

2 | Een geneste structuur

2.1 Inleiding

Om verschillen in prestaties van leerlingen te verklaren zijn kenmerken op verschillende niveaus nodig. Deze verschillende niveaus worden in dit onderzoek in een geneste structuur weergegeven, namelijk leerlingen binnen klassen binnen scholen.

Op soortgelijke wijze kan voor de verklaring van verschillen in ziekteverzuim van docenten bijvoorbeeld worden gekeken naar kenmerken op verschillende niveaus. Voor de verklaring zijn bedrijfskenmerken, zoals het schooltype, het personeelsbeleid of de gemiddelde leeftijd van het docententeam, niet genoeg. Een deel van de verschillen kan namelijk worden toegeschreven aan persoonskenmerken, zoals de sekse, de leeftijd, het aantal uren of de motivatie van docenten. Om de vraag naar het ziekteverzuim van docenten te beantwoorden kan dus het beste een geneste structuur worden gehanteerd, namelijk docenten genest binnen scholen. In onderzoek naar verschillen in het gemiddeld aantal dagen dat artsen patiënten na een bepaalde operatie in het ziekenhuis laten verblijven, is de multiniveau structuur patiënten genest binnen artsen binnen ziekenhuizen. Om te achterhalen in hoeverre het loon van werknemers verschilt per bedrijfstak en per land is een structuur op te stellen met werknemers genest binnen bedrijfstakken binnen landen. Voor talrijke vraagstukken in de sociale wetenschappen is het van belang om met de hiërarchie van gegevens rekening te houden.

Door de ontwikkeling van technieken voor multiniveau analyse is het mogelijk geworden om data afkomstig van geneste structuren te analyseren.¹ In de jaren tachtig kwam deze ontwikkeling in een stroomversnelling door het werk van onder andere Aitkin en Longford (1986), Goldstein (1995²) en Raudenbush en Bryk (1986). Tot die tijd analyseerden wetenschappers hiërarchische gegevens gewoonlijk op één niveau. Om op één niveau te analyseren, moest er een keuze gemaakt worden tussen *desaggregeren* of *aggregeren*. Aan beide methoden zijn echter grote nadelen verbonden. Bij desaggregatie worden gegevens van een hoger op een lager niveau geanalyseerd. In een onderzoek met gegevens van 7.000 leerlingen en 150 scholen, worden de schoolgegevens dan op individueel niveau geanalyseerd. Het aantal onderzoekselementen op schoolniveau wordt zo

op miraculeuze wijze vermenigvuldigd, want er wordt gedaan alsof er over leerlingen en scholen evenveel verschillende, onafhankelijke gegevens zijn. De effecten van gegevens op een hoger niveau worden door desaggregatie opgeblazen.

Door individuele gegevens naar een hoger niveau te aggregeren kunnen een verlies aan onderscheidingsvermogen (*power*), ecologische fouten en interpretatieproblemen ontstaan. In het voorbeeld met 7.000 leerlingen en 150 scholen wordt bij aggregatie het aantal onderzoekselementen teruggebracht tot 150, waardoor het moeilijker wordt om een onderscheid te maken tussen verbanden die wel en die niet ‘waar’ zijn. Door de verkleining van het aantal onderzoekseenheden neemt de kans op de zogeheten *fout van de tweede soort* toe. Dit is de kans op het ten onrechte handhaven van de nulhypothese. Ecologische fouten en interpretatieproblemen kunnen voorkomen doordat individuele gegevens op hoger niveau worden geanalyseerd. Het klassieke voorbeeld van een ecologische fout is de relatie tussen het percentage zwarten en het stemgedrag in Amerikaanse wijken (Robinson, 1950). Naarmate er meer zwarten in een wijk wonen, wordt er meer op de Republikeinse partij gestemd. Uit dit verband zou de verkeerde afleiding kunnen worden gemaakt dat zwarten vaak Republikeins stemmen. Met gegevens op wijkniveau kunnen echter geen uitspraken op individueel niveau worden gedaan. Uit deze geaggregeerde gegevens kan dus niet worden opgemaakt wat blanken of zwarten stemmen.

Een voorbeeld van een interpretatieprobleem kan worden toegelicht met het Trouw-onderzoek. Uit dat onderzoek blijkt dat er een verband is tussen brede scholen en het rendement van de havo en vwo. Brede scholen blijken goed uit te pakken voor de hogere schooltypen van voortgezet onderwijs. De vraag die een ouder zich zou willen stellen is echter of brede scholen ook goed uitpakken voor potentiële havo- en vwo-leerlingen. Die vraag kan niet met de gegevens van Trouw worden beantwoord.ⁱⁱ Ook kan niet worden nagegaan waarom scholen met veel allochtone leerlingen nu slechter presteren in het Trouw-onderzoek. Komen deze slechtere prestaties doordat allochtone leerlingen (een kenmerk op individueel niveau) een achterstand hebben? Of komt dit doordat het percentage allochtone leerlingen (een kenmerk op schoolniveau) van invloed is op prestaties?

Voor de analyse van geneste structuren is multiniveau analyse dus aangewezen (Bryk & Raudenbush, 1992; Goldstein, 1995²; Hox, 1994; Snijders & Bosker, 1999; Woodhouse, Rasbash, Goldstein & Yang, 1996). Multiniveau analyse heeft een aantal voordelen. Door rekening te houden met de hiërarchie van gegevens zijn de standaardfouten, betrouwbaarheidsintervallen en significantietoetsen nauwkeuriger. Hierbij geldt dat deze statistieken voor een analyse van een geneste

structuur doorgaans ‘conservatiever’ zijn in een multiniveau analyse dan in een gedesaggregeerde analyse. In een gedesaggregeerde analyse zijn de standaardfouten vaak kleiner dan in een multiniveau analyse. Als de regressiecoëfficiënten in beide methoden even groot zijn, dan is het significantieniveau van een effect in een analyse op één niveau lager dan in een analyse op meer niveaus. Het significantieniveau wordt immers bepaald door de verhouding tussen de regressiecoëfficiënt en de standaardfout. Deze sterkere significantie is echter schijn, omdat hij berust op de onjuiste assumptie van onafhankelijke onderzoekseenheden (in dit geval leerlingen).

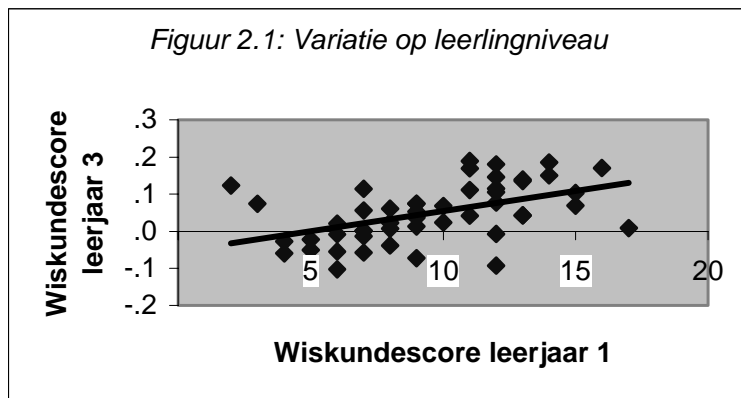
In de tweede plaats geldt dat er met multiniveau analyse onderscheid gemaakt kan worden tussen regressie-effecten binnen en tussen scholen, waardoor er minder interpretatieproblemen optreden. De vraag of etniciteit niet alleen op individueel maar ook op schoolniveau invloed heeft op prestaties, kan met multiniveau analyse worden beantwoord. In de derde plaats kan met multiniveau analyse naar differentiële effecten worden gekeken: is een school voor alle leerlingen even goed? Er kan bijvoorbeeld worden nagegaan of scholen meer verschil uitmaken voor leerlingen uit lagere sociale milieus dan voor leerlingen uit hogere sociale milieus. Ten vierde geldt dat het met multiniveau analyse eenvoudig is om scholen te rangordenen naar hun netto prestatie. Hiervoor kan gebruik worden gemaakt van wat genoemd wordt *posterior means*. In paragraaf 2.3 worden de scholen op deze manier geordend. Verder geldt dat met multiniveau analyse veelal recht wordt gedaan aan de hiërarchische structuur van de *steekproefselectie*. Zo is de steekproef van dit onderzoek in drie stappen geselecteerd. Eerst zijn er scholen getrokken, vervolgens klassen en uit die klassen zijn alle leerlingen genomen. In dit onderzoek is er dus sprake van een drietraps-steekproef.

Voor de analyse van geneste structuren is multiniveau analyse dus noodzakelijk en kan er niet worden volstaan met eenvoudigere, maar ongeschiktere analyses op één niveau. In paragraaf 2.2 van dit hoofdstuk behandel ik de basis van multiniveau analyse. Deze paragraaf biedt de technische achtergrond voor de komende hoofdstukken. In paragraaf 2.3 laat ik zien dat het voor een oordeel over de prestaties van scholen veel uitmaakt of er op één of meer niveaus wordt geanalyseerd.

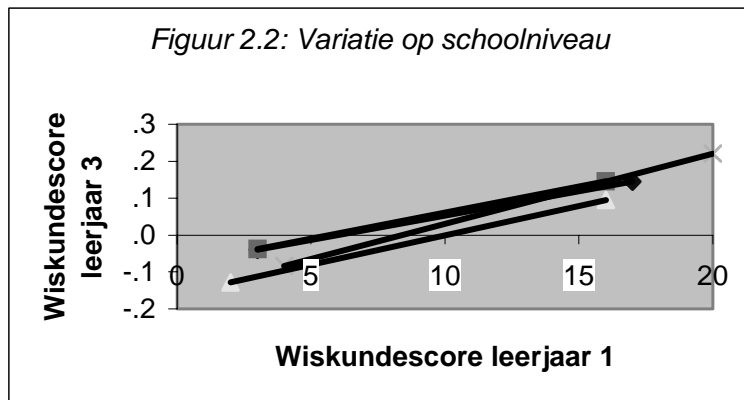
2.2 Multiniveau analyse

Verschillende niveaus

Multiniveau analyse zal ik in de komende hoofdstukken gebruiken voor het verklaren van de invloed van school-, docent-, leerling- en gezinskenmerken op de prestaties en vorderingen van leerlingen. In deze analyses maak ik gebruik van vier niveaus. Niveau nul is een *pseudo-niveau*, dat nodig is voor een analyse met verschillende afhankelijke variabelen tegelijk. Op dat niveau bevinden zich de scores op de toetsen Nederlands en wiskunde. Het eerste niveau is het *individuele niveau* met leerling- en gezinskenmerken. Op het tweede en derde niveau, het *klas-* en *schoolniveau*, zijn de docent- en schoolkenmerken gemodelleerd. De multiniveau structuur voor dit onderzoek is dus afhankelijke variabelen genest binnen leerlingen binnen klassen binnen scholen.



In multiniveau analyse gaat het om het vinden van een formule waarmee de scores op de afhankelijke variabelen zo goed mogelijk geschat kunnen worden vanuit de scores op de onafhankelijke variabelen. In Figuur 2.1 is elke score als een ruit weergegeven. Om in een lineaire regressieanalyse op één niveau de afhankelijke variabele Y te voorspellen uit de onafhankelijke variabele X is het nodig om een vergelijking van een rechte lijn op te stellen, waarbij Y wordt voorspeld uit het *intercept*, de waarde van Y als X gelijk is aan nul, en de *helling* (of: regressiecoëfficiënt; in het Engels: *slope*) van de lijn met de X -as. Wanneer de voorspelling van Y uit X niet perfect is, dan liggen niet alle scores van Y op de regressielijn, zoals in Figuur 2.1 het geval is. De afwijkingen van deze scores tot de regressielijn is de variatie die niet verklaard is, de onverklaarde of residuele *variantie*. Tussen de scores op Y en de voorspellingen van Y door X zijn er dan verschillen. Deze verschillen worden aangeduid met de term *residu* (of: fout), omdat het dat gedeelte van de score van Y is dat niet wordt voorspeld door de regressievergelijking.



