

Vergaderjaar 2005–2006

29 700

Wijziging van de Vreemdelingenwet 2000 in verband met het stellen van een inburgeringsvereiste bij het toelaten van bepaalde categorieën vreemdelingen (Wet inburgering in het buitenland)

Nr. 33

LIJST VAN VRAGEN EN ANTWOORDEN

Vastgesteld 17 januari 2006

De algemene commissie voor het Integratiebeleid¹ heeft over de brief van de minister voor Vreemdelingenzaken en Integratiebeleid d.d. 11 november 2005 over de Toets Gesproken Nederlands (TGN) en de toets Kennis van de Nederlandse Samenleving (KNS) (29 700, nr. 30) een aantal vragen ter beantwoording aan de minister voorgelegd. Bij brief van 16 januari 2006 heeft de minister de vragen beantwoord. Vragen en antwoorden zijn hieronder afgedrukt.

De voorzitter van de commissie,
Kalsbeek

De waarnemend griffier van de commissie,
Coenen

¹ Samenstelling:

Leden: De Vries (PvdA), Kalsbeek (PvdA), voorzitter, Vos (GL), Hofstra (VVD), Lambrechts (D66), Hamer (PvdA), Arib (PvdA), Bussemaker (PvdA), Kant (SP), Wilders (Groep Wilders), Örgü (VVD), Balemans (VVD), Dijsselbloem (PvdA), Çörüz (CDA), ondervoorzitter, Depla (PvdA), Vergeer (SP), Lazrak (Groep Lazrak), Ferrier (CDA), Huizinga-Heringa (CU), Van Haersma Buma (CDA), Lid, Sterk (CDA), Varela (LPF), Algra (CDA), Eski (CDA), Nawijn (Groep Nawijn), Hirsi Ali (VVD), Visser (VVD), Azough (GL) en Jonker (CDA).

Plv. Leden: Stuurman (PvdA), Tjon-A-Ten (PvdA), Van Gent (GL), Luchtveld (VVD), Dittrich (D66), Leerdam, MFA (PvdA), Wolfsen (PvdA), Van Heemst (PvdA), Gerkens (SP), Van Miltenburg (VVD), Van der Sande (VVD), Albayrak (PvdA), Koopmans (CDA), Eijssink (PvdA), Van Velzen (SP), De Vries (CDA), Van der Staaij (SGP), Mastwijk (CDA), Van Bochove (CDA), Kraneveldt (LPF), Van de Camp (CDA), Rambocus (CDA), Eerdmans (LPF), Blok (VVD), Weekers (VVD), Halsema (GL) en Vacature (algemeen).

1

De resonansgroep adviseert de minister voorafgaand aan de invoering een onderzoek uit te voeren naar de resultaten die de Toets Gesproken Nederlands (TGN) in de praktijk levert. TNO adviseert de toetsen in te voeren en aan de hand van de praktijk te verbeteren. Acht TNO het onmogelijk om de toets te verbeteren aan de hand van onderzoek buiten de daadwerkelijke examensituatie? Hoe acht de resonansgroep dit mogelijk?

TNO acht het niet onmogelijk om de toets te verbeteren aan de hand van een onderzoek dat voorafgaat aan de invoering van de toets, maar vindt het beter om de toets te evalueren en zo mogelijk te verbeteren in de daadwerkelijke examensituatie. Daarvoor geeft TNO drie redenen: In de eerste plaats is de taalvaardigheid van de kandidaten van invloed op het aantal fouten dat de toets maakt. Op grond van de gegevens verzameld in de «laboratorium»-onderzoeken, vermoedt TNO dat de toets rond de A1-min cesuur waarschijnlijk meer fouten zal maken dan ruim boven of ruim onder de A1-min cesuur. Juist op basis van een steekproef met de juiste samenstelling in taalvaardigheidsniveaus van werkelijke examen-kandidaten kan hierover een uitspraak gedaan worden. Ten tweede kan de instelling van de zak/slaag-grens het best worden geverifieerd, dan wel geoptimaliseerd op basis van praktijkdata, omdat de praktijksituatie altijd afwijkt van de laboratoriumsituatie. Zo zullen kandidaten in de praktijk waarschijnlijk beter gemotiveerd zijn en hierdoor beter presteren dan in een onderzoek waarbij het behalen van een voldoende uitslag voor hen geen betekenis heeft. En in de derde plaats is het zo dat alleen beproeving in de praktijk de meeste zekerheid geeft tot het snel opsporen en corrigeren van eventuele afwijkingen tussen laboratorium- en praktijksituatie. Ik concludeer dat TNO van opvatting is dat een onderzoek dat wordt uitgevoerd vóór de invoering van de toetsen vragen onbeantwoord zal laten, omdat dat onderzoek niet in reële examencondities zal kunnen worden uitgevoerd. CINOP heeft dezelfde opvatting. In de brief van 10 november 2005 van CINOP is wel aangegeven dat de resonansgroep adviseert «om voorafgaand aan de invoering een onderzoek uit te voeren naar de resultaten die de TGN in de praktijk oplevert, ondermeer ter beantwoording van de vraag in welke mate de afstestgrens op de toets daadwerkelijk op het beoogde niveau ligt».

2

Kan de conclusie worden getrokken dat de spraakherkenner als examineringsstelsel een objectiever meetinstrument is dan wanneer gebruik wordt gemaakt van een menselijke beoordelaar?

Ja, het stelsel met de spraakherkenner is zowel meer consistent als meer objectief. Het is meer consistent omdat, een tweede, derde, en volgende beoordeling van hetzelfde antwoord van dezelfde kandidaat door de spraakherkenner altijd op dezelfde manier zal worden beoordeeld. Dit is bij menselijke beoordeling niet het geval. Mensen kunnen vermoeid raken of kunnen even niet opletten. Daarnaast is het stelsel meer objectief omdat een spraakherkenner niet wordt beïnvloed door subjectieve factoren, zoals persoonlijke voorkeur en non-verbale communicatie en emoties.

3

Wat is de juridische houdbaarheid van de TGN nu de resonansgroep kritiek heeft geuit op de wijze waarop de toets wordt afgenomen? In hoeverre kan de kritiek van de resonans-groep door een geëxamineerde worden aangegrepen om de positieve uitslag van een test te betwisten teneinde een hogere niveauvaststelling te verkrijgen?

Het basisexamen inburgering, waarvan de TGN deel uitmaakt, heeft slechts de beoordeling van de vraag of de potentiële nieuwkomer beschikt over basiskennis van de Nederlandse taal en de Nederlandse samenleving tot doel, en derhalve de vaststelling of men voldoet aan één van de voorwaarden voor de verkrijging van een machtiging tot voorlopig verblijf. Het behalen van het basisexamen inburgering is ook de enige wijze waarop de potentiële nieuwkomer die in aanmerking wenst te komen voor een machtiging tot voorlopig verblijf, kan aantonen dat hij beschikt over basiskennis van de Nederlandse taal. Andere vormen van «bewijs», bijvoorbeeld in de vorm van een verklaring van een al dan niet gerenommeerd en al dan niet buitenlands taleninstituut, worden niet aanvaard. De potentiële nieuwkomer die het basisexamen inburgering heeft behaald, kan met het instellen van beroep tegen de positieve examenuitslag geen erkenning verkrijgen van het feit dat hij beschikt over taalvaardigheden op een hoger niveau dan A1-min van het Europese Raamwerk voor Moderne Vreemde Talen. Waar het bij de TGN om gaat, is de vraag of de potentiële nieuwkomer beschikt over mondelinge taalvaardigheden op het niveau A1-min en derhalve voldoet aan een van de voorwaarden om in aanmerking te kunnen komen voor een machtiging tot voorlopig verblijf (mvv). Deze vreemdeling die in het kader van de mvv-procedure heeft gevraagd om erkenning van dat feit en die erkenning ook heeft gekregen in de vorm van een positieve uitslag van het basisexamen inburgering, heeft derhalve die beschikking gekregen waarom hij heeft verzocht. Aangezien de uitslag van het basisexamen inburgering voor hem alleen van belang is voor het verkrijgen van een mvv en hem, eenmaal in Nederland aangekomen, niet meer van enig nut is, heeft hij geen belang bij het aanwenden van rechtsmiddelen tegen de (positieve) examenuitslag. Met het aanwenden van rechtsmiddelen tegen het besluit kan hij niet bewerkstelligen dat hij in een betere positie komt te verkeren. Bovendien volgt uit artikel 8:4, onder e, van de Algemene wet bestuursrecht dat geen beroep kan worden ingesteld tegen een besluit, inhoudende een beoordeling van het kennen of kunnen van een kandidaat die ter zake is geëxamineerd of op enigerlei andere wijze is getoetst, dan wel inhoudende de vaststelling van opgaven, beoordelingsnormen of nadere regels voor die examinering of toetsing. Dat de resonansgroep kanttekeningen heeft geplaatst bij de opzet en rapportage van het valideringsonderzoek door CINOP, staat hier geheel los van. Die kanttekeningen hebben, voorzover hier relevant, betrekking op de mate waarin binnen een bepaalde marge rond het A1-min-niveau de computeruitslag gelijk is aan de uitslag van menselijke beoordelaars, en hebben geen invloed op de juridische houdbaarheid van het basisexamen inburgering. De uiteindelijke uitslag in die gevallen waarin de TGN niet is behaald, wordt tijdelijk vastgesteld na een tweede beoordeling door getrainde menselijke examinatoren die niet bekend zijn met de TGN-uitslag en de geleverde prestatie nogmaals beoordelen. Als die examinatoren tot de conclusie komen, dat de geleverde prestatie wel op of boven het niveau A1-min is, krijgt de vreemdeling te horen dat hij is geslaagd. Als ook zij tot de conclusie komen dat de prestatie van de potentiële nieuwkomer onder de maat (A1-min) is, bevestigen zij de TGN-uitslag. De prestatie is in dat geval beoordeeld door zowel de computer als door vier examinatoren. Dat is zeer zorgvuldig en daarmee wordt niemand te kort gedaan door de onzekerheid waar de resonansgroep kanttekeningen bij heeft geplaatst.

4

Klopt het dat de kritiek die door de resonansgroep op de valideringswijze van de TGN geuit wordt, niet te maken heeft met de keuze voor een spraakherkenningsysteem als meetinstrument?

Ja, dat klopt. De resonansgroep heeft geen principieel bezwaar geuit tegen het gebruik van een spraakherkenningsysteem.

5

Welke maatregelen zullen genomen worden om, conform het advies van TNO, de TGN rond de A1-min cesuur te verbeteren?

Op advies van TNO zal in de praktijksituatie nader onderzoek worden gedaan op twee specifieke punten. Ten eerste wordt onderzocht of de kwaliteit van de spraakherkenner verbetert wanneer deze getraind wordt met een groter aantal personen met een zeer laag niveau van taalvaardigheid. TNO verwacht dat de kwaliteit van de spraakherkenner zal kunnen toenemen als deze is getraind met meer data van kandidaten met een taalvaardigheidsniveau rond de A1-min cesuur. Ten tweede zal worden onderzocht of de berekende cesuur, zoals in het CINOP-rapport voorgesteld, ook in de praktijk optimaal de zak- en slaaggrens van het niveau A1-min weergeeft. Van dit onderdeel van het onderzoek wordt met name verwacht dat dit méér argumenten oplevert om de nog bestaande twijfel over de plaats van de cesuur weg te nemen.

