

'Multilevel' als techniek en optiek

Roel Bosker en Peter van Laarhoven

- Goldstein, H. (1987). *Multilevel models in educational and social research*. London, Charles Griffin & Co.
- Kreft, G.G. (1987). *Models and methods for the measurement of school effects*. Dissertatie Universiteit van Amsterdam.
- Mooij, T. (1987). *Interactional multi-level investigation into pupil behaviour, achievement, competence, and orientation in educational situations*. Nijmegen, ITS.

Een meerderheid onder degenen die zich met onderzoek bezighouden lijkt multilevel-analyse – een inmiddels gangbaar anglicisme – te beschouwen als een esoterische techniek die voor een beperkt aantal onderzoeksvragen relevant is. Naar onze mening wordt het tijd dat dit beeld op de helling gaat. Er bestaat zoiets als het 'multilevel-probleem', waar weinig onderzoek in veldsituaties omheen kan (Cronbach, 1976). Multilevel-analyse is op de eerste plaats een methodische optiek waarin dat probleem naar voren wordt gehaald en vervolgens technieken worden ontwikkeld om het op te lossen.

Dit is ook het uitgangspunt van Ita Kreft in haar eind 1987 verschenen proefschrift *'Models and methods for the measurement of school effects'*. Kreft werkt een en ander uit aan de hand van een controversieel onderzoeksgebied, effecten van schoolkenmerken op leerlinggedrag. In het eerste deel van haar boek laat zij een reeks valkuilen de revue passeren waar nog teveel onderzoekers in vallen. Wij kunnen het betoog niet beter samenvatten dan met een resumé van deze valkuilen:

- *Schooleffecten stellen in vergelijking met de effecten van individuele achtergrondkenmerken niet veel voor.*
Kreft verwerpt deze vergelijking, omdat geen onderscheid gemaakt wordt tussen manipuleerbare en niet-manipuleerbare factoren. De lengte van mensen bijvoorbeeld verklaart meer variantie in het gewicht dan hun suikerconsumptie, maar toch is de laatste veel interessanter als we iets aan het gewicht willen doen. Schoolkenmerken zijn alleen al relevant omdat zij, in tegenstelling tot veel individuele kenmerken, manipuleerbaar zijn.
- *Schoolkenmerken beïnvloeden elke leerling in gelijke mate.*
Deze assumptie stamt uit de zogenaamde 'input-output'-traditie. Door het optre-

- den van tal van geplande en ongeplande vormen van differentiatie in het onderwijs, is de assumptie niet plausibel. Differentiatie leidt waarschijnlijk tot verschillende effecten bij verschillende groepen leerlingen.
- *Schoolkenmerken zijn een autonome causale factor, die los staat van de kenmerken van de leerlingenpopulatie van de school.*
Kreft laat echter zien dat beide vaak verweven zijn. De schoolorganisatie wordt beïnvloed door het feit dat bepaalde groepen ouders en leerlingen de school kiezen, bijvoorbeeld groepen met een hoog aspiratieniveau en een duidelijke verwachting van het onderwijs. Anderzijds zal zo'n organisatie juist bepaalde groepen aantrekken. Deze kluwen is onontwarbaar.
 - *Schooleffectiviteitsonderzoek kan een monodiscipline zijn.*
Dit is niet houdbaar, omdat de kring van potentiële controle- en verklarende variabelen niet af te grenzen is. Alleen als we ons op de hoogte stellen van de ontwikkelingen in andere vakgebieden weten we welke covariaten interessant zijn voor het onderzoeksontwerp, zoals Kreft laat zien bij het 'status attainment'-onderzoek.
 - *Een contextueel effect is vaak het gevolg van een onjuiste specificatie van relaties op het individuele niveau.*
Deze stelling, die teruggaat op Hauser (1970), zal soms opgaan, maar het is even aannemelijk dat een individueel effect wordt gevonden door een onjuiste specificatie op groepsniveau. Hier zal de theorie ons moeten helpen.
 - *Op groepsniveau gevonden verbanden mogen naar het individuele niveau vertaald worden.*
Als bijv. blijkt dat op scholen met minder zittenblijvers gemiddeld betere prestaties geleverd worden, zegt dat nog niet dat zittenblijvers slechtere schoolprestaties leveren. Het kan heel goed zijn, dat achter deze samenhang op groepsniveau een ontbrekende of tegengestelde samenhang op individueel niveau schuilgaat. Zo is het niet ondenkbaar, dat zittenblijvers zelf niet slechter of beter presteren, maar naarmate zij een groter gewicht in de klas hebben de prestaties van hun klasgenoten drukken. Het risico van halve waarheden van dit type blijft levensgroot aanwezig als alleen data op groepsniveau beschikbaar zijn, zoals de data van de Periodieke Peiling van het Onderwijsniveau (PPON), die als bron voor secundaire analyse ongetwijfeld populair gaan worden (zie bijv. Blok & Eiting, 1988).