Het multiniveau model bouwt voor een deel voort op het lineaire regressiemodel. In Figuur 2.2 is voor vier scholen een regressielijn weergegeven waarin de relatie tussen de wiskundetoetsen in leerjaar één en drie wordt getoond. Omwille van de presentatie ontbreken de scores van leerlingen in Figuur 2.2. Voor alle vier de scholen is de relatie tussen de beide toetsen positief. Naarmate leerlingen in het eerste leerjaar beter scoren op wiskunde behalen ze twee jaar later gemiddeld genomen een hogere score op dit vak. De regressielijnen zijn echter niet identiek. Scholen verschillen namelijk in intercept en in helling. Het intercept geeft de score aan die leerlingen in leerjaar drie behalen, wanneer ze in leerjaar één nul punten scoorden op de wiskundetoets. In dit voorbeeld ligt het punt $x=0$ niet in het bereik van x . Een verschil in helling houdt in dat de relatie tussen de scores in beide leerjaren van school tot school verschilt. Op scholen waar de relatie sterker (een steilere helling) is, gaan de prestaties van leerlingen meer uiteenlopen. Deze scholen worden dan ook wel *selectief* genoemd. Op scholen waar de relatie zwakker (een vlakkere helling) is, treedt een nivellering in prestaties van leerlingen op. Deze scholen worden daarom *egalitair* genoemd. Met verschillen in zowel het intercept als de helling kan de variatie tussen scholen complex zijn.

Multiniveau modellen

Het meest eenvoudige model voor een multiniveau analyse met drie niveaus en één afhankelijke variabele is het *lege model*. Het lege model voor de verklaring van Y bestaat uit de volgende vier termen: het intercept en de residuen op school-, klas- en individueel niveau.

$$(2.1) Y_{ijk} = \beta_{000} + V_{00k} + U_{0jk} + R_{ijk}$$

In het model is i de index voor leerlingen, j de index voor klassen en k de index voor scholen. Wanneer een variabele de index ijk heeft dan varieert de score op leerling-, klas- en schoolniveau. Wanneer een variabele de index $00k$ heeft dan varieert het alleen op schoolniveau. In de vergelijking zijn V_{00k} , U_{0jk} en R_{ijk} random variabelen met een gemiddelde van nul en, zo wordt aangenomen, een normale verdeling en respectievelijk als residuele variantie Φ_v^2 , Φ_u^2 en Φ_r^2 . Door de opname van de *random effects* (V_{00k} , U_{0jk} en R_{ijk}) in de analysestructuur wordt multiniveau analyse gekenmerkt.

Het lege model is nuttig om de residuele varianties tussen scholen, klassen en leerlingen uit te rekenen. In hoofdstuk 1 werd in dit verband gesproken over relatieve binnenschoolse en tussenschoolse variantie. Hierbij is de *intra-klasse correlatiecoëfficiënt* de maat om de verhouding van de residuele varianties aan te geven. De intra-klasse correlatiecoëfficiënt voor het klas- en schoolniveau ten opzichte van het individueel niveau is de variantie *tussen* scholen en klassen gedeeld door de totale variantie: $(\Phi_v^2 + \Phi_u^2) / (\Phi_v^2 + \Phi_u^2 + \Phi_r^2)$. De intra-klasse correlatiecoëfficiënt voor het schoolniveau ten opzichte van het klasniveau is: $\Phi_v^2 / (\Phi_v^2 + \Phi_u^2)$. De intra-klasse correlatiecoëfficiënt is dus de proportie variantie die op hoger niveau wordt aangetroffen. Het is een correlatiecoëfficiënt, omdat deze gelijk is aan de correlatie tussen de scores van twee random getrokken onderzoekselementen (micro-units) in dezelfde, random getrokken macro-unit. In de volgende hoofdstukken wordt deze correlatiecoëfficiënt veelvuldig gehanteerd en zal de formule van voorbeelden worden voorzien.

De tweede stap is om het lege model uit te breiden door *vaste effecten* van onafhankelijke variabelen op het *individuele niveau* op te nemen, bijvoorbeeld de intelligentie van een leerling. Door de opname van deze variabele wordt getracht om de variatie in Y te verklaren. Met één onafhankelijke variabele X ziet het model er als volgt uit:

$$(2.2) Y_{ijk} = \beta_{000} + \beta_{100} x_{ijk} + V_{00k} + U_{0jk} + R_{ijk}$$

Behalve uit een intercept en drie residuele variantiecomponenten bestaat het model uit een regressiecoëfficiënt β_{100} . Deze coëfficiënt kan op de gebruikelijke manier worden geïnterpreteerd: een toename van een eenheid X leidt tot een toename in Y van β_{100} eenheden. Het intercept en de regressiecoëfficiënt vormen het vaste gedeelte van het model. De drie variantiecomponenten vormen het random gedeelte. Voor het model kan opnieuw de verhouding tussen de variantiecomponenten worden berekend. Deze maat wordt nu de *residuele intra-klasse correlatiecoëfficiënt* genoemd. De correlatie is gelijk aan de correlatie tussen de scores van twee micro-units in dezelfde macro-unit, *gegeven variabele X*.

De onafhankelijke variabele X kan op een hoger niveau differentiële effecten hebben. Vandaar dat er als derde stap moet worden gekeken of er op klas- of schoolniveau *random effects* zijn. Een model met een random intercept en een random helling van de predictorvariabele ziet er als volgt uit:

$$(2.3) Y_{ijk} = \beta_{0jk} + \beta_{1jk} x_{ijk} + R_{ijk}$$

met de volgende uitwerkingen:

$$(2.4) \beta_{0jk} = \gamma_{00k} + U_{0jk}$$

$$(2.5) \beta_{1jk} = \gamma_{10k} + U_{1jk}$$

$$(2.6) \gamma_{00k} = \delta_{000} + V_{00k}$$

$$(2.7) \gamma_{10k} = \delta_{100} + V_{10k}$$

Het *random intercept* op klasniveau (2.4) is opgebouwd uit een schaalgemiddelde en een klasspecifieke afwijking. Op schoolniveau bestaat het random intercept (2.6) eveneens uit een schaalgemiddelde en een afwijking. De *random helling* (2.5) duidt erop dat er voor bepaalde groepen leerlingen, bijvoorbeeld leerlingen met een hoge intelligentie in vergelijking met leerlingen met een lage intelligentie, differentiële effecten zijn. Naarmate \bar{x}_{ijk} groter is, zijn er in een klas grotere verschillen tussen leerlingen naar gelang hun intelligentie. Klassen kunnen dus verschillen in het gemiddelde prestatieniveau en ook naar de relatie van intelligentie met prestaties. Op schoolniveau houdt de *random helling* (2.7) in dat de relatie van intelligentie met prestaties verschilt van school tot school. Beide random hellingen hebben ook een component voor afwijking, zoals in de formule te zien is. Substitutie van de vier formules in 2.3 leidt tot een vergelijking met een intercept, een vast effect en vijf random effecten:

$$(2.8) Y_{ijk} = \delta_{000} + \delta_{100} x_{ijk} + V_{00k} + V_{10k} x_{ijk} + U_{0jk} + U_{1jk} x_{ijk} + R_{ijk}$$

Als vierde stap kunnen *effects* van variabelen op *hoger niveau* worden toegevoegd, bijvoorbeeld een vast effect van de grootte van een klas of het aantal allochtone leerlingen op een school. Aan de indices kan worden gezien op welk niveau de variabelen variëren.

$$(2.9) Y_{ijk} = \delta_{000} + \delta_{100} x_{ijk} + \gamma_{010} x_{0jk} + \delta_{001} x_{00k} + V_{00k} + V_{10k} x_{ijk} + U_{0jk} + U_{1jk} x_{ijk} + R_{ijk}$$

Het model bestaat nu uit een vast gedeelte met een intercept en drie vaste effecten en een random gedeelte met vijf effecten. Dit model is vooral ter illustratie. In de komende hoofdstukken zal blijken dat de gehanteerde multiniveau modellen meer

vaste dan random effecten bevatten. De random effecten bevinden zich in die modellen met name op het klasniveau en niet op het schoolniveau.

Dit model kan tot slot nog worden uitgebreid door *multiniveau interacties* (cross-level interactions) op te nemen, bijvoorbeeld tussen een variabele op klasniveau (x_{0jk}) en een variabele op individueel niveau (x_{ijk}):

$$(2.10) Y_{ijk} = \delta_{000} + \delta_{100}x_{ijk} + \delta_{010}x_{0jk} + \delta_{001}x_{00k} + \delta_{110}x_{0jk}x_{ijk} \\ + V_{00k} + V_{10k}x_{ijk} + U_{0jk} + U_{1jk}x_{ijk} + R_{ijk}$$

Het vaste gedeelte van het model bestaat uit een intercept, drie hoofdeffecten en een interactie-effect. Het random gedeelte bestaat wederom uit vijf effecten.

Multivariate analyse

Een multivariaat model is een model met twee of meer afhankelijke variabelen per onderzoekselement. In dit onderzoek gebruik ik twee criteriumvariabelen, te weten de toetsscores Nederlands en wiskunde. Deze afhankelijke variabelen bevinden zich op niveau nul van het multiniveau model, het pseudo-niveau.

Hoewel het eenvoudiger is om voor elke afhankelijke variabele een aparte analyse uit te voeren, is er een aantal voordelen van een multivariate aanpak (Duncan, Jones & Moon, 1996; Goldstein, 1995²; Snijders & Bosker, 1999). Een belangrijk pluspunt van multivariate modellen is dat kan worden nagegaan of variabelen een *vergelijkbaar of verschillend effect* hebben op de afhankelijke variabelen. Er kunnen significantietoetsen worden uitgevoerd om na te gaan of de effecten verschillen. In de volgende hoofdstukken zal dit voordeel veelvuldig worden benut. In de tweede plaats kan de *correlatie* tussen de afhankelijke variabelen *op verschillende niveaus* worden bepaald. Behalen dezelfde leerlingen goede prestaties op Nederlands en wiskunde? Zijn klassen waar de prestaties gemiddeld op het ene vak hoog zijn ook de klassen waar de prestaties op het andere vak hoog zijn? En hoe ziet deze relatie eruit op schoolniveau? Deze correlaties kunnen niet worden uitgerekend wanneer er aparte analyses voor Nederlands en wiskunde worden uitgevoerd. Tot slot hoeven de respondenten niet op alle afhankelijke variabelen een score te hebben. *Ontbrekende waarden* spelen dus minder een probleem bij multivariate analyse. Voor dit onderzoek geldt dus dat leerlingen in de analyse kunnen worden meegenomen als ze een score hebben op ten minste één van de twee toetsen. Hierbij geldt dat leerlingen die maar op één van de toetsen een score hebben alleen maar meedoen in de analyse van die toets. Door de covariantie tussen beide toetsen wordt de analyse van de andere toets iets nauwkeuriger in de multivariate aanpak.

