

Aap, Noot, Mustafa...

Het effect van taalafstand en koloniaal verleden op leesprestaties

Jan-Paul Kooistra, Wout Ultee en Ben Pelzer¹

Summary

A is for apple, B is for Bushra: The influence of language distance and colonial experience on reading

This study examines the effect of the distance between origin language and destination language on the reading skills of immigrant students, by building upon insights on the development of world languages, obtained from human population genetics. Lower-level research units are 1759 first-generation immigrants from PISA 2003, all aged fifteen or sixteen, and living in a country which has a Germanic national language. Higher-level research units are 13 destination countries and 46 origin countries. Cross-classified multilevel models test for effects of language distance, and those of the modernization of origin countries and their colonial past, as well as for influences of individual characteristics. Results show that the more an origin language differs from a Germanic destination language, the lower a student's reading score. Migrants from countries colonized by a country with a Germanic language have higher reading scores.

1. Inleiding

De ongelijkheden tussen migranten en autochtonen in westerse industrielanden leiden aanvankelijk tot onderzoek naar de invloed van individuele factoren zoals lage scholing, geringe beheersing van de taal in het bestemmingsland, en een klein vriendennetwerk. Er verschijnen tegenwoordig evenwel studies die tevens de invloed van macrofactoren nagaan. Baanbrekend zijn Van Tubergen en Kalmijn (2005). Om de taalvaardigheid van migranten te verklaren, onderscheiden ze maar liefst drie soorten maatschappijenmerken. De sociale voorzieningen voor vreemdelingen en de mate van discriminatie jegens migranten in een land zijn bestemmingsfactoren. Herkomsteffecten zijn onder meer de religie en de moderniteit van het herkomstland. Dan zijn er nog effecten die vergelijkingen tussen het herkomstland van migranten en hun bestemmingsland behelzen. Voorbeelden zijn het aantal migranten uit een bepaald her-

komstland in verhouding tot de omvang van de inheemse bevolking en verschillen tussen de herkomsttaal en de bestemmingstaal, kortweg taalafstand genoemd.

Dit artikel volgt de door Van Tubergen en Kalmijn (2005) ingeslagen weg en bevat vier nieuwigheden. Terwijl de genoemde auteurs afgingen op de beweringen van migranten over hun taalvaardigheden, benutten wij scores van migrantenleerlingen op toetsen ontworpen voor de vaststelling van leesvaardigheid. Verder achtten we het minder gewenst om, zoals Van Tubergen en Kalmijn, de moderniteit van een land af te lezen aan het gemiddelde inkomen per hoofd van de bevolking. Om leesvaardigheid te verklaren, is het gepast moderniteit te meten als het onderwijspeil in een land. Tevens speelden we uitspraken over de huidige toestand in een land zelf uit tegen de hypothese dat migranten uit de voormalige koloniën van wereldrijken beter lezen. Wanneer, zoals de Amsterdamse socioloog De Swaan (2001: 37) aantoonde, 'een halve eeuw na de onafhankelijkheid de ex-koloniale talen nog altijd een heersende rol spelen in veel postkoloniale samenlevingen', kunnen migranten uit ex-koloniën, vergeleken met migranten uit nooit gekoloniseerde landen, daarmee hun voordeel doen.

Ten slotte pasten we een minder grove operationalisering van taalafstand toe. Van Tubergen en Kalmijn bepaalden de afstand tussen twee talen door herkomsttalen te plaatsen in een van drie aan taalstambomen ontleende categorieën. Wanneer de herkomst- en bestemmingstaal behoren tot één subfamilie van de Indo-Europese taalfamilie, wordt de taalafstand het geringst genoemd. Hierop volgt de categorie waarvoor de herkomst- en bestemmingstaal beide Europese talen zijn, maar geen deel uitmaken van dezelfde subfamilie. De taalafstand is het grootst wanneer de herkomsttaal niet behoort tot de Indo-Europese taalfamilie en de bestemmingstaal wel. Van Tubergen en Kalmijn (2005) vonden dat migranten de bestemmingstaal meer beheersen wanneer de afstand tussen de herkomst- en de bestemmingstaal kleiner is. Deze trichotome operationalisering van taalverwantschap stoelt op bevindingen uit de historische linguïstiek. De populatiegenetica beschikt sinds kort echter over fijnere metingen van taalafstand. Cavalli-Sforza, Menozzi en Piazza (1996) laten een hoge overeenstemming zien tussen de scores van verschillende bevolkingen wat betreft hun taalfamilie en wat betreft hun genen. Terwijl de taalkunde tot nu toe geen bruikbare instrumenten verschaft voor de mate van overeenkomst tussen bijvoorbeeld Engels en Vietnamees, kunnen de genen van Engelsen en Vietnamezen wel worden vergeleken. Al doende ontstaat een schaal voor de afstand tussen onverschillig welke twee talen. Zo'n instrument pasten we toe. De hoofdvraag van dit artikel is hiermee: *Beïnvloedt niet alleen de afstand tussen talen volgens de historische linguïstiek de leesvaardigheid van migrantenleerlingen, maar ook taalafstand zoals gemeten in de genetika?*

Bovenstaande vraag beantwoorden we met leesscores van leerlingen die in 46 herkomstlanden zijn geboren en naar een van 13 Germaanstalige bestemmingslanden zijn gemigreerd, zoals Australië, Duitsland, Nederland en het Verenigd Koninkrijk. Andere bestemmingslanden blijven buiten beschouwing. De gegevens zijn in 2003 verzameld onder 15- en 16-jarigen. Om de onderzoeken eenheden te laten lijken op die uit eerder onderzoek, richtten we ons op enkel Germaanstalige bestemmingslanden. Van Tubergen en Kalmijn onderzochten volwassen migranten uit negen bestemmingslanden, waarvan acht Germaans- en een Romaanstalig. We beperkten ons ook tot Germaanstalige bestemmingslanden, omdat taalafstandscores volgens de

populatiegenetica alleen beschikbaar zijn in de vorm van een vergelijking met het Engels, de meest gesproken taal van de Germaanstalige subfamilie.

2. Theoretische afleidingen

Voor migranten is de beheersing van de bestemmingstaal van groot belang. Het effect van de herkomsttaal op de beheersing van de inheemse taal kan per herkomstland verschillen. Deze verschillen komen voort uit de afstand tussen herkomsttaal en bestemmingstaal. Die afstand bepalen we op twee manieren: met metingen stoeiend op de historische linguïstiek en met een schaal uit de populatiegenetica. Voor elke methode formuleerden we een hypothese. Daarnaast toetsten we hypothesen over de moderniteit en het koloniale verleden van het herkomstland. De macrohypothesen werden getoetst onder controle van individuele karakteristieken.

2.1 De moderniteit van herkomstlanden

Van Tubergen en Kalmijn (2005) gingen na of de moderniteit van herkomstlanden de taalbeheersing van migranten in West-Europese landen vergroot. Ze gebruikten daartoe het bruto nationaal product (bnp) per hoofd van de bevolking. De verwachting was dat migranten uit een land met een laag bnp per capita ervoor kiezen langer te blijven en deze beslissing de inspanning versterkt om de bestemmingstaal onder de knie te krijgen. De auteurs vonden dat de moderniseringsgraad van het herkomstland de verwerving van de bestemmingstaal niet beïnvloedt. In de onderhavige studie deed het opleidingspeil van een land dienst als indicator voor moderniteit. Onder het opleidingspeil van het herkomstland wordt verstaan de geletterdheid onder volwassenen en de onderwijsdeelname in het lager, middelbaar en hoger onderwijs. De keuze viel op deze aanwijzing, omdat bij macroverklaringen voor de leesvaardigheid van migrantenleerlingen een maat voor de hand ligt die op het onderwijsstelsel van een land betrekking heeft – en niet op een gevolg daarvan, dat ook van andere factoren afhangt. Natuurlijk is het de vraag of de taalvaardigheid van een migrant, behalve door het opleidingspeil van haar of zijn ouders, ook wordt bepaald door de gemiddelde opleiding in haar of zijn land van herkomst. Dat is de hypothese die we wilden toetsen. Met deze stelling over een contextueel effect volgden we niet alleen een lange onderzoekslijn, de stelling is ook aannemelijk als ervan wordt uitgegaan dat iemands vaardigheid in een andere taal afhangt van meer personen in diens omgeving dan diens ouders.