Vervolgens gaat Kreft in het tweede deel van haar boek over tot het ontwikkelen en toepassen van een adequaat statistisch model voor multilevel schooleffecten-onderzoek. Dat daar behoefte aan is, kan geïllustreerd aan de hand van een schooleffectenstudie die ongeveer in dezelfde tijd is verschenen. Het gaat om Ton Mooij's dissertatie *Interactional multi-level investigation into pupil behaviour, achievement, competence, and orientation in educational situations*. Deze getuigt weliswaar van een scherp inzicht in de multilevel-problematiek, maar mist nog de middelen om dat in zijn analyses voldoende tot uitdrukking te brengen. Mooij komt als het ware tot het punt waar Kreft verder wil gaan. Alvorens we Kreft verder volgen, gaan we daarom uitgebreid in op het werk van Mooij.

Mooij's boek is de vrucht van tien jaar onderzoek naar 'leerlingproblemen', problemen van en met leerlingen in het voortgezet onderwijs. Het geeft een erg compacte en gestroomlijnde reconstructie van diverse brokken onderzoek, die op de eerste plaats tot het wetenschappelijke forum is gericht. De doelstelling is tweeledig. Ten eerste wil Mooij een bijdrage leveren aan de uitbouw van de 'interactionele' benadering van onderwijsprocessen, door deze ook daadwerkelijk in onderzoek in te zetten. Interactioneel onderzoek moet daarbij volgens hem longitudinaal worden opgezet en plaatsvinden in veldsituaties. Ten tweede wil hij laten zien dat de statistische analyse van relaties uit een interactioneel model, waarin wederzijdse invloeden tussen een leerling en diens schoolse omgeving centraal staan, vraagt om een *multilevel*-aanpak.

Mooij onderscheidt vijf niveaus in het onderwijssysteem, te weten leerling, klas, school, regio en land. Alleen de eerste drie worden in de analyses betrokken. Het databestand komt voort uit omvangrijk onderzoek op scholen met Lager Technisch Onderwijs (LTO). Op de meeste punten is het onderzoek toegespitst op de schoolvakken Nederlands, Wiskunde en Handvaardigheid/Algemene Technieken.

Het hart van de studie wordt gevormd door een flink aantal analyses met telkens wisselende analyse-eenheden en -niveaus, die Mooij groepeerd rond drie tijdsperspectieven:

1. De afzonderlijke *les*. Hier wordt in analyses per vak en leerjaar nagegaan wat de invloed is van leskenmerken als werkvormen, managementgedrag van de leerkracht en de mate van interne differentiatie op de mate waarin leerlingen 'bij de les zijn'. De analyses spelen zich op één niveau af, doordat globale leskenmerken worden gerelateerd aan geaggregeerde – zoals het gemiddelde 'academic on-task behaviour'. Mooij doet dat om de problemen van niveau-verwarring waaraan met name de contextuele analyse mank gaat, te vermijden. Interne differentiatie blijkt dan niet en managementgedrag wel een positief effect te hebben op het 'on-task' gedrag.
2. Het *schooljaar* in een bepaald vak. Binnen dit tijdsperspectief worden twee typen relaties onderzocht:
 - a. De invloed van de *rapportcijfers* in de loop van het jaar op prestaties, motivatie en zelfbeeld van een leerling. Mooij spreekt hier van een intraniveau-proces op het niveau van de individuele leerling. Ook hier blijft Mooij zich bewust op één analyseniveau bewegen, zij het dit keer dat van de individuele leerling.
 - b. De invloed van *klaskenmerken* op de resultaten van leerlingen. Daarbij verstaat Mooij onder klaskenmerken zowel gemeten kenmerken van de samenstelling van klassen (zoals het percentage zittenblijvers), als kenmerken van de 'treatment' door de leerkracht, waarvan Mooij *aanneemt* dat zij processen tussen leerlingen beïnvloeden. Ook hier weer kiest hij achtereenvolgens verschillende invalshoeken:
 - De eerste ontleent hij aan de theorie van het '*frog pond*'-effect, een theorie over het effect van sociale vergelijkingsprocessen in de klas. Deze hypothese over de effecten van relatieve beoordeling in klassen – wat wij hier te lande de Wet van Posthumus zijn gaan noemen – moet Mooij verwerpen.
 - Vervolgens voert Mooij een ander type analyse uit, dat aansluit bij de tra-