Om de multivariate analyse uit te kunnen voeren, moeten de afhankelijke variabelen genest worden binnen leerlingen. Per leerling moet dan voor elke criteriumvariabele een record worden gemaakt. In een analyse met één afhankelijke variabele ziet een datamatrix er als volgt uit:

Tabel 2.1: Datamatrix voor multiniveau analyse

NUMMER	KLAS	SCHOOL	NEDERLANDS	WISKUNDE	ETNICITEIT	SES
1	1	1	115	.	0	3
2	1	1	100	85	1	2
.						
.						
6896	452	149	112	130	0	4

In dit onderzoek doen 6.896 leerlingen mee. Deze leerlingen hebben een score op ten minste een van de beide prestatietoetsen. Leerling 1 heeft een ontbrekende waarde op wiskunde. Voor multivariate analyse moet deze matrix worden omgezet naar de volgende structuur:

Tabel 2.2: Datamatrix voor multivariate multiniveau analyse

NUMMER	KLAS	SCHOOL	PRESTATIE	INDICATOR	ETNICITEIT	SES
1	1	1	115	1	0	3
2	1	1	100	1	1	2
2	1	1	85	2	1	2
.						
.						
6896	452	149	112	1	0	4
6896	452	149	130	2	0	4

In deze structuur zijn de scores op Nederlands en wiskunde samengevoegd. De scores kunnen van elkaar worden onderscheiden door een indicator die 1 is voor Nederlands en 2 voor wiskunde. Het pseudo-niveau wordt dus geschapen door gebruik te maken van een dummy. Voor de multivariate analyse zijn er maximaal 13.792 onderzoekselementen, dat wil zeggen twee voor elk van 6.896 leerlingen. Aangezien sommige leerlingen niet beide toetsen hebben gemaakt, zal het aantal onderzoekselementen wat lager uitvallen.

In een multivariate multiniveau analyse met twee afhankelijke variabelen genest binnen leerlingen binnen klassen binnen scholen, is Y_{hijk} de afhankelijke va-

riabele. Het is de meting van afhankelijke variabele h , voor persoon i binnen klas j en school k . Wanneer het random intercept model (formule 2.2) wordt uitgebreid voor de multivariate aanpak, dan is het resultaat:

$$(2.11) Y_{hijk} = \beta_{h000} + \beta_{h100} x_{1hijk} + V_{h00k} + U_{h0jk} + R_{hijk}$$

Vergelijking 2.11 bestaat dus voor elk van de afhankelijke variabelen uit een intercept, een regressiecoëfficiënt en drie residuele variantiecomponenten. Er is een random gedeelte op het leerling-, klas- en schoolniveau, maar niet op het pseudo-niveau. Stel dat de onafhankelijke variabele X in dit multivariate model het aantal dagen is dat een leerling huiswerk maakt, dan worden er voor deze predictor twee parameters geschat: één voor het effect op Nederlands en één voor het effect op wiskunde. Door een chi-kwadraattoets uit te voeren kan worden nagegaan of deze effecten significant verschillen (Goldstein, 1995²). Wanneer dat niet het geval is, kan worden volstaan met één parameter voor het effect van het maken van huiswerk op beide afhankelijke variabelen.

Op de drie niveaus met een residuele variantiecomponent bestaat het random gedeelte uit de variantie van elk van de afhankelijke variabelen plus de covariantie tussen de afhankelijke variabelen. Met deze drie schattingen kan op elk niveau de correlatie tussen de criteriumvariabelen worden berekend. Op het schoolniveau is dit bijvoorbeeld:

$$(2.12) \rho(U_{100k}, U_{200k}) = \frac{\text{cov}(U_{100k}, U_{200k})}{sd(U_{100k})sd(U_{200k})}$$

Gegeven de opgenomen verklarende variabelen, geeft deze correlatie aan of scholen die gemiddeld genomen goed scoren op Nederlands ook de scholen zijn die gemiddeld genomen goed scoren op wiskunde.

Significantietoetsen

Om na te gaan of de *vaste effecten* significant van nul afwijken wordt de parameterwaarde gedeeld door de bijbehorende standaardfout. Dit quotiënt is een t-waarde. Hoewel er duidelijke hypothesen zijn over de richting van effecten, wordt in dit onderzoek voor het gemak tweezijdig getoetst. Op die manier vormt het voor de weergave van de significantie van effecten geen probleem dat een aantal effecten een richting heeft die tegengesteld is aan de verwachte richting. De toetsing is hierdoor aan de voorzichtige kant voor de meeste effecten. Voor een eenzijdige toetsing is een effect dat kleiner is dan $-1,645$ of groter is dan $1,645$ significant bij een alfa van $0,05$ en meer dan duizend vrijheidsgraden; bij een tweezijdige toetsing moet de t-waarde absoluut gezien groter dan $1,96$ zijn.

Voor de *random effects* kan er geen gebruik worden gemaakt van t-toetsen. Hier kan echter het principe van de deviance-toets worden gehanteerd. Om twee modellen te vergelijken, moet van beide modellen de deviance worden uitgerekend. De deviance van een model is gelijk aan -2 maal de natuurlijke logaritme van de gezamenlijke kansdichtheid van de waarnemingen, waarbij de via de maximum-likelihood-methode geschatte parameters worden ingevuld. Om na te gaan of de opname van één of meer variabelen tot een verbetering van een model leidt, kan worden gekeken naar het *verschil in deviance* tussen de modellen. Dit verschil is chi-kwadraat verdeeld.

Verklaarde variantie

De variantie in Y kan worden gesplitst in een deel dat door de vaste effecten wordt verklaard en een deel dat onverklaard blijft. De verklaarde variantie kan worden berekend door de proportionele reductie in de onverklaarde variantie te berekenen (Snijders & Bosker, 1999). Voor niveau één tot en met drie is dit in de multivariate aanpak:

$$(2.13) R_{h1}^2 = 1 - \frac{\text{var}(Y_{hijk} - \sum_g \delta_g X_{ghijk})}{\text{var}(Y_{hijk})}$$

$$(2.14) R_{h2}^2 = 1 - \frac{\text{var}(\bar{Y}_{h.jk} - \sum_g \delta_g \bar{X}_{gh.jk})}{\text{var}(\bar{Y}_{h.jk})}$$

$$(2.15) R_{h3}^2 = 1 - \frac{\text{var}(\bar{Y}_{h..k} - \sum_g \delta_g \bar{X}_{gh..k})}{\text{var}(\bar{Y}_{h..k})}$$

Voor niveau één is de verklaarde variantie de proportionele reductie in de som van Φ_v^2 , Φ_u^2 en Φ_r^2 door de opname van één of meer onafhankelijke variabelen. Voor niveau twee is het de proportionele reductie in de som van Φ_v^2 , Φ_u^2 en Φ_r^2/n . De variantie op het individuele niveau wordt gedeeld door het aantal onderzoekselementen binnen niveau twee. In dit onderzoek verschilt het aantal leerlingen van klas tot klas. Bij dergelijke ongebalanceerde data is het verstandig om een representatieve waarde voor het aantal leerlingen te nemen, bijvoorbeeld $n=20$. De verklaarde variantie voor niveau drie is de proportionele reductie in de som van Φ_v^2 , Φ_u^2/j en Φ_r^2/nj , waarbij j het totaal aantal klassen en nj het totaal aantal leerlingen per school is. Aangezien het aantal klassen van school tot school verschilt, moet opnieuw een representatieve waarde worden gekozen, bijvoorbeeld $j=3$ (Snijders & Bosker, 1999).

2.3 Scholen op rapport

In de vorige paragraaf heb ik uitgelegd wat multiniveau analyse in grote lijn inhoudt. Deze uitleg vormt de technische achtergrond voor dit onderzoek. In de hoofdstukken 3, 4 en 5 gebruik ik multivariate multiniveau analyse om de verschillende invloeden op de prestaties en vorderingen van leerlingen te verklaren. In deze paragraaf zal ik uiteenzetten dat in het Trouw-onderzoek multiniveau analyse had moet worden gebruikt, omdat er sprake is van een hiërarchie van gegevens. Hierbij gebruik ik een hiërarchische structuur met een individueel en een schoolniveau. In de volgende hoofdstukken neem ik ook het klasniveau op. De probleemstelling is: *In hoeverre verschillen de schattingen van de netto prestaties van scholen in een analyse met geaggregeerde gegevens en een multiniveau analyse?*ⁱⁱⁱ

Rapportcijfer

Het rapportcijfer uit het Trouw-onderzoek is een cijfer voor het totale rendement van een school, of eventueel van een schooltype binnen een school (Dronkers, 1997b). Dit rendement is een gewogen optelsom van examencijfers, het slagingspercentage, het uitvalpercentage (twee typen) en het percentage zittenblijvers. Voor vwo-, havo-, mavo- en vbo-afdelingen worden respectievelijk vijftien, dertien, tien en negen kenmerken gehanteerd.

Per schooltype is een rapportcijfer samengesteld, waarbij geldt dat het rapportcijfer hoger is naarmate de eindexamencijfers hoger zijn, het slagingspercentage hoger is en de percentages zittenblijvers en uitvallers lager zijn. Het rapportcijfer is daarbij gecorrigeerd voor het percentage allochtone leerlingen op een school. Indien de rendementcijfers gelijk blijven, neemt voor bijvoorbeeld het vbo het rapportcijfer met één punt toe bij 25 procent meer allochtone leerlingen. Bij 12 procent meer allochtone leerlingen neemt het vwo-rapportcijfer met dezelfde hoeveelheid toe.

Voor de berekening van het rapportcijfer is onder meer factoranalyse en regressieanalyse gebruikt. In het onderstaande kom ik op deze analysemethoden terug. Uit de uiteindelijke analyses komt per school een score. Bij die score worden zes punten opgeteld, zodat de gemiddelde school een zes krijgt. Scholen met een vijf of lager behoren tot het kwart laagst scorende scholen. Scholen met een zeven of meer maken deel uit van het kwart hoogst scorende scholen. Als laatste stap zijn de rapportcijfers afgerond op hele cijfers.^{iv} Wanneer er binnen een bepaald schooltype minder dan vijftig eindexamenkandidaten zijn, heeft die school geen rapportcijfer gekregen.

