



Directoraat-Generaal Internationale Aangelegenheden en Vreemdelingenzaken

Postadres: Postbus 20301, 2500 EH Den Haag

Aan de Voorzitter van de Tweede Kamer der Staten-Generaal
Postbus 20018
2500 EA Den Haag

Bezoekadres

Juliana van Stolberglaan 10
2595 CL Den Haag
Telefoon (070) 3 70 91 57
Fax (070) 3 70 62 74
www.justitie.nl

Onderdeel	Projectdirectie Modernisering Inburgering
Datum	11 november 2005
Ons kenmerk	5381314/05
Onderwerp	Wetenschappelijke rapportages examen buitenland

Bij beantwoording de datum en ons kenmerk vermelden. Wilt u slechts één zaak in uw brief behandelen.

1. Inleiding

Hierbij zend ik u, zoals ik u in april jl. heb toegezegd, de wetenschappelijke verantwoording over de totstandkoming en de kwaliteit van de Toets Gesproken Nederlands (TGN) en van de toets Kennis van de Nederlandse Samenleving (KNS). Deze verantwoordingen zijn vervaardigd door het consortium (CINOP, LTS, Ordinate) dat de toetsen heeft ontworpen. Beide toetsen vormen tezamen het inburgeringsexamen in het buitenland.

In het hierna volgende informeer ik u ook over de oordelen over deze rapportages van de leden van de door CINOP ingestelde begeleidingsgroepen voor beide toetsen. Daarnaast stuur ik u een rapport van TNO, die op mijn verzoek een "second opinion" over deze toetsen heeft vervaardigd. Beide rapporten zijn voorzien van een samenvatting, waarnaar ik u hier gaarne korthedshalve verwijs. Tenslotte wil ik u informeren over de conclusies die ik wil verbinden aan deze rapporten en de commentaren daarop voor de invoering van het inburgeringsexamen in het buitenland. Daarmee geef ik ook antwoord op de vragen van de Algemene Commissie voor het Integratiebeleid, die zijn gesteld in haar brief van 29 september jl.

2. De totstandkoming van het examen buitenland

Bij de ontwikkeling van het inburgeringsexamen in het buitenland is stap-voor-stap gewerkt aan de bouw van een kader en instrumentarium voor een zorgvuldige afname van het examen. Over het juridische kader voor het inburgeringsexamen in het buitenland heb ik mij laten adviseren door de Adviescommissie voor Vreemdelingenzaken¹. Voor de inhoud en normering van de examens heb ik

¹ ACVZ, Advies over inburgering als vereiste voor toelating tot Nederland en voor de vergunning voor

mij laten adviseren door de Tijdelijke Adviescommissie Normering Inburgeringsexamens onder leiding van de heer Franssen. Het advies van de Commissie Franssen heeft mij richting gegeven voor het vaststellen van niveaus van kennis en vaardigheden². Ik heb de Tweede Kamer op 30 september 2004 in een technische briefing op de hoogte gesteld van de hoofdlijnen van de invulling van het examen buitenland. Vervolgens is er bij de operationalisering van de door de Commissie Franssen voorgestelde normering van de eisen ten aanzien van de Kennis van de Nederlandse Samenleving (KNS) gesproken met zowel autochtone als allochtone Nederlanders. Dit heeft geresulteerd in het examenprogramma op video "Naar Nederland", waarvan ik op 3 februari 2005 een verkorte versie aan leden van de Tweede Kamer heb gepresenteerd. Bij die gelegenheid heb ik deze leden ook in staat gesteld om zelf kennis te maken met de Toets Gesproken Nederlands.

Het CINOP-consortium dat de instrumenten voor het examen buitenland heeft ontwikkeld, heeft om de kwaliteit van het ontwikkelproces te bevorderen, ervoor gekozen twee groepen van onafhankelijke deskundigen in te stellen. Deze "resonansgroepen" zijn ingesteld als interne adviesgroepen voor het consortium, en dragen als zodanig geen verantwoordelijkheid voor het eindproduct van de toetsconstructeurs. Deze wetenschappers hebben zich respectievelijk gebogen over het ontwikkelproces en de kwaliteit van de TGN en van de toets KNS.

Als kwaliteitsborging voor hun producten hanteren toetsontwikkelaars en onderwijsdeskundigen veelal codes voor goed gedrag. LTS, één van de drie participanten in het consortium, is lid van een aantal vooraanstaande (internationale) organisaties, die "Codes of Good Practice" hanteren. Ook de andere partners in het consortium leven deze codes na. Dat draagt bij aan het vertrouwen in de kwaliteit van de beide ontwikkelde toetsen.