In westerse samenlevingen draagt scholing sterk bij aan de allocatie van mensen op de arbeidsmarkt. Algemene kennis en op school aangeleerde vaardigheden worden ingezet op de werkvloer. De verwachting is dat migranten uit moderne landen zich sneller aanpassen aan de omstandigheden in het bestemmingsland. Gevolg hiervan is dat zij minder moeite hebben met het onderwijs aldaar en de taal van dit land beter beheersen. Hypothese (1) is tegengesteld aan die van Van Tubergen en Kalmijn (2005) en luidt: *Hoe hoger het opleidingspeil in het herkomstland, des te beter lezen migranten.*

2.2 Het koloniale verleden van herkomstlanden

Moderniseringstheorieën doen alsof landen autonome eenheden zijn, wereldsysteemtheorieën postuleren effecten van het ene op het andere land (Ultee, Arts & Flap, 1992: 485). Zo kan het koloniale verleden van herkomstlanden invloed hebben op migranten uit die landen. Het Westen – waar onder meer Germaanse talen worden gesproken – koloniseerde gebieden ten behoeve van economisch gewin. Tevens legde het zijn leefstijl op aan de inwoners van deze gebieden. Zo ‘beschaafden’ christelijke missionarissen ‘barbaarse stammen’. De inheemse bevolking werd veelal onderwezen in de taal van de kolonisator. Na de dekolonisatie bleef vaak de taal van de kolonisator in gebruik (De Swaan, 2001). Een uitzondering is Indonesië, het voormalige Nederlands-Indië, waar sinds de uitroeping van de onafhankelijkheid Bahasa Indonesia, een vorm van Maleis, de taal van het land is (Anderson, 1990; Groeneboer, 1998). Verwacht wordt dat migranten uit voormalige koloniën zich makkelijker de levenswijze in moderne bestemmingslanden eigen maken. Ze hebben niet alleen een grotere bekendheid met de taal van hun bestemmingsland, maar ook zijn ze wellicht meer bekend met de aard van de scholen aldaar en met schoolgang. Hypothese (2) zegt: *Migranten naar Germaanstalige bestemmingslanden uit herkomstlanden zonder een Germaanstalig koloniaal verleden zullen minder goed lezen dan migranten uit herkomstlanden mét een Germaanstalig koloniaal verleden.*

2.3 De verbondenheid van talen in de wereld: de historische linguïstiek

De historische linguïstiek beantwoordt vragen over taalverwantschap. Als talen in uitspraak, vervoeging, woordenschat en zinsbouw sterk op elkaar lijken, duidt dit op eenzelfde oorsprong. De oorsprongtaal is in veel gevallen een dode taal, bijvoorbeeld het Latijn. In 1786 veronderstelde Sir William Jones het bestaan van een oertaal, waar de Europese talen zich van hebben afgescheiden (Cannon, 1993). Deze oertaal staat tegenwoordig bekend als het Indo-Europees en bracht een van de grootste taalfamilies uit de wereld voort. De Indo-Europese taalfamilie omvat 449 levende talen, waaronder het Nederlands.

De Indo-Europese taalfamilie bestaat uit subfamilies. Het Nederlands behoort met het Engels en het Duits tot de Germaanse subfamilie. De verwantschap tussen Nederlands en Duits is groter dan die tussen Frans en Nederlands, omdat het Frans tot de Romaanse subfamilie behoort (Störig, 1988). De Indo-Europese familie is een van vele taalfamilies.

De indeling van een taal bij een bepaalde familie is binnen de historische linguïstiek soms omstreden. Andere methoden leidden tot afwijkende resultaten. Begin 19e eeuw werd het Armeens bij de Indo-Europese taalfamilie ingedeeld. Deze taal zou een zustertaal van het Perzisch zijn, omdat de woordenschatten van de twee talen sterk op elkaar leken. Eind 19e eeuw werd het Armeens echter op grond van oudere documenten als een subfamilie binnen de Indo-Europese taalfamilie beschouwd (Hübschmann, 1897). Het grootste probleem bij de indeling van talen is de schaarste aan oude teksten. Het eerste schrift dateert van 3000 voor Christus, wat onderzoek naar de verwantschap van talen bemoeilijkt. Door het gebrek aan oude bronnen hanteren veel taalkundigen een taalmodel dat uitgaat van meerdere oorsprongen. De verschil-

lende taalfamilies zouden zich los van elkaar hebben ontwikkeld, taalfamilies zijn autonome entiteiten die niet met elkaar in verband kunnen worden gebracht. Een vergelijking tussen de talen van verschillende taalfamilies zou vrijwel onmogelijk zijn. Gezien dit alles formuleerden we hypothese (3): *Migranten in Germaanstalige landen waarvan de herkomsttaal behoort tot de Germaanse subfamilie kunnen beter lezen dan migranten met een herkomsttaal die niet behoort tot de Germaanse subfamilie.* We opperden eveneens hypothese (4): *Migranten in Germaanstalige landen waarvan de herkomsttaal niet behoort tot de Germaanse subfamilie, maar wel behoort tot de Indo-Europese taalfamilie, lezen beter dan migranten waarvan de herkomsttaal niet behoort tot de Indo-Europese taalfamilie.*

2.4 De verbondenheid van de talen in de wereld: de populatiegenetica

Er zijn taalwetenschappers die beweren dat alle talen zich hebben ontwikkeld uit één oertaal (Greenberg, 2000; Ruhlen, 1987). Door het gebrek aan oude geschreven bronnen lijkt een dergelijke uitspraak een slag in de lucht. Recente ontwikkelingen in de populatiegenetica wijzen evenwel op een sterke overeenkomst tussen de genetische en de linguïstieke afstand tussen bevolkingsgroepen, wat kan leiden tot beteugeling van wilde uitspraken. Cavalli-Sforza, Menozzi en Piazza (1996) onderscheiden met behulp van genenvergelijkingen 42 bevolkingsgroepen. Die komen sterk overeen met de takken van taalbomen. Over het algemeen geldt: hoe eerder zich een bevolkingsgroep bij de uitstroom uit Afrika afscheidde, des te groter is de linguïstieke en genetische differentiatie in deze populatie. De taaleigenheid en de genetische samenstelling van een volk wordt voor een groot deel bepaald door de ruimtelijke ligging en het daarbij behorende isolement. Ouders dragen in deze gebieden niet alleen genen aan hun kinderen over, maar ook in sterke mate cultuur. Een belangrijk onderdeel van cultuur is taal. De samenhang tussen genetische bomen en taalfamilies is bepaald door Cavalli-Sforza, Menozzi en Piazza en blijkt nauw. Doordat de twee evoluties sterk op elkaar lijken, vallen hypothesen af te leiden die meer informatieve voorspellingen opleveren over de invloed van taalafstand op de beheersing van een bestemmingstaal door migranten. De genetische boom kan onduidelijkheden in de taalkunde over taalverwantschap verhelderen en leemten in de kennis van dit vak over de evolutie van talen opvullen. Hypothese (5): *Hoe groter de afstand tussen herkomst- en bestemmingstaal, des te minder de leesvaardigheid.*

2.5 Individuele karakteristieken

Migranten van dezelfde leeftijd die al jong naar hun bestemmingsland kwamen, zullen de taal aldaar beter spreken omdat ze daartoe meer gelegenheid hebben gehad. Van Tubergen en Kalmijn (2005) laten zien dat een langer verblijf de beheersing van de bestemmingstaal vergroot. In deze studie werd een dergelijk effect tevens verwacht voor de leesvaardigheid van jonge migranten. Het effect van toenemende op school opgedane kennis wordt zichtbaar met leesvaardigheidstesten. Hypothese (6) luidt: *Migranten die zich op jongere leeftijd in een land vestigden, lezen beter dan migranten die zich daar korter geleden vestigden.*

De Franse socioloog Bourdieu (1991) verwoordde gedachten over de (onbedoelde) gevolgen van ouderlijk kapitaal voor onderwijsongelijkheden. Zijn conflictsociologisch model zegt dat een school als een slagveld is. Om hun kinderen te laten winnen, passen ouders strategieën toe. Voorbeelden zijn verzorgd taalgebruik, culturele conversaties, museabezoek en muziekles. Zo krijgen kinderen uit hogere milieus een voorsprong en worden in het onderwijs ongelijkheden cultureel gereproduceerd. Ook doet zich zelfselectie voor. Kinderen uit lagere milieus voelen zich niet thuis in kringen waar vaardigheden en wetenswaardigheden op cultureel gebied tellen. Wij onderzochten de invloed van cultureel kapitaal op onderwijsprestaties bij migranten. Deze groep wordt in de desbetreffende literatuur verwaarloosd. De hypothese dat ongelijkheid door cultureel kapitaal wordt gereproduceerd, gaat evenwel ook voor migranten op. Migranten die in contact zijn gebracht met de 'haute culture' voelen zich meer thuis op school, wat een positief effect heeft. Hypothese (7): *Hoe groter het culturele kapitaal van een migrantenleerling, des beter leest deze persoon.*