Om de onderzoeksvraag naar de betrouwbaarheid van het rapportcijfer van het Trouw-onderzoek te beantwoorden volg ik een tweevoudige strategie. In eerste instantie volg ik de door Trouw gehanteerde analysemethode, een combinatie van factor- en regressieanalyse, en ga ik na wat de effecten zijn van verschillende manieren om te controleren voor de kenmerken van de leerlingenpopulatie. In tweede instantie wordt gebruik gemaakt van multiniveau analyse, eerst met de variabelen waarop het rapportcijfer van Trouw is gebaseerd en vervolgens met een alternatieve correctie voor instroomverschillen tussen scholen. De analyse omvat derhalve vier stappen:

1A: methode Trouw - correctiefactor Trouw

1B: methode Trouw - alternatieve correcties

2A: alternatieve methode - correctiefactor Trouw

2B: alternatieve methode - alternatieve correcties

Cohortonderzoek VOCL'89

In deze replicatie maak ik geen gebruik van de onderwijsindicatoren die de inspectie van alle scholen in Nederland verzamelt en waarop het onderzoek van Trouw is gebaseerd. Een belangrijk nadeel van die gegevens is de beperkte beschikbaarheid van informatie over de achtergrondkenmerken van leerlingen. Bij het berekenen van een rapportcijfer kon Trouw namelijk alleen voor het percentage allochtone leerlingen corrigeren. Een recente dataset met informatie over schoolloopbanen in het voortgezet onderwijs waarin wel uitgebreide leerlinggegevens beschikbaar zijn, is het cohortonderzoek VOCL'89. Aangezien ik niet zozeer geïnteresseerd ben in de rapportcijfers van te identificeren scholen, maar in de vraag in hoeverre de in het Trouw-onderzoek gekozen analysemethode en gebruikte correctie effect hebben op de beoordeling van scholen, is de anonimiteit van de scholen in het VOCL-bestand geen bezwaar. De navolgende analyses zijn derhalve op dit cohort gebaseerd. Hierbij gebruik ik gegevens van 19.260 leerlingen van 381 scholen.^v

Voor dit cohortonderzoek hebben het Centraal Bureau voor de Statistiek (CBS), het Gronings Instituut voor Onderwijsonderzoek (GION) en het Nijmeegs Instituut voor Toegepast Sociaal-wetenschappelijk onderzoek (ITS) gegevens verzameld over scholen, docenten, leerlingen en ouders. Deze gegevens worden in eerste instantie gebruikt voor de evaluatie van de basisvorming en het onderwijsvoorrangsbeleid, maar kunnen ook voor secundaire analyses worden gebruikt. Driessen en Van der Werf (1992) bespreken de selectie en de representativiteit van de steekproef, waarbij het van belang is dat de steekproef representatief is voor alle scholen voor voortgezet onderwijs in Nederland.

De 19.260 leerlingen zaten in het schooljaar 1989-'90 in de eerste klas van het voortgezet onderwijs. In het schooljaar 1996-'97 volgde 42 procent van deze leerlingen nog onderwijs. Deze leerlingen volgden voornamelijk middelbaar beroeps-, hoger beroeps- of wetenschappelijk onderwijs. Met een diploma heeft 48 procent het onderwijs verlaten. Eentiende van het cohort is zonder diploma uit het onderwijs vertrokken. In Tabel 2.3 vat ik deze gegevens samen.

Tabel 2.3: Positie leerlingen uit VOCL'89 in het schooljaar 1996-'97 (N=19.260)

Leerling:

- volgt momenteel onderwijs	42.2%	8.136
- heeft onderwijs met diploma verlaten	48.0%	9.254
- heeft onderwijs zonder diploma verlaten	9.7%	1.870

Prestatiematen

Er worden vier maten voor prestaties gehanteerd, te weten of een leerling onvertraagd het examen heeft gehaald, het aantal keren dat een leerling is blijven zitten, of een leerling al dan niet voortijdig de school heeft verlaten en de positie van een leerling op een zogeheten leerjarenladder. Deze maten bereken ik per schooltype, waarbij de richtlijn is dat er minimaal tien leerlingen binnen één schooltype vallen. De scores op deze maten zijn op individueel niveau in Tabel 2.4 weergegeven. In Bijlage 2A zijn deze scores op schoolniveau afgebeeld.

Om te bepalen of een leerling al dan niet *onvertraagd geslaagd* is, ben ik voor vwo, havo, mavo en vbo nagegaan of een leerling op dat niveau examen heeft afgelegd.^{vi} Leerlingen zijn onvertraagd geslaagd, wanneer zij het voorlaatste en laatste leerjaar in één keer hebben doorlopen. Leerlingen zijn vertraagd geslaagd, wanneer zij in het voorlaatste leerjaar zijn blijven zitten of wanneer zij voor hun examen zijn gezakt. Een leerling die na zijn mavo-examen verder gaat op de havo en daar ook examen doet, krijgt voor beide niveaus een score op deze prestatie maat. Het percentage onvertraagd geslaagden is het hoogste bij mavo-leerlingen, 86 procent, en het laagste bij havo-leerlingen, 67 procent.

Tabel 2.4: Prestaties op individueel niveau

VBO	GEM.	SD	MIN.	MAX.	N
onvertraagd geslaagd	0,75	0,43	0	1	6.357
zittenblijvers leerjaar 2	0,10	0,30	0	2	7.191
zittenblijvers leerjaar 3	0,13	0,35	0	3	7.191
uitvallers leerjaar 2	0,03	0,17	0	1	7.191

uitvallers leerjaar 3	0,08	0,26	0	1	7.191
positie leerjarenladder	6,3	1,9	2	11	7.191
MAVO					
onvertraagd geslaagd	0,86	0,34	0	1	6.348
zittenblijvers leerjaar 2	0,11	0,32	0	2	6.131
zittenblijvers leerjaar 3	0,15	0,36	0	2	6.131
uitvallers leerjaar 2	0,02	0,16	0	1	6.131
uitvallers leerjaar 3	0,03	0,18	0	1	6.131
positie leerjarenladder	7,9	1,5	4	11	6.131
HAVO					
onvertraagd geslaagd	0,67	0,47	0	1	3.237
zittenblijvers leerjaar 2	0,08	0,26	0	2	3.264
zittenblijvers leerjaar 3	0,10	0,30	0	2	3.264
zittenblijvers leerjaar 4	0,49	0,59	0	2	3.264
uitvallers leerjaar 2	0,01	0,09	0	1	3.264
uitvallers leerjaar 3	0,02	0,12	0	1	3.264
uitvallers leerjaar 4	0,05	0,22	0	1	3.264
positie leerjarenladder	9,0	1,2	5	12	3.264
Vwo					
onvertraagd geslaagd	0,82	0,39	0	1	2.519
zittenblijvers leerjaar 2	0,03	0,16	0	1	2.674
zittenblijvers leerjaar 3	0,03	0,17	0	2	2.674
zittenblijvers leerjaar 4	0,08	0,27	0	1	2.674
zittenblijvers leerjaar 5	0,24	0,44	0	2	2.674
uitvallers leerjaar 2	0,00	0,06	0	1	2.674
uitvallers leerjaar 3	0,01	0,10	0	1	2.674
uitvallers leerjaar 4	0,01	0,12	0	1	2.674
uitvallers leerjaar 5	0,04	0,19	0	1	2.674
positie leerjarenladder	10,4	1,1	6	12	2.674

Voor elke leerling is het aantal keren *zittenblijven* berekend door na te gaan hoe vaak hij een bepaald leerjaar meer dan één keer heeft doorlopen. In navolging van het Trouw-onderzoek telt bij de bepaling van deze prestatie maat doubleren in het eerste of laatste leerjaar niet mee. Een overgang van bijvoorbeeld mavo-4 naar havo-4 of havo-5 naar vwo-5 beschouw ik wel als zittenblijven. Met een voorbeeld licht ik deze keuze toe. In het eerste leerjaar zitten twee leerlingen in een havo/vwo-brugklas. De ene leerling krijgt een havo-advies. Deze leerling haalt in vijf jaar tijd zijn diploma en stroomt door naar vwo-5. De andere leerling gaat na de brugklas naar de vwo-afdeling en blijft zitten in het vierde leerjaar. Het

jaar daarop gaat zij wel over naar vwo-5. In hun zesde jaar van voortgezet onderwijs belanden beide leerlingen dus weer bij elkaar in de klas. Vandaar dat ik hen een zelfde score op deze prestatie maat toeken, namelijk één keer blijven zitten. Van alle leerlingen uit VOCL'89 heeft 62 procent geen enkele klas overgedaan, 32 procent is een keer blijven zitten, 6 procent heeft tweemaal gedoubleerd en een gering aantal leerlingen (76 van de 19.260) heeft drie maal een leerjaar overgedaan. In Tabel 2.4 valt te zien dat het percentage zittenblijvers hoog is in vwo-5 (24 procent) en vooral in havo-4 (49 procent).

Als derde prestatie maat heb ik berekend of een leerling al dan niet tot de *uitvallers* van het voortgezet onderwijs behoort. In navolging van het Trouwonderzoek beschouw ik leerlingen als uitvallers, wanneer ze de school na het eerste en voor het laatste leerjaar hebben verlaten. Van alle leerlingen uit VOCL'89 heeft 10 procent de school voortijdig verlaten. Tot deze groep behoren ook leerlingen die bijvoorbeeld doorstromen van havo-4 naar mbo-1. Voor de meeste uitvallers geldt dat ze geen diploma voor voortgezet onderwijs hebben, maar een klein deel (152 van de 2.022) valt gediplomeerd uit. Dit gaat bijvoorbeeld op voor leerlingen met een mavo-diploma die in havo-4 uitvallen. Deze leerlingen gelden niet als mavo- maar wel als havo-uitvaller. De uitval is hoog in vbo-3 (8 procent), havo-4 (5 procent) en vwo-5 (4 procent), zie Tabel 2.4.

Als vierde prestatie maat heb ik de *leerjarenladder* (Bosker, Van der Velden en Hofman, 1985) gebruikt, zie Tabel 2.5. Met deze leerjarenladder kan ik nagaan welk niveau leerlingen sinds het schooljaar 1989-'90 hebben bereikt. De leerjarenladder drukt de afstand in leerjaren uit tot de top van het onderwijssysteem. Het maximale niveau voor deze groep blijkt twaalf te zijn, zie Tabel 2.4. Dit niveau kan worden bereikt door leerlingen die zonder doubleren het vwo hebben doorlopen en die inmiddels in het tweede leerjaar van het wetenschappelijk onderwijs zitten.^{vii} Bij de opstelling van de leerjarenladder hebben Bosker, Van der Velden en Hofman rekening gehouden met zittenblijven en doorstromen. Een leerling die blijft zitten verliest een jaar op de overigen. Bij doorstroming van bijvoorbeeld mavo-4 naar havo-4 doet een leerling een jaar langer over het bereiken van dat niveau dan leerlingen die de havo regulier doorlopen. Bij de bepaling van de positie op de leerjarenladder valt jammer genoeg geen onderscheid te maken tussen de verschillende niveaus binnen de mavo en het vbo. Leerlingen die de mavo afsluiten met alle vakken op D-niveau krijgen dezelfde positie op de leerjarenladder als leerlingen die vakken op C-niveau afleggen. In Tabel 2.4 valt te zien dat naarmate het schooltype hoger is, de positie op de leerjarenladder gemiddeld genomen ook hoger is. Vwo-leerlingen scoren gemiddeld 10,4 en vbo-leerlingen gemiddeld 6,3.