6

Tijdens de ontwikkeling van de TGN heeft de minister besloten op meer niveaus dan A1-min te beoordelen, zodat de geslaagde examinandus direct weet op welk niveau hij, bij aankomst in Nederland, een vervolgcursus kan volgen. Is de kritiek van de resonansgroep op de ontwikkeling van de TGN terug te leiden op die keuze en de consequenties die die keuze tijdens de ontwikkelingsfase van de TGN hebben gehad. Zou het weglaten van die meerdere niveaus en teruggaan tot A1-min de kritiek van de resonansgroep ondervangen?

Of de kritiek van de resonansgroep als geheel op de ontwikkeling van de TGN is terug te leiden op de keuze om een toets te ontwikkelen die het mogelijk maakt op meer niveaus te beoordelen dan het niveau A1-min, is mij niet bekend. Wel heeft één lid van de resonans-groep gemeld conceptueel bezwaar te hebben tegen een toets met een bereik van 0 tot C2. Naar de mening van dit lid zou het beter geweest zijn om zich bij de ontwikkeling van de toets te concentreren op de beoogde functies met verschillende toetsen op A1-min niveau voor het buitenland en op A2-niveau binnen Nederland.

Overigens zijn, juist om de kwaliteit van de toets op het A1-min niveau te onderbouwen op verzoek van ondermeer de resonansgroep, na het gereedkomen van een eerste versie van het verantwoordingsrapport, aanvullende experimenten uitgevoerd. In die experimenten is nadrukkelijk gestreefd naar het samenstellen van steekproeven met een relatief grote vertegenwoordiging van proefpersonen met lage taalvaardigheidsniveaus. Uit de in het rapport van CINOP c.s. gepresenteerde resultaten blijkt niet dat het onderscheidend vermogen van de toets te gering wordt, indien uitsluitend personen met lage taalvaardigheidsniveaus aan de toets deelnemen. Ook de resonansgroep heeft van de resultaten van deze aanvullende experimenten kennis genomen, zoals ook van de eindversie van het gehele rapport. Ik kan dan ook geen uitspraak doen over de vraag of het weglaten van die meerdere niveaus de kritiek van de resonans-groep zou ondervangen. Ook TNO heeft expliciet naar de kwaliteit van het examensysteem op deze twee niveaus gekeken. Daarvan is verslag gedaan in het onderzoeksrapport van TNO.

Het genoemde bezwaar geldt volgens CINOP, in de algemene termen waarin het is gesteld, zeker voor meerkeuzetoetsen gericht op het meten van receptieve vaardigheden. Bij dit soort toetsen hebben kandidaten die het antwoord niet kennen, een kans om door het antwoord te raden een goed antwoord te geven. We hebben hier echter te maken hebben met een toets waarbij de kandidaat zelf het antwoord moet formuleren en

waarbij een belangrijk deel van de score op uitingen van taalvaardigheid van de kandidaat is gericht. Kandidaten kunnen op hun eigen niveau antwoorden en worden ook op hun eigen niveau beoordeeld.

Voorts is het bij internationale taaltoetsing heel gebruikelijk om toetsen in te zetten die over een brede range van vaardigheidsniveaus meten. Ik noem bijvoorbeeld de toetsen van de grootste testinstellingen in de wereld TOEFL van ETS en IELTS van UCLES, waaraan jaarlijks honderdduizenden kandidaten deelnemen. Deze toetsen stellen gebruikers in staat om voor verschillende niveaus – afhankelijk van hun doel – cesuren vast te stellen.

Tenslotte wijs ik erop dat het gebruikelijk is bij de ontwikkeling van toetsen ten minste drie eigenschappen tegelijk te optimaliseren, te weten: de betrouwbaarheid, de validiteit en de uitvoerbaarheid. Vanwege de uitvoerbaarheid worden bijvoorbeeld bij de eindexamens havo geen aparte toetsen ontwikkeld voor mensen die zwaar onvoldoende scores, mensen die bijna voldoende scores en mensen die zeer goede resultaten behalen, ook al zouden daarmee hogere betrouwbaarheden kunnen worden gerealiseerd. Toch is er sprake van grote niveauverschillen tussen deze groepen kandidaten, zelfs groter dan tussen de gemiddelde havo- en vwo-kandidaat. Bij het optimaliseren van de drie genoemde eigenschappen dient men af te wegen in hoeverre men één van de eigenschappen kan opvoeren zonder de beide andere in gevaar te brengen. Ik beschouw de inzet van de TGN voor de wereldwijde afname van examens wereldwijd als voordeel voor de uitvoerbaarheid: het is kostenbesparend en effectief. Dat hierbij de consistentie niet in gevaar is gebracht blijkt uit het feit dat de toets consistent meet, ook in de aanvullende experimenten waaraan overwegend kandidaten van alleen de laagste taalniveaus hebben deelgenomen.

7

Zowel de resonansgroep ingesteld door CINOP als TNO, dat later een contra expertise heeft uitgevoerd, blijken van oordeel te zijn dat de validiteit van de TGN door het onderzoek van CINOP niet is aangetoond. Hoe beoordeelt de minister de deskundigheid van beide groepen deskundigen?

Ik zie geen reden om aan de deskundigheid de toetsconstructeurs van het consortium onder leiding van CINOP, noch aan die van de leden van de resonansgroep, noch aan die van de onderzoekers van TNO te twijfelen. Het feit dat deskundigen vanuit verschillende disciplines en invalshoeken naar dezelfde verschijnselen kijken en daarbij tot verschillende conclusies komen, is een gegeven dat we vaker tegen komen, en behoeft mijns inziens dan ook niet noodzakelijk te leiden tot vragen over de deskundigheid van de betrokken wetenschappers.

8

In de brief van 11 november 2005 bericht de minister aan de Tweede Kamer dat het oordeel van de door de minister en CINOP ingestelde resonansgroep luidt dat «de rapportage over de valideringsstudie van de TGN 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.» Ook de onderzoekers van TNO die een contra-expertise uitvoerden, hebben «noch bewijs gevonden dat de toets voldoende precies meet wat de toets moet meten, noch dat de toets dit in onvoldoende mate doet.» «Aangezien een goede «human benchmark» ontbreekt, achtten de onderzoekers het niet mogelijk om te

concluderen of de fouten die rond de A1min grens voor slagen of zakken worden gemaakt acceptabel zijn.» Hoe beoordeelt de minister de conclusie dat alleen CINOP overtuigd is dat de toets voldoende valide en betrouwbaar is?

Deze conclusie laat ik graag voor rekening van de vragensteller. Zowel de resonansgroep van CINOP als ook TNO behouden zich in feite een oordeel over de prestaties van de TGN voor. TNO en CINOP adviseren om onderzoek uit te voeren naar de resultaten die de toets in de praktijk oplevert. Zoals ik in mijn brief van 11 november 2005 heb aangekondigd, is dat ook precies mijn bedoeling.

9

Waarom accepteert de minister dat het CINOP de forse kritiek van zowel resonansgroep als TNO naast zich neerlegt? Immers, tijdens de mondelinge behandeling van de wet stelde de minister «De Kamer weet dat ik steeds naar de resonansgroep luister». Waarom accepteert de minister dat het CINOP de forse kritiek van zowel resonansgroep als TNO naast zich neerlegt?

Vooropgesteld zij dat de resonansgroep voor de TGN heeft gefunctioneerd als een interne kwaliteitsborging voor de productie van het consortium van CINOP, LTS en Ordinate, en dat het dus de verantwoordelijkheid is van het consortium is om kritische kanttekeningen van de zijde van de resonansgroep te waarderen. Dat gezegd zijnde, moet worden geconstateerd dat het consortium de kritiek zeker niet naast zich heeft neergelegd maar er juist dankbaar gebruik van heeft gemaakt, zoals ook is vermeld in de brief van CINOP van 10 november 2005. In de loop van het traject hebben verschillende vergaderingen met de resonansgroep plaatsgevonden. Tijdens die vergaderingen is door de resonansgroep gevraagd om aanvullend onderzoek. In de brief van CINOP geeft de bestuursvoorzitter aan dat op initiatief van de resonansgroep aanvullende gegevens zijn verzameld. Ook zijn er tijdens de fasen van het «laboratoriumonderzoek», de schaling en normering, én ten behoeve van de betrouwbaarheid en validiteit op verzoek van de resonansgroep aanvullende analyses uitgevoerd. Bovendien wordt er voor een deel tegemoet gekomen aan het advies van de resonansgroep: zij adviseren immers om een onderzoek uit te voeren naar de resultaten die de TGN in de praktijk oplevert? Dat is precies wat er zal gebeuren. Er is dus geen sprake van dat CINOP de kritiek naast zich neer heeft gelegd. Het heeft juist naar een oplossing gezocht om aan de kritiek tegemoet te komen. Die oplossing ligt ook in lijn met de aanbevelingen van TNO. Zoals hiervoor in het antwoord op vraag 3 vermeld, ben ik van plan die aanbeveling over te nemen.

10

Tijdens de mondelinge behandeling van de wet in de Tweede Kamer stelde de minister «Als de onderzoeksresultaten onverhoopt een kant opgaan die wij niet verwachten – daar is geen enkele aanleiding toe, want alle onderzoeksresultaten zijn zeer positief – zal ik de eerste zijn om te zeggen dat de invoering van het wetsvoorstel moet worden uitgesteld.» Toch stelt de minister dat het computersysteem en de TGN per 1 januari 2006 ingevoerd dient te worden. Hoe is dit te rijmen met haar eerdere standpunt, gezien de conclusie van zowel de resonansgroep als TNO dat de toets onvoldoende valide is?

Tijdens het ontwikkeltraject ben ik op de hoogte gesteld van de adviezen van de resonansgroep, en heb ik nauwlettend gevolgd of aan die adviezen gehoor werd gegeven. Noch de resonansgroep noch TNO beweren overigens dat de toets onvoldoende valide is. In hun eindconclusies melden zij dat naar hun maatstaven het bewijs voor validiteit hen niet

kan overtuigen. Beide adviseren om nog een aanvullend onderzoek te verrichten naar de resultaten die de toetsen in de praktijk opleveren.

11

De Tweede Kamer heeft nog niet met de minister kunnen spreken over de wetenschappelijke rapportages over de TGN. Dat brengt met zich mee dat de invoering van de wet en het examensysteem uitgesteld dient te worden, aangezien er door de wetenschappers betrokken als externe deskundigen, fundamentele kritiek geuit is op het onderzoek door CINOP? Is de minister niet de mening toegedaan dat dit traject zorgvuldig en wetenschappelijk onderbouwd dient te worden afgelegd?

De Tweede Kamer heeft op 5 april 2005 ingestemd met het wetsvoorstel Inburgering in het buitenland. De Eerste Kamer heeft op 20 december 2005 eveneens ingestemd met het wetsvoorstel. Ik heb op 20 december, bij gelegenheid van de plenaire behandeling van het wetsvoorstel in de Eerste Kamer aangegeven, dat ik nu streef naar publicatie van de wet in het Staatsblad op 1 februari 2006. Zo spoedig mogelijk daarna zal de wet in werking treden. Daarbij heb ik ook opgemerkt dat ik, naar verwachting op 19 januari 2006, nog een overleg met de Algemene Commissie Integratiebeleid zal hebben. Mocht tijdens dit overleg blijken dat de Kamer in meerderheid besluit dat deze wet kan doorgaan, omdat er voldoende waarborgen zijn, dan ga ik ermee door. Zegt de Kamer in meerderheid dat ik dat niet moet doen, dan zal ik mij verder beraden. De geplande invoering van de wet ligt dus na het overleg dat ik op de 19e januari 2006 met uw Kamer zal hebben. Uiteraard zal ik de invoering van de wet met even veel zorgvuldigheid ter hand nemen, als ik betracht heb bij de voorbereiding van dit wetsvoorstel.