Als een extra borg op de kwaliteit van de examens heb ik TNO gevraagd om een "second opinion" ten aanzien van de gebruikte technologie in relatie tot het doel te vervaardigen. De belangrijkste uitkomsten van deze onderzoeken treft u in de hierna volgende paragrafen aan.

3. De verantwoording van CINOP van de Toets Gesproken Nederlands

3.1 De kwaliteit van de Toets Gesproken Nederlands (TGN)

CINOP heeft ten behoeve van zijn rapportage een aantal onderzoeken uitgevoerd, waarmee de kwaliteit van het instrument beoordeeld kan worden. Uit deze pretests is gebleken dat de betrouwbaarheid van de toets voor het gehele niveaubereik van de toets vergelijkbaar is met die van toetsen die thans in het NT2-onderwijs gebruikt worden. De betrouwbaarheid komt volgens de onderzoekers overeen met die van toetsen die in de Angelsaksische landen ingezet worden als onderdeel van toelatingsprocedures (TSE en TOEFL, beide van ETS, en de IELTS van Cambridge University).

Een belangrijke vraag is of de toets ook op het niveau A1min een voldoende mate van

onbepaalde tijd, 2004.

² Adviescommissie Normering Inburgeringseisen (2004) Inburgering getoetst. Advies over het inburgeringsexamen in het buitenland.

betrouwbaarheid heeft. Daar wordt immers de zak-/slaag-beslissing voor het examen buitenland genomen. Mede om deze vraag te beantwoorden zijn twee aanvullende onderzoeken uitgevoerd onder circa 800 personen. Daarbij is ook gekeken naar de zogenoemde meetfout, een maat om vast te stellen in hoeverre men op de toetsresultaten staat kan maken. Uit deze analyses blijkt volgens CINOP, dat de TGN op de lagere niveaus, waaronder A1min en A2, preciezer meet dan op de hogere niveaus.

Een toets moet geen andere dingen meten dan dat wat men wil beoordelen. Een taaltoets moet dus taalvaardigheid meten en geen kennis van andere zaken. Om dit aspect van de kwaliteit van het taalexamen Nederlands te meten is daarom bestudeerd of er geen andere variabelen zijn die met de toets worden gemeten dan de spreek- en luistervaardigheid in het Nederlands. Dit is onderzocht voor de variabelen leeftijd, geslacht, land van herkomst, verblijfsduur in Nederland, geletterdheid, hoogstgenoten opleiding, en thuistaal. Geen van deze variabelen blijkt de toetsresultaten te beïnvloeden, met uitzondering van de verblijfsduur in Nederland. CINOP concludeert op grond van deze gegevens dat mag worden aangenomen dat de toets meet wat gemeten moet worden.

De kwaliteit van de TGN is verder onderzocht door de uitkomsten van de TGN te vergelijken met de oordelen van mensen. Uit het rapport blijkt dat de correlaties tussen het oordeel van de spraakherkenner en de oordelen door mensen positief en zeer hoog zijn. Naarmate de beoordelaars beter getraind zijn om deze toets af te nemen, wordt de overeenstemming tussen menselijke oordelen en de computer hoger.

De kwaliteit van de beslissingen is ook onderzocht door vergelijkingen te maken tussen zak- en slaagbeslissingen die gebaseerd zijn op de Toets Gesproken Nederlands en zak- en slaagbeslissingen die genomen zouden worden op basis van de oordelen van getrainde beoordelaars.

CINOP concludeert op grond van zijn onderzoek dat het automatische spraakherkennings- en scoringssysteem van de TGN voldoende precies functioneert om oordelen te geven die vergelijkbaar zijn met die van goed getrainde beoordelaars.

3.2 Het oordeel van de resonansgroep over de TGN

Deze resonansgroep bestond uit een viertal onafhankelijke deskundigen met een wetenschappelijke achtergrond in de disciplines taalkunde, methodologie, statistiek en spraakherkenning. De resonansgroep heeft belangrijke bijdragen geleverd tijdens het traject van ontwikkeling en verantwoording van de TGN, zoals ook blijkt uit de eindverantwoording. Op advies van de resonansgroep is een aantal aanvullende studies verricht en zijn additionele gegevens verzameld. Eveneens op hun voorstel zijn er tijdens verschillende fasen van het onderzoek, èn bij de vaststelling van de kwaliteit van de toets aanvullende analyses uitgevoerd. De resonansgroep heeft tot slot substantiële bijdragen geleverd aan het tot stand komen van de eindverantwoording.