Tabel 2.5: *Leerjarenladder en bijbehorende onderwijsposities*
(gebaseerd op Bosker, Van der Velden & Hofman, 1985)

WAARDE	ONDERWIJSPOSITIE				
14	wo-4				
13	wo-3				
12	wo-2	hbo-4			
11	wo-1	hbo-3			
10	vwo-6	hbo-2	mbo-4		
9	vwo-5	hbo-1	mbo-3		
8	vwo-4	havo-5	mbo-2		
7	vwo-3	havo-4	mbo-1	kmbo-2	
6	vwo-2	havo-3	mavo-4	kmbo-1	
5	vwo-1	havo-2	mavo-3	vbo-4	
4		havo-1	mavo-2	vbo-3	ivbo-4
3			mavo-1	vbo-2	ivbo-3
2				vbo-1	ivbo-2
1					ivbo-1

Netto prestaties

Voor de beoordeling van scholen mogen verschillen in leerlingenpopulatie niet uitmaken. Vandaar dat de bruto rapportcijfers moeten worden gecorrigeerd voor enkele kenmerken van de leerlingenpopulatie. In deze replicatie doe ik dat voor de sociaal-economische status van het gezin, de etniciteit en het beginniveau van een leerling. Voor de sociaal-economische status hanteer ik drie kenmerken, te weten de opleiding van de vader, de opleiding van de moeder en het beroep van de kostwinner. Bij de *opleiding van ouders* onderscheid ik zes eindniveaus, te weten: (1) minder dan basisonderwijs; (2) basisonderwijs; (3) lagere vormen van voortgezet onderwijs (lager beroepsonderwijs, mavo of de onderbouw van havo en vwo); (4) hogere vormen van voortgezet onderwijs en middelbaar beroepsonderwijs; (5) universitair en hoger beroepsonderwijs; (6) postdoctorale universitaire opleidingen (inclusief doctoraal/ingenieursexamen oude stijl). Bij het *beroepsniveau van de kostwinner* maak ik gebruik van een indeling in vijf categorieën: (1) mensen zonder betaald werk; (2) arbeiders; (3) zelfstandigen, landbouwers en lagere employés; (4) middelbare employés; (5) hogere employés en vrije beroepen. Om de *eticiteit* van leerlingen te bepalen heb ik gekeken naar het geboorteland van leerlingen en hun ouders. Hierbij onderscheid ik de volgende groepen: (1) Nederlands; (2) Westers en overig (bijvoorbeeld Duits, Amerikaans,

Pools, Chinees); (3) Turks; (4) Marokkaans; (5) Postkoloniale immigranten (Antilliaans, Arubaans, Surinaams en Moluks). Wanneer leerlingen tot een van de laatste drie groepen behoren, noem ik hen allochtoon. Hierbij hanteer ik als regel dat leerlingen tot een van die groepen behoren als zij zelf of één van hun ouders in Turkije, Marokko of een (voormalige) kolonie is geboren. Dit betekent dus dat ik ook kinderen uit gemengde gezinnen (bijvoorbeeld een Turkse man en een Nederlandse vrouw) als allochtoon beschouw.^{viii} Om de *eerdere prestaties* van leerlingen te bepalen, gebruik ik hun scores op de toetsen rekenen, taal en informatieverwerking die zijn afgenomen in het eerste leerjaar. De maximale score bedraagt zestig punten. Driessen en Van der Werf (1992) beschrijven deze toetsen. In Bijlagen 2B en 2C geef ik enkele beschrijvende gegevens over de kenmerken van de leerlingenpopulatie.

De onderlinge samenhang tussen de kenmerken van de leerlingenpopulatie is op schoolniveau hoog, zie Bijlagen 2D en 2E. De correlaties tussen het gemiddelde beroepsniveau van de kostwinner en het opleidingsniveau van beide ouders is 0,90 of hoger. Met de gemiddelde score op de begintoetsen correleren deze kenmerken van de sociaal-economische status ongeveer 0,80. De kenmerken hebben dus een groot deel van de variantie gemeenschappelijk. Er kan hier dan ook van *multicollineariteit* worden gesproken. Het gevolg van deze sterke intercorrelaties is dat de unieke bijdrage van de afzonderlijke kenmerken gering is. Wanneer ik één van de kenmerken van de leerlingenpopulatie in een regressieanalyse opneem en vervolgens een andere toevoeg, dan zal het extra kenmerk nauwelijks extra variantie kunnen binden. De samenhang tussen het percentage allochtone leerlingen en de overige vier achtergrondkenmerken is minder sterk. De correlaties variëren van $-0,34$ tot $-0,53$.

Analyse: stappen 1A en 1B

Als eerste stap repliceer ik het rapportcijfer van Trouw door zowel de methode als de correctie voor het percentage allochtone leerlingen over te nemen (*stap 1A*). Vervolgens pas ik ook een correctie voor andere kenmerken van de leerlingenpopulatie toe (*stap 1B*).

Analoog aan Trouw heb ik met factoranalyse per schooltype een score berekend die staat voor de prestaties van de school. Het doel van de factoranalyse is enerzijds om het latente begrip schoolprestaties te meten door de scores van school op verschillende metingen (percentage onvertraagd geslaagden, percentage zittenblijvers, percentage uitvallers en gemiddelde score op de leerjarenladder) en anderzijds om die verschillende metingen bondig te beschrijven in één maat. Uit de vier factoranalyses blijkt dat de eerste factor steeds het beste het succes van een school weergeeft. Op deze factor laden de percentages zittenblijvers en uitvallers

positief. Het percentage onvertraagd geslaagden en de gemiddelde score op de leerjarenladder hebben een negatief gewicht op de factor.^{ix} In Bijlage 2F geef ik de ladingen van deze kenmerken op deze factor weer. De eerste factor bindt respectievelijk 35,6; 41,5; 27,4 en 27,6 procent aan variantie voor vbo, mavo, havo en vwo.^x De scores op de factor, die ik voortaan als *succesmaat* aanduid, voor de scholen uit VOCL'89 variëren overigens van een drie tot een acht. Het gemiddelde is een zes en de standaarddeviatie is één. Scholen die extreem hoog of laag scoren komen in deze steekproef niet voor.

Tabel 2.6: Regressieanalyse van de succesmaat op het percentage allochtone leerlingen (1A) en de gemiddelde score op de begintoetsen (1B)

	CONSTANTE	% ALLOCHTONEN	BEGINTOETS	R ²
vbo (N=176)	0,124 (0,091)	-0,014 (0,007) *		0,02
	-1,049 (0,358)	-0,007 (0,007)	0,042 (0,013) **	0,07
mavo (N=208)	0,226 (0,073)	-0,046 (0,007) **		0,17
	-0,391 (0,425)	-0,043 (0,008) **	0,017 (0,012)	0,17
havo (N=74)	0,268 (0,108)	-0,052 (0,012) **		0,20
	-1,113 (0,886)	-0,043 (0,013) **	0,033 (0,021)	0,22
vwo (N=75)	0,236 (0,117)	-0,060 (0,016) **		0,15
	-0,968 (1,054)	-0,050 (0,018) **	0,027 (0,023)	0,16

Met regressieanalyse ben ik per schooltype allereerst nagegaan wat de invloed is van het percentage allochtone leerlingen op de succesmaat. Uit de regressieanalyses komt naar voren dat het percentage allochtone leerlingen steeds significant van invloed is op de schoolprestatie, zie Tabel 2.6. Hierbij blijkt dat deze invloed het zwakste is voor het voorbereidend beroepsonderwijs. Voor alle schooltypen geldt dat naarmate het percentage allochtone leerlingen toeneemt de succesmaat gemiddeld genomen lager is. Het is dan ook eerlijk om de score te corrigeren van scholen met veel leerlingen met een andere etnische afkomst. De verklaarde variantie varieert van 2 procent voor vbo tot 17 procent voor mavo.

In de volgende analyse kijk ik niet alleen naar de invloed van het percentage allochtone leerlingen, maar ook naar die van de eerdere prestaties. De toevoeging van de eerdere prestaties sorteert effect voor het vbo. In de scores op de succesmaat voor vbo-scholen treden veranderingen op. Deze veranderingen zijn maximaal zeventiende punt, wat afgerond neerkomt op mogelijke verschillen van een punt per school. De verklaarde variantie neemt hierbij toe tot 7 procent. Voor de andere schooltypen volstaat het om alleen rekening te houden met het

percentage allochtone leerlingen. De eerdere prestaties van leerlingen hebben geen significante invloed op de schoolprestatie.

Toevoeging van een van de kenmerken van de sociaal-economische status, de opleiding van de vader, leidt alleen voor de havo tot een verbetering van de verklaring van de schoolprestatie, zie Tabel 2.7. De verklaarde variantie neemt toe tot 36 procent en de regressievergelijking verandert behoorlijk. Opmerkelijk genoeg is het effect van de gemiddelde opleiding van de vader op de succesmaat negatief. Gegeven de overige kenmerken, scoren scholen slechter op de succesmaat naarmate ze meer leerlingen hebben met een goed opgeleide vader. Dit verband wordt alleen gevonden, wanneer er wordt gecontroleerd voor de twee andere variabelen.

Tabel 2.7: Regressieanalyse van de succesmaat voor de havo (N=74) op het percentage allochtone leerlingen (stap 1A), de gemiddelde score op de begintoetsen en de gemiddelde opleiding van de vader (stap 1B)

	MODEL 1	MODEL 2
Percentage allochtone leerlingen	-0,052 (0,012) **	-0,070 (0,014) **
Gemiddelde score begintoetsen		0,074 (0,021) **
Gemiddelde opleiding vader		-1,204 (0,290) **
Constante	0,268 (0,108)	1,941 (1,087)
Aangepaste R^2	0,17	0,36

Uit deze uitkomsten blijkt dat de effecten van een meer uitgebreide correctie voor populatieverschillen tussen scholen *voorlopig* meevallen. Door de extra correcties veranderen de scores op de succesmaat voor vbo- en havo-scholen maximaal met een punt, zie de vergelijking tussen de stappen 1A en 1B in Tabel 2.9. De conclusie is dat op schoolniveau de samenhang tussen het percentage allochtone leerlingen, de gemiddelde sociaal-economische status en het gemiddelde prestatieniveau bij intrede zo hoog is dat extra correcties de rapportcijfers nauwelijks veranderen. Kleine verschillen tussen de twee stappen zijn er alleen voor vbo- en havo-scholen.

Prestatie en sturing

Wat kan een school doen om goede examenresultaten te behalen? Er zijn drie mogelijkheden: goed onderwijs geven, goede leerlingen werven of door zitten blijven, afstroom en uitstroom slechte leerlingen wegwerken. Zo beschouwd hebben de resultaten in de klassen vóór de examenklas een dubbel karakter: het

zijn tussenliggende prestatie-maten, maar ook sturingsvariabelen voor de scholen. Een factoranalyse van examenresultaten samen met maten voor uitstroom en zittenblijven geeft de empirische samenhang aan tussen al deze gebruikte variabelen. Maar de mate waarin scholen een streng beleid voor de overgang voeren is op zich niet relevant voor een prestatie-maat. Een scholier die zeven jaar over het vwo doet heeft deze verblijfsduur op school ongeacht of hij een keer is blijven zitten of een keer voor zijn eindexamen is gezakt. In een factoranalyse van examenresultaten met maten voor uitstroom en zittenblijven wordt succes dus gecontamineerd met het door de school gehanteerde sturingsmechanisme. Dit wordt ondersteund door de lage hoeveelheid door de eerste factor verklaarde variantie (zie noot 10). De conclusie moet dan ook zijn dat de in de factoranalyse samengenomen variabelen niet één dimensie vormen. Ze bevatten elementen van zowel echte prestatie als sturingsbeleid.

Het verdient dan ook voorkeur om een succesmaat te gebruiken die niet ten dele een sturingsvariabele is. Voorbeelden van succesmaten die niet deze nadelen hebben zijn het aantal leerlingen dat onvertraagd het eindexamen haalt in het aan het eind van de brugperiode gekozen schooltype en de leerjarenladder. Deze maten behoren dan niet gecombineerd te worden met maten voor uitstroom en zittenblijven. Tegen deze achtergrond heb ik voor de tweede replicatie van het Trouw-rapportcijfer de scores op een leerjarenladder als indicator van de prestaties van scholen gekozen.