12

CINOP stelt dat de kwaliteit van de toets alleen getest kan worden bij afname in het buitenland, bij echte examenkandidaten, onder reguliere examencondities. Is de minister het hiermee eens? Betekent dit dat CINOP stelt dat valide onderzoek alleen in de «praktijk» gedaan kan worden? Kan daaruit niet de conclusie getrokken worden dat het eerdere onderzoek door CINOP dan niet valide is?

Kwaliteit is geen absoluut begrip. Onderzoek naar de kwaliteit van een toets is een proces waarbij de vele factoren die de kwaliteit kunnen bedreigen zorgvuldig en stelselmatig, stuk voor stuk worden onderzocht. Het gaat uiteraard niet aan om een instrument waarover nog geen enkele informatie omtrent de kwaliteit bekend is in de praktijk in te zetten. Daarom is eerst een zogenaamd «laboratoriumonderzoek» gedaan. Dit onderzoek, verricht door het consortium van CINOP, LTS en Ordinate tijdens het ontwikkeltraject, heeft informatie over een groot aantal facetten met betrekking tot de validiteit opgeleverd. Op basis van de pre-testgegevens kon niet worden aangetoond dat de toets onterecht onderscheid maakt op grond van irrelevante aspecten zoals leeftijd, geslacht, land van herkomst. Op enig moment in het proces van validering moeten dan besluiten dat voldoende «laboratorium» gegevens zijn verzameld om het gebruik in de praktijk te starten. Dat betekent niet dat dan het onderzoek naar de validiteit ophoudt. Ook andere, algemeen erkende toetsen worden op grond van in de praktijk van het examen verzamelde gegevens geëvalueerd. Ook bij bijvoorbeeld de eindexamens leidt dit soms tot een bijstelling van de tevoren vastgestelde cesuur of de verwerping van één of meer opgaven. Ook TNO beveelt aan om de kwaliteit van de toets vast te stellen met data «die ontstaan gedurende het gebruik van de TGN» in de praktijk met de echte doelgroep. En dat ben ik ook van plan te doen.

13

Is er een rapport over de TGN opgesteld door de resonansgroep? Indien ja, kunt u de Kamer daar exemplaren van doen toekomen? Indien dit niet het geval is, zijn er dan andere documenten waarin de resonansgroep haar oordeel over de TGN heeft toegelicht? Is het mogelijk de Kamer deze informatie te verstrekken?

Nee, er is geen rapport van de Resonansgroep. Het commentaar van de Resonansgroep heb ik bijna letterlijk in mijn brief aan de Kamer opgenomen. Voor de volledige tekst verwijs ik u naar de brief van CINOP over de «Verantwoording Toets Gesproken Nederlands en advies van de resonansgroep» van 10 november 2005 die ik u bij de stukken heb doen toekomen.

14

Is de minister van mening dat onderzoek dat aan moet tonen dat de toets TGN valide is, wel degelijk voorafgaand aan invoering gedaan kan worden? En dient dergelijk onderzoek dan niet alsnog te worden uitgevoerd, aangezien het voorliggende onderzoek door CINOP niet tot de conclusie leidt dat de toets valide is?

Zowel de resonansgroep als TNO wijten het feit dat zij niet overtuigd zijn door de uitkomsten van het onderzoek naar de kwaliteit van de TGN onder andere aan het feit dat nog geen onderzoek op de echte doelgroep is gedaan. Overigens is ook door het CINOP aangedrongen op onderzoek in de praktijk. Ik ben daarom van plan dat te doen.

15

Is de minister de mening toegedaan dat de toets TGN en KNS meet wat zij dient te meten? Hoe kunt u dan de conclusie van zowel de resonansgroep als TNO verklaren?

Ik stel vast dat TNO concludeert dat de validatiestudies niet bewijzen dat de kwaliteit van de toets voldoende is, noch dat de kwaliteit onvoldoende is. Ik constateer evenwel ook dat TNO wèl aanwijzingen heeft gevonden dat de toets op een zinnige manier het niveau van taalvaardigheid meet:

- Er wordt een redelijke tot hoge correlatie gevonden tussen de TGN toetsscores en verschillende menselijke beoordelingen over het mondelinge taalvaardigheidsniveau.
- De toets maakt onderscheid tussen moedertaal- en niet-moedertaalsprekers.
- Er is geen ongewenst verband gevonden met achtergrondvariabelen die niet van belang zijn voor mondelinge taalvaardigheid.

De resonansgroep meent, zoals blijkt uit de brief van CINOP van 10 november 2005. Dat «de rapportage van de valideringsstudie van de Toets Gesproken Nederlands onvoldoende evidentie bevat voor de stelling dat de TGN bruikbaar is (...). Anders gezegd, de rapportage geeft onvoldoende garantie dat de toets adequaat werkt voor het doel waarvoor de toets ontwikkeld is». De resonansgroep spreekt hier tot tweemaal toe over de kwaliteit van de rapportage van CINOP die onvoldoende evidentie, c.q. onvoldoende garantie zou bieden over de kwaliteit van de TGN. Zij beveelt daarom aan om «een onderzoek uit te voeren naar de resultaten die de TGN in de praktijk oplevert». De resonansgroep adviseert dit onderzoek te laten uitvoeren, voorafgaand aan de invoering van de toets. In de brief van CINOP wordt beargumenteerd dat dit onderzoek alleen in de praktijk kan worden gedaan en stelt voor dit nà invoering van de TGN te doen. Ook TNO komt tot die conclusie en adviseert om een proefondervindelijk onderzoek in de praktijk te laten uitvoeren om de kwaliteit van de toetsen meer definitief te kunnen vaststellen.

16

Hebben LTS en van Ordinate een oordeel gegeven over de betrouwbaarheid en validiteit van de TGN- en de KNS-toets en zo ja, komt dat overeen met het oordeel van CINOP?

De verantwoording voor beide toetsen is met volledige instemming van de drie leden van het consortium aan mij uitgebracht.

17

In hoeverre zijn de door Ordinate ontwikkelde technologie voor de toetsing van Engels en Spaans en de daarmee te bedienen doelgroepen vergelijkbaar met de technologie en de doelgroepen van de Wet inburgering in het buitenland?

De technologie waarmee de spraak van kandidaten wordt herkend en vervolgens wordt beoordeeld is voor de Engelse en Spaanse toetsen gelijk aan die voor de TGN. De doelgroep van de Wet inburgering in het buitenland verschilt van de meest gebruikelijke doelgroepen bij de Engelse en Spaanse versies in die zin dat lagere eisen aan de taalvaardigheid worden gesteld en daarom ook personen met een lager niveau van taalvaardigheid zullen deelnemen. Daarom is de schaal van de TGN naar onder toe uitgebreid en is de verzameling opgaven afgestemd op dit te verwachten lagere niveau.

18

Is de minister bereid te bevorderen en te garanderen dat Ordinate inzage geeft in de wijze waarop individuele toetsen worden samengesteld en hoe men rekent met de beschikbare data, zodat de kwaliteit van de toets kan worden vastgesteld?

De wijze waarop individuele toetsen worden samengesteld (pp. 26–27) en de gehanteerde rekenmethoden (pp. 34–40) staan vermeld in de mij toegezonden verantwoording, waarvan u ook een exemplaar heeft ontvangen. In deze passages is door CINOP c.s. verantwoord in hoeverre de uit de itembank individueel samengestelde toetsen vergelijkbaar zijn met elkaar en dus of elk individueel samengesteld examen kenmerken heeft die overeenkomen met de kenmerken van een andere uit de itembank samengestelde toets. Het gaat hierbij om een verantwoording die belangrijk is voor het vaststellen van de interne consistentie van het examensysteem.

De kwaliteit van het examensysteem kan echter ook op een geheel andere wijze worden benaderd, namelijk vanuit de idee dat «the proof of the pudding is in the eating». In dat geval wordt even afgezien van de interne werking van de toets. Deze wordt als «black box» beschouwd en er wordt alleen gekeken naar de uitkomst en die uitkomst wordt vergeleken met een extern oordeel. Dit is de methode die in een deel van het onderzoek van CINOP is toegepast, namelijk daar waar menselijke oordelen met de oordelen van het geautomatiseerde examensysteem zijn vergeleken. En dit is ook de methode die door TNO wordt aanbevolen.

19

Hoe verklaart de minister dat CINOP concludeert dat de TGN op de lagere niveaus preciezer meet dan op de hogere niveaus, terwijl TNO concludeert dat de kwaliteit van het machinale oordeel minder goed is rond de A1-min cesuur dan rond de A2 cesuur?

Bij de ontwikkeling van een nieuwe toets streeft men ernaar om gegevens te verzamelen op grond waarvan men een schatting kan maken van de betrouwbaarheid, dan wel de consistentie waarmee de toets meet. Er zijn

verschillende methoden om de betrouwbaarheid te schatten. Het in uw vraag genoemde verschil tussen de verantwoording van de toets en de conclusie van TNO is gelegen in het feit dat zij op verschillende schattingsmethoden zijn gebaseerd. CINOP gebruikt een methodologie die in psychometrisch onderzoek gebruikelijk is, en maakt de schatting op basis van een statistisch model. TNO gebruikt een methodologie die gebruikelijk is in spraaktechnologisch onderzoek en baseert de schatting op observaties bij een bepaalde dataset. Hierin schuilt een belangrijk verschil. De schatting die TNO geeft is afhankelijk van de gebruikte dataset, terwijl dit niet het geval is voor de schatting die het consortium maakt. Van belang voor mij is dat TNO de conclusie deelt uit het verantwoordingsrapport van het consortium, namelijk dat er evidentie is gevonden dat de toets voldoende consistente oordelen geeft.

20

Op welke manier is de representativiteit van de onderzoeken door het CINOP gegarandeerd, ten opzichte van de situatie waarin de examens afgenomen zullen gaan worden?

Uiteraard is de situatie waarin de onderzoeken van het consortium zijn gedaan niet volledig vergelijkbaar met de situatie waarin straks het examen wordt afgenomen. Denk alleen al aan het feit dat straks in de examensituatie de uitslag voor de kandidaten van werkelijke betekenis is, terwijl in de vooronderzoeken – de «pretesten» – de kandidaten geen enkel belang hadden bij de uitslag. Het is een bekend gegeven dat kandidaten tijdens een examensituatie beter presteren dan in een pretest-situatie. Overigens worden afnamen op andere populaties dan de eigenlijke wel vaker georganiseerd bij toetsontwikkeling, bijvoorbeeld om een betere geheimhouding van examenopgaven voor de eigenlijke doelgroep te waarborgen. Zoals bekend is wel getracht om de werkelijke examen situatie zoveel mogelijk te benaderen, door middel van het aanvullende experiment dat in Amsterdam is uitgevoerd, en waarvan in het rapport verslag is gedaan.