De leden van de resonansgroep menen over het eindresultaat dat "de rapportage over de valideringsstudie van de Toets Gesproken Nederlands onvoldoende aannemelijk maakt dat de TGN

bruikbaar is. Door gebrekkige dataverzameling, gebrekkige analyses en veel onduidelijkheden in de rapportage meent de resonansgroep dat de voorliggende stukken de kwaliteit van de toets niet aantonen. Anders gezegd, de rapportage geeft onvoldoende garantie dat de toets adequaat werkt voor het doel waarvoor de toets is ontwikkeld. Indien de Minister besluit de toets in te voeren, adviseren wij daaraan voorafgaande een onderzoek uit te voeren naar de resultaten die de TGN in de praktijk oplevert, onder meer ter beantwoording van de vraag in welke mate de afstegrens op de toets daadwerkelijk op het beoogde niveau ligt".

CINOP c.s. blijft ook na dit commentaar van de resonansgroep overtuigd van de kwaliteit van het geleverde werk. Zij menen dat er in Nederland geen enkele NT2-toets of NT2-examen bestaat waarover vóór de ingebruikname meer onderzoek is gedaan dan de Toets Gesproken Nederlands. De toetsconstructeurs sluiten zich ten dele aan bij het advies van de leden van de resonansgroep om *voorafgaand* aan de invoering van de toets een onderzoek uit te voeren naar de resultaten van de TGN in de praktijk. Het echte functioneren van de toets kan alleen *in de praktijk* plaatsvinden, dat wil zeggen in het buitenland en onder reële en echte examencondities. Om definitief vast te stellen hoe de toets in de praktijk functioneert en om vast te stellen in welke mate de afstegrens op het beoogde niveau ligt, moet gestart worden met de afname in het buitenland. Een echte examensituatie is niet te simuleren met vrijwilligers of met een situatie waar iedere deelnemer slaagt voor de toets, ongeacht het resultaat. Het moet een afname zijn bij echte examenkandidaten, die een echt examen afleggen onder reguliere examencondities.

3.3 *De second opinion van TNO*

TNO kijkt door een andere bril naar het onderzoeksmateriaal dan de toets-ontwikkelaars. De laatsten zoeken in hun onderzoek naar zoveel mogelijk aanwijzingen voor de hypothese dat de toetsen op een kwalitatief voldoende wijze de Nederlandse taalvaardigheid en de kennis van de Nederlandse samenleving meten. De onderzoekers van TNO bezien het toetsstelsel door een meer natuurwetenschappelijke en wiskundige bril en stellen een hard criterium aan de kwaliteit van de oordelen door de machine in relatie tot die van menselijke beoordelaars.

Op basis van de studie van CINOP heeft TNO de indruk dat de TGN voldoende consistente oordelen geeft. Ook constateert TNO dat kandidaten met een sterk buitenlands accent door de spraakherkenner niet worden benadeeld met een lagere deelscore. Gebleken is verder dat het trainingsmateriaal waarmee de spraakherkenner is getraind, voor de meest frequente taalachtergronden representatief is voor de praktijk. Men heeft noch bewijs gevonden dat de toets voldoende precies meet wat de toets moet meten, noch dat de toets dit in ónvoldoende mate doet. TNO vond aanwijzingen dat de toets op een goede manier taalvaardigheid meet, maar meent dat hiermee niet is gegarandeerd dat de toets dit op alle niveaus voldoende adequaat doet.

Ook vond men aanwijzingen dat de toets minder precies zou meten rond de A1min cesuur. Aangezien een goede "human benchmark" ontbreekt, achten de onderzoekers het niet mogelijk om te concluderen of de fouten die rond de A1min grens voor slagen of zakken worden gemaakt, acceptabel zijn.

4. **Het inburgeringsexamen Kennis van de Nederlandse Samenleving**

4.1 *De kwaliteit van de toets KNS*

Ook bij de beoordeling van de kwaliteit van de toets KNS gaat het, hoewel het hier om een ander type toets gaat, uiteraard om de kwaliteit van de oordelen. Als maat voor de kwaliteit van het machinale oordeel hanteert CINOP wederom de overeenkomst tussen de uitkomst gebaseerd op menselijke beoordeling, en de score die door de machine wordt gegenereerd. Ook hier blijkt de overeenstemming tussen menselijke beoordelaars en de spraakherkenner over de gehele set van opgaven zeer hoog te zijn.