Bij het presteren op school spelen klasgenoten een grote rol. Zoals volwassenen door vrienden betere banen krijgen, zo kunnen leerlingen door banden met klasgenoten beter presteren. Zonder die banden zullen leerlingen – en vooral migrantenleerlingen – lager scoren op leesvaardigheid. Hypothese (8): *Migranten met een positiever oordeel over het gemak waarmee vrienden op school vallen te maken, hebben een hoger niveau van leesvaardigheid.*

Ten slotte enige opmerkingen over individuele kenmerken die wellicht ten onrechte niet aan bod komen. Bij die variabelen valt te denken aan het niveau van de school dat door leerlingen wordt bezocht. Het hier gebruikte databestand betreft echter leerlingen van 15 of 16 jaar. Omdat zij nog geen diploma hebben behaald en meerdere landen niet onderscheiden tussen lager en hoger middelbaar onderwijs, zal een maat voor het niveau van de bezochte school weinig toevoegen. Verder lijkt het minder gewenst deze variabele op te nemen omdat zo een te groot deel van de effecten van migratie wordt uitgezuiverd. Een ander individueel kenmerk dat buiten beschouwing blijft, betreft het onderscheid tussen jongeren die met hun ouders uit hun herkomstland zijn vertrokken en baby's die zijn geadopteerd door volwassenen in hun bestemmingsland. Het hier gebruikte bestand bevat daarover geen gegevens. Wat dit betreft zou de samenstelling van de groep leerlingen uit bepaalde herkomstlanden sterk kunnen afwijken. Tegenwoordig is China een bekend adoptieland, eerder was dat Vietnam. Onder de 15- of 16-jarige leerlingen in het in 2003 vergaarde PISA-bestand (PISA: *Programme for International Student Assessment*) zullen zich echter nauwelijks wegens de één-kind-politiek te vondeling gelegde Chinese kinderen bevinden, omdat adoptie uit dit land de afgelopen tien jaar op gang kwam. De gegevens zullen ook nauwelijks Vietnamese adoptiekinderen bevatten. De baby's die Amerikaanse mannen rond 1970, tijdens de Vietnam-oorlog, bij Vietnamese vrouwen verwekten en die deze vrouwen voor adoptie afstonden, zijn ondertussen te oud voor PISA.

3. Eenhedenselectie en databewerking

3.1 Data

Voor de toetsing van de acht hypothesen gebruikten we in onderhavig artikel de gegevens verzameld in 2003 onder 276.165 leerlingen, in 41 landen, voor het *Programme for International Student Assessment* (OECD/PISA, 2003). Voor dit artikel werden eerst de 21.472 respondenten uitgekozen die zeiden buiten hun bestemmingsland te zijn geboren. Daarna bepaalden we hun precieze herkomstland. Sommige aan PISA deelnemende landen gaven alleen aan dat migranten uit een bepaald werelddeel kwamen, wat bijvoorbeeld de gegevens verzameld in de Verenigde Staten onbruikbaar maakte. Als leerlingen zelf geen geboorteland aangaven, werd het herkomstland van beide ouders tot hun geboorteland gemaakt. Wanneer ouders uit verschillende landen kwamen, maakten we het land van de moeder tot het herkomstland. Hiervoor is gekozen omdat over het algemeen moeders zich meer bezighouden met de opvoeding van kinderen. Als het herkomstland van één ouder ontbrak, werd dat van de andere ouder het geboorteland van een kind. Leerlingen vielen buiten de analyse als het herkomstland na deze stappen nog onbekend was. Er bleven nu 10.893 leerlingen over. Uit die groep kozen we de leerlingen die in Australië, België, Denemarken, Duitsland, Ierland, Liechtenstein, Nederland, Nieuw-Zeeland, Noorwegen, Oostenrijk, het Verenigd Koninkrijk, Zweden of Zwitserland woonden, dat wil zeggen: Germaanstalige bestemmingslanden. Een moeilijkheid vormde de meertaligheid van sommige landen, bijvoorbeeld België en Zwitserland. Door te letten op de taal waarin de leestoets is afgenomen, zijn niet-Germaanse bestemmingen uit het onderzoek verwijderd. Zo hielden we 2332 leerlingen over. Na de verwijdering van respondenten met ongeldige scores op de hieronder te behandelen individuele variabelen, bleven er 1759 respondenten uit 46 herkomstlanden over.

3.2 Databewerking

Leesvaardigheid

Het PISA-bestand bevat items voor het leesvaardigheidsniveau van leerlingen, in het bijzonder hun vermogen de intentie van teksten te vatten. Er is uitgegaan van vier typen teksten: teksten voor persoonlijk gebruik, teksten voor publiek gebruik, teksten uit werksituaties en leerteksten. Bij persoonlijke teksten gaat het om levensverhalen, belletrie, familiebrieven en dergelijke. Bij teksten voor publiek gebruik stond het lezen van officiële documenten centraal. Teksten van het derde type hebben betrekking op direct uitvoerbare taken. Bij teksten van het laatste type werden leerlingen getoetst op hun vermogen om begrijpend te lezen.

De scores op de items zijn door PISA gestandaardiseerd. Er zijn vijf niveaus met elk een eigen scorebereik. Hierbij staat het vijfde niveau voor het hoogst haalbare. De niveaus zijn per teksttype vastgesteld. Om tot scores voor elk van de vier teksttypen te komen, heeft PISA gemiddelden uitgerekend, weer bestaande uit vijf niveaus. Voor deze studie is van deze gemiddelden één intervalschaal gemaakt. De maximumscore onder de leerlingen is gelijk aan

754,79, de laagste bedraagt 195,11. Het gemiddelde is een score van 495,69, met een standaardafwijking van 105,42. De Cronbachs Alpha van de leesschaal bedraagt 0,974.

Het opleidingspeil en het Germaanstalig koloniale verleden van het herkomstland

Voor het opleidingspeil in een herkomstland is de *education-index* uit het *Human Development Report 2005* van het *United Nations Development Programme* gebruikt. De index heeft betrekking op 2003, net als de data voor de leerlingen in het PISA-bestand. De scores zijn berekend door de geletterdheid onder volwassenen van het land en de onderwijsparticipatie in het primaire, secundaire en tertiaire niveau van het land te standaardiseren. De schaal loopt van 0 tot 10, waarbij 10 het maximum vormt. Het gemiddelde opleidingspeil van de herkomstlanden bedraagt 9,07, met een standaardafwijking van 1,07. In tabel 1 zijn de scores per herkomstland weergegeven. De laagst scorende landen zijn China, India, Libanon, Maleisië, Nigeria, Pakistan, Turkije en Zuid-Afrika.

Voor de variabele 'Germaanstalig koloniaal verleden' zijn de 46 herkomstlanden gerekend tot een van drie categorieën. De eerste bevat alle Germaanstalige herkomstlanden, deze categorie geldt als referentie. De niet-Germaanstalige herkomstlanden zonder een Germaanstalig koloniaal verleden bevinden zich in de tweede categorie. Merk op dat deze landen niet gekoloniseerd behoeven te zijn. Voorbeelden zijn Frankrijk en Italië. In de derde staan de niet-Germaanstalige herkomstlanden mét een Germaanstalig koloniaal verleden. Landen zijn ingedeeld aan de hand van *Encyclopædia Britannica* (1992). Deze indeling staat eveneens in tabel 1. Herkomstlanden die als gekoloniseerd worden beschouwd, hebben veelal behoord tot een Germaanstalige kolonisator. De uitzonderingen zijn Brazilië, de Filipijnen en Vietnam. Fiji, de Cook Eilanden, India, Maleisië, Nigeria, Pakistan, Samoa en Tonga zijn allemaal koloniën geweest van het Britse Rijk. Zuid-Afrika heeft onder Brits bestuur gestaan, maar viel eerder gedeeltelijk tevens onder Nederlands bestuur. Australië, Canada, Nieuw-Zeeland en de Verenigde Staten hebben we tot de Germaanstalige herkomstlanden gerekend. Deze landen waren Britse koloniën, maar hun hedendaagse inwoners stammen veelal af van de kolonisten en zelden van de oorspronkelijke bevolking.