21

Moet uit de opmerking dat naarmate de beoordelaars beter getraind zijn om deze toets af te nemen, de overeenstemming tussen menselijke oordelen en de computer hoger wordt, worden afgeleid dat de computer in beginsel beter scoort dan de mens? Moet uit deze opmerking tevens worden afgeleid dat in de praktijk de zakslaagkans bij menselijke beoordeling in belangrijke mate afhankelijk is van de mate van getraindheid van de beoordelaars?

Een computer werkt zowel consistent als objectief. U zie ook het antwoord op vraag 2. Voor beoordeling door mensen geldt dat naarmate zij beter zijn getraind in het hanteren van een beoordelingsprocedure de consistentie van hun oordelen toenemen. Wanneer hun oordelen daardoor meer op die van de computer gaan lijken, kan dat alleen wanneer de computer en de menselijke beoordelaars overeenstemmen in wat zij beoordelen.

22

Wat is de verklaring voor enerzijds de mening van de leden van de resonansgroep dat door gebrekkige dataverzameling, gebrekkige analyses en veel onduidelijkheden in de rapportage de voorliggende stukken de kwaliteit van de toets niet aantonen, terwijl anderzijds op advies van diezelfde resonansgroep allerlei aanvullende studies zijn verricht, aanvullende gegevens zijn verzameld en aanvullende analyses zijn uitgevoerd? Betekent dit dat de resonansgroep van mening is dat de kwaliteit van die aanvullende informatie onder de maat is, dan wel dat de wijze waarop daarover is gerapporteerd de toets der kritiek niet kan doorstaan?

Er leefde na de oplevering van het eerste concept van het onderzoeksrapport naar de ontwikkeling en kwaliteit van de TGN onder meer bij de resonansgroep van CINOP nog een aantal vragen. In overleg met de resonansgroep is een experiment ontworpen om die vragen duidelijk beantwoord te krijgen. Over dit experiment is uitvoerig gerapporteerd in de studie van CINOP c.s. De uiteindelijke conclusie van de resonansgroep, zoals vermeld in de brief van de directievoorzitter van CINOP van 10 november 2005, laat zien dat de resonansgroep klaarblijkelijk niet is overtuigd van de kwaliteit (van de rapportage) van het onderzoek. Wel constateer ik dat CINOP in het onderzoek een basis vindt om te adviseren de toets in de praktijk nader te laten onderzoeken. Ook TNO komt tot dit advies, en ik ben van plan dit advies over te nemen.

23

Door de resonansgroep wordt geconcludeerd dat niet is aangetoond in welke mate de aftestgrens van de TGN daadwerkelijk op het beoogde niveau ligt en door TNO wordt gesteld dat niet kan worden geconcludeerd of de fouten die rond de A1min grens voor slagen of zakken worden gemaakt, acceptabel zijn. Hoe kan de minister dan na invoering bepalen wanneer een herbeoordeling door meerdere beoordelaars moet plaatsvinden? Op grond van het oordeel van de resonansgroep en van TNO mag je namelijk concluderen dat niet goed is vast te stellen of uitslagen zich in een bepaalde marge rond de A1-min norm bevinden dan wel welke marge acceptabel is.

Dit zijn twee geheel verschillende punten, die bovendien enigszins afwijkend van de beide bronnen zijn geciteerd. Beide commentaren zijn vanuit zeer formeel wetenschappelijke standpunten geformuleerd. In de formulering van de resonansgroep wordt gevraagd naar meer onderzoek om de vraag te beantwoorden «in welke mate de aftestgrens op de toets daadwerkelijk op het beoogde niveau ligt». Dat is niet hetzelfde als de bewering dat de aftestgrens niet op het beoogde niveau zou liggen. Officieel geldt in de testwetenschap – zoals ook in de meeste wetenschappelijke disciplines – het adagium dat validiteit nooit kan worden aangetoond. Men kan slechts de aanname van non-validiteit vanwege geringe waarschijnlijkheid voorlopig verwerpen totdat eventueel meer informatie is verzameld ten bewijze waarvan alsnog de aanname van validiteit moet worden verworpen. De bewering van TNO berust op een ander principe, namelijk dat iedere beslissing een risico inhoudt en dat onderzoekers geen uitspraak kunnen doen over de aanvaardbaarheid van dat risico. Omdat geen enkele toets of examen 100% garantie kan bieden dat alle uitslagen corresponderen met de ware vaardigheid van de kandidaten zijn er ook aan het gebruik van toetsen risico's verbonden, namelijk dat iemand onterecht zal slagen of onterecht zal zakken. Onderzoeksinstellingen zoals TNO kunnen in feite alleen uitspraken doen over de grootte van dit risico. Het is dan aan de gebruiker om te bepalen of deze dit risico aanvaardbaar acht. TNO gaat er vervolgens van uit dat het risico bij gebruik van de TGN aanvaardbaar zou moeten zijn wanneer het niet groter zou zijn dan het risico dat bij door mensen beoordeelde toetsen aanvaardbaar wordt geacht. Volgens TNO, weten we wel hoe hoog het risico bij de TGN is, maar weten we niet welke norm bij door mensen beoordeelde toetsen wordt aangelegd, dus kunnen we ook die vergelijking niet maken. Daarom zal er ook een aanvullend onderzoek plaatsvinden. Daarin zullen in de aanlooperperiode alle kandidaten aan een herbeoordeling worden onderworpen. Een eventueel niet optimale cesuur kan gedurende die periode niet tot gevolg hebben dat kandidaten worden benadeeld. U zie voorts mijn antwoord op de vragen 60 en 66.

24

Waarom is de TGN niet volgens de standaardkwaliteitscriteria (van Cotan) beoordeeld? Is de minister bereid dat alsnog te bewerkstelligen?

De Commissie voor Testaangelegenheden (COTAN) van het NIP vormt een mogelijkheid om toetsen aan een onafhankelijk oordeel te onderwerpen en aan kwaliteitsborging te doen. De COTAN stelt een aantal voorwaarden voor beoordeling van tests. De COTAN stelt bijvoorbeeld het vereiste dat een door COTAN te beoordelen instrument in Nederland verkrijgbaar moet zijn en officieel moet zijn uitgegeven. Daarom zijn van de honderden tests die jaarlijks worden geproduceerd ook slechts enkele door de COTAN beoordeeld. Als kwaliteitsborging voor hun producten hanteren toetsontwikkelaars en onderwijsdeskundigen veelal codes voor goed gedrag. LTS, een van 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. Ik geef nu voorrang aan het onderzoek zoals dat door CINOP en TNO is aanbevolen, en waaruit de kwaliteit van de toetsen meer definitief zal moeten blijken. Voordien acht ik een beoordeling door de COTAN niet aan de orde. Als extra borg op de kwaliteit van heb ik de TGN op deelaspecten voorgelegd aan een ander onafhankelijk instituut, namelijk TNO. TNO heeft zich zelfstandig een oordeel kunnen vormen over de spraaktechnologische aspecten van de TGN en heeft ook onderzoeksdata ter beschikking gekregen waarmee zij zelf opnieuw berekeningen over de kwaliteit van de toets hebben kunnen uitvoeren. Daarnaast heb ik de onderzoeksrapporten van zowel CINOP als TNO openbaar gemaakt, waardoor ieder die dat wil de uitgevoerde onderzoeken kan controleren en op zijn merites beoordelen.

25

Deelt de minister het oordeel dat het noodzakelijk is dat de willekeurig samengestelde toetsen vergelijkbaar zijn en hoe wil zij dat vaststellen?

Het mag niet zo zijn dat de scores van kandidaten afhankelijk zijn van de selectie van opgaven die zij krijgen voorgelegd. Daartoe is door CINOP in de pretestfase een psychometrisch model gehanteerd waarbij de onderlinge verschillen in moeilijkheid tussen opgaven worden gewogen. In aanvullende experimenten is evidentie beschikbaar gekomen op basis waarvan de aanname van vergelijkbaarheid van willekeurig samengestelde toetsen niet kon worden verworpen.

26

Wat is het oordeel van de minister over de stelling van de resonansgroep dat de door het CINOP gebruikte steekproef niet representatief is en daardoor van een te hoog niveau?

Mede naar aanleiding van de kritiek op de oorspronkelijke steekproef ten behoeve van de pretests, zijn voor de aanvullende experimenten aparte steekproeven getrokken. Bij twee van deze steekproeven is nadrukkelijk *gestreefd* naar vertegenwoordiging van personen op de laagste taalvaardigheidsniveaus. De steekproef bij het experiment Amsterdam bestond volgens de betrokken getrainde beoordelaars voor een derde uit deelnemers nog onder het niveau A1-min en voor een derde op het niveau A1-min en ruim driekwart was korter dan een jaar in Nederland. De steekproef bij het experiment MFA-Fit bevatte volgens getrainde beoordelaars 30% deelnemers onder A1-min en 34% op het niveau A1-min. Overigens kunnen geen absolute uitspraken over representativiteit worden gedaan omdat niemand kan weten op welk niveau de deelnemers in het buitenland zich zullen bevinden.

27

Is de minister van oordeel dat het formuleren van kwaliteitseisen en het verbeteren van de kwaliteit van de toetsen alleen mogelijk is door een volledige invoering van de toetsen, waarbij aan de uitslagen ook formele gevolgen zijn verbonden? Waarom zou het niet mogelijk zijn om met de toetsen in de praktijk te experimenteren zonder dat er formele gevolgen zijn verbonden aan de uitslag? Kan de minister haar oordeel wetenschappelijk onderbouwen of is dit een politieke keuze zoals de externe onderzoekers stellen?

De adviezen van TNO en CINOP geven aan dat dit de beste oplossing is. Uit ervaring met toetsontwikkeling is gebleken dat kandidaten bij toetsen waaraan voor hen geen gevolgen zijn verbonden, andere en overigens veelal lagere uitslagen behalen dan bij toetsen waarbij dat wel het geval is. U zie ook antwoord op vraag 30.

28

De resonansgroep meent dat door gebrekkige dataverzameling, gebrekkige analyses en veel onduidelijkheden in de rapportage de stukken de kwaliteit van de TGN niet aantonen. Welke data ontbreken, welke analyses zijn gebrekkig en op welke punten en op welke punten zou de rapportage dan duidelijker kunnen zijn volgens de resonansgroep?

De brief van CINOP bevat geen nadere toelichting op de stelling van de resonansgroep dat de dataverzameling en de analyses gebrekkig zouden zijn, noch welke onduidelijkheden het rapport zou bevatten. Wel is bekend dat leden van de resonansgroep er bij herhaling op hebben gewezen dat informatie ontbreekt over personen «die het inburgeringsexamen echt doen». Daar wil ik dus door middel van het aanvullende praktijkonderzoek in voorzien.

29

Is het mogelijk om door onderzoek aan gegevens over de kwaliteit van vergelijkbare menselijke taalvaardigheidsbeoordelingen («human benchmark») te komen alvorens de TGN in de praktijk in te voeren?

Ja, dat is in zekere mate mogelijk. In mijn brief van 11 november 2005 ben ik uitvoerig ingegaan op de vraag waarom extra onderzoek, voorafgaand aan de invoering van het inburgeringsexamen, niet méér informatie oplevert over de kwaliteit van het examen: een dergelijk onderzoek leidt niet naar een situatie waarbij de werkelijke examensituatie in voldoende mate kan worden benaderd. De werkelijke examensituatie kan slechts worden bereikt door de examens daadwerkelijk in te voeren en eventuele gevolgen ervan voor de kandidaten tijdens de duur van het nadere onderzoek zoveel mogelijk te elimineren.