Belang van multiniveau analyse

De prestatie-maten van het Trouw-onderzoek zijn gebaseerd op aggregaties van leerlingresultaten. In een onderzoek naar prestaties van leerlingen moet echter rekening worden gehouden met de geneste structuur van de gegevens, in dit geval leerlingen binnen scholen. Aan het geen rekening houden met de hiërarchische structuur van de gegevens is een aantal nadelen verbonden. Het onderscheid tussen de analyse van geaggregeerde data en multiniveau analyse is hier op drie punten relevant, te weten de *nauwkeurigheid* van de resultaten, het onderscheid tussen *binnenschoolse en tussenschoolse regressie-effecten* en het berekenen van *posterior means* (zie ook: Dijkstra, Veenstra, Snijders & Peschar, 1999).^{xi}

Voor variabelen zoals de intelligentie en de culturele en etnische achtergrond van leerlingen kan het effect op individueel niveau, het *binnenschoolse effect*, worden onderscheiden van het effect op schoolniveau, het *tussenschoolse effect*. Bij de analyse van schoolgemiddelden zijn alleen de tussenschoolse effecten te zien. Nu is het mogelijk dat scholen met een 'moeilijke' leerlingenpopulatie, bijvoorbeeld met gemiddeld laag opgeleide ouders of een hoog percentage allochtonen, dat compenseren door goed beleid, zodat het tussenschoolse effect

van hiermee overeenkomende variabelen kleiner is dan het binnenschoolse effect. Wanneer bijvoorbeeld het effect van de opleiding van de ouders op individueel niveau positief is en wanneer scholen met gemiddeld laag opgeleide ouders perfect voor de invloed van dat kenmerk compenseren, dan is bij de analyse van schoolgemiddelden, het effect van opleiding van de ouders nihil. Een kind van laag opgeleide ouders kan dan echter beter op een school zitten waar de gemiddelde opleiding van de ouders laag is! Met andere woorden bij een regressieanalyse op schoolniveau en ervan uitgaande dat sommige scholen goed compenseren voor een 'moeilijke' samenstelling van de leerlingenpopulatie, worden dergelijke scholen onvoldoende voor hun prestatie beloond.

Voor ouders is het dus belangrijk om binnen- en tussenschoolse effecten te kunnen onderscheiden. Dit laat ik in het onderstaande zien voor één variabele X , maar het principe is toepasbaar op een willekeurig aantal variabelen. Wanneer de uitkomstvariabele voor leerling i in school j wordt aangegeven met Y_{ij} , kan ik het effect op Y_{ij} van een variabele X met individuele waarden x_{ij} en schoolgemiddelden $\bar{x}_{.j}$ uitdrukken door de formule:

$$(2.16) Y_{ij} = \beta_{00} + \beta_B (x_{ij} - \bar{x}_{.j}) + \beta_T \bar{x}_{.j} + U_{0j} + R_{ij}$$

Hierin is β_B de binnenschoolse en β_T de tussenschoolse regressiecoëfficiënt. De *onverklaarde* bijdrage van de school, die gecorrigeerd is voor de binnen- en tussenschoolse effecten van X , is weergegeven met U_{0j} . De *onverklaarde* eigen bijdrage van de leerling, voor hetzelfde gecorrigeerd, is R_{ij} . Deze formule komt overeen met de situatie waarin het schoolgemiddelde en de binnenschoolse deviatiescore als variabelen in de analyse worden opgenomen. Het is mogelijk de formule te herschrijven door de oorspronkelijke leerlingvariabele x_{ij} en het schoolgemiddelde $\bar{x}_{.j}$ als variabelen op te nemen:

$$(2.17) Y_{ij} = \beta_{00} + \beta_B x_{ij} + (\beta_T - \beta_B) \bar{x}_{.j} + U_{0j} + R_{ij}$$

Bij een regressie binnen een enkele school geldt op populatieniveau de binnenschoolse regressiecoëfficiënt. Indien de gegevens (zowel X als Y) worden geaggregeerd tot schoolgemiddelden en vervolgens op schoolniveau een regressie wordt uitgevoerd dan geldt de tussenschoolse regressiecoëfficiënt. In dat geval gaat het dus om de grootte van het effect van het schoolgemiddelde X op het schoolgemiddelde Y .

De betekenis van deze coëfficiënten kan als volgt worden toegelicht. Stel bijvoorbeeld dat X de etnische afkomst van een leerling aangeeft: $x_{ij} = 1$ voor allochtone en 0 voor autochtone leerlingen. Dan fungeert de proportie allochtone

leerlingen in school j als het schoolgemiddelde in formule 2.17. Binnen een gegeven school, met een gegeven proportie allochtone leerlingen, is de bijdrage van de etnische achtergrond van een leerling gemiddeld gelijk aan $\bar{\beta}_B$, waarbij de verwachting is dat het effect van etniciteit negatief is (Driessen, 1995; De Jong, 1989). De proportie allochtone leerlingen op een school kan ook een eigen effect hebben op de prestaties van allochtone en autochtone leerlingen. Dergelijke effecten kunnen bijvoorbeeld het gevolg zijn van de invloed van de samenstelling van de leerlingenpopulatie op het groepsproces binnen een klas en de school, bijvoorbeeld door een veranderde tijdsverdeling van docenten en door vergelijkings- en beïnvloedingsprocessen tussen leerlingen. Ook kunnen effecten van het percentage allochtone leerlingen het gevolg zijn van de inspanningen die de school zich getroost om de problemen ontstaan door een 'moeilijke' leerlingenpopulatie het hoofd te bieden. Indien op schoolniveau dit soort effecten zich inderdaad voordoen dan zal de binnenschoolse regressiecoëfficiënt verschillen van de tussenschoolse regressiecoëfficiënt, dus $(\beta_T - \beta_B) \neq 0$. Hierbij kan het overigens zo zijn dat deze effecten elkaar opheffen. De 'moeilijke' samenstelling van de leerlingenpopulatie zal een negatief effect hebben op $(\beta_T - \beta_B)$, het beleid van de scholen zal, zo is de bedoeling, een positief effect hebben op deze coëfficiënt.

Wat een rapportcijfer voor scholen voor voortgezet onderwijs betreft, zijn er verschillen tussen de standpunten van *ouders* en de *overheid*. Voor ouders is de eigen etnische achtergrond een gegeven. Indien ouders de verwachte schoolprestaties van hun eigen kind willen maximaliseren, zullen zij de scholen willen ordenen op grond van de waarde van:

$$(2.18) (\beta_T - \beta_B) \bar{x}_{.j} + U_{0j}$$

Dit is de schoolbijdrage waarbij wordt gecorrigeerd voor de samenstelling van de leerlingenpopulatie door de binnenschoolse en niet door de tussenschoolse regressiecoëfficiënt. Voor ouders wordt dus geen correctie uitgevoerd voor het effect van de leerlingenpopulatie op schoolniveau. Dit is overeenkomstig aan de wensen van ouders. Als hun kind veel leert, omdat de samenstelling van de school gunstig is, dan is dat voor hen mooi meegenomen.

De overheid is niet zo zeer in de prestaties van individuele leerlingen, maar van een groep leerlingen, bijvoorbeeld een school, geïnteresseerd. Voor de overheid is de leerlingenpopulatie van een school dan een gegeven. Indien de overheid de gemiddelde prestaties van de leerlingen op een school wil belonen, zal zij de scholen willen ordenen volgens de waarde van U_{0j} . Dit is immers de

eigen bijdrage van een school, die niet kan worden toegeschreven aan de samenstelling van variabele X . Op deze manier wordt rekening gehouden met zowel de moeilijkheden op schoolniveau door variabele X als de op schoolniveau gemiddelde gevolgen van het beleid waarmee scholen deze moeilijkheden het hoofd bieden. Dat kan ook niet anders, want zonder gegevens over schoolbeleid te gebruiken zijn deze twee invloeden niet van elkaar te isoleren. De scholen met een 'moeilijke' leerlingenpopulatie worden zo dus als groep niet beloond voor hun inspanningen om hier goed mee om te gaan. Alleen wanneer een school het beter of slechter doet dan andere scholen met een zelfde soort leerlingenpopulatie, dan wordt een school beloond of bestraft. Het zou voor de overheid *eerlijker* zijn om wel voor succes op groepsniveau te belonen, maar dit zou hoge eisen stellen aan de te gebruiken data en aan inzichten in de aspecten van het beleid die er toe doen.

Met wat voor analysemethoden kunnen de hier genoemde getallen om de scholen te ordenen worden geproduceerd? Wat de bovenstaande formule vanuit het gezichtspunt van de ouders betreft, is een multiniveau analyse noodzakelijk om afzonderlijk de binnenschoolse regressiecoëfficiënt $\bar{\epsilon}_B$ en de tussenschoolse regressiecoëfficiënt $\bar{\epsilon}_T$ te schatten. Wat het schatten van alleen de voor de overheid relevante waarde U_{0j} betreft zijn er twee mogelijkheden. De gemakkelijkste methode is een regressie van schoolgemiddelden, waaruit U_{0j} als residu kan worden berekend. Dit is analoog aan een onderdeel van de door Trouw uitgevoerde analyse. Een alternatieve methode is weer het schatten van dit residu uit een multiniveau analyse. Dit heeft als voordeel dat met het aantal in de analyse gebruikte leerlingen rekening wordt gehouden, waarbij voor scholen met weinig data over leerlingen een meer voorzichtige schatting wordt gebruikt dan voor scholen met veel data. Bij scholen met een klein aantal leerlingen speelt het toeval namelijk een grotere rol bij de gemiddelde prestaties dan bij scholen met een groot aantal leerlingen, omdat toevalsinvloeden bij de eerste scholen in mindere mate worden weggemiddeld. Dit betekent dat scholen met een relatief klein aantal leerlingen eerder toevallig hoge of lage rapportcijfers zullen krijgen dan scholen met veel leerlingen.

In een multiniveau analyse wordt pas duidelijk hoe groot of klein de nauwkeurigheid van de schoolcijfers is. Wanneer een school bijvoorbeeld een zes als rapportcijfer heeft, dan kan worden bepaald of dat cijfer in het hypothetische geval van een volgende meting, en dus met een andere leerlingengroep, wel eens twee punten hoger of lager kan uitkomen of dat het cijfer waarschijnlijk tussen een 5,5 en een 6,5 zal schommelen. Voor een Engelse hitlijst van scholen laten Goldstein en Spiegelhalter (1996) zien dat de nauwkeurigheid in sommige gevallen tamelijk gering is.^{xiii}

Voor zowel ouders als de overheid verdient multiniveau analyse de voorkeur boven een simpele regressieanalyse op geaggregeerde data. Door multiniveau analyse te gebruiken kan voor ouders het noodzakelijke onderscheid tussen binnen- en tussenschoolse regressie-effecten worden gemaakt. Voor de overheid is dit onderscheid van minder gewicht. Een ander voordeel dat wel belangrijk is voor de overheid, betreft de grotere nauwkeurigheid van de standaardfouten en betrouwbaarheidsintervallen in een multiniveau analyse. Verder geldt dat ook scholen met een gering aantal leerlingen een rapportcijfer kunnen krijgen in een multiniveau analyse, omdat dan een meer voorzichtige schatting kan worden gegeven.

Analyse: stappen 2A en 2B

Multiniveau analyse verdient dus de voorkeur boven een op schoolniveau geaggregeerde regressieanalyse. Om na te gaan in hoeverre gebruik van deze in statistisch opzicht meer geëigende methode effect heeft op de succesmaat van scholen, zijn de stappen 2A en 2B gebaseerd op een multiniveau analyse, uitgevoerd met het programma HLM (Bryk, Raudenbush & Congdon, 1993). De score op de leerjarenladder is als succesmaat gekozen. De analyse wordt opnieuw per schooltype uitgevoerd. Om scholen te rangordenen is gebruik gemaakt van de *posterior means*, dat wil zeggen de residuen die voor het intercept op schoolniveau zijn berekend. Met deze empirisch Bayesiaanse schatting kan worden nagegaan welke scholen onverwacht hoog of laag scores op de succesmaat, gegeven de waarden op de onafhankelijke variabelen. Voor ouders is dit de maat die hen informatie geeft over welke scholen voor de gemiddelde leerling succesvol zijn. In stap 2A en 2B gebruik ik deze 'posterior means', waarbij deze schattingen zijn gestandaardiseerd.