ditie van onderzoek naar 'Aptitude-Treatment Interaction' (*ATI*). Zijn aandacht richt zich hier op de mate waarin het eindniveau van een leerling (na een schooljaar) wordt bepaald door diens aanvangsniveau. De voornaamste hypothese is, dat interne differentiatie via individualisering van het leerproces leidt tot een sterkere relatie tussen begin- en eindniveau van een leerling, ofwel ook hier: dat de verschillen in de loop van een jaar groter worden. Hij past dan een zogenaamde tweestapsprocedure ('slopes as outcomes') toe. Daarbij wordt in de eerste stap met data van de individuele leerlingen de relatie tussen begin- en eindniveau berekend in de vorm van een regressie-coëfficiënt (de hellingshoek of 'slope'). In de tweede stap wordt de analyse verplaatst naar het klasniveau en nagegaan hoe deze coëfficiënt, nu als *klaskenmerk*, beïnvloed wordt door andere kenmerken van de klas, zoals hier: de mate van interne differentiatie.

3. Het derde en laatste perspectief van waaruit Mooij analyseert is dat van de *schoolloopbaan* over een periode van vier jaar na de instroom in het voortgezet onderwijs. Daarbinnen plaatst hij twee analyses, een op individueel- en een op schoolniveau:
 - a. In de eerste analyse gaat Mooij na welke factoren van invloed zijn op een aantal aspecten van de schoolloopbaan, zoals de kans op zittenblijven of onvertraagd slagen voor het examen. Aan de rapportcijfers in de eerste twee leerjaren wordt daarbij een centrale rol toebedacht. De analyses zijn multiële regressies op uitsluitend *individueel* niveau, met wisselende resultaten.
 - b. Tenslotte onderzoekt Mooij op het niveau van de *school* de invloed van schoolkenmerken op de relaties tussen begin- en eindniveau van leerlingen. Als schoolkenmerk neemt hij een schaal die de verhoudingen tussen schoolleiding, leerkrachten en leerlingen karakteriseert en tot de categorie 'schoolklimaat' gerekend kan worden. De analyse wordt uitgevoerd met behulp van de genoemde tweestapsprocedure.

Tot zover dit overzicht, dat noodgedwongen nogal uitvoerig is door de veelheid aan invalshoeken die Mooij kiest. Dat is een van de sterke punten van zijn studie. Hij stelt zich een ambitieus doel, maar gaat dat vervolgens stap voor stap te lijf. Daarbij worden steeds wisselende analyseperspectieven gekozen, beredeneerd vanuit de vereisten van een bepaalde deelvraag. Ondanks de omvang van het boek is Mooij's weergave van de analyses daardoor eerder aan de beknopte kant. Bovendien is voor de meting van een aantal schoolkenmerken teruggegrepen op observatie in de klas, een stap die vaak wordt aanbevolen maar in onderzoek op deze schaal niet vaak gezet.

Daar staat tegenover dat de uitwerking van deze opzet op een aantal punten te wensen overlaat. Voor we ons concentreren op de multilevel-aspecten gaan we kort in op een tweetal zwakke punten.

Ten eerste zijn de lijnen tussen theorie en operationalisering soms wel erg dun. Zo worden de schoolvorderingen-toetsen, overgenomen uit CBS-onderzoek, enerzijds rigoureus uitgedund, maar anderzijds drie maal herhaald afgenomen, terwijl zij ontworpen zijn voor de peiling van het instroomniveau en na twee jaar voortgezet onder-

wijs niet zonder meer curriculair relevant kunnen worden geacht.