Taalafstand: de taalkundige meting

Allereerst werd de herkomsttaal van de respondent bepaald. Het vertrekpunt van de hercoderingen was het item '*Welke taal spreekt U thuis?*' uit het PISA-bestand. Een groot deel van de eerste generatie allochtone leerlingen bleek de taal van het bestemmingsland niet in de eigen leef sfeer te spreken. In deze studie werd de taal die dan wél thuis werd gesproken als herkomsttaal beschouwd. Wanneer de respondent aangaf de bestemmingstaal thuis te spreken, werd de herkomsttaal aan de hand van het land van herkomst bepaald. De herkomsttalen zijn vervolgens gecategoriseerd. In de eerste categorie bevinden zich alle Germaanse talen. Deze categorie fungeert als referentiecategorie bij de analyses. De talen in de tweede categorie zijn talen uit de Indo-Europese taalfamilie, met uitzondering van de Germaanse subfamilie. In de derde categorie bevinden zich de talen die niet behoren tot de Indo-Europese taalfamilie. De drie categorieën zijn als dichotomieën opgenomen. Ze kunnen als maat voor taalafstand gelden, omdat alle

leerlingen in een Germaanstalig land wonen en er dus steeds met eenzelfde punt wordt vergeleken. Volgens deze meting is de taalafstand tussen iemand die bijvoorbeeld in Nederland is geboren en in Engeland woont per definitie nul. Dat is, zo blijkt hieronder, ietwat anders bij de populatiegenetische meting.

In totaal zijn er 1063 respondenten met een Germaanse taal, 465 respondenten met een andere Indo-Europese taal en 231 migrantenleerlingen met een taal die niet tot de Europese taalfamilie behoort. In tabel 1 staat per herkomstland de categorisering van de daar veelal gesproken taal volgens de historische linguïstiek.

Taalafstand: de populatiegenetische meting

Terwijl de taalkundige meting van taalverwantschap in het midden laat of alle talen één oorsprong hebben, gaat de genetische meting uit van één oertaal. De afstand tussen de talen wordt bepaald aan de hand van de mate waarin de genen van een bevolkingsgroep verschillen van die van een andere populatie. Cavalli-Sforza, Menozzi en Piazza (1996) bepaalden deze mate voor 42 bevolkingsoorten van de wereld, daarbij steeds de Engelsen als vergelijkingspunt gebruikend. Voor elke bevolkingsgroep gingen ze daarbij uit van 120 bepaalde genen (Cavalli-Sforza, Menozzi & Piazza, 1996: 75). Voor de volken van Europa stelden zij deze maat vast op basis van 88 genen (Cavalli-Sforza, Menozzi & Piazza, 1996: 270). De resulterende maat wordt door hen de 'Fst-distance' genoemd (Cavalli-Sforza, Menozzi & Piazza, 1996: 29). Deze maat voor genetische distantie werd in dit artikel als indicator gebruikt voor de taalafstand.

Deze studie handelt over migranten naar Germaanstalige landen. Cavalli-Sforza, Menozzi en Piazza geven echter niet één maat voor de afstand tussen bijvoorbeeld Vietnamees en alle Germaanse bevolkingsgroepen bij elkaar, maar alleen voor de afstand tussen de bevolking van Vietnam en die van Engeland. Daarom zijn alle genetische distanties ten opzichte van de bevolking van Engeland in kaart gebracht. Overigens bleken de bevolkingen van Engeland en bijvoorbeeld Nederland genetisch vrijwel niet van elkaar te verschillen. Bij deze meting is de feitelijke afstand tussen bijvoorbeeld het Engels en het Nederlands dus (nagenoeg) nul en verschilt de afstand tussen het Engels en het Vietnamees niet van die tussen het Nederlands en het Vietnamees. De inwoners van landen die thuis mogelijk een niet-Germaanse taal spreken maar in een land wonen met een officiële taal uit de Germaanse taalgroep, krijgen ook een taalafstand van 'nul'. Voorbeelden zijn Fiji, Nigeria, Zuid-Afrika en de Cook Eilanden.

Voor sommige van de 46 herkomstlanden van de door ons uit het PISA-bestand gelichte migrantenleerlingen, kon niet onmiddellijk bij Cavalli-Sforza, Menozzi en Piazza (1996) een taaldistantiemaat worden gevonden. Om het aantal respondenten niet te verkleinen, is ervoor gekozen deze landen langs een omweg op te nemen in het onderzoek.

De Balkanlanden zijn door Cavalli-Sforza, Menozzi en Piazza minder gedetailleerd benoemd dan in het PISA-bestand. Terwijl de genetische data werden verzameld tussen 1978 en 1986, toen Joegoslavië nog bestond, spreekt PISA over Bosnië-Herzegovina, Kroatië, Montenegro en Servië. Deze landen gaven we de score voor Joegoslavië. Dat deden we ook voor Albanië. Een tweede probleem deed zich voor bij Libanon als herkomstland. Hier wordt Arabisch gesproken, hetgeen een taal is uit de Afro-Aziatische taalfamilie. Cavalli-Sforza,

Menozzi en Piazza hebben de Libanese bevolking niet onderzocht. Van alle volken met een taal behorend tot de Afro-Aziatische taalfamilie geven ze alleen voor de Berbers een maat voor genetische distantie ten opzichte van een Germaanse bevolkingsgroep. Bij gebrek aan beter werd in deze studie die distantiescore toegekend aan leerlingen uit Libanon.

De genetische samenstelling van Litouwen, Slovenië en Slowakije werd gelijkgesteld aan die van Polen. Al deze landen behoren tot de Baltisch-Slavische taalgroep, welke weer een aftakking is binnen de Indo-Europese taalfamilie. De genetische afstand tussen de Brazilianen en de Germaanse groep is gelijkgesteld aan de afstand tussen de Portugezen en de Germaanse groep. De nationale taal van Brazilië is het Portugees. In deze studie werd verondersteld dat de migranten uit Brazilië geen inheemse Zuid-Amerikaanse herkomsttaal spreken, deze is vrijwel geheel uitgestorven. Cavalli-Sforza, Menozzi en Piazza geven geen taalafstandscore voor Georgië. In deze studie werden de Georgiërs als een Russisch volk beschouwd. Verondersteld werd dat, door de eeuwenlange Russische bezetting, de genetische samenstelling van de Georgiërs overeenkomt met die van de Russen en dat de Georgiërs ook de Russische taal beheersen. Estland is ondergebracht in de 'Oeraalse' categorie van Cavalli-Sforza, Menozzi en Piazza, Samoa en Tonga zijn gerekend tot hun Polynesische categorie.

De Fst-distanceschaal voor genetische distantie heeft een bereik van 0 tot 1,275. Tabel 1 geeft per herkomstland de taalafstand ten opzichte van de Germaanse talen. De genetische afstand tussen de Maleisische bevolking en de Germaanse volken blijkt het grootst te zijn (Fst-distance = 1,275). Hierop volgen de Chinezen (1,152), de Filippijnen (1,117) en de Vietnamezen (1,100). Van de Europese volken wijken de Grieken, genetisch gezien, het sterkst af van de Germaanse volken (0,204). De Fransen verschillen, met betrekking tot de genetische distantie, ten opzichte van de Germaanse volken het minst (0,024). De scores voor taalafstand werden gestandaardiseerd (als z-score) opgenomen in de analyse.

De taalkundige en de populatiegenetische meting vergeleken

Uit het bovenstaande blijkt dat de taalkundige meting een nominale schaal behelst en de populatiegenetische een intervalschaal. Daarom ligt het wellicht voor de hand dat de populatiegenetische meting een grotere verklaringskracht heeft dan de nominale meting. Twee opmerkingen zijn hier op hun plaats. Ten eerste dat het, gezien de stand van het historisch taalonderzoek, onmogelijk is om op wereldniveau tot een intervalmeting te komen. Ten tweede dat er geen logische noodzaak is dat de populatiegenetische meting iets toevoegt aan de verklaringskracht van de taalkundige meting. Het is zeer wel mogelijk dat de populatie-genetische meting onnodig fijne differentiaties aanbrengt. In onze bespreking van de onderzoekresultaten komen we terug op de vraag van de verhouding tussen de verklaringskracht van beide metingen van taalafstand.

De individuele karakteristieken

De leeftijd waarop de leerlingen uit het PISA-bestand in hun bestemmingsland aankwamen, is opgenomen als intervalvariabele. De gemiddelde leeftijd bij aankomst bedraagt acht jaar.

