respons beschreven wordt (0=CR, 1=RR).

Mogelijke verschillen in het responsproces tussen completion rate en response rate kunnen worden geanalyseerd door het toevoegen van interactietermen tussen deze dummy variabele en de andere verklarende variabelen. Om na te gaan welke interacties in aanmerking komen zijn twee pre-analyses uitgevoerd, één op alle beschikbare gegevens voor de

completion rate, en één op alle beschikbare gegevens voor de response rate. Drie van de verklarende variabelen lijken zich anders te gedragen bij de twee responstypen: het soort steekproef (select vs random), het belang (saliency) van het onderwerp, en de onderzoeksorganisatie (overheid of universiteit vs. marktonderzoek). In de analyse op de gehele dataset is daarom voor deze drie verklarende variabelen een interactieterm toegevoegd met de dummy variabele 'type respons'.¹ In verband met de interpreteerbaarheid van de interactietermen zijn de verklarende variabelen op studieniveau gecentreerd rond het algemene gemiddelde; de interactietermen zijn niet opnieuw gecentreerd (Jaccard et al., 1990; Aiken & West, 1991).

In de analyses is het multiniveau model opgebouwd conform de volgende stapsgewijze procedure:

1. Nulmodel: d.w.z. een model zonder verklarende variabelen.
2. Conditie-model: inbreng van de beide verklarende variabelen op het conditieniveau (telefoon- en post-dummy), als gefixeerde effecten.
3. Studie-model: inbreng van alle verklarende variabelen op studie-niveau, voorzover significant. Als significantiekriterium is genomen dat de parameterschatting (absoluut) groter is dan twee maal de bijbehorende standaardfout.
4. Random-model: specificeren van random effecten voor de beide conditie-variabelen, voorzover significant.
5. Cross-level model: inbreng van interacties tussen de random variabelen op conditieniveau en studieniveau variabelen, voorzover significant.

Tabel 2 op de volgende bladzijde geeft de resultaten weer voor bovenstaande vijf modellen:

¹Voor de verdere analyse is gekozen voor bovenstaande analyse op het gehele bestand (completion rate en respons rate, totaal 45 cases) in plaats van parallelle analyses op de beide deelbestanden in verband met de kleine steekproefgrootte bij de deelanalyses.

Tabel 2. Resultaten multiniveau modellen voor respons.

	1	2	3	4	5
<u>Fixed Eff.</u>					
intercept	.81	1.10	1.16	1.06	1.05
tel	-.32	-.34	-.19 ^{ns}	-.14 ^{ns}	
post		-.59	-.62	-.39	-.39
soort stkprf			-.45	-.40	-.23
saliency			.50	.53	.47
onderz. organ.			-.49	-.44	-.45
respons=RR			.61	.59	.56
jaar (1947=0)				-.02	
interactie					
soort stkprf*respons			-.48	-.47	
cross-level interactie					
tel*soort stkprf					-.27
post*jaar					.02
<u>Random Eff.</u>					
σ^2	1.00	1.00	1.00	1.00	1.00
σ^2_{interc}	.61	.57	.30	.21	.26
σ^2_{tel}				.39	.35
σ^2_{post}				.44	.31
deviantie	105625	103087	99835	98248	98602

Alle effecten in Tabel 2 zijn tenminste twee maal zo groot als de bijbehorende standaardfout (tenzij uitdrukkelijk is vermeld dat het effect niet significant is). We merken nogmaals op dat de waarde van 1.00 voor de schaalfactor σ^2 geen substantieve betekenis heeft.

In het nulmodel wordt de te verklaren intercept-variantie tussen de studies geschat als 0.61. In het conditie-model (model 2) blijkt dat beide conditievariabelen een significante invloed hebben in de verwachte richting: in de telefonische conditie is de respons lager dan in de face-to-face conditie, en in de postconditie is de respons nog weer lager.