30

Waarom kiest de minister er niet voor om het advies van de resonansgroep (gedeeld door CINOP) te volgen om voorafgaand aan de invoering een onderzoek uit te voeren naar de resultaten die praktijk oplevert, onder meer ter beantwoording van de vraag in welke mate de aftestgrens op de toets daadwerkelijk op het beoogde niveau ligt?

Men kan de praktijk niet voorafgaand aan de invoering realistisch toetsen, omdat een wezenlijk onderdeel van de praktijk is dat een examen betekenis heeft – in de zin van gevolgen – voor de kandidaat. Dit is dan ook niet door het consortium van CINOP, LTS en Ordinate voorgesteld. Ook TNO spreekt van invoering gevolgd door onderzoek in de praktijksituatie.

31

Om de betrouwbaarheid van de TGN op het niveau A1min te toetsen, zijn twee aanvullende onderzoeken uitgevoerd. Wat waren de omstandigheden waaronder deze twee aanvullende onderzoeken zijn uitgevoerd? In hoeverre zijn de omstandigheden waaronder deze onderzoeken zijn uitgevoerd representatief voor de situatie waarin de examens in de praktijk afgenomen zullen worden?

De omstandigheden waaronder deze onderzoeken hebben plaatsgevonden staan uitvoerig in de verantwoording bij de toetsen beschreven. Ik verwijs u hiervoor graag naar Hoofdstuk 4. De belangrijkste inbreuk op de representativiteit van de bij deze onderzoeken betrokken doelgroepen wordt gevormd door het feit dat voor de proefpersonen de toetscores geen gevolgen hadden. Daarom zal de werking van de toets, na invoering ervan, in de werkelijke praktijksituatie nader worden onderzocht, zoals mij van verschillende zijden is geadviseerd.

32

Kan de conclusie van de resonansgroep, «dat de voorliggende stukken de kwaliteit van de toets niet aantonen», (ook) zo worden uitgelegd dat evenmin is aangetoond dat de kwaliteit onvoldoende is? Is dit ook niet wat TNO eigenlijk zegt?

Volgens TNO bewijzen de validatiestudies niet dat de kwaliteit van de toets voldoende is, noch dat de kwaliteit onvoldoende is. Een belangrijke reden hiervoor is dat voorafgaande aan de validatiestudies geen zekerheidsmarge of criterium is gesteld waarboven of waaronder men wetenschappelijk gefundeerd tot één van beide uitspraken «wel voldoende bewijs» of «geen voldoende bewijs» kan besluiten. Of de resonansgroep dat ook vindt is niet in haar eindoordeel opgenomen.

33

Waarmee hangt de vraag samen of de toets (zowel de TNG- als de KNS-toets) meet wat hij moet meten? In verband waarmee zou de toets niet meten wat hij moet meten?

TNO heeft geconcludeerd dat de validatiestudies niet bewijzen dat de toets (zowel de TGN- als de KNS-toets) meet wat hij moet meten, echter evenmin dat de toets dat *niet* doet. Maar op grond van de pretestdata kon geen verband worden aangetoond tussen de toetscores en enkele van de meest voor de hand liggende mogelijk storende variabelen zoals leeftijd, geslacht, land van herkomst. Om te bewijzen dat dit verband in de praktijk niet bestaat vindt TNO het beter om deze validatiestudies te herhalen met echte praktijkdata.

34

Kan de minister reageren op de stelling van TNO (blz. 20) dat de Engelse en Spaanse versie van de toets niet vergelijkbaar zijn, doordat deze toetsen uitgaan van een «human benchmark» die bestaat uit globale oordelen over de taalvaardigheid, terwijl met de door de minister voorgestelde toets met voldoende nauwkeurigheid bepaald moet worden of aan een zeker (laag) taalvaardigheidsniveau wordt voldaan?

U zie ook het antwoord op vraag 17. Evenals de Nederlandse toets rapporteren de toetsen voor Engels en Spaans op een continue schaal lopend van zeer lage tot zeer hoge taalvaardigheid. De Nederlandse toets biedt echter ten opzichte van de toetsen voor de beide andere talen de mogelijkheid om op een nog lager beginniveau te meten. Dit is bewerkstelligd door gemakkelijkere opgaven in de toets op te nemen en door bepaalde typen opgaven uit te sluiten. Evenals bij de Nederlandse toets zijn de

scores bij de Engelse en de Spaanse toets gerelateerd aan de schaal van de Raad van Europa. De hierbij gevolgde procedures zijn nagenoeg identiek: er zijn menselijke oordelen over de kandidaten verzameld op grond van prestaties van deze kandidaten, onafhankelijk van hun prestaties op de toetsen. De relatie tussen deze menselijke oordelen en de toetscores vormen de grondslag waarop de voorspellende waarde van de scores met betrekking tot de niveaus van de Raad van Europa is vastgesteld. Evenals de Nederlandse toets worden de Engelse en de Spaanse toets gebruikt om in examensituaties na te gaan of kandidaten op een bepaald minimumniveau kunnen functioneren. De Engelse toets wordt bijvoorbeeld gebruikt in Hong Kong om te achterhalen of werknemers voldoende Engels beheersen om bepaalde beroepsfuncties te kunnen uitoefenen, waarbij scores overeenkomend met A1 voldoende worden geacht voor assistent functies, terwijl voor receptionisten minimaal A2 wordt vereist, voor secretaresses B1, enz. De Spaanse toets wordt onder andere gebruikt in het Amerikaanse leger, waarbij de verschillende niveaus mede bepalend zijn voor salariëring en of bevordering in rang.

35

Hoe kun je echte examencondities creëren waaronder je de toets gesproken Nederlands verder wil ontwikkelen terwijl je tegelijkertijd deelnemers niet als proefkonijnen wil gebruiken?

Dit kan door de antwoorden van de kandidaten ook te laten beoordelen door vier getrainde menselijke beoordelaars. Voor iedere kandidaat worden op deze manier twee toetsuitslagen verkregen, één toetsuitslag gebaseerd op vier menselijke oordelen en één toets-uitslag verkregen met de automatische toets. Wanneer als uitgangspunt wordt genomen dat de kandidaten in elk geval niet benadeeld mogen worden in deze fase, dan kan de meest gunstige van beide toetsuitslagen als doorslaggevend worden aangemerkt. De belasting voor de kandidaat blijft gelijk, hij of zij hoeft de toets maar één keer te doen.

36

Naar aanleiding van het door hen uitgevoerde onderzoek vond TNO aanwijzingen dat de toets minder precies zou meten rond de A1min cesuur. Op welke manier is deze opmerking van TNO meegenomen in het advies van CINOP? Waarom wil de minister TGN invoeren, terwijl uit verschillend onderzoek blijkt dat er geen duidelijkheid bestaat of de toets daadwerkelijk meet wat hij hoort te meten, voordat er garanties zijn dat de toets op alle niveaus goed meet? Kan de minister motiveren welke fouten die gemaakt kunnen worden rondom de A1min grens acceptabel zijn?

U zie mijn antwoord op vraag 19. Het vermoeden van TNO dat de toets minder precies zou meten rond de A1-min cesuur dan op de A2 cesuur kon niet worden verwerkt in het valideringsrapport omdat dit eerder gereed kwam dan het rapport van TNO. Overigens betekent dit nog niet dat er met behulp van de toets op het A1-min niveau onvoldoende precies zou worden gemeten.

Zowel CINOP als TNO zijn op basis van hun eigen onderzoek en analyses – en CINOP mede op basis van de oordelen van zijn resonansgroep – tot een gelijklopend advies gekomen. Namelijk het aanvullend onderzoek uit te voeren parallel aan de invoering van het examen in het buitenland. Bij mijn voorstel om nu een aanvang te maken met de invoering van het examenstelsel in het buitenland, baseer ik me dus op deze twee adviezen, die beide ook aangeven dat verder onderzoek voorafgaand aan invoering van het nieuwe stelsel niet tot een complete beantwoording van de voorliggende onderzoeksvragen zullen leiden.

Bij het vervolgonderzoek zal het in hoge mate gaan om het beantwoorden van de vraag in hoeverre menselijke oordelen en oordelen door het geautomatiseerde scoringsstelsel van de TGN met elkaar corresponderen. Voor wat betreft de mate waarin fouten acceptabel zijn, moge duidelijk zijn dat ik daar streef naar een maximale optimalisering van het stelsel. Hoe minder fouten, hoe beter. Daarbij moet wel opgemerkt worden dat geen enkele toets in staat is om in 100% van de gevallen de juiste score te bepalen. Geen enkele toets is onfeilbaar. Alle toetsen die in Nederland en daarbuiten gebruikt worden en waarbij een zak-slaaggrens vastgesteld is, kent beslissingen die leiden tot mogelijk onterecht zakken of tot mogelijk onterecht slagen. Dat geldt overigens net zo sterk voor toetsen die door menselijke examinatoren afgenomen worden. Bij het vaststellen van en onderbouwen van zak-slaaggrenzen gaat het erom om die onterechte beslissingen en de gevolgen van die beslissingen te minimaliseren. Na ommekomst van de onderzoeksresultaten, zal ik u daarover uiteraard nader berichten, waarbij ik nog eens nadrukkelijk wil aangeven dat het er mij om gaat dat menselijke en automatisch gescoorde oordelen zo goed mogelijk corresponderen.

37

Welke conclusies verbindt de minister nu precies aan de bevindingen betreffende de toets KNS met betrekking tot de invoering van die toets?

Het CINOP heeft mij aangeraden om de cesuur voor het KNS-examen op een beheersingsniveau van 70% vast te leggen, dit op basis van het onderzoek van CINOP en van de conclusies van de resonansgroep die het KNS-examen heeft begeleid. Ik zal die aanbeveling van CINOP en de resonansgroep overnemen. Daarnaast zal in het onderzoek dat na de invoering van examens buitenland zal plaatsvinden ook onderzocht worden of de metingen van de KNS-toets op basis van dat onderzoek nog geoptimaliseerd kunnen worden.

38

Deelt de minister de opvatting van degenen die stellen dat het examen KNS meer woorden bevat dan die welke voor (het slagen voor) de TGN nodig zijn en dat dit een onnodige verzwaring van de exameneisen betekent?

Neen, die opvatting deel ik niet. Om te kunnen spreken op het niveau A1-min moet men zich verstaanbaar kunnen maken met een beperkt aantal zeer frequente losse woorden en standaardformuleringen. Om te kunnen luisteren op hetzelfde niveau moet men een beperkt aantal vertrouwde woorden en basiszinnen die betrekking hebben op de directe persoonlijke levenssfeer en ervaringswereld begrijpen. Het gaat dus om zeer frequente woorden, en om vertrouwde woorden, die betrekking hebben op de directe persoonlijke levenssfeer en ervaringswereld. Een groot deel van de woorden in de toets KNS voldoet aan de eerste voorwaarde: ze zijn zeer frequent. Een aantal andere woorden voldoet hier niet aan: bijvoorbeeld «Grondwet» behoort in het Nederlands niet tot de 500 meest frequente woorden, en moet dus voldoen aan het criterium «vertrouwde woorden die betrekking hebben op de directe persoonlijke levenssfeer». Juist omdat de directe persoonlijke levenssfeer en ervaringswereld voor een ieder weer anders is, is ervoor gekozen om de film «Naar Nederland», en oefenmateriaal voor de toets te laten vervaardigen. Ik ga ervan uit dat iedereen die zich serieus wil voorbereiden op de KNS toets dit materiaal aanschafte. In de film komt het woord Grondwet wel tien keer voor. Ook in het oefenpakket met 100 vragen komen kandidaten die zich voorbereiden op het examen het woord «Grondwet» een aantal malen tegen. Door middel van de film en het oefenmateriaal wordt dus aan kandidaten duidelijk gemaakt dat zij vertrouwd moeten zijn met het

woord «Grondwet». Na kennisneming van de film en het oefenmateriaal behoort dit woord tot de hen «vertrouwde woorden».