The index of possessions related to 'classical' culture in the family home, te vinden in het PISA-bestand, is gebruikt als variabele voor het culturele kapitaal van de respondent. Deze variabele is

Tabel 1: Voor 46 herkomstlanden: de genetische afstand (*Fst-distance*) ten opzichte van de Germaanse bevolking, hun opleidingspeil en hun Germaanstalig koloniaal verleden (waarbij 1 = referentiecategorie (Germaanstalige herkomstlanden), 2 = niet-Germaanstalig herkomstland zonder Germaans koloniaal verleden, 3 = niet-Germaanstalig herkomstland met Germaans koloniaal verleden)

Herkomst-land	Taal-afstand (taalkundig)	Taal-afstand (genetisch)	Herkomst-land	Germaanstalig koloniaal verleden	Opleidingspeil	Germaanstalig koloniaal verleden	Taal-afstand (taalkundig)	Taal-afstand (genetisch)	Opleidingspeil	Germaanstalig koloniaal verleden
Albanie	2	0,160	Montenegro	2	8,9	2	2	0,160	8,7	2
Australië	1	0,000	Nederland	1	9,9	1	1	0,000	9,9	1
Bosnië-Herzegovina	2	0,160	Nigeria	1	8,6	2	3	0,000	6,6	3
Brazilië	2	0,046	Nieuw-Zeeland	1	8,9	2	1	0,000	9,9	1
Canada	1	0,000	Pakistan	2	9,7	1	3	0,280	4,4	3
China	3	1,152	Roemenië	2	8,4	2	2	0,160	8,9	2
Cook Eilanden	1	0,000	Polen	2	9,9	3	2	0,070	9,6	2
Denemarken	1	0,000	Portugal	2	9,9	1	2	0,046	9,7	2
Duitsland	1	0,000	Rusland	2	9,6	1	2	0,079	9,6	2
Eeland	3	1,023	Samoa	2	9,7	2	3	0,991	8,9	3
Fiji	1	0,000	Schotland	1	8,6	3	1	0,000	9,9	1
Filipijnen	3	1,117	Servië	2	8,9	2	2	0,160	8,7	2
Frankrijk	2	0,024	Slovenië	2	9,9	2	2	0,070	9,8	2
Georgië	3	0,079	Slowakije	2	9,0	2	2	0,070	9,1	2
Griekenland	2	0,204	Spanje	2	9,7	2	2	0,047	9,7	2
Hongarije	3	0,070	Tonga	3	9,6	2	3	0,993	9,3	3
India	2	0,280	Tsjechië	2	6,1	3	2	0,060	9,3	2
Italië	2	0,051	Turkije	3	9,5	2	2	0,886	8,2	2
Joegoslavië	2	0,160	Vietnam	3	8,7	2	2	1,100	8,2	2
Kroatië	2	0,160	Verenigd Koninkrijk	1	9,0	2	1	0,000	9,9	1
Libanon	3	0,273	Verenigde Staten	1	8,4	2	1	0,000	9,4	1
Litouwen	3	0,070	Zuid-Afrika	1	9,7	2	3	0,000	8,1	3
Maleisië	3	1,275	Zweden	1	8,3	3	1	0,000	9,9	1

Bronnen: Cavalli-Sforza, Menozzi & Piazza (1996), United Nations Development Programme (2005) en Encyclopædia Britannica (1992)

geconstrueerd uit items die betrekking hebben op het culturele bezit van de ouders, zoals de aanwezigheid van klassieke literatuur, poëzie en kunstwerken in het ouderlijk huis. De variabele vormt een intervallschaal, waarbij een hogere score voor meer cultureel kapitaal staat. De minimale score op deze variabele bedraagt -1,28, de maximale 1,35. Het gemiddelde ligt op -0,15. De variabele werd gestandaardiseerd opgenomen in de analyse.

Helaas bevat het PISA-bestand geen directe meting voor het hebben van vrienden in de klas, vandaar de keuze voor een minder rechtstreekse aanwijzing. Het PISA-bestand bevat het antwoord van leerlingen op de stelling *‘Op mijn school maak ik gemakkelijk vrienden’*. Dit item gebruikten we omdat we de veronderstelling gepast achtten dat leerlingen die het met deze stelling eens zijn meer vrienden in de klas hebben. De leerling werden vier antwoordcategorieën aangeboden: 1 = ‘helemaal mee eens’, 2 = ‘mee eens’, 3 = ‘niet mee eens’, 4 = ‘helemaal niet mee eens’. In dit onderzoek is ervoor gekozen om een hoge score gelijk te stellen aan een positieve houding ten opzichte van het vrienden maken. De vier categorieën bleven gehandhaafd, alleen geldt de volgende interpretatie: 1 = ‘helemaal niet mee eens’, 2 = ‘niet mee eens’, 3 = ‘mee eens’, 4 = ‘helemaal mee eens’. De gemiddelde score op deze variabele bedraagt 3,15, het oordeel van migrantenleerlingen over het maken van vrienden op school is overwegend positief. Deze variabele is ook gestandaardiseerd. Het geslacht van de leerling is een dummy met vrouwelijk als referentiecategorie. Ook het opleidingsniveau van de vader en dat van de moeder is in het onderzoek betrokken.

Het opleidingsniveau van de ouders is gecodeerd volgens de ISCED, de *International Standard Classification of Education*, van de Unesco. De schaal telt zes niveaus, lopend van laag naar hoog.

4. ‘Cross-classified’ multiniveaumodellen

In de uitgevoerde analyses werden een of meer onafhankelijke variabelen in een regressiemodel opgenomen en werd gekeken op welke wijze en in welke mate deze de afhankelijke variabele verklaren. Er is voor multiniveaumodellen gekozen om ervoor te zorgen dat passend rekening wordt gehouden met de afhankelijkheid van leesvaardigheidscores van leerlingen die het herkomstland, het bestemmingsland of beide met elkaar gemeen hebben.

In de door ons gehanteerde multiniveaumodellen werden geen variabelen voor de bestemmingslanden van migranten opgenomen. De reden hiervoor was dat onderhavig onderzoek betrekking heeft op migrantenleerlingen in Germaanstalige bestemmingslanden. Door de homogeniteit van de bestemmingslanden werden de individuele, de herkomstland- en de taalafstandkenmerken het meest relevant geacht voor de prestaties op het gebied van lezen.

Data voor multiniveaumodellen hebben veelal een structuur waarin lagere eenheden te ordenen zijn onder hogere eenheden, die op hun beurt te ordenen zijn onder nog hogere eenheden, enzovoort. In dit onderzoek is geen sprake van zo’n hiërarchische structuur. Er zijn weliswaar lagere eenheden (leerlingen) en hogere (herkomst- en bestemmingslanden), maar de hogere zijn onderling niet hiërarchisch te ordenen: onder één herkomstland zijn (leerlingen wonend

in) meerdere bestemmingslanden te ordenen, maar andersom zijn ook onder één bestemmingsland (leerlingen afkomstig uit) meerdere herkomstlanden te ordenen. De leerlingen zijn bijgevolg onder twee verschillende eenheden van een hoger niveau te ordenen. Op dergelijke gegevens kunnen 'cross-classified' modellen worden toegepast (Browne, Goldstein & Rasbash, 2001). Daarbij werden in deze studie de invloeden van herkomst- en bestemmingsland als randomeffecten in de regressiemodellen opgenomen.

Voorts is ervoor gekozen om de regressiecoëfficiënten van de onafhankelijke variabelen *fixed* op te nemen. Dit is gebeurd omdat voor geen enkele onafhankelijke variabele werd verwacht dat de invloed op de afhankelijke variabele in sterkte verschilt per herkomst- of bestemmingsland. Ter verduidelijking een voorbeeld: verwacht werd dat de invloed van het ouderlijk opleidingsniveau op de schoolprestaties voor migranten in Australië even sterk is als voor migranten in Nederland, ongeacht hun herkomst.

Tabel 2 toont de resultaten van vier modellen voor de leesvaardigheid onder eerste generatie migranten in Germaanstalige bestemmingslanden. In het basismodel, model 1, wordt alleen het intercept en de variantie tussen individuen, herkomstlanden en bestemmingslanden geschat. De totale variantie bedraagt 10.371, waarvan 72 procent toe te schrijven valt aan het individuele niveau, 23 procent aan het niveau van de herkomstlanden en 5 procent aan dat van de bestemmingslanden.