Toevoeging van de achtergrondvariabelen op studieniveau in model (3) verandert dit beeld niet. Deze variabelen verklaren een aanmerkelijk deel van de intercept-variantie tussen

de studies; de residuele intercept variantie is in dit model 0.30. De effecten zijn alle in de verwachte richting. Het soort steekproef heeft invloed; hoe meer het gaat om een landelijk aselechte steekproef, des te lager wordt de verwachte respons. Hoe groter de saliency van het onderwerp voor de respondenten, des te hoger de respons. Onderzoeken afkomstig uit het marktonderzoek mogen een lagere respons verwachten dan onderzoeken van de overheid of universiteiten. Tenslotte, wanneer een respons rate vermeld wordt mag een grotere gerapporteerde respons verwacht worden dan wanneer een completion rate gerapporteerd wordt. De negatieve interactieterm van 'type respons' met 'soort steekproef' geeft aan dat het voordeel van het vermelden van de respons rate bij aselechte steekproeven kleiner is. Dit kan verklaard worden indien we aannemen dat met name bij aselechte steekproeven met een slechte respons de onderzoekers wat frequenter kiezen voor het rapporteren van de respons rate in plaats van de completion rate. De verklarende variabelen soort en vorm publicatie, jaartal, indringendheid vragenlijst, en kwaliteit onderzoek zijn in dit model niet significant.

Het toelaten van variërende regressiecoëfficiënten voor de verschillende condities in model (4) levert een iets ander beeld op. De variantie van de regressielijnen voor de telefoon- en postconditie blijkt vrij groot te zijn, terwijl de regressiecoëfficiënten zelf nu beduidend lager worden geschat. De regressiecoëfficiënt van 'telefoon' is nu niet langer significant volgens ons criterium (de regressiecoëfficiënt is -0.19, met een standaardfout van .11).

In model (5) wordt gepoogd door toevoeging van cross-level interacties de variatie van de hellingen van de regressielijnen voor 'telefoon' en 'post' te verklaren. Er blijken slechts twee significante interacties te zijn: één tussen 'jaar' en 'postconditie' en één tussen 'soort steekproef' en 'telefoonconditie': in de loop van de tijd wordt de achterstand van de postconditie op de andere twee dataverzamelingsmethoden kleiner, en de telefonische enquête doet het bij lokale selecte steekproeven relatief beter dan bij landelijke aselechte steekproeven. In model (5) is ook het jaar van publicatie van de studie opgenomen. Deze variabele heeft in dit model een merkbaar effect; latere publikaties vermelden in het algemeen een wat lagere respons. De interactie tussen de soort steekproef en het vermelden van de respons rate in plaats van de completion rate is niet langer significant en daarom uit het model verwijderd (de regressiecoëfficiënt is -.35 bij een standaardfout van .24); we nemen aan dat dit effect nu door de opgenomen cross-level interacties wordt verklaard. Het opnemen van cross-level interacties is echter niet geheel succesvol. De residuele varianties van de intercept en de regressiehellingen zijn nog duidelijk groter dan nul; er zijn dus nog onverklaarde effecten.

In het laatste model (model 5) zijn twee cross-level interacties opgenomen: de interactie tussen dataverzamelingsmethode (specifiek de telefonische enquêtering) en soort steekproef, en de

interactie tussen dataverzamelmethode (specifiek de postenquête) en het jaartal. Om een duidelijk beeld te krijgen van de invloed van verklarende variabelen bij interacties is het inzichtelijk om de resultaten uit te zetten in een grafiek. Dit is met name van belang voor de interactie tussen dataverzamelmethode en jaartal, omdat bij deze variabelen het effect optreedt dat de verklarende variabele 'jaartal' in modellen zonder deze interactieterm niet significant is, en in het model met deze interactieterm wel.

Figuur 1. geeft de interactie tussen 'dataverzamelmethode' en 'jaartal':

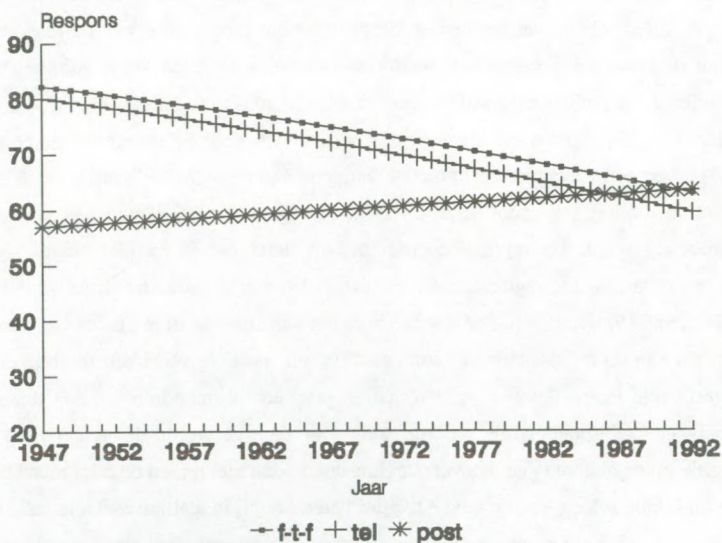


Fig. 1. Respons per Conditie per Jaar

Figuur 1 laat zien dat het verschil tussen de postenquête en de andere twee dataverzamelmethoden in de loop van de tijd kleiner is geworden, terwijl de globale respons daarbij licht is afgenomen. In model 5 is het verschil tussen de drie dataverzamelmethoden in het laatste jaar van observatie, 1992, vrijwel verdwenen.