Uit de gehele set opgaven zijn tien boekjes samengesteld met elk 30 overlappende opgaven. Iedere kandidaat krijgt één van deze boekjes voorgelegd op het examen. De kwaliteit van de tien toetsen voldoet volgens CINOP aan de vooraf gestelde eis op dit punt.

Aanwijzingen voor het antwoord op de vraag of de KNS-toets meet wat deze moet meten zijn onderzocht door de relaties te analyseren tussen de toetsuitslagen en enkele achtergrondgegevens van de deelnemers aan de pretests. Uit deze analyses blijkt dat géén van de achtergrondvariabelen als sexe, leeftijd, taalvaardigheid, opleiding, duur van het verblijf in Nederland een rol spelen bij de behaalde uitslagen op de toets. Gelet op het feit dat de toets moet meten of iemand zich voldoende relevante informatie over de Nederlandse samenleving heeft verworven, en niet of iemand kan lezen of schrijven, is geanalyseerd of geletterdheid al dan niet een rol speelt bij de examinering. De onderzoekers constateren dat geletterden gemiddeld wel een duidelijk hogere uitslag behalen dan niet-geletterden, maar dat de laatsten ook wel degelijk kans maken een hoge uitslag op de toets te bereiken.

4.2 *Het oordeel van de resonansgroep KNS*

Deze resonansgroep bestond uit vijf leden; twee ervaringsdeskundigen uit de doelgroep, een NT2-deskundige en twee deskundigen op het gebied van alfabetisering. De leden van de resonansgroep KNS hebben vastgesteld dat het examen toetst of kandidaten kennis hebben genomen van de informatie die wordt gepresenteerd in het examenmateriaal in de vorm van de film "Naar Nederland" en of en in hoeverre kandidaten van deze kennis blij kunnen geven. Succesvolle kandidaten hebben, volgens de resonansgroep laten zien dat zij voldoende kennis hebben genomen van het voorbereidingsmateriaal, dat zij gemotiveerd zijn en dat zij voldoende inspanningen hebben geleverd ter voorbereiding op hun komst naar Nederland.

De resultaten uit de pretestgegevens bieden de resonansgroep voldoende vertrouwen ten aanzien van de kwaliteit van de toets en van de werking van de spraakherkenner. Of kandidaten ook daadwerkelijk de inhoudelijke betekenis van alle vragen begrijpen, is volgens de leden van de resonansgroep niet vast te stellen, aangezien het examen uitvoerbaar moet zijn voor kandidaten met een taalvaardigheid op A1min-niveau. Zij wijzen er bovendien op dat de condities en de doelgroep in het buitenland niet geheel vergelijkbaar zijn met de condities en de doelgroep van het onderzoek. Daarom beveelt de resonansgroep aan om het vereiste niveau niet hoger te leggen dan 70% goede antwoorden en deze norm bij de evaluatie van het examen te betrekken.

Eén lid van de resonansgroep kon zich niet vinden in dit advies en heeft aan de andere leden laten

weten dat zij zich terugtrekt met als motief dat haar

“..... bezorgdheid over de doelmatigheid en kwaliteit van de toets onvoldoende tot uitdrukking komt, en het gegeven dat niet uitgesloten kan worden dat een deel van de doelgroep (de analfabeten en de laaggeletterden) een lagere kans heeft dan anderen om voor de toets te slagen”.

4.3. *De second opinion van TNO over de toets KNS*

TNO concludeert op basis van statistische analyse van een beperkte testset dat de toetsuitslag van naar schatting 10 à 15% van de kandidaten niet terecht is, en zowel naar boven als naar beneden kan afwijken. Voor een nauwkeuriger schatting van het aantal foute beslissingen zijn gegevens uit een een grotere steekproef nodig. Omdat gegevens van vergelijkbare menselijke beoordelaars ontbreken en een kwaliteitsnorm niet gesteld is, kunnen de onderzoekers niet concluderen dat 10 à 15% (te) veel of (te) weinig is.

5. **Aanbevelingen TNO**

TNO is van mening dat er wel degelijk mogelijkheden zijn om zowel de toets, de instelling van de beide cesuren, als de wetenschappelijke onderbouwing van beide toetsen verder te verbeteren, en dat dit het beste kan geschieden door het examen in de praktijk in te voeren. TNO adviseert, net als CINOP om over te gaan tot invoering van de toetsen en parallel daaraan de volgende stappen te ondernemen:

1. Verzamel voor de TGN toetsgegevens uit de praktijk
2. Formuleer kwaliteitsnormen voor de toets TGN en de toets KNS door vast te leggen in welke mate de fouten die de automatische toets maakt, mag afwijken van de mate waarin menselijke examinatoren fouten zouden maken. Hierbij wordt de automatische toets als alternatief beschouwd voor beoordeling door mensen.
3. Controleer met de toetsgegevens uit de praktijk hoeveel onterechte beslissingen de automatische toets maakt en hoeveel onterechte beslissingen menselijke examinatoren zouden maken. Op basis van het verschil in percentages onterechte beslissingen wordt bepaald in hoeverre aan de kwaliteitsnorm wordt voldaan.
4. Verbeter de kwaliteit van de TGN rond de A1 min cesuur.