In model 2 staat de taalkundige meting voor taalverwantschap centraal. Door de toevoeging van de relevant geachte variabelen daalt de variantie tussen herkomstlanden naar 18 procent, die op het niveau van de bestemmingslanden blijft 5 procent. De geschatte parameters voor de taalkundige meting wijzen in de richting van hypothesen (3) en (4), onderling verschillen de taalcategorieën echter niet. Wat de individuele karakteristieken betreft, blijken migrantenleerlingen die langer in hun bestemmingsland verblijven, beter te lezen. Deze bevinding bekrachtigt hypothese (6). Het model laat verder een positief effect zien van het culturele kapitaal van de leerling op het niveau van leesvaardigheid. Hoe groter het culturele kapitaal van de leerling, des te beter leest deze (hypothese 7). Het effect van de opleiding van de vader en de moeder is eveneens positief. Mannen blijken significant over een geringer niveau van leesvaardigheid te beschikken dan vrouwen. De -2 Log likelihood van model 2 is significant lager dan die van model 1, wat zegt dat model 2 beter bij de data past dan model 1.

In model 3 blijkt de populatiegenetisch vastgestelde taalafstand van het herkomstland van immigranten in tegenstelling tot onze verwachtingen (hypothese 5) een positieve, zij het niet significante, invloed te hebben op hun leesvaardigheid. Omdat in de analyse slechts 46 herkomstlanden zijn betrokken, onderzochten we in hoeverre dit resultaat veroorzaakt kan zijn door één of enkele herkomstlanden, die het als 'invloedrijke gevallen' tot stand brachten. Hiervoor bestaat echter geen standaardprocedure bij 'cross-classified' multiniveaumodellen.

Vertrekpunt voor de door ons ontwikkelde werkwijze voor deze multiniveaumodellen is de maat $|\hat{b} - \hat{b}_{-j}| / \hat{s}_{-j}$. Deze maat wordt bij regressiemodellen die zich tot eenheden van één enkel niveau beperken, wel aangeduid als 'sdfbeta'. In deze maat staat $\hat{b}$ voor het geschatte effect van taalafstand wanneer de gegevens van *alle* respondenten worden gebruikt en $\hat{b}_{-j}$ voor het geschatte effect na verwijdering van respondenten afkomstig uit herkomstland j . Het absolute

verschil tussen deze twee effecten wordt gestandaardiseerd via deling door de geschatte standaardfout $\hat{s}_{-j}$ van $\hat{b}_{-j}$. Zo geeft een *sdfbeta* waarde van 1 aan dat door toevoeging van de respondenten uit herkomstland j , de waarde van het geschatte effect van taalafstand verandert met 1 standaardfout. Als criterium voor te sterke invloed van één geval (hier: een bepaald herkomstland) op de geschatte waarde van een parameter hanteert men wel de waarde van $2/\sqrt{n}$ (hier: 0,29 met $n = 46$). Dat deden wij ook. De *sdfbeta* als indicator van de invloed van eenheden van een hoger niveau op het regressie-effect van één predictor in een multiniveaumodel, is sterk verwant aan Cook's Distance C_j^{OF} (Snijders & Bosker, 1999: 134), die de invloed weergeeft van een hogere eenheid op *alle* fixed regressie-effecten tezamen.

Bij eenniveauregressiemodellen is het niet ongebruikelijk om enkele invloedrijke gevallen te verwijderen uit het gegevensbestand en de analyse opnieuw uit te voeren op de beperkte dataset. Bij multiniveaumodellen ligt dat minder voor de hand. In de onderhavige analyse gaat het om de invloed van specifieke hogere eenheden (herkomstlanden) op het geschatte effect van het hogerniveaumenek 'taalafstand'. Voor het schatten van de regressie-effecten van de individuele karakteristieken is verwijdering van (alle respondenten afkomstig uit) een herkomstland niet gewenst. Daarom pasten we het voorstel van Langford en Lewis (1998) toe: een invloedrijk herkomstland wordt alleen uit het *random* deel van het multiniveauregressiemodel verwijderd en krijgt een eigen *fixed* deel (intercept c.q. dummy) toegewezen. Daarmee verdwijnt de invloed van het betreffende herkomstland op het effect van taalafstand nagenoeg volledig.

Bij de nadere analyse van het verband tussen genetische taalafstand en leesvaardigheid pasten we bovenstaande procedure herhaald toe. We bepaalden daarbij steeds het volgens ons criterium meest invloedrijke herkomstland, waarna we dit herkomstland uit het random deel verwijderden en een dummy voor dat herkomstland aan het model toevoegden.² Vier herkomstlanden bleken een te grote invloed te hebben gehad op het met model 3 gevonden effect van genetische taalafstand. Dat waren achtereenvolgens, met tussen haakjes de waarde van *sdfbeta*: Maleisië (1,24), China (0,78), Vietnam (0,89) en de Filippijnen (1,09). Het effect van taalafstand in model 4 van tabel 2 is vrij van de invloed van die herkomstlanden.

De -2 Log likelihood van model 4 daalt significant ten aanzien van model 3. Nu blijkt dat de populatiegenetische bepaalde taalafstandsschaal een significante invloed uitoefent op het niveau van lezen onder migrantenleerlingen, en dat de desbetreffende parameter het verwachte negatieve teken heeft: hoe groter de afstand tussen de herkomsttaal en de bestemmingstaal, des te lager het leesniveau. Hypothese (5): *hoe groter de afstand tussen de herkomsttaal en de bestemmingstaal, des te lager het niveau van leesvaardigheid*, wordt geverifieerd, althans voor migranten uit landen buiten Oost-Azië naar Germaanstalige landen. Dit resultaat houdt in dat de taalkundige meting voor de verwantschap van wereldtalen onvoldoende gedetailleerd is. Uit de bevindingen blijkt verder dat de leerlingen uit de Oost-Aziatische landen allen een hogere leesscore behalen dan de gemiddelde migrantenleerling. Leerlingen uit Maleisië scoren het hoogst, die uit de Filippijnen het laagst. In de discussie opperen we verklaringen voor de uitzonderlijke Oost-Aziatische landen.

Tabel 2: Cross-classified multiniveaumodellen voor leesvaardigheid onder migranten van de eerste generatie in 13 Germaanstalige bestemming-landen

	Model 1		Model 2		Model 3		Model 4	
	B	S.E.	B	S.E.	B	S.E.	B	S.E.
Taalkundige meting								
<i>Germaanse subgroep (ref.)</i>	-	-	-	-	-	-	-	-
Indo-Europese taalfamilie			-15,78*	7,264	-15,139*	7,349	-16,323*	7,259
Niet Indo-Europese taalfamilie			-25,512*	10,671	-27,549*	11,213	-23,771*	11,221
Populatiegenetische meting								
Taalafstandschaal					5,051	8,128	-33,078**	10,322
Individuele karakteristieken								
Leeftijd bij aankomst in bestemmingsland			-4,764**	0,592	-4,760**	0,592	-4,772**	0,589
Cultureel kapitaal van de leerling			14,327**	2,049	14,341**	2,049	14,404**	2,045
Houding t.o.v. vrienden maken in de klas			-6,247**	1,942	-6,283**	1,942	-6,043**	1,938
Herkomsteffecten								
Moderniseringsgraad van het herkomstland			-10,131	8,329	-9,549	8,319	-8,938	6,867
Germaanstalig koloniaal verleden								
<i>Germaanstalig herkomstland (ref.)</i>	-	-	-	-	-	-	-	-
Niet Germaanstalig gekolonialiseerd			-41,736*	19,046	-46,013*	20,128	-36,096*	16,023
Wel Germaanstalig gekolonialiseerd			-57,860*	28,018	-61,538*	28,789	-39,114	24,828
China							765,519**	68,907
Maleisië							803,262**	82,238
Filippijnen							689,725**	72,745
Vietnam							753,194**	70,122

Tabel 2: Cross-classified multiniveaumodellen voor lesvaardigheid onder migranten van de eerste generatie in 13 Germaanstalige bestemmingslanden (vervolg)

	<i>Model 1</i>		<i>Model 2</i>		<i>Model 3</i>		<i>Model 4</i>	
	B	S.E.	B	S.E.	B	S.E.	B	S.E.
Controlevariabelen								
Geslacht (vrouw = ref.)			-29,644**	3,900	-29,658**	3,901	-29,490**	3,892
Opleiding van de vader			5,079**	1,369	5,089**	1,369	5,114**	1,367
Opleiding van de moeder			4,596**	1,437	4,602**	1,437	4,636**	1,430
Constante	478,091**	11,277	629,959**	83,470	627,643**	83,074	599,894**	68,306
Variantie tussen individuen	7421,806**	254,109	6381,966**	218,627	6384,043**	218,765	6373,948**	217,879
Variantie tussen herkomstlanden	2400,260**	714,875	1481,484**	478,019	1464,108**	473,369	674,761*	252,254
Variantie tussen bestemmingslanden	548,885	345,186	392,888	274,966	364,909	265,221	371,793	281,578
-2 Log likelihood	20768,419		20491,379		20491,006		20461,031	
Vershil -2 Log likelihood (-2)			277,040**		0,373		29,975*	

$N=1759$; * = $p<0,05$; ** = $p<0,01$.