39

Is het mogelijk om door onderzoek met een grotere dataset een nauwkeurigere schatting van de foutenpercentages m.b.t. de toets KNS te geven? Is het mogelijk aan zo'n grotere dataset te komen alvorens de toets KNS in de praktijk in te voeren?

Het aanvullende onderzoek waartoe TNO en CINOP c.s. adviseren is ook gericht op de toets KNS. De grotere dataset ontstaat dus tijdens dit onderzoek. De argumenten om dit onderzoek niet voorafgaand, maar na invoering van de toets uit te voeren heb ik onder andere gegeven in de antwoorden op de vragen 1, 12 en 20.

40

Is het mogelijk om onderzoeksgegevens te verwerven betreffende vergelijkbare menselijk beoordeelde examens alvorens de toets KNS in de praktijk in te voeren?

Er zijn al onderzoeken in zogenaamde laboratoriumsituaties uitgevoerd en daarover is door het Consortium CINOP/LTS/Ordinate gerapporteerd. Hieraan namen in totaal ruim 1000 niet-Nederlandstaligen deel. TNO en CINOP zijn van mening dat aanvullend onderzoek het best uitgevoerd kan worden op basis van praktijkgegevens. U zie onder andere de argumenten bij het antwoord op vraag 1.

41

Waarom heeft het CINOP er voor gekozen om de validatie van de TGN en de toets KNS op dezelfde wijze te benaderen als elke willekeurige toets zonder spraakherkenning?

Het is onjuist dat het consortium de validatie van beide toetsen op dezelfde wijze heeft benaderd als die voor een toets zonder spraakherkenning. Het consortium heeft in de verantwoording veel aandacht besteed aan de beschrijving (u zie pp. 34 – 40) en validatie (u zie pp. 89 – 94) van de met gebruikmaking van spraakherkenning gegenereerde scores. Vanaf het moment dat deze scores beschikbaar zijn, wijkt de onderzoeksmethodologie niet af van de methodologie die gebruikelijk is bij het onderzoek van toetsen waarbij de scores op andere wijze zijn verkregen.

42, 43

Het CINOP stelt dat een cesuur van 0.8 voor de toets KNS behoorlijk hoog is voor de inspanningsverplichting die van aspirant nieuwkomers wordt gevraagd. Waarop is deze stelling gebaseerd? Het CINOP stelt dat, gelet op een aantal factoren, voorzichtigheid in dit stadium geboden is met de toets KNS. Daarom stelt het CINOP voor de cesuur in ieder geval niet hoger te leggen dan 0.70 voor de toets KNS. Is de minister het met dit voorstel van CINOP eens? Waarop is het door de resonansgroep aanbevolen percentage van 70 gebaseerd?

Op het KNS-examen krijgen kandidaten een selectie van 30 vragen uit de totale set van 100 vragen. Doordat de moeilijkheid van alle vragen tevoren is onderzocht kan op grond van de antwoorden op deze 30 vragen worden voorspeld hoeveel vragen van de totale set van 100 een kandidaat goed zou kunnen beantwoorden. CINOP geeft aan dat het in de test-wetenschappen gebruikelijk is om bij kennisdomeinen 80% als «mastery» te beschouwen. Dat wil zeggen dat als men 80% van de vragen goed beantwoordt men dat domein voldoende beheerst. De regering vindt het van

groot belang dat potentiële nieuwkomers kennis hebben van de Nederlandse samenleving opdat zij beter voorbereid naar Nederland komen en daardoor een sneller inburgeringsproces in Nederland doorlopen. De inburgering in Nederland moet er toe leiden dat nieuwkomers snel zullen beschikken over voldoende kennis van de Nederlandse samenleving. Voor de inburgering in het buitenland kan dit nog niet van mensen verwacht worden. Het is wel de bedoeling dat mensen zich goed voorbereiden op hun komst naar Nederland. En dat zij gemotiveerd zijn om aan hun toekomst in Nederland te werken. Dat moeten zij tijdens het examen op de ambassades laten zien. Om een onderscheid te kunnen maken tussen mensen die zich goed hebben voorbereid op het examen en mensen die dat niet hebben gedaan, adviseren CINOP en de resonansgroep op basis van de onderzoeksgegevens om de cesuur van het KNS examen op 70% vast te stellen. Hierdoor wordt uitgesloten dat mensen bij toeval of door raden een voldoende score behalen. Anderzijds wordt ook uitgesloten dat mensen worden afgewezen wanneer zij een redelijke inspanning hebben gepleegd om zich op dit deel van het examen voor te bereiden.

44

Geletterden hebben een hogere kans om de KNS te halen dan niet-geletterden, waardoor een discriminatoir onderscheid wordt gemaakt. Wat is de mening van de minister over de lagere kans voor niet-geletterden om te slagen voor de KNS? Is de minister van mening dat zowel geletterden als niet-geletterden gelijke kansen behoren te hebben op het behalen van de KNS, dat niet-geletterdheid daarbij geen belemmering mag vormen?

Aan de toets KNS is ondermeer als eis gesteld dat geletterdheid geen rol mocht spelen bij de examinering. Of mensen slagen voor het KNS deel van het examen hangt af van wat zij weten over de Nederlandse samenleving. Het wetenschappelijke rapport vermeldt dat het totale percentage geslaagden onder de niet-geletterde deelnemers lager was dan dat onder de geletterden. Het is wel mogelijk dat niet-geletterden gemiddeld meer tijd nodig hebben om de vereiste kennis te verwerven. Uit de rapportage blijkt evenwel ook dat het voor analfabeten mogelijk is om voldoende en zelfs hoge uitslagen te behalen. Dat niet-geletterden gemiddeld lagere scores haalden hangt samen met het feit dat onder niet-geletterden een groter percentage deelnemers het niveau A1-min volgens hun docenten niet beheersten. U zie ook mijn antwoord op vraag 46.

45

Kandidaten moeten voor de KNS kennis nemen van de film «Naar Nederland». Wat is de samenstelling van deze film? Bevat deze film naast gesproken tekst ook geschreven tekst?

Op basis van het advies van de Commissie Franssen om voorlichtingsmateriaal voor potentiële nieuwkomers te ontwikkelen, heb ik besloten om een film over de Nederlandse samenleving te laten maken die zij vooraf kunnen bekijken. In deze film wordt de kijker in 105 minuten wegwijs gemaakt in Nederland: geografie, vervoer en wonen, geschiedenis, staatsinrichting, politiek en wetgeving, de Nederlandse taal en het belang ervan om deze te leren, opvoeding en onderwijs, gezondheidszorg, en werk en inkomen. Een presentatrice geeft de kijker voor een potentiële nieuwkomer nuttige informatie over Nederland. Daarbij wordt haar verhaal ondersteund door beelden over de diverse onderwerpen. Sommige beelden bevatten geschreven tekst, die door de presentatrice mondeling wordt uitgelegd. Hierdoor kunnen ook niet-geletterden de informatie uit de film begrijpen.

Wat zijn de eisen ten aanzien van de kansen voor geletterden en niet-geletterden? Is tijdens het opstellen van het wetsvoorstel niet de eis geuit dat niet-geletterdheid geen belemmering mag vormen voor het behalen van de toets, zowel de TGN als de toets KNS?

Ik ben van mening dat iedereen die bereid en gemotiveerd is een succesvol inburgerings-proces te doorlopen in staat moet zijn het inburgeringsexamen in het buitenland te halen. Op advies van de tijdelijke Adviescommissie Normering Inburgeringsexamens onder leiding van de heer Franssen wordt van potentiële nieuwkomers gevraagd over kennis van de Nederlandse taal te beschikken op minimaal A1-min niveau. De commissie geeft aan dat er ook met dit relatief lage niveau sprake zal zijn van selectie. Die selectie zal volgens de commissie echter niet noodzakelijkerwijs samenhangen met verschillen tussen laag- en hoogopgeleiden. Het gaat dan om selectie van personen die kunnen en willen investeren in hun eigen inburgering. Uit het wetenschappelijke onderzoek naar de toetsen is gebleken dat de examens KNS en TGN niet selecteren op diverse achtergrondkenmerken van examinandi, bijvoorbeeld leeftijd of geslacht. De mate van geletterdheid heeft volgens dit onderzoek geen significante invloed op de slagingskansen voor het taalexamen. Bij het examen kennis van de Nederlandse samenleving halen geletterden gemiddeld wel betere resultaten op het KNS examen dan niet-geletterden. Wellicht werd dit veroorzaakt doordat een kwart van de niet-geletterden het Nederlands niet op het vereiste A1min-niveau beheerste, terwijl onder de geletterden slechts 2 procent onder A1-min functioneerde. Uit het onderzoek is immers gebleken dat ook niet-geletterden in staat zijn om goed te presteren op het examen. Het feit dat iemand niet-geletterd of geletterd is, man of vrouw, jong of oud is bepaalt dan ook niet of iemand succesvol zal presteren op het examen, maar de mate waarin iemand bereid is zich in te spannen voor een succesvolle inburgering en integratie in Nederland.

Als wordt gesteld dat een kandidaat het examen nogmaals kosteloos mag afleggen als zich technische problemen voordoen die niet aan hem of haar kunnen worden verweten, aan wat voor technische problemen moet hier dan worden gedacht? Liggen die in hetzelfde vlak als die welke door de resonansgroep en TNO naar voren zijn gebracht? Hoe en door wie wordt vastgesteld of er sprake is van «die niet aan hem of haar kunnen worden verweten»?

De resonansgroep en TNO hebben geadviseerd over het wetenschappelijke onderzoek naar de toets. Dit is een wetenschappelijk onderzoek en geen technisch probleem. Onder technische problemen wordt bijvoorbeeld verstaan stroomuitval, het wegvallen van de telefoonverbinding of storingen/uitval van de apparatuur waardoor afname van het examen wordt bemoeilijkt of onmogelijk gemaakt. In het geval er zich technische problemen voordoen dan mogen kandidaten indien het mogelijk is en de voorkeur heeft van de kandidaat, het examen direct nog een keer doen. Het is immers denkbaar dat er zich, ondanks een zorgvuldige uitrol van de apparatuur naar de posten, toch technische problemen voordoen tijdens het examen, zoals dat met alle technische apparatuur ook in onze dagelijkse werkomgeving mogelijk is. Deze problemen zijn objectief vaststelbaar door de toezichthouder (medewerker van Buitenlandse Zaken) op de ambassade. Indien deze problemen niet door de kandidaat veroorzaakt zijn, kan een kandidaat kosteloos opnieuw examen doen.

48

Als wordt gesteld dat men kosteloos nogmaals het examen mag afleggen, wat wordt hiermee dan precies bedoeld? Kan/moet het examen op dezelfde dag nogmaals worden afgelegd? Krijgt men reiskosten vergoed als men op een andere dag moet terugkomen om het examen af te leggen? Hoe wordt rekening gehouden met de eventuele negatieve invloed van technische problemen op de examenprestaties van examinandí?