Ten tweede is Mooij soms te zeer op zoek naar bevestiging van zijn interactionele optiek als zodanig. Dat leidt niet alleen tot het nagenoeg ontbreken van inhoudelijke interpretatie van de resultaten in termen van onderwijsprocessen — wat gezien zijn doelstellingen te verdedigen is —, maar ook tot ongefundeerde causale interpretaties van gevonden verbanden. Dat betreft vooral de rol van rapportcijfers in de analyses op individueel niveau. Mooij lijkt die hoofdzakelijk te interpreteren als beoordelingsmoment dat van invloed is op motivatie en zelfbeeld van een leerling, bijna zelfs als uitdrukking van een evaluatie-strategie van de leerkracht, en negeert het beschrijvende moment dat ze als indicatie van het prestatieniveau natuurlijk ook hebben, zeker gezien de zwakte van de toetsen die in het onderzoek gebruikt zijn. Op één punt gaat hij zover om op aanbeveling van LISREL effecten van kenmerken aan het eind van het schooljaar op het rapportcijfer met Pasen te introduceren, wat een dubieus element van niet-recursie met zich brengt en causaal onhoudbaar is. Hij interpreteert dat dan als ondersteuning bij uitstek van het interactionele model, terwijl een interpretatie in termen van meetfouten en ongemeten gemeenschappelijke oorzaken meer voor de hand ligt.

Afgezien van een theoretisch pleidooi bestaat de multilevel-kant van Mooij's analyses bij nader inzien hoofdzakelijk in het *vermijden* van niveau-problemen, wat naar ons idee ook wel lukt. Daarbij blijft hij, zoals we al eerder lieten uitkomen, behoedzaam op telkens één niveau analyseren. Hij heeft echter de neiging om slechts delen van de voorgestelde procedures over te nemen, waardoor een aantal analyses onvolledig is.

Ten aanzien van de ATI-hypothese bijvoorbeeld voert Mooij alleen een binnengroepsanalyse uit. Dat leidt echter tot een onvolledig beeld ten aanzien van de onderzochte relaties, zoals de effecten van differentiatie binnen klasverband. Bij de evaluatie zijn we immers niet alleen geïnteresseerd in de vraag of door interne differentiatie de verschillen in de klas groter worden — de vraag waarop een binnengroepsanalyse antwoord kan geven — maar ook in de vraag welke categorieën leerlingen erbij gebaat zijn en welke niet — waarvoor een tussengroepsanalyse van klasgemiddelden of -intercepten nodig is. Een 'vlakke' relatie tussen begin- en eindniveau zal anders gewaardeerd worden als die samengaat met een hoog gemiddeld niveau (het ideaal van de Middenschool) dan wanneer die samengaat met een laag gemiddelde (het gevreesde 'niveauverlies' van de tegenstanders).

Een tweede gevolg van de nadruk op het vermijden van niveauproblemen is dat de vele analyses los van elkaar blijven staan. Integratie van de uitkomsten wordt door Mooij achterwege gelaten. Daarvoor zijn de analysetechnieken die hij gebruikt ook ontoereikend.

Inmiddels staan ons echter nieuwe analysetechnieken ter beschikking waarmee de multilevel-beloofte beter kan worden ingelost. Daarmee zijn we terug bij Kreft. In het tweede deel van haar studie laat zij zien in welke richting een goed statistisch model voor multilevel-analyse gezocht moet worden. Zij introduceert daartoe het zogenaamde 'random coefficient model'. Dat lijkt op het model dat we door Mooij toege-

past zagen (dus met statistische grootheden uit de ene analysestap als afhankelijke variabelen in een volgende stap), maar is wel statistisch verantwoord. In feite gaat het om een uitbreiding van het conventionele regressiemodel tot een meer algemeen lineair model, één met minder assumpties. Zo wordt geen onafhankelijkheid van waarnemingen binnen groepen verondersteld en mag zowel het intercept als de regressie van de criteriumvariabele op de onafhankelijken van groep tot groep verschillen. Bovendien hoeft niet de assumptie gemaakt te worden dat deze verschillen vast (dat wil zeggen niet-stochastisch) zijn, zoals Mooij noodgedwongen moest veronderstellen, maar er wordt juist van uitgegaan dat zij gesplitst kunnen worden in twee componenten: verschillen ten gevolge van 'echte' parametervariantie en verschillen ten gevolge van steekproevenvariantie. Kortom, de totale variantie van een effectvariabele kan in drie componenten ontleed worden:

- variantie tussen leerlingen binnen scholen;
- variantie tussen scholen, die is toe te schrijven aan parametervariantie in de scholenpopulatie;
- variantie tussen scholen, die is toe te schrijven aan de toevallige samenstelling van de steekproef van scholen.