Bron: eigen bewerkingen OECD/PISA, 2003

Hypothese (1), *hoe hoger het opleidingspeil van het herkomstland, des te beter lezen de migranten*, wordt door de modellen 3 en 4 in tabel 2 weersproken. Het effect van de moderniteit blijkt niet significant. Dat blijft het ook in een model met de bij deze variabele mogelijk invloedrijke gevallen, dat we nu niet weergeven.

Volgens de modellen 3 en 4 in tabel 2 blijkt het koloniale verleden van een herkomstland wel effect te hebben op de leesvaardigheid van migranten. Hypothese (2), *volgens welke migranten uit herkomstlanden zonder een Germaanstalig koloniaal verleden minder goed lezen dan migranten uit herkomstlanden mét een Germaanstalig koloniaal verleden*, wordt bevestigd. Opvallend is verder dat leerlingen uit gebieden die gekolonialiseerd zijn door een Germaanstalige mogendheid in leesvaardigheid niet significant afwijken van leerlingen uit Germaanstalige herkomstlanden (de referentiecategorie). De laatste verwachting van dit artikel was dat migranten met veel vrienden in de klas beter zouden lezen. Deze gedachte wordt niet ondersteund door de resultaten.

5. Conclusie en discussie

Deze studie handelde over leerlingen uit de eerste generatie migranten in Germaanstalige bestemmingslanden en de invloed van de afstand tussen hun herkomst- en bestemmingstaal op hun vaardigheid om de bestemmingstaal te lezen. Van Tubergen en Kalmijn (2005) stellen dat grotere verschillen tussen herkomst- en bestemmingstaal een geringere beheersing van de bestemmingstaal tot gevolg hebben. Door leesvaardigheid volgens specifieke toetsen als indicator voor de beheersing van een bestemmingstaal te nemen, richtte onderhavig onderzoek zich op een nauwkeurig vastgesteld gevolg. Naast de klassieke, taalkundige, indeling van de wereldtalen, benutte dit onderzoek een schaal die stoelt op recente bevindingen uit de populatiegenetica (Cavalli-Sforza, Menozzi & Piazza, 1996). Het gebruik van beide schalen leidde tot twee opmerkelijke resultaten. Verwacht werd dat een grotere genetische taalafstand samengaat met geringere leesprestaties. Deze verwachting vond steun. Vier Oost-Aziatische landen – China, Maleisië, de Filipijnen en Vietnam – verstoorde dit verband echter. Leerlingen uit deze landen scoren buitengewoon hoog wat betreft leesvaardigheid.

Deze studie is niet de enige waarin migranten uit Aziatische landen uitzonderlijk goed op school blijken te presteren. Zo verscheen er in de Verenigde Staten, door de opvallend goede onderwijsresultaten van Aziatische leerlingen aldaar, een boek dat aan Europeanen ‘het geheim van het goede leren’ bij Aziaten onthult (Abboud & Kim, 2005). China vormt trouwens tevens een moeilijkheid voor de theorie dat economische welvaart tot democratische politiek leidt (Lynn & Vanhanen, 2002).

Eén verklaring voor het hoge leesniveau van Oost-Aziatische leerlingen grijpt terug op de redenen voor vertrek van hun ouders uit deze gebieden naar het Westen. In Oost-Azië is het Confucianisme de belangrijkste godsdienst, dat een betrekkelijk activistisch wereldbeeld bevat (Ultee, Arts & Flap, 1992). De aanhangers van deze godsdienst kunnen dus een hoge arbeidsmoraal hebben en een reden om te vertrekken naar landen waar deze moraal zich ontvouwen

kan \en hun nazaten zetten zich mogelijk meer in om hogere diploma's te behalen. Morrison (1990) vindt dat in de Verenigde Staten de druk om te presteren onder Aziatische ouders veel malen groter is dan onder Europese. Aziatische kinderen zouden hierdoor veel langer bezig zijn met hun huiswerk en veel minder met buitenschoolse activiteiten.

Bovenstaande verklaring doet een beroep op de motivatie van personen. Daarnaast is echter een verklaring mogelijk die gelegenheden in het spel brengt, in het bijzonder belemmeringen waarmee taalverschillen in meer of mindere mate gepaard gaan. We passen hier dus de algemene uitspraak toe dat het gedrag van mensen door motivaties en gelegenheden wordt bepaald. Een belemmeringenverklaring gaat uit van de complexiteit van het Chinese schrift. Naast de spreektaal is daar de schrijftaal. Het Chinese schrift kent ongeveer 5000 tekens en het Latijnse alfabet waarin de Germaanse talen worden geschreven maar 26. Daarnaast hebben Chinese tekens vaak betrekking op lettergrepen en woorden. Migranten die een complex schrijfsysteem beheersen, zouden een veel simpeler systeem met gemak kunnen hanteren.

Deze twee verklaringen kunnen worden gebruikt om de gevallen China én Maleisië te verklaren. De wetgeving van Maleisië probeert de Maleissprekende Islamitische bevolking op te trekken tot het niveau van de Chineestalige Confucianistische bevolking. Daartoe legt de Maleisische wet quota's vast, wat heeft geleid tot vertrek van Chinezen uit Maleisië. Hier doet met een verklaring op die ook van toepassing kan zijn op Vietnamezen en op Chinezen uit China: voor deze migranten zijn de omstandigheden in hun land van herkomst zodanig dat terugkeren niet meer wordt overwogen.

Wat betreft de Filippijno's als 'invloedrijk geval' is er een andere mogelijke verklaring. De mensen uit dit land zijn door ons geklasseerd als gewoond hebbend in een land met een kolonisator die geen Germaanse taal oplegde. Dat is gebeurd omdat Spanje vanaf het einde van de zestiende eeuw de Filippijnse eilanden in bezit heeft gehad. In 1898 vertrokken de Spanjaarden echter, en werden de Filippijnen *de facto* een kolonie van de Verenigde Staten. De codering van het koloniale verleden van de Filippijnen was dus mogelijk verkeerd. Iets dergelijks geldt wellicht ook voor de Vietnamezen. De lange Franse heerschappij aldaar werd in de jaren vijftig van de vorige eeuw beëindigd, maar in de jaren zestig en zeventig verbleven er veel Amerikaanse soldaten en andere Amerikanen.

Met het hier gebruikte bestand kan de verklaring voor de Filippijno's als 'invloedrijk geval' worden getoetst. Dat deden we door de code voor dit herkomstland van 2 in 3 te veranderen en model 4 opnieuw te schatten. De dummy voor de Filippijnen blijkt dan insignificant te zijn, de andere parameters van het model veranderen nauwelijks van waarde. Dit resultaat is bemoedigend. Helaas kunnen met het hier gebruikte bestand de andere geopperde verklaringen niet worden getoetst. Over het algemeen geldt echter dat de negatieve invloed van taalafstand op leesprestaties, bepaald aan de hand van de genetische distantie tussen de wereldvolkeren, niet wordt verworpen. Deze hypothese ging ook op toen de historisch-linguïstieke tezamen met de genetische maat voor taalafstand als voorspeller in de regressiemodellen werd opgenomen.

Naast de invloed van herkomsttaal op leesvaardigheid zijn in deze studie tevens andere herkomstfactoren bestudeerd. De mate van moderniteit van een herkomstland bleek geen gevolgen te hebben voor de onderwijsprestaties van de migrantenleerlingen. Tot deze conclusie kwa-

men van Tubergen en Kalmijn (2005) ook, toen ze de moderniteit van een herkomstland bepaalden aan de hand van het gemiddelde inkomen in een land. In onderhavige studie deed echter het gemiddelde opleidingspeil in een land dienst als maatstaf voor moderniteit. De resultaten wijzen uit dat het effect van opleiding op leesvaardigheid vooral voorkomt op individueel niveau, of liever gezegd gezinsniveau. Zowel migrantenleerlingen met een hoger opgeleide vader als migrantenleerlingen met een hoger opgeleide moeder lezen beter.