Kandidaten mogen kosteloos een nieuw examen doen indien er zich bij het afnemen van het examen technische problemen voordoen die niet aan de kandidaat te wijten zijn, of er zich onverwachte verstoringen zoals geluidsoverlast voordoen. Nadat door de toezichthouder dergelijke problemen zijn geconstateerd zal in eerste instantie getracht worden deze problemen ter plaatse te verhelpen waarna de kandidaat direct kosteloos een nieuw examen kan doen. Mocht de kandidaat er de voorkeur aan geven het examen op een later tijdstip opnieuw te doen, dan kan de kandidaat hiervoor een nieuwe afspraak maken. Indien de problemen van langdurige aard blijken te zijn zal de kandidaat verzocht worden op een ander tijdstip terug te komen. In deze, naar verwachting zeldzame, situatie komen de eventuele extra reis- of verblijfkosten voor rekening van de kandidaat.

49

In hoeverre verschillen de toetsomstandigheden in de verschillende (vooral de «minder gangbare») herkomstlanden van elkaar, bijvoorbeeld bezien naar beschikbaarheid van oefenmateriaal in de eigen taal, de getraindheid van de spraakherkenningcomputer in de meest frequente taalachtergronden, de examenomstandigheden op de ambassades/consulaten, e.d.? Hebben kandidaten in «minder gangbare» herkomstlanden en met «minder frequente taalachtergronden» naar het oordeel van de minister dezelfde kansen om het examen te halen?

De regering vindt dat potentiële nieuwkomers zelf verantwoordelijk zijn voor hun inburgering. Dit impliceert onder meer dat zij zelf kunnen beslissen op welke manier zij zich voor willen bereiden op het inburgeringsexamen in het buitenland. Hierbij staan zij er uiteraard niet alleen voor. Zij hebben immers een partner, een familielid of een andere relatie in Nederland die kan helpen bij de komst naar Nederland. Uiteraard kunnen de omstandigheden in de landen van herkomst verschillen. Dat kan betekenen dat examenkandidaten uit bepaalde landen en hun relatie in Nederland harder hun best moeten doen om een goede voorbereiding te bewerkstelligen dan in een ander land. Zodra kandidaten goed zijn voorbereid kunnen zij examen gaan doen op de ambassade. Ambassadegebouwen verschillen onderling, en daarmee ook de examenruimten. Overall is echter voorzien in een aparte ruimte waarin het examen ongestoord kan worden afgelegd. Voor wat betreft het examen zijn er in de pretesten van TGN en KNS geen aanwijzingen gevonden dat de resultaten van kandidaten samenhangen met het land van herkomst. Gelet op het grote aantal landen waaruit deelnemers aan de pretesten afkomstig waren kan niet worden aangenomen dat sprekers van «minder gangbare» talen worden benadeeld. Van belang is de inspanning die kandidaten zich willen getroosten voor het verwerven van de voor succesvolle deelname aan het examen noodzakelijk kennis en vaardigheden.

50

Kan de minister toelichten waarom zij van oordeel is dat er slechts sprake is van «kinderziekten», wanneer zowel de resonansgroep als TNO concluderen dat onvoldoende de bruikbaarheid en kwaliteit van de TGN is aangetoond?

U zie ook het antwoord op vraag 12. Wanneer er sprake zou zijn van evidentie dat de TGN niet valide was, of niet consistent zou meten, zou ik zeker niet spreken van «kinderziekten». Dit is echter ook volgens TNO niet het geval. TNO vraagt wel om meer bewijs voor de validiteit, met name in de praktijk en bij de echte doelgroep. Ik heb er vertrouwen in dat deze evidentie voldoende zal worden verkregen met het onderzoek naar de praktijk van de uitvoering van de TGN. «Kinderziekten» kunnen bij geen enkel nieuw proces worden uitgesloten, en dus ook niet bij de invoering van deze TGN. Zoals ik de Kamer reeds heb gezegd bij de behandeling van het wetsvoorstel, zal ik zo goed mogelijk proberen te voorkomen dat kandidaten van dergelijke kinderziekten negatieve gevolgen ondervinden.

51

De TGN heeft een bereik van A1 tot en met B2 of hoger. Dat wordt ook duidelijk in de aard en inhoud van de vraagstelling en het daarbij gebruikte taalniveau. Kan de minister wetenschappelijk aantonen dat dit diegenen die examens doen op A1-min niveau en niet meer kunnen niet onnodig benadeelt?

Allereerst: de TGN heeft een bereik al vanaf *onder* A1-min tot aan zelfs niveau C2. Kandidaten verdienen hun score niet met wat ze niet kunnen, maar met wat ze aantonen wel te kunnen. Kandidaten kunnen opgaven op hun eigen niveau beantwoorden. De beheersingsniveaus worden conform de definities van de Raad van Europa bepaald door *wat* kandidaten zeggen en door de kwaliteit *waarmee* ze dat doen. Antwoorden van kandidaten kunnen verschillen in de mate waarin zij correct zijn en door de mate waarin zij verstaanbaar zijn. Het antwoord op een vraag kan slechts gedeeltelijk correct zijn en slechts met moeite verstaanbaar, dan spreken we van een laag niveau. Maar niemand kan of wil verhinderen dat sommige kandidaten in staat zijn op dezelfde vraag een volledig correct, zeer duidelijk verstaanbaar en vlot gesproken antwoord te geven. Dan spreken we van een hoog niveau. Er is geen reden te veronderstellen dat kandidaten met een laag beheersingsniveau worden benadeeld door het feit dat andere kandidaten beter Nederlands beheersen. De vragen verschillen weliswaar onderling in moeilijkheidsgraad. Wanneer echter deze verschillen zo groot zouden zijn, dat kandidaten met een laag niveau zouden worden benadeeld ten opzichte van kandidaten met een hoger niveau, dan zou men verwachten dat de meetfout toeneemt bij de lagere niveaus. CINOP heeft juist om de kwaliteit van de toets op het A1-min niveau te onderbouwen aanvullende experimenten uitgevoerd. In die experimenten is nadrukkelijk gestreefd naar het samenstellen van steekproeven met een relatief grote vertegenwoordiging van proefpersonen met lage taalvaardigheidsniveaus. Uit de in het rapport van CINOP c.s. gepresenteerde resultaten blijkt niet dat diegene die het examen doen op A1-min niveau en niet meer kunnen onnodig wordt benadeeld. Het tegendeel blijkt waar, de meetfout neemt geleidelijk toe naarmate het gemeten niveau toeneemt (zie Tabel 6.3 van het Verantwoordingsrapport).

52

Zou de betrouwbaarheid van de TGN en de KNS gediend zijn met een toets die een bereik heeft van A1 tot en met A2 in plaats van tot en met B2 of hoger? Zo nee, kan de minister dat wetenschappelijk aantonen?

Allereerst wil ik u erop wijzen dat voor succesvolle deelname aan de KNS het niveau A1-min volstaat en dat er daar dus geen sprake is van niveaus hoger dan A1-min en hoeft het niveau dus niet te worden beperkt. Voor de TGN wil ik u erop wijzen dat de toetsscore voor een groot deel is gebaseerd op een productieve taalvaardigheid, namelijk spreken. Iedere kandidaat kan op eigen wijze op een vraag antwoorden. Sommige kandidaten geven een antwoord op niveau A1-min, met losse woorden en gebrekkige

uitspraak. Andere kandidaten geven bij diezelfde vraag een antwoord in perfect, goedlopend en vloeiend Nederlands. Er is geen reden om te veronderstellen dat de kandidaten die een antwoord op A1-min niveau geven worden benadeeld door het feit dat sommige anderen het veel beter kunnen doen. Beperking van de scoreschaal bij de TGN tot scores die hooguit tot en met niveau A2 voorspellen heeft dus geen zin.

53

Ontvangt de Kamer voorafgaand van het algemeen overleg nog de gewijzigde algemene maatregel van bestuur, die al bij brief van 15 februari 2005 aan de Kamer is aangeboden?

Voorafgaande aan verzending van het ontwerp voor een Besluit inburgering in het buitenland op 15 februari 2005 aan uw Kamer (Kamerstukken II 2004/2005, 29 700, nr. 8) heb ik in de nota naar aanleiding van het verslag van uw Kamer op 7 december 2004 aangegeven dat dat ontwerp nog kan worden gewijzigd, bijvoorbeeld naar aanleiding van de parlementaire behandeling van het wetsvoorstel of naar aanleiding van het advies van de Raad van State (Kamerstukken II 2004/2005, 29 700, nr. 6, p. 42). Nadien is advies van de Raad van State gevraagd en verkregen en is het ontwerpbesluit aangepast. Op grond van artikel 25a van de Wet op de Raad van State is de tekst op dit moment niet openbaar. De tekst van het besluit en het advies van de Raad van State zullen pas openbaar worden nadat ik de Koningin het nader rapport hierop heb aangeboden, en zij vervolgens het besluit heeft afgekondigd. Naar ik aanneem wordt met deze vraag geïnformeerd naar de juridisch-technische wijze waarop ik voornemens ben de reeds aangekondigde herbeoordeling door menselijke examinatoren in het Besluit inburgering te verankeren. Daarover wil ik uw Kamer nu reeds meedelen dat mij een overgangsrechtelijke bepaling voor ogen staat: in afwijking van artikel 3.98c, derde lid, van het Vreemdelingenbesluit 2000 worden de resultaten van het basisexamen inburgering, die door middel van het geautomatiseerde systeem, bedoeld in artikel 3.98c, derde lid, van dat besluit als onvoldoende zijn beoordeeld, nogmaals beoordeeld door examinatoren, indien het basisexamen is afgelegd voor een bij koninklijk besluit te bepalen tijdstip.

54

Zijn er onderdelen van het TNO-rapport of het oordeel van de resonansgroep waarvan de minister afstand neemt? Zo ja, welke en op welke wetenschappelijke gronden?

Zoals uit mijn brief van 11 november 2005 duidelijk moge zijn geworden, ben ik voornemens om de kern van de adviezen te volgen, namelijk het invoeren van de toetsen in de praktijk en daarbij door proefondervindelijk onderzoek de kwaliteit van de toetsen en de juiste plaats van de zak- en slaaggrens te verifiëren.

55

Is de minister bereid om na het verbeteren van de toetsen in de praktijk TNO opnieuw om een second opinion te vragen?

Op basis van de adviezen van CINOP en TNO bereid ik thans een aanvullend onderzoek voor, dat gelijktijdig met de invoering van het basisexamen inburgering gestart zal worden. Ik sluit niet uit dat TNO daarbij een rol zal vervullen in de begeleiding of uitvoering van dit onderzoek. Omwille van een zorgvuldige aanbestedingsprocedure kan ik echter niet vooruitlopen op de uiteindelijke samenstelling van de onderzoeksgroep.

56

Hoe beoordeelt de minister het feit dat CINOP, in tegenstelling tot TNO, de validatie van de toetsen op dezelfde wijze benadert als elke willekeurige toets zonder spraakherkenning? Acht de minister dit acceptabel?

U zie het antwoord op vraag 41. Van belang is dat de onderzoeken op een wetenschappelijk verantwoorde manier plaatsvinden. CINOP en TNO hebben dat mijn inziens binnen de onderzoeksopdracht en volgens hun eigen onderzoekstraditie gedaan.