In stap 2A hanteer ik dezelfde correctie als in het Trouw-onderzoek. Analooq aan de werkwijze van Trouw corrigeer ik alleen voor het percentage allochtone leerlingen van de school. Het verschil is dus enkel de analysemethode. Ik gebruik namelijk multiniveau analyse en als succesmaat de leerjarenladder. De correctie voor het percentage allochtone leerlingen voer ik uit op schoolniveau, terwijl ik op individueel niveau een leeg model schat.

Uit de analyses komt naar voren dat het percentage allochtone leerlingen in drie van de vier gevallen van invloed is op de score op de leerjarenladder, zie Tabel 2.8. Voor havo-leerlingen blijkt er geen invloed te zijn van deze correctiefactor. Voor de overige schooltypen geldt dat naarmate het percentage allochtone leerlingen toeneemt de score op de leerjarenladder gemiddeld genomen lager is. De verklaarde variantie op schoolniveau is 3 procent voor vbo en mavo, en 10 procent voor vwo.

Tabel 2.8: Multiniveau analyse op de leerjarenladder-score voor afzonderlijke schooltypen

	VBO	MAVO	HAVO	VWO
STAP 2A (met correctie Trouw)				
Constante	-0,005 (0,016)	-0,015 (0,021)	-0,057 (0,040)	-0,007 (0,050)
<i>Schoolniveau</i>				
Percentage allochtone leerlingen	-0,003 (0,001) *	-0,007 (0,002) *	0,001 (0,009)	-0,052 (0,022) *
TUSSENSTAP (uitgebreid model)				
Constante	0,205 (0,064)	-0,417 (0,042)	0,076 (0,044)	-0,007 (0,042)
<i>Individueel niveau</i>				
Etniciteit (1=allochtoon)	-0,119 (0,044) *	-0,023 (0,039)	-0,189 (0,147)	-0,361 (0,208)
Opleidingsniveau vader	0,002 (0,009)	0,023 (0,009) *	0,092 (0,015) *	0,062 (0,022) *
Prestaties bij instroom	0,009 (0,001) *	0,026 (0,002) *	0,039 (0,003) *	0,041 (0,005) *
<i>Schoolniveau</i>				
Percentage allochtone leerlingen	-0,002 (0,002)	-0,002 (0,002)	0,004 (0,007)	-0,042 (0,018) *
Gemiddelde opleidingsniveau vader	0,033 (0,059)	0,089 (0,056)	-0,148 (0,101)	0,104 (0,108)
Gemiddelde prestaties bij instroom	-0,002 (0,004)	-0,001 (0,004)	-0,013 (0,008)	-0,014 (0,012)
STAP 2B (spaarzaam model)				
Constante	0,194 (0,024)	-0,347 (0,031)	0,102 (0,042)	-0,012 (0,039)
<i>Individueel niveau</i>				
Etniciteit (1=allochtoon) *	-0,130 (0,042) *			
Opleidingsniveau vader		0,026 (0,009) *	0,083 (0,015) *	0,063 (0,021) *
Prestaties bij instroom	0,009 (0,001) *	0,026 (0,002) *	0,038 (0,003) *	0,041 (0,005) *
<i>Schoolniveau</i>				
Percentage allochtone leerlingen				-0,032 (0,013) *

In de volgende analyse voeg ik de variabelen toe waarmee ook andere aspecten van verschillen in leerlingenpopulatie in de succesmaat kunnen worden verdisconteerd. Om een maat te verkrijgen die goed vergelijkbaar is met die uit stap 1B beperk ik me tot de daar meegenomen variabelen. De toevoeging van de etniciteit, het opleidingsniveau van de vader, de prestaties bij instroom (kenmerken op *individueel niveau*), het gemiddelde opleidingsniveau van de vaders, de gemiddelde prestaties bij instroom (kenmerken op *schoolniveau*), blijkt voor alle schooltypen effect te sorteren voor de scores op de leerjarenladder. Het volstaat niet om alleen rekening te houden met het percentage allochtone leerlingen. Op de succesmaat hebben de prestaties bij instroom in alle gevallen een effect. Het opleidingsniveau van de vader blijkt niet van belang voor vbo-leerlingen, maar wel voor leerlingen in het algemeen vormend onderwijs. Voor vbo-leerlingen blijkt het juist van gewicht om te corrigeren voor de etniciteit van leerlingen. De drie kenmerken op individueel niveau hebben in één of meer van de analyses een effect. Van de kenmerken op schoolniveau heeft alleen het percentage allochtone leerlingen een effect en dan alleen voor vwo-leerlingen, zie Tabel 2.8. Net als in de regressieanalyse op geaggregeerde gegevens heeft het gemiddelde opleidingsniveau van de vader een negatief effect op de succesmaat. De regressiecoëfficiënt van het opleidingsniveau op schoolniveau blijkt $-0,148$ te zijn. Op individueel niveau is het effect van de opleiding van de vader $0,092$. Wanneer echter de standaardfouten in beschouwing worden genomen, dan blijkt dat het effect op individueel niveau wel en op schoolniveau niet significant is. Dit wijst erop dat het negatieve effect van het gemiddelde opleidingsniveau van de vader in Tabel 2.7 het gevolg kan zijn van een samenloop van binnen- en tussenschoolse regressie-effecten.

Als tussenstap is voor alle schooltypen een uitgebreid model opgesteld met alle mogelijke effecten op individueel en schoolniveau. Uit deze uitkomsten blijkt dat vooral met kenmerken op individueel niveau en niet zo zeer op schoolniveau rekening moet worden gehouden. Vervolgens zijn in stap 2B alleen de significante effecten meegenomen. De verschillen tussen de in stap 2A en 2B berekende scores op de succesmaat geef ik weer in Tabel 2.9. Het spaarzame model in stap 2B leidt tot aanzienlijke verschillen in de scores van scholen. Voor veel scholen treden verschillen van één of meer punten op. De verschillen zijn het grootst voor mavo- en vwo-scholen. De verklaarde variantie in het spaarzame model varieert op leerlingniveau van 5 procent voor vbo-leerlingen tot 13 procent voor vwo-leerlingen. De verschillen in verklaarde variantie op schoolniveau zijn groot. Op schoolniveau wordt voor havo-leerlingen niets verklaard, maar voor vwo-leerlingen blijkt de helft van de variantie verklaard te worden.

Tabel 2.9: Percentage scholen met verschillende scores op de vier succesmaten

VERSCHIL IN:	VARIABLEN		METHODE	
	1A-1B	2A-2B	1A-2A	1B-2B
vbo				
geen	92	85	53	50
één cijfer	8	15	41	46
twee cijfers	0	0	5	3
meer dan twee cijfers	0	0	1	1
mavo				
geen	100	51	45	34
één cijfers	0	44	44	47
twee cijfers	0	5	6	13
meer dan twee cijfers	0	0	5	6
havo				
geen	99	78	27	22
één cijfer	1	22	52	44
twee cijfer	0	0	21	32
meer dan twee cijfers	0	0	0	2
vwo				
geen	100	55	43	38
één cijfer	0	39	45	36
twee cijfers	0	7	10	26
meer dan twee cijfers	0	0	2	0

Legenda:

1A: methode Trouw - correctiefactor Trouw

1B: methode Trouw - alternatieve correcties

2A: alternatieve methode - correctiefactor Trouw

2B: alternatieve methode - alternatieve correcties

Vergelijking van de vier rapportcijfers

Behalve de verschillen tussen de stappen 2A en 2B, geef ik in Tabel 2.9 de verschillen tussen nog drie stappen weer. De in de tabel weergegeven rapportcijfers zijn de naar z-scores getransformeerde uitkomsten van de analyses uit de stappen 1A tot 2B. Om reden van presentatie heb ik de uiteindelijke verschillen afgerond. Voor de correlaties tussen de scores op de vier succesmaten verwijs ik naar Bijlage 2G. Uit die correlatiematrices valt opnieuw op te maken dat de rapportcijfers uit de stappen 1A en 1B vrijwel gelijk aan elkaar zijn. De correlaties zijn voor alle schooltypen vrijwel gelijk aan 1. De rapportcijfers uit de stappen 2A en 2B hangen sterk met elkaar samen. De correlaties variëren tussen 0,70 en 0,94. Deze samenhang is echter niet perfect. Het maakt voor de rapportcijfers dus wel uit of er met meer of minder factoren wordt rekening gehouden.

De variabelenvergelijking maakt duidelijk dat effecten van verschillen in correctie voor een meer of minder uitgebreide correctie zich vooral voordoen bij de op multiniveau analyse gebaseerde succesmaten (2A-2B). Dergelijke verschillen treden, zoals eerder geconcludeerd, nauwelijks op bij een op schoolniveau geaggregeerde regressieanalyse (1A-1B). Vergelijking van beide methoden (1A-2A, 1B-2B) laat zien dat gebruik van een multiniveau analyse met de leerjarenladder als succesmaat tot een aanzienlijke verschuiving in de scores van scholen leidt.

In Tabel 2.10 vergelijk ik per schooltype de rapportcijfers van enerzijds de op schoolniveau geaggregeerde regressieanalyse met beperkte correctie (stap 1A) en anderzijds de multiniveau analyse met uitgebreide correctie (stap 2B). Voor het vwo geldt bijvoorbeeld dat 39 procent van de scholen in beide stappen hetzelfde cijfer heeft. Deze scholen zijn weergegeven op de diagonaal in Tabel 2.10. De overige vwo-scholen hebben verschillende cijfers in beide stappen, waarbij 19 procent van de scholen een onvoldoende behaalt in de stap die analoog is aan het Trouw-onderzoek en een voldoende scoort in de door mij bepleite stap. Omgekeerd scoort 12 procent van de vwo-scholen in stap 1A voldoende en in stap 2B onvoldoende.

Tabel 2.10: Samenhang in scores tussen de succesmaten uit stap 1A en stap 2B

VBO		CIJFER STAP 2B					MAVO		CIJFER STAP 2B				
		4	5	6	7	8			4	5	6	7	8
	4	3	3	2	0	0		4	0	3	3	0	2
CIJFER	5	1	11	10	2	0	CIJFER	5	1	7	9	3	1
STAP 1A	6	2	3	23	7	0	STAP 1A	6	0	8	20	6	3
	7	0	1	12	13	1		7	0	5	18	6	1
	8	0	0	1	5	0		8	1	1	1	1	1
HAVO		CIJFER STAP 2B					VWO		CIJFER STAP 2B				
		4	5	6	7	8			4	5	6	7	8
	4	0	2	2	2	0		4	0	5	2	2	0
CIJFER	5	0	7	8	13	0	CIJFER	5	2	12	10	5	0
STAP 1A	6	3	21	2	8	0	STAP 1A	6	2	5	17	2	0
	7	0	5	7	15	2		7	0	5	12	10	5
	8	0	0	5	0	0		8	0	0	2	0	0

Legenda:

1A: methode Trouw - correctiefactor Trouw

2B: alternatieve methode - alternatieve correcties

De keuze van de succesmaat en de analysemethode heeft duidelijke gevolgen voor de scores van scholen. Scholen die met een op schoolniveau geaggregeerde regressieanalyse, de methode van het Trouw-onderzoek, bijvoorbeeld een zes scoren, kunnen met een multiniveau analyse zo maar één of twee punten meer behalen. Deze verschillen zijn het gevolg van een meer adequate controle voor de samenstelling van de leerlingpopulatie. Een voorzichtige interpretatie van rapportcijfers gebaseerd op een berekeningswijze zoals in het dagblad *Trouw* is dan ook geboden.