3.3. Respons per conditie op grond van het model

In Tabel 1 (sectie 3.1) is de completion rate en de respons rate vermeld voor de drie verschillende dataverzamelmethodeën. Hierbij werd het gemiddelde per conditie berekend zonder rekening te houden met mogelijke systematische verschillen tussen de face-to-face, interview- en postcondities met betrekking tot invloedrijke achtergrondvariabelen. Uit de daarop volgende analyses (sectie 3.3, Tabel 2) blijkt dat dergelijke achtergrondvariabelen er inderdaad zijn. Bij een vergelijking van de respons bij verschillende dataverzamelmethodeën dient met deze effecten rekening te worden gehouden. Daartoe is de completion rate en de respons rate opnieuw berekend, maar nu op basis van de regressiegewichten van het definitieve multiniveau model (model 5). Voor de verklarende variabelen in het model is daarbij steeds de gemiddelde waarde ingevuld (voor 'jaartal' afgerond op het gehele jaartal 1977), waarna de voorspellingen op de logit schaal zijn teruggerekend naar percentages. Dit levert de volgende resultaten op:

Tabel 3. Voorspelde completion rate (CR) en respons rate (RR) per conditie (percentages, voorspellingen op basis van model 5 in Tabel 2).

	F-t-F	Tel.	Post	Totaal
CR	70.3	67.2	61.3	66.9
RR	80.5	78.2	73.5	78.0

Vergeleken met de ongecorrigeerde Tabel 1 zijn er twee opvallende verschillen. Allereerst zijn de verschillen tussen de dataverzamelmethodeën nu wat kleiner. Kennelijk blijft de respons bij telefonische en postenquête voor een deel achter omdat deze methodeën wat vaker worden ingezet in (voor de respons) ongunstige omstandigheden, zoals een landelijke en aselechte steekproeftrekking. Verder is het verschil tussen de completion rate en de respons rate aanmerkelijk groter geworden. Ook dit kan verklaard worden doordat de response rate kennelijk met name vermeld wordt in die (ongunstige) omstandigheden waarin de respons laag is.

4. Discussie

De gevonden directe effecten in het multiniveau model zijn alle conform de bevindingen in de onderzoeksliteratuur. Wanneer het onderwerp voor de respondenten belangrijk is (hoge saliency) dan is de respons hoger. Aselecte landelijke steekproeven behalen een lagere respons dan selecte lokale steekproeven, onderzoekers uit het marktonderzoek moeten een wat lagere respons verwachten dan onderzoekers uit de sfeer van overheid/universiteit. Deze resultaten zijn ook in Nederland eerder gevonden, zij het dan specifiek voor postenquêtes (De Leeuw & Hox, 1985; Van Rooij, 1987).

Wanneer we teruggaan naar de centrale onderzoeksvraag dan concluderen we dat er significante verschillen in respons zijn tussen het face-to-face interview, het telefonisch interview, en de postenquête. Bovendien blijkt dat er sprake is van interactie-effecten; een aantal achtergrondvariabelen (soort steekproef en jaar van onderzoek) heeft een differentiële invloed op de respons.