Het tweede bestudeerde kenmerk van een herkomstland is zijn koloniale verleden. Verwacht werd dat migranten uit landen die gekolonialiseerd werden door een Germaanstalige mogendheid meer te maken hebben gehad met de taal en de overige cultuur van een Germaanstalig bestemmingsland dan migranten uit landen met een anderstalige kolonisator of geen kolonisator. Mensen uit de voormalige koloniën van de Germaanstalige bestemmingslanden bleken inderdaad een hogere score voor leesvaardigheid te hebben. Dat is wellicht niet verwonderlijk. Zo zijn de banden tussen kolonisator en ex-kolonie, zoals Groot-Brittannië en India, in meerdere opzichten nog steeds hecht. Op de Nederlandse Antillen krijgen vandaag de dag kinderen op school les in het Nederlands. Oomens, Driessen en Scheepers (2003) concludeerden al dat in Nederland de opname van migranten uit de voormalige Nederlandse koloniën (zoals Suriname en de Nederlandse Antillen) voorspoediger verloopt dan die van andere migrantengroepen. Volgens onderhavige studie geldt deze bevinding mogelijk voor meer combinaties van herkomst- en bestemmingsland.

Opvallend was ten slotte dat migrantenleerlingen die zeiden dat op hun school gemakkelijk vrienden kunnen worden gemaakt, minder presteerden in lezen. Het idee dat contact met andere leerlingen een positieve uitwerking heeft op de onderwijsprestaties, gaat voorbij aan de vraag *met wie* de leerling is bevriend. Integratie in een vriendenkring die geen belang hecht aan presteren op school, kan leerlingen negatief beïnvloeden. Vervolgonderzoek naar de buitenschoolse activiteiten van migranten dient over die kwestie te gaan. Daarnaast is de vraag naar het effect van taalafstand op de resultaten bij andere vakken dan lezen, zoals het rekenen, een belangwekkende vervolgvraag. Rekenen mag abstracter zijn dan lezen, rekenen leren mensen steeds van een docent die de bestemmingstaal spreekt en uit een boek in die taal.

Ons vertrekpunt was Van Tubergen en Kalmijn (2005). We gingen niet in op hun hypothesen over de gevolgen van de politiek in het bestemmingsland van migranten. Dat lieten we achterwege wegens de hypothesen van De Swaan (2001) over de blijvende invloed van koloniale mogelijkheden in de herkomstlanden van migranten. Die hypothesen gaan ook over een politieke factor, en het leek ons minder gewenst – gezien het kleine aantal bestemmingslanden – ook een variabele voor de hedendaagse politiek in een bestemmingsland op te nemen. Tevens bestudeerden we niet, zoals Van Tubergen en Kalmijn (2005), de taalvaardigheid van alle migranten boven de achttien jaar. Deze heterogeniteit in leeftijd leidde bij hen overigens tot een aanvechtbaar vastpinnen van de meting voor de aard van de politiek in een land op één bepaald jaartal. Wij onderzochten uitsluitend de leesvaardigheid van migrantenleerlingen. De onderzoekpopulatie in onderhavige studie, 15- en 16-jarige leerlingen uit 2003, vormt echter een nuttig ijkpunt voor het effect van herkomsttaal op de beheersing van de bestemmingstaal. Nieuw onderzoek kan zich richten op leerlingen van dezelfde leeftijd uit eerdere of latere steekjaren.

Al met al roept dit onderzoek interessante inhoudelijke en methodologische vragen over migratie op. De vraag of bevindingen uit de populatiegenetica dienst kunnen doen om het verschijnsel taalafstand te meten, is met een voorzichtig ja beantwoord, maar vergt natuurlijk nader onderzoek. Dat taalafstand populatiegenetisch gemeten invloed heeft op leesprestaties bij migranten was misschien niet verbazingwekkend, wel dat taalafstand historisch-linguïstisch gemeten invloed blijft houden nadat rekening is gehouden met de genetische meting van taalafstand. Een verklaring voor deze bevinding hebben we niet, in elk geval hebben maten uit de historische linguïstiek geenszins afgedaan. Daarnaast rees de inhoudelijke kwestie hoe de hoge leesprestaties van kinderen uit Oost-Aziatische landen te verklaren vallen. Deze bevindingen zijn onverwacht gezien de hypothese dat een grote afstand tussen talen, historisch-linguïstisch dan wel populatiegenetisch gemeten, tot slechtere leesprestaties leidt. De voornaamste bevinding uit dit onderzoek vatten we, als voorlopig antwoord op de centrale vraagstelling van dit artikel, als volgt samen:

Onder migranten in Germaanstalige bestemmingslanden verklaren niet alleen metingen van taalafstand uit de historische linguïstiek verschillen tussen migranten in leesvaardigheid, nauwkeurigere metingen van taalafstand uit de populatiegenetica doen dat ook.

Noten

1. De eerste auteur bedankt in het bijzonder Mark Levels voor de verwijzing naar het PISA-bestand en voor andere hulp. Jan-Paul Kooistra behaalde de master Sociologie aan de Radboud Universiteit Nijmegen en volgt nu een opleiding Internationaal recht aan de Universiteit van Amsterdam. Wout Ultee is hoogleraar Sociologie aan de Radboud Universiteit en Ben Pelzer is docent bij Methoden van maatschappij-wetenschappelijk onderzoek aldaar. Correspondentieadres: w.ultee@maw.ru.nl.
2. Voor de bepaling van de *sdfbeta* gebruikten we binnen *spss* een zelfgeschreven syntax, die bij de auteurs kan worden opgevraagd.

Literatuur

- Abboud, S. K. & Kim, J. Y. (2005). *Top of the Class: How Asian Parents Raise High Achievers and How You Can Too*. New York: Berkeley Publishing Group.
- Anderson, B. (1990). *Language and Power*. Ithaca: Cornell University Press.
- Bourdieu, P. (1991). *Language and Symbolic Power*. Cambridge: Cambridge University Press.
- Browne, W. J., Goldstein, H. & Rasbash, J. (2001). 'Multiple Membership Multiple Classification (MMMC) Models'. *Statistical Modelling*, 1, 103-124.
- Cannon, G. H. (1993). *The Collected Works of Sir William Jones*. New York: NYU Press.
- Cavalli-Sforza L., Menozzi, P. & Piazza, A. (1996). *The History and Geography of Human Genes*. Abridged Paperback Edition. Princeton: Princeton University Press.
- Encyclopædia Britannica (1992). *1992 Two Britannica Book of the Year*. Chicago: Encyclopædia Britannica Inc.

- Greenberg J. H. (2000). *Indo-European and Its Closest Relatives*. Stanford: Stanford University Press.
- Groeneboer, K. (1998). *Gateway to the West*. Amsterdam: Amsterdam University Press.
- Hübschmann, H. (1897). *Armenische Grammatik*. Leipzig: Breitkopf & Härtel.
- Langford, I. H. & Lewis, T. (1998). 'Outliers in Multilevel Data'. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 161, 121-160.
- Lynn, R. & Vanhanen, T. (2002). *IQ and the Wealth of Nations*. Westport: Praeger.
- Morrison, G. W. (1990). 'The Education of White, Chinese, Filipino, and Japanese Students: A Look at "High School and Beyond"'. *Sociological Perspectives*, 3, 355-374.
- OECD/PISA (2003). Pisa-codebook 2003/Pisa 2003 SPSS data. http://pisaweb.acer.edu.au/oecd_2003/oecd_pisa_data_s1.html. Bezocht 01-11-2006.
- Oomens, S., Driessen, G. & Scheepers, P. (2003). 'De Integratie van allochtone ouders en onderwijsprestaties van hun kinderen'. *Tijdschrift voor Sociologie*, 24, 289-312.
- Ruhlen, M. (1987). *A Guide to the World's Languages*. Stanford: Stanford University Press.
- Snijders, T. & Bosker, R. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling*. London: SAGE Publications.
- Störrig, H. J. (1988). *Taal - Het grote avontuur*. Utrecht: Het Spectrum.
- Swaan, A. de (2001). *Woorden van de wereld*. Amsterdam: Prometheus.
- Tubergen, F. van & Kalmijn, M. (2005). 'Destination-language Proficiency in Cross-National Perspective: A Study of Immigrant Groups in Nine Western Countries'. *American Journal of Sociology*, 110, 1412-1457.
- Ultee, W., Arts, W. & Flap, H. (1992). *Sociologie: vragen, uitspraken, bevindingen*. Groningen: Wolters-Noordhoff.
- United Nations Development Programme (2005). *Human Development Report 2005*. New York: Oxford University Press. Voor de berekening van het opleidingspeil zie: *Technical Note 1: Calculating the Human Development Indices*. http://hdr.undp.org/docs/statistics/indices/technote_1.pdf Bezocht op 22-10-2006.