Deze splitsing in variantiecomponenten past bij uitstek bij de gangbare wijze van steekproeftrekking in schooleffectenonderzoek: tweetraps steekproeven (cluster samples) waarbij scholen met leerlingen worden getrokken en waarbij men ook op het niveau van de scholen wil kunnen generaliseren.

Terugkomend op de eerste van de genoemde valkuilen – scholen verklaren zo weinig variantie – kunnen we stellen: de variantie die door scholen verklaard zou moeten worden, zal in het algemeen kleiner zijn dan met de gangbare analyseprocedure impliciet wordt aangenomen. Het gaat alleen om de tweede variantiecomponent, de parametervariantie tussen scholen (in de nu beschikbare toepassingen blijkt bij de intercepten 50% tot 80% van de variantie tussen scholen echte variantie te zijn, terwijl dit bij de hellingen maar 20% tot 30% is). Ook de tweede valkuil – dat scholen elke leerling in gelijke mate beïnvloeden – kan nu onderzocht worden: de variantie in de random regressiecoëfficiënt (bijvoorbeeld die van de schoolprestaties op het milieu van herkomst) mag in dat geval niet significant van nul afwijken. Als dat wel het geval is, kunnen we door het opnemen van geschikte interactietermen (bijvoorbeeld: schooltype * milieu) trachten dit verschijnsel te voorspellen.

In het laatste hoofdstuk van haar proefschrift demonstreert Kreft op de SMVO-data de toepassing van het 'random coefficient'-model, aan de hand van een analyse van het effect van scholengemeenschappen en categoriale scholen op de loopbaan van leerlingen met verschillende cognitieve beginvaardigheden. Na correctie voor deze individuele verschillen blijkt een fors effect het resultaat: een gemiddelde leerling komt op een scholengemeenschap een half 'niveau' hoger uit. Een leerling bijvoorbeeld die in het categoriale systeem de MAVO haalt, eindigt – ceteris paribus – op een scholengemeenschap ergens tussen MAVO en HAVO. Met plaatjes laat Kreft zien hoe een en ander kan worden voorgesteld.

Op de uitkomsten valt overigens wel wat af te dingen. De loopbaanvariabele die

Kreft hanteert, is opgezet als een procesvariabele. Door de brugperiode zal het begin van de loopbaan op een scholengemeenschap in veel gevallen soepeler verlopen dan op een categoriale school, wat automatisch leidt tot een hogere score. Het resultaat van Kreft kan een artefact van de gebruikte loopbaanvariabele zijn.

Plaatjes zijn bij deze analyses niet alleen nuttig om een en ander inzichtelijk te maken voor minder statistisch geschoolden, maar ook omdat ze ons wijzen op een aantal problemen rond het gebruik van intercepten in multilevel-analyses. Om te beginnen is het intercept in veel gevallen inhoudelijk niet te interpreteren, doordat de gebruikte variabele geen nulpunt heeft. Het intercept kan bijvoorbeeld staan voor de prestatiescore van een leerling met een IQ van '0'. Bovendien is het intercept in bijna alle gevallen een extrapolatie, die minder betrouwbaar is naarmate zij verder buiten het waarnemingsinterval ligt. Ten slotte mogen in het 'random coefficient'-model helling en intercept correleren, waardoor een afzonderlijke interpretatie van een van beide riskant is. De beste manier om deze bezwaren te ondervangen, is het toepassen van centrering (geïntroduceerd door Boyd & Iversen, 1980). Daarbij worden per school de individuele scores omgezet in afwijkingen van het schoolgemiddelde en met die afwijkingen wordt dan verder geanalyseerd. Feitelijk is dit de techniek van variantie-pooling die ook Mooij gebruikt. Dat leidt tot verschuivingen in de interpretatie: het schoolgemiddelde wordt het nulpunt, er wordt niet meer geëxtrapoleerd en de helling kan nu afzonderlijk worden beschouwd.