De grote en significante variatie van de regressielijnen voor de telefoon- en de postconditie heeft een interessante implicatie. Op de logit-schaal is de regressiecoëfficiënt voor 'post' gelijk aan -0.39 (model 4 Tabel 2). Deze coëfficiënt is significant, dat wil zeggen dat we mogen aannemen dat face-to-face interviews *gemiddeld* een hogere respons bereiken dan postenquêtes. Echter, σ^2_{post} , de variantie van de regressiecoëfficiënten over de studies, wordt geschat als 0.44, hetgeen overeenkomt met een standaard afwijking van 0.63. Dit betekent dat we mogen verwachten dat in 27 procent van soortgelijke studies de betreffende regressiecoëfficiënt positief uitvalt! Hoewel *gemiddeld* de postenquête een lagere respons heeft dan het face to face interview, is het toch heel goed mogelijk dat in *een specifiek onderzoek* de resultaten precies omgekeerd zijn. De interactieterm tussen methode en jaartal in model 5 verklaart een gedeelte van deze variantie: in model 5 blijft de regressiecoëfficiënt voor 'post' -0.39, maar de variantieterm σ^2_{post} is kleiner geworden. De geschatte waarde van 0.31 is nog steeds significant. Een voor de hand liggende verklaring is bijvoorbeeld het aantal contactpogingen of herhalingsronden (Heberlein & Baumgartner, 1978); variatie tussen studies in het aantal contactpogingen en een groter aantal contactpogingen bij recente postenquêtes zou zowel de grote variantie van de regressiehellingsen als de convergentie tussen de verschillende dataverzamelingsmethoden kunnen verklaren. De oorspronkelijke onderzoeksartikelen bevatten helaas niet voldoende details om aanvullende verklarende variabelen te coderen en hiermee de variantie verder te modelleren.

De afnemende bereidheid van het publiek om mee te doen aan survey onderzoek in het algemeen is in de inleiding al besproken. De bevinding dat face to face surveys niet langer

per definitie de hoogste respons opleveren is ook elders gevonden. Zo merkt Goyder (1987, p.67) op dat de respons voor het face-to-face interview over de jaren heen lijkt te dalen, terwijl de respons bij postenquêtes stabiel lijkt te zijn, een opmerking die goed aansluit op de resultaten bij model (5) in Tabel 2. In Nederland rapporteren Snijders (1991) en De Leeuw (1992) eveneens een lagere respons voor de face-to-face survey.

Dat de postenquête tegen de trend lijkt in te gaan, is een hoopvol teken. Het is onwaarschijnlijk dat dit het gevolg is van een in de loop van de tijd toenemende *a priori* bereidheid van het publiek om te responderen op postenquêtes. Aannemelijker is dat de tegendraadse trend bij postenquêtes de resultante is van de grote hoeveelheid onderzoek gericht op het verhogen van de respons bij deze methode van dataverzameling. Het werk van Dillman (1978, 1991) in de VS is daarbij van groot belang geweest; voor Nederlands onderzoek zie o.a. Nederhof (1982) en Hox, De Leeuw & Duijx (1984). De omvang van het onderzoek naar responsverhogende maatregelen bij andere vormen van dataverzameling is daarbij vergeleken duidelijk minder (zoals blijkt uit de aantallen studies in de meta-analyses van Goyder, 1987). Het is aannemelijk dat een vergelijkbare inspanning gericht op het verhogen van de respons ook bij de andere methoden van dataverzameling een positief effect zal hebben. Een investering in ontwikkelingsonderzoek, specifiek gericht op het terugdringen van de nonrespons bij interviews, zou wel eens een zichtbaar effect kunnen hebben.

In de hier gerapporteerde meta-analyse is gebruik gemaakt van een gegeneraliseerd lineair multiniveau model. Het grote voordeel van een multiniveau model is dat het mogelijk is te onderzoeken in hoeverre de geanalyseerde studies heterogeen zijn, eenvoudig door de regressiecoëfficiënten op studieniveau random te maken en ze daarmee te laten variëren over verschillende studies. Volstrekt vanzelfsprekend is deze aanpak echter niet. Men zou kunnen aanvoeren dat het niet zinvol is de regressiegewichten van 'telefoon' en 'post' te laten variëren over verschillende studies, omdat het hier gaat om dummy variabelen die refereren naar experimentele condities die door de onderzoekers zijn gemanipuleerd. Dit sluit echter niet uit dat de effecten van de experimentele manipulatie van studie tot studie kunnen verschillen. Voorzover de studies zijn uitgevoerd in verschillende jaren, bij verschillende populaties, en dergelijke, zijn verschillen tussen studies in principe mogelijk; deze zijn in onze meta-analyse ook daadwerkelijk gevonden.