57

Is de minister van oordeel dat CINOP de opdracht al naar behoren en naar tevredenheid heeft uitgevoerd? Wat betekent dit voor de overeengekomen vergoeding?

De opdracht is nog niet volledig uitgevoerd en een eindafrekening heeft ook nog niet plaatsgevonden.

58

De wetenschappelijke rapportages zijn laat ontvangen. Is getracht de rapportage van CINOP nog in overeenstemming te brengen met die van TNO?

De wetenschappelijke rapportages zijn inderdaad later dan ik had gehoopt ontvangen en aan de Kamer aangeboden. Na de behandeling in tweede en derde termijn van het wetsvoorstel in uw Kamer en mede naar aanleiding van de commentaren van de leden van de resonansgroep van CINOP ten behoeve van de TGN heb ik CINOP gevraagd om aanvullende onderzoeken te verrichten. Ik had verwacht deze rapporten nog voor de zomer te ontvangen. Het consortium onder leiding van CINOP is door de grote hoeveelheid werk die verbonden was aan de aanvullende onderzoeksvragen, niet in staat geweest om al voor de zomer de onderzoeken af te ronden. Het eindrapport over de toets KNS is ontvangen op 31 mei 2005, voorzien van een oordeel van de resonansgroep KNS. Het eindrapport van CINOP over de TGN heb ik ontvangen op 19 september 2005. Na de oplevering van deze rapporten moest de resonansgroep van CINOP nog zijn oordeel vellen over de rapportage over de TGN. Het oordeel van de resonansgroep heb ik door middel van de brief van de bestuursvoorzitter van CINOP op 10 november 2005 ontvangen. Voor het verrichten van een second opinion door TNO was uiteraard ook de definitieve rapportage van CINOP van groot belang. TNO had daarvoor de uitkomsten uit de aanvullende onderzoeken van CINOP nodig om de analyses te vervolmaken. Ik heb het eindrapport van TNO op 24 oktober 2005 ontvangen. Er is niet getracht om het valideringsrapport van CINOP na het uitkomen van het TNO rapport aan te passen.

59

Welke rol vervult CINOP in het in de praktijk vervolmaken van de toetsen?

Het ligt voor de hand dat het consortium een rol zal spelen bij de uitvoering van het vervolgonderzoek. Welke rol dat precies zal zijn, en wie daar verder bij zullen worden betrokken is op dit moment een kwestie van overleg en onderhandeling.

60

Kan de minister preciezer aangeven hoe zij de herbeoordeling door meerdere beoordelaars in de praktijk denkt vorm te geven? Zal dit geschieden in het land van herkomst of in Nederland en welke gevolgen heeft dat?

Op advies van TNO en de resonansgroep zal ik nader onderzoek laten doen om de voorgestelde cesuur voor het A1-min niveau te verifiëren. Dit betekent dat vooralsnog – gedurende de looptijd van dit onderzoek – de examens van alle kandidaten tevens worden voorgelegd aan een panel van vier mensen die het examen afzonderlijk beoordelen. De herbeoordeling zal plaatsvinden op basis van het digitale bestand van het (opgenomen) examen dat de kandidaat heeft afgelegd op de ambassade. Dit bestand zal naar Nederland worden verstuurd en hier door vier, getrainde mensen, worden beoordeeld. De herbeoordeling heeft gevolgen voor kandidaten die volgens de computer zouden moeten zakken, maar door de menselijke beoordelaars zouden moeten slagen. Voor hen geldt het meest gunstige resultaat waardoor zij alsnog slagen. Zodra over 500 kandidaten per examen menselijke oordelen zijn verzameld, kunnen deze worden geanalyseerd en kunnen de resterende onderzoeksvragen van de resonansgroep en TNO worden beantwoord. De herbeoordeling door mensen kan worden beëindigd zodra de resultaten van dit onderzoek bekend zijn en de eventueel daaruit voortvloeiende aanpassingen zijn geïmplementeerd.

61

Is er bekend of toetsen die gebruik maken van spraaktechnologie minder of meer fraudegevoelig zijn dan toetsen die afgenomen worden door menselijke beoordelaars?

Ik ken geen onderzoek naar de relatieve fraudegevoeligheid van door computers of door mensen beoordeelde toetsen. Wel kan men constateren dat geen gelegenheid tot omkopen of beïnvloeden van de beoordelaar mogelijk is. De voornaamste voordelen van beoordeling met de computer zijn wel bekend: de beoordeling is in ieder geval objectief en kan veel sneller dan door mensen worden uitgevoerd.

62

Hoeveel tijd is er nog nodig om de toets gebruiksklaar te maken en de ambassades voor te bereiden en in te richten op het afnemen van de toetsen?

In de afgelopen periode zijn de diverse voorbereidingen voor het basisexamen inburgering getroffen; op alle posten waar een mvv kan worden aangevraagd en waar straks een examen wordt afgenomen is voorzien in een ruimte waar het examen in alle stilte kan worden afgenomen, de apparatuur die nodig is voor het examen is aangeschaft en verzonden, het personeel dat straks toezicht zal houden bij het examen is opgeleid; er is een ICT helpdesk opgericht om posten te ondersteunen met technische problemen en ook hebben er testen op alle posten plaatsgevonden of de examenapparatuur goed werkt. Volgens de Minister van Buitenlandse Zaken zijn de posten dan voorbereid en ingericht. Ook de ontwikkeling het examen is met de afronding van het wetenschappelijke onderzoek afgerond. Op dit moment wordt de laatste hand gelegd aan de voorbereidingen voor de uitvoering van de aanbevelingen van TNO. Zodra dit is afgerond kan het basisexamen inburgering in het buitenland worden ingevoerd.

63

Mensen die kosteloos het examen opnieuw kunnen afleggen moeten soms enorme afstanden afleggen om dat examen te kunnen doen. Worden ook deze reiskosten vergoed?

Wie in het geval van technische problemen in de gelegenheid wordt gesteld om kosteloos het examen opnieuw af te leggen, bevindt zich in de regel al in de buurt van de ambassade waar het examen wordt afge-

nomen. Vergoeding van reiskosten, waarvan overigens ook in andere gevallen geen sprake is, is hier dus niet aan de orde. Het is echter wel denkbaar dat het examen niet op dezelfde dag opnieuw afgenomen kan worden omdat het technische probleem niet direct verholpen kan worden of de kandidaat er zelf voor kiest om later in de week een nieuwe afspraak te maken. In deze, naar verwachting sporadisch voorkomende situatie waarbij sprake is van overmacht, zullen de eventuele verblijfskosten voor rekening blijven van de kandidaat. Uiteraard zal getracht worden de kandidaat zo spoedig mogelijk in de gelegenheid te stellen het examen opnieuw af te laten leggen.

64

De onafhankelijke wetenschappers die de invoering van de toetsen in de praktijk zullen begeleiden, zijn dat geheel nieuwe personen of zijn dit wetenschappers die ook deel hebben uitgemaakt van de resonansgroep?

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. Deze begeleidingscommissie zal niet bestaan uit wetenschappers die een rol hebben gehad in de voorbereiding van de wet of die op enige wijze in verband staan met de resonansgroep, CINOP of TNO. Ik nodig deskundigen uit op het terrein van internationaal onderwijs, diplomatie en innovatieve technologie. De commissie zal zich laten informeren door middel van de gegevens uit de monitor.

65

Wat verstaat u onder technische storingen die ertoe kunnen leiden dat de kandidaat het examen opnieuw kosteloos mag afleggen?

U zie het antwoord op vraag 47.

66

Hoe ziet de herbeoordeling er in praktijk uit? Krijgt een deelnemer ter plekke te horen dat hij niet is geslaagd en wordt dan direct een nieuw examen afgenomen door mensen te beoordelen of vindt deze herbeoordeling later plaats?

U zie het antwoord op vraag 60. De herbeoordeling heeft geen directe gevolgen voor kandidaten die volgens de beoordeling van de computer zijn geslaagd. Zij horen spoedig na het afleggen van het examen op de ambassade dat zij geslaagd zijn. Voor kandidaten die gezakt zijn betekent dit dat zij mogelijk alsnog slagen indien het panel een andere mening heeft dan de computer. Kandidaten die gezakt zijn krijgen op de ambassade te horen dat er nog geen uitslag kan worden gegeven. De uitslag ontvangen zij zo spoedig mogelijk nadat het panel klaar is met de beoordeling van het examen. Mocht blijken dat de kandidaat door het panel wel op A1- min niveau wordt geschat dan ontvangt de kandidaat bericht dat hij of zij geslaagd is voor het examen. Mocht het panel dezelfde mening zijn toegedaan als de computer dan moet een kandidaat zich beter voorbereiden en kan hij of zij na betaling een nieuwe afspraak maken met de ambassade om het examen te komen afleggen.

67

Worden in dat geval reiskosten die gemaakt worden vergoed?

Reiskosten worden in geen enkel geval vergoed.

68, 69

Volgens TNO wijkt de toetsuitslag van de KNS 10 à 15% af, zowel naar boven als naar beneden. Waarom heeft er geen vervolgonderzoek plaatsgevonden om meer zekerheid te krijgen over de afwijking in toetsuitslag? Zijn er maatregelen ondernomen om deze afwijking terug te dringen dan wel weg te nemen?

TNO constateert niet dat de toetsuitslag 10 à 15% naar boven of naar beneden kan afwijken, maar dat de totale fout, oftewel de afwijking naar boven én naar beneden, in totaal 10 à 15% kan zijn. TNO adviseert nader onderzoek te doen naar de KNS met meer data die in de praktijk zijn verzameld. Zoals ik u reeds in mijn brief van 11 november 2005 heb geïnformeerd zal ik nader onderzoek in de praktijk laten doen om onder realistische examencondities meer informatie te vergaren over de TGN en de KNS toets. Overigens is het onmogelijk een toets te ontwikkelen waarbij dit percentage nul is. Het praktijk onderzoek heeft ook ten doel te bezien of het mogelijk is dit percentage nog verder te verkleinen.

70

Een kandidaat mag, wanneer er zich technische problemen voordoen die niet aan hem/haar kunnen worden verweten, kosteloos het examen nogmaals voorleggen. Hoe moet aangetoond worden dat het de kandidaat niet kan worden verweten, en wie moet dat doen (dus: bij wie ligt de bewijslast)?

Bij technische problemen is de kandidaat zelf uiteraard de meest gereede partij om dat direct en ter plaatse aan te geven bij de aanwezige medewerker van het ministerie van Buitenlandse Zaken. De kandidaat wordt voor aanvang van het basisexamen door de toezichthouder van het ministerie van Buitenlandse Zaken geïnstrueerd dat hij of zij direct moet aangeven dat hij of zij problemen ervaart met het examen. De toezichthouder kan de technische problemen direct ter plaatse beoordelen en zal de klacht van de examen-kandidaat als een klacht in de zin van de Algemene wet bestuursrecht opvatten. Het spreekt voor zich dat technische problemen de kandidaat niet kunnen worden verweten; die heeft immers geen enkele invloed op het technisch functioneren van de examen-apparatuur. Ook spreekt het voor zich dat de kandidaat die zelf het examen onderbreekt en het basisexamen niet afrondt, daarvoor zelf verantwoordelijk is. In die gevallen heeft de kandidaat geen recht op een kosteloos nieuw examen.

71

Het lijkt hierbij alleen te gaan om technische problemen, maar wat wordt daar eigenlijk onder verstaan? Worden er ook eisen gesteld aan de omgeving waar