2.4 Besluit

Het blijkt dat de keuze van de analysemethode grote gevolgen heeft voor de scores van scholen. Scholen die met een op schoolniveau geaggregeerde regressie-analyse, de methode van het Trouw-onderzoek, bijvoorbeeld een zes halen, kunnen met een multiniveau analyse zo maar een punt lager of hoger scoren.

In de verdere ontwikkeling van methoden om de kwaliteit van scholen inzichtelijk te maken, moet worden gepleit voor een multiniveau analyse, omdat alleen met een dergelijke methode adequaat kan worden gecontroleerd voor de samenstelling van de leerlingenpopulatie. Voor zowel ouders als de overheid blijkt multiniveau analyse de voorkeur te verdienen boven een simpele regressieanalyse op geaggregeerde data. Door multiniveau analyse te gebruiken kan voor ouders het noodzakelijke onderscheid tussen binnen- en tussenschoolse regressie-effecten worden gemaakt. Voor de overheid is dit onderscheid van minder gewicht. Een ander voordeel dat wel belangrijk is voor de overheid, betreft de grotere nauwkeurigheid van de resultaten in een multiniveau analyse. Verder geldt dat ook scholen met een gering aantal leerlingen een rapportcijfer kunnen krijgen in een multiniveau analyse, omdat dan een meer voorzichtige schatting kan worden gegeven.

Met multiniveau analyse scholen zijn gemakkelijk te rangordenen door gebruik te maken van zogeheten posterior means. Hierbij verdient het de voorkeur om een leerjarenladder of een soortgelijke maat als indicator van de prestaties van scholen te gebruiken. Een dergelijke maat geeft alleen het succes van scholen weer. In het Trouw-onderzoek zijn variabelen samengenomen die enerzijds het succes en anderzijds de sturingsmechanismen van scholen weergeven. Deze variabelen vormen niet één dimensie. Door de lage hoeveelheid verklaarde variantie van de eerste factor wordt deze conclusie ondersteund.

Wanneer wordt gekozen voor een multiniveau analyse dan blijkt dat de correctie voor alleen het percentage allochtone leerlingen mager is. Het rapportcijfer dient te worden verfijnd met meer gegevens over de kwaliteit van de instroom van scholen. Deze uitgebreide controle heeft eigenlijk vooral zin in een multiniveau analyse.

Bosker e.a. (1998) wijzen ook op het belang van een analyse op individueel en schoolniveau. Ook verwerpt Bosker de factoranalyse, omdat de computer dan bepaalt *in welke mate* factoren, zoals het percentage geslaagden, zittenblijvers, uitvallers en bèta-vakken bijdragen aan de beoordeling van een school (Bosker in Jager, 1998, p. 14). "Dat is alsof je bij een WK schaatsen alle tijden op vier afstanden in de computer stopt. Vervolgens zeg je tegen de deelnemers: 'Wacht even, de computer bepaalt nu hoe zwaar elke afstand voor de uitslag meetelt, en dan weten we zo wie wereldkampioen is.' Als daar dan de 5000 meter als

belangrijkste uitrolt is dat vreemd, want bij het vorige kampioenschap was dat nog de 1500 meter. Aan zo'n koffiemolentechniek kun je een dergelijke beslissing niet toevertrouwen."

In een reactie op Veenstra e.a. (1998) vindt Dronkers (1998a) echter dat de voorkeur voor multiniveau analyse om twee redenen onjuist is. Ten eerste vindt hij dat met de keuze voor die analysemethode te hoog wordt ingezet wat de kwaliteit van data betreft. Aangezien er voorlopig geen gegevens zijn over een voldoende aantal leerlingen van alle scholen voor voortgezet onderwijs, is multiniveau analyse volgens Dronkers (1998a, p. 149) niet toepasbaar. "Als wetenschappers dan toch staan op multilevel analyses, sluiten zij zichzelf uit van het verdere maatschappelijke debat over de meting van 'toegevoegde waarde' van scholen. De wetenschappelijke opvattingen over de favoriete methode hebben dan de bovenhand gekregen, waardoor de maatschappelijke dienstverlening door de wetenschap gehinderd zou worden. In dat geval is het mogelijk betere (een superieur geachte multilevel methode) de vijand van het goede (een min of meer verantwoorde meting van kwaliteit van scholen)." Als aanzet tot een rangordening van scholen mag de maatschappelijke dienstverlening van Dronkers geslaagd worden genoemd. Wetenschappelijk blijft het de vraag of deze manier van rangordenen goed genoeg is. In een vergelijking tussen verschillende methoden blijkt dat in de rapportcijfers van scholen geregeld een verandering van één punt optreedt en dat ook veranderingen van twee punten voorkomen. De resultaten zijn zodanig onnauwkeurig dat het nemen van beslissingen over individuele scholen leidt tot onterechte benadeling (of bevoordeling) van sommige scholen. Bij het opstellen van rangordeningen dient van deze onnauwkeurigheid melding te worden gemaakt.

Ten tweede stelt Dronkers (1998a) dat met multiniveau analyse een andere vraag wordt beantwoord. Inderdaad moet er een verschil worden gemaakt tussen het standpunt van de overheid en dat van ouders. Vanuit het standpunt van de overheid zou het eventueel volstaan om de prestaties van scholen in kaart te brengen met een regressie van schoolgemiddelden, omdat voor de overheid het onderscheid tussen binnen- en tussenschoolse regressie-effecten minder belangrijk is. Om de gegevens van uit het gezichtspunt van ouders te bekijken is het echter noodzakelijk om een multiniveau analyse uit te voeren. Indien ouders de verwachte schoolprestaties van hun eigen kind willen maximaliseren, zullen zij voor de schoolkeuze de scholen willen ordenen op grond van een score van scholen waarbij is gecorrigeerd voor de samenstelling van de leerlingenpopulatie door de binnenschoolse en niet door de tussenschoolse regressiecoëfficiënten. Met het kenmerk etniciteit is dit onderscheid in dit hoofdstuk toegelicht. Voor ouders is het alleen nodig om te controleren voor etniciteit als leerlingkenmerk. In

een op schoolniveau geaggregeerde analyse wordt niet gecontroleerd voor etniciteit op individueel niveau, maar voor het percentage allochtone leerlingen. Dat percentage kan echter ook een eigen effect hebben. Deze effecten kunnen bijvoorbeeld komen door de invloed van de samenstelling van de leerlingenpopulatie op de tijdsverdeling van docenten of de groepsprocessen binnen een school. Ook kunnen effecten het gevolg zijn van de inspanningen die een school verricht om te compenseren voor de 'moeilijke leerlingenpopulatie'. Indien dit soort effecten zich voordoen, dan verschillen de binnen- en tussenschoolse regressie coëfficiënten en worden sommige scholen ten onrechte te veel of te weinig beloond voor hun presteren. Het Trouw-onderzoek bekijkt de prestaties dus niet vanuit het gezichtspunt van ouders, hoewel de redactie dat wel duidelijk voor ogen had.

Eindnoten

ⁱ Voor een multiniveau model worden ook andere termen gebruikt, zoals een random-coëfficiëntmodel, een hiërarchisch lineair model of een variantiecomponentmodel.

ⁱⁱ Als leerlingen in zeven jaar tijd een vwo-diploma behalen, dan doet hun school het volgens het Trouw-onderzoek slechter dan wanneer leerlingen in vier jaar voor hun mavo-examen slagen. Prick (1997) zegt hierover: "De school die streng selecteert, die vindt dat alleen een smalle top in staat is een vwo-diploma te halen, krijgt hoge cijfers; de school die ook minder getalenteerden, wellicht moeizamer, maar toch aan zo'n diploma helpt wordt gediskwalificeerd. Dat is natuurlijk de omgekeerde wereld."

ⁱⁱⁱ De onderstaande tekst is een bewerking van twee artikelen (Dijkstra e.a., 1999; Veenstra e.a., 1998).

^{iv} Dronkers heeft de cijfers naar eigen zeggen afgerond om de indruk van exactheid te voorkomen. Die indruk moet met deze cijfers inderdaad niet worden gewekt. De keuze om dan maar af te ronden is niet gelukkig, omdat kleine verschillen tussen scholen dan groter kunnen lijken dan ze werkelijk zijn.

^v Het Trouw-onderzoek laat zitten blijven of uitvallen in het eerste leerjaar van het voortgezet onderwijs buiten beschouwing. In lijn hiermee neem ik de 262 leerlingen die in het eerste leerjaar zijn uitgevallen niet in de analyses mee.

^{vi} Behalve de gebruikelijke categorieën van vbo, mavo, havo en vwo worden er in VOCL'93 ook enkele andere categorieën gehanteerd, te weten voortgezet speciaal onderwijs, brugklas smal, brugklas breed en middenschool. In het vierde leerjaar worden die categorieën voor 1 procent van de leerlingen gebruikt. De categorie 'brugklas smal' is gelijkgesteld aan havo. De categorieën 'brugklas breed' en 'middenschool' zijn gelijkgesteld aan mavo. De categorie 'voortgezet speciaal onderwijs' is gelijkgesteld aan vbo. Bij het bepalen van de positie op de leerjarenladder worden leerlingen uit het voortgezet speciaal onderwijs overigens als ivbo-leerlingen beschouwd.

^{vii} Dronkers (1999) merkt op dat de leerjarenladder beter 'stilgezet' had kunnen worden wanneer leerlingen hun school voor voortgezet onderwijs verlaten. De vervolgloopbaan van leerlingen is immers het *extern rendement* van scholen.

^{viii} Het effect van etniciteit wordt hierdoor mogelijk enigszins onderschat. Deze operationalisatie is echter gebruikelijk bij de rapportages over VOCL'89 en VOCL'93. Het zou echter goed zijn om de gevolgen van deze keuze na te gaan.

^{ix} Ik heb als tussenstap het teken van de factorscore veranderd, zodat een hoge score inhoudt dat er veel onvertraagd geslaagden, weinig zittenblijvers en weinig uitvallers zijn en de gemiddelde positie op de leerjarenladder hoog is.

^x Het percentage verklaarde variantie is bij de inspectie-gegevens overigens lager dan bij de VOCL-data. De factor bindt respectievelijk 23,5, 25,1, 20,3 en 14,9 procent aan variantie voor vbo, mavo, havo en vwo.

^{xi} De verschillen tussen de analyse van geaggregeerde data en multiniveau analyse zullen vooral aan het licht komen, wanneer de scores van scholen zijn gebaseerd op de gegevens van weinig leerlingen. Snijders en Bosker (1999) geven aan dat de verschillen tussen de methoden gering zijn als er meer dan honderd leerlingen van elke school in het onderzoek zijn betrokken.

^{xii} Dijkstra e.a. (1999) gaan in op de onnauwkeurigheid in de ranglijsten voor Nederlandse scholen voor voortgezet onderwijs. Uitgaande van honderd eindexamenleerlingen binnen een schooltype geldt dat slechts een op de vier scholen significant afwijkt van het gemiddelde. Van deze 25 procent wijkt de helft positief en de andere helft negatief af.