Een andere bron van verschillen tussen de verschillende studies is de variatie in de operaties die de experimentele condities (face-to-face, telefoon, post) implementeren. De invloed van een groot aantal onbelangrijk lijkende details op de respons is in de literatuur overtuigend aangetoond (cf. Goyder, 1987). De in de analyse opgenomen publikaties

vermelden weinig details over de precieze uitvoering van de dataverzamelingen, maar het is uiterst waarschijnlijk dat deze in de 45 opgenomen onderzoeken niet identiek zijn geweest. Dit resulteert in variatie van de regressiecoëfficiënten, en de (significante) residuele variantie in het uiteindelijke multiniveau-model (model 5) is naar alle waarschijnlijkheid terug te voeren op niet gedocumenteerde verschillen in de implementatie.

De indringendheid van de vragenlijst, soort en vorm van publikatie en de kwaliteit van het onderzoek blijken niet gerelateerd te zijn met de respons. De bevinding dat de publikatiesoort (intern dan wel extern), publikatievorm (boek, rapport, of artikel), en onderzoekskwaliteit niet significant zijn is geruststellend, want het vermindert de waarschijnlijkheid dat de resultaten vertekend zijn door publikatiebias (een oververtegenwoordiging van significante uitkomsten afkomstig van 'officieel' gepubliceerd onderzoek). Het land (waarin het onderzoek plaats vond) is niet in onze analyse opgenomen, omdat dit met zich mee brengt dat een categoriale variabele met een groot aantal spaarzaam gevulde categorieën wordt toegevoegd. Een analyse van de residuen van het laatste model naar land van herkomst levert echter geen significante verschillen op. De hier vermelde resultaten gelden kennelijk zowel voor Noord-Amerika als West-Europa.

5. Literatuur

ontvangen 16-11-1992
geaccepteerd 28-2-1993

- Aiken, L.S. & West, S.G. (1991). **Multiple regression: testing and interpreting interactions**. Newbury Park, CA: Sage.
- Bangert-Drowns R.L. (1986). 'Review of developments in meta-analytic method'. **Psychological Bulletin**, 99: 388-399.
- Betlehem, J.G., & Kersten, H.M.P. (1981). 'The nonresponse problem'. **Survey Methodology**, 7, 130-156.
- Betlehem, J.G., & Kersten, H.M.P. (1986). **Werken met non-respons**. Amsterdam: Universiteit van Amsterdam (dissertatie).
- Bosker, R.J. & Snijders, T.A.B. (1990). 'Statistische aspecten van multiniveau onderzoek'. **Tijdschrift voor Onderwijsresearch**, 15, 317-329.
- De Bie, S.E. (1987). 'Naar een standaard-definitie van de responsfractie in personen steekproeven'. **Kwantitatieve Methoden**, 24, 67-73.
- De Leeuw, E.D. (1992). **Data quality in mail, telephone and face- to-face surveys**. Amsterdam: Vrije Universiteit (dissertatie).
- De Leeuw, E.D. & Hox, J.J. (1985). Recente ontwikkelingen bij schriftelijk onderzoek per post. **Jaarboek van de Nederlandse Vereniging van Marktonderzoekers**. Haarlem: De Vrieseborgh.
- Dillman, D.A. (1978). **Mail and telephone surveys: The Total Design Method**. New York: Wiley.
- Dillman, D.A. (1991). 'The design and administration of mail surveys'. **Annual Review of Sociology**, 17, 225-249.