6. **Overwegingen en conclusies voor de invoering van het inburgeringsexamen in het buitenland**

Het veranderen van systemen, en het invoeren van nieuwe maatregelen brengt altijd een zekere mate van onzekerheid mee. In dit geval – het gaat hier immers om een voor Nederland nieuwe techniek - betreft dit zowel onzekerheden voor degenen die het inburgeringsexamen moeten afleggen, als onzekerheden voor de overheid over de mate van correctheid van de uitslagen die het beoordelingssysteem geeft. Aan die onzekerheid valt nooit helemaal te ontkomen, maar ze valt wel zoveel mogelijk te verminderen.

Zoals ik in de Kamer al heb gezegd, staat voor mij de zorgvuldigheid bij dit ontwikkelings- en invoeringsproces voorop. Daarom heb ik bij de behandeling van het wetsvoorstel gezegd kinderziekten bij de uitvoering van het inburgeringsexamen in het buitenland weliswaar niet volledig te kunnen voorkomen, maar er wel voor te kunnen zorgen dat kandidaten hiervan geen nadelige gevolgen zullen ondervinden. Daarom mag een kandidaat, als zich technische problemen voordoen die niet aan hem of haar kunnen worden verweten, kosteloos het examen nogmaals afleggen. Bij die zorgvuldigheid past ook dat de prestaties van het examensysteem voortdurend nauwlettend worden gevolgd en beoordeeld, waarop ook door de resonansgroep voor de TGN van Cinop is aangedrongen. Dat zal dan ook gebeuren: met behulp van een nauwkeurige monitor zal het uitvoeringsproces op een fors aantal relevante aspecten worden gevolgd. Een begeleidingscommissie van onafhankelijke personen zal mij adviseren over de wijze waarop kinderziekten of andere problemen die met behulp van de monitor worden geconstateerd, zo snel mogelijk kunnen worden opgelost.

De grote zorgvuldigheid die ik bij de invoering van dit examen noodzakelijk vind, de uitkomsten van de onderzoeken én het commentaar van de resonansgroepen brengen mij tot de keus om het voorstel van TNO te volgen. Ik streef er naar om op 1 januari 2006 te starten met de invoering van het inburgeringsexamen in het buitenland. De invoering zal worden begeleid door wetenschappelijk onderzoek naar de werking van de spraakherkenner in de praktijk en van de kwaliteit van de beoordeling door middel van deze geautomatiseerde examens. Tijdens dit onderzoek zal een herbeoordeling plaatsvinden van het oordeel van de computer door meerdere beoordelaars van alle uitslagen die zich in een bepaalde marge rond de A1min norm bevinden. Die marge zal in overleg met de onderzoekers worden bepaald. Deze herbeoordeling moet ertoe leiden dat de nog bestaande onzekerheid over de examens en van de machinale beoordeling van de examens definitief wordt weggenomen. Op basis van de gegevens die tijdens de uitvoering van dit onderzoek worden verzameld, zal met name de zak/slaaggrens zo nauwkeurig mogelijk worden bepaald.

De facto betekent dit dat het geautomatiseerde examensysteem daadwerkelijk in de praktijk verder zal worden ontwikkeld. Het betekent uiteraard ook dat gedurende de looptijd van het onderzoek niemand, die op het examen rond de A1min norm presteert en bij de menselijke herbeoordeling een andere – negatieve of positieve - uitslag krijgt, gevolgen zal ondervinden van het feit dat het geautomatiseerde examen nog onvoldoende heeft kunnen aantonen dat het even goed of vrijwel even goed als menselijke beoordelaars beslissingen kan nemen. Uiteraard zal dit ook leiden tot een aanpassing van de Algemene Maatregel van Bestuur (Kamerstukken II, 2004/05, 29700, nr. 8), die ik bij brief van 15 februari aan de Tweede Kamer heb aangeboden. Over deze aanpassing zal ik u zo spoedig mogelijk informeren.

De Minister voor Vreemdelingenzaken en Integratie,