De opbrengst van beide dissertaties laat zich als volgt samenvatten. Onderzoek in veldsituaties vraagt zowel theoretisch als statistisch om een multilevel-benadering. Mooij weet overtuigend te betogen dat onderwijsprocessen het best opgevat kunnen worden als interacties die verlopen in de tijd (longitudinaal) en in de gelede sociale ruimte (multiniveau). Kreft laat zien dat dit geen luchtkastelen behoeven te blijven, maar dat het goed mogelijk is dit soort theorieën daadwerkelijk te onderzoeken.

Een indruk van de vele situaties waarin het multilevel-model toepasbaar is, kunnen we ontleen aan Goldstein (1987). Deze heeft onlangs zijn werk op het gebied van 'random coefficient'-modellen gebundeld in *Multilevel models in educational and social research*. Hij laat daarbij diverse toepassingsgebieden de revue passeren. Zo kunnen herhaalde metingen worden behandeld als waarnemingen die zijn gegroepeerd onder het 'niveau' van individuen. Ook multivariate analyses kunnen zo worden opgevat, met meerdere afhankelijke variabelen als waarnemingen onder het niveau van de persoon waarop zij betrekking hebben. Verder kan met overlappende niveaus gewerkt worden, zoals wanneer leerlingen worden opgevat als gegroepeerd onder zowel scholen als woonwijken. Ten slotte kunnen ook drie of meer hiërarchische niveaus onderscheiden worden, zoals in analyses van de mate waarin scholen leerkrachteffecten op leerlingen bewerkstelligen of de ontwikkeling van leerlingen beïnvloeden. Het model is zelfs toegepast voor meta-analyse (Raudenbush & Bryk, 1985). De beschikbaarheid van een instrumentarium voor dergelijke analyses kan op zijn beurt een stimulans voor theorie-ontwikkeling vormen.

Inmiddels begint ook programmatuur te verschijnen die helemaal op multilevel-ana-

lyse is toegesneden: VARCL-9 (Longford, 1988), HLM (Bryk et al., 1986) en ML.2 (Goldstein, 1988). Deze programma's hebben tal van geavanceerde mogelijkheden, zoals analyse met maximaal negen niveaus, centreer-opties ter vermindering van intercept-problemen, etcetera. Het hoeft niet lang te duren voor deze programma's als modules in de grote standaardpakketten beschikbaar zijn. Reden te meer om in het M&T-onderwijs het multilevel-model een ruimere plaats te geven.

Literatuur

- Blok, H. & H.M. Eiting (1988). Openbare en bijzondere lagere scholen vergeleken op de taalvaardigheid van hun zesdeklassers. *Mens en Maatschappij*, 63, 260-276.
- Boyd, L.H. & G.R. Iversen, (1979). *Contextual analysis: concepts and statistical techniques*. Belmont, Wadsworth.
- Bryk, A.S., S.W. Raudenbush, M. Seltzer & R.T. Congdon (1986). *An introduction to HLM. Computer program and user's guide, version 1.0*. Chicago, University of Chicago.
- Cronbach, L.J. (1976). *Research on classrooms and schools: formulation of questions, design and analysis*. Occasional paper. Stanford, Stanford Evaluation Consortium.
- Goldstein, H. (1987). *Multilevel models in educational and social research*. London, Charles Griffin & Co.
- Goldstein, H. & J. Rasbash (1988). *Multilevel software, version 88.3*. London, University of London, Institute of Education.
- Hauser, R.M. (1970). Context and consex: a cautionary tale. *American Journal of Sociology*, 75, 645-663.
- Kreft, G.G. (1987). *Models and methods for the measurement of school effects*. Dissertatie Universiteit van Amsterdam.
- Longford, N. (1988). *Variance component analysis (VARCL3 & VARCL9): manual*. Princeton, Educational Testing Service.
- Mooij, T. (1987). *Interactional multi-level investigation into pupil behaviour, achievement, competence, and orientation in educational situations*. Nijmegen, ITS.
- Raudenbush, S.W. (1987). *Educational applications of hierarchical linear models. A review*. Michigan State University.
- Raudenbush, S.W. & A.S. Bryk (1985). Empirical Bayes meta-analysis. *Journal of Educational Statistics*, 10, 75-98.
- Raudenbush, S.W. & A.S. Bryk (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59, 1-17.