- Fricke, R. & Treinies, G. (1985). **Einführung in die Metaanalyse**. Bern: Huber.
- Goyder, J. (1982). 'Further evidence on factors affecting response rates to mailed questionnaires'. **American Sociological Review**, 47, 550-3.
- Goyder, J. (1987). **The silent minority**. Cambridge: Blackwell.
- Goldstein (1991). 'Nonlinear multilevel models, with an application to discrete response data'. **Biometrika**, 78, 45-51.
- Groves, R.M. & Kahn, R.L. (1979). **Surveys by telephone; A national comparison with personal interviews**. New York: Academic Press.
- Glass, McGaw, & Smith (1981). **Meta-analysis in social research** Beverly Hills: Sage.
- Groves, R.M. & Lyberg, L.E. (1988). 'An overview of nonresponse issues in telephone surveys'. In: R.M. Groves, P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls II, & J. Waksberg (eds). **Telephone Survey Methodology** (Pp. 191-212). New York: Wiley.
- Heberlein T.A. & Baumgartner, R. (1978). 'Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature'. **American Sociological Review**, 43, 447-462.
- Hedges, L.V. & Olkin, I. (1985). **Statistical methods for meta-analysis**. Orlando: Academic Press.
- Hox, J.J., De Leeuw, E.D. & Duijx, A.W.M. (1984). 'The postman rings thrice. Een onderzoek naar twee kenmerken van Dillman's Total Design Method voor postenquêtes'. **Mens en Maatschappij**, 59, 189-194.
- Hox, J.J. (1991). 'Multiniveau analyse voor markt- en opinieonderzoek'. In: **Recente ontwikkelingen in het marktonderzoek**. NvM Jaarboek '91-'92. Haarlem: De Vrieseborgh.
- Hunter, J.E., Schmidt, F.L., & Jackson, G.B. (1982). **Meta-analysis: Cumulating research findings across studies**. Beverly Hills: Sage.
- Hunter, J.E., & Schmidt, F.L. (1990). **Methods of meta-analysis**. Beverly Hills: Sage.
- Jaccard, J., Turrisi, R., & Wan, C.K. (1990). **Interaction effects in multiple regression**. Newbury Park, CA: Sage.
- Kerssemakers, F.A.M. (1985). 'Telefonisch enquêteren'. In: **CBS- select 3** (pp. 211-230). Voorburg/Heerlen: Centraal Bureau voor de Statistiek.
- Kviz, F.J. (1977). 'Towards a standard definition of response rate'. **Public Opinion Quarterly**, 41, 265-267.
- Light, R.J. & Pillemer, D.B. (1984). **Summing up: The science of reviewing research**. Cambridge, MA: Harvard University Press.
- Longford, N.T. (1988). **A quasi-likelihood adaptation for variance component analysis**. Princeton, NJ: Educational Testing Service.
- Longford, N.T. (1990). **VARCL. Software for variance component analysis of data with nested random effects (maximum likelihood)**. Princeton, NJ: Educational Testing Service.
- McCullagh, P. & Nelder, J.A. (1989). **Generalized Linear Models**. London: Chapman & Hall.
- Mislevy, R.J. & Bock, R.D. (1989). 'A hierarchical item-response model for educational testing'. In R.D. Bock (ed.) **Multilevel analysis of educational data**. San Diego, CA: Academic Press.
- Nederhof, A.J. (1982). 'An analysis of response patterns in mail surveys: a comparison of Dutch and North-American results'. In: **Sociometric research 1982**. Amsterdam: Dutch Sociometric Society.

- Raudenbush, S.W. & Bryk, A.S. (1985). 'Empirical Bayes meta-analysis'. **Journal of Educational Statistics**, 10, 75-98.
- Raudenbush, S.W. & Bryk, A.S. (1987). 'Examining correlates of diversity'. **Journal of Educational Statistics**, 12, 241-269.
- Rosenthal, R. (1984). **Meta-analytic procedures for social research**. Beverly Hills: Sage.
- Rosenthal, R. & Rubin, D.B. (1986). 'Meta-analytic procedures for combining studies with multiple effect sizes'. **Psychological Bulletin**, 99, 400-406.
- Snijders, T.A.B. (1991). 'Some remarks on multilevel models for binary dependent variables'. Ongepubliceerd manuscript.
- Snijders, G.J.M.E. (1991). **Computer gestuurd enquêteren: Telefonisch of persoonlijk?**. Voorburg: Centraal Bureau voor de Statistiek.
- Steeh, C.G. (1981). 'Trends in nonresponse rates, 1952-1979'. **Public Opinion Quarterly**, 45. Herdrukt in: E. Singer, & S. Presser (1989), *Survey Research Methods, A Reader*. Chicago: university of Chicago Press.
- Sugiyama, M. (1991). 'Response and non-response'. **Proceedings of the seminar on quality of information in sample surveys** (pp. 213- 224). Paris, June 13-15.
- Van Rooij, C. (1987). 'Responsvoorspellingen: toverformules of realisme?' In: **Recente ontwikkelingen in het marktonderzoek. NvM Jaarboek '86-'87**. Haarlem: De VRIeseborgh.
- Walberg, H.J. & Haertel, E.H. (1980). 'Research integration: an introduction and an overview'. **Evaluation in Education**, 4, 5-10.
- Wolf, F.M. (1986). **Meta-analysis: Quantitative methods for research synthesis**. Beverly Hills: Sage.
- Wong, G.Y. & Mason, W.M. (1985). 'The hierarchical logistic regression model for multilevel analysis. Extensions of the hierarchical normal linear model for multilevel analysis'. **Journal of the American Statistical Association**, 80, 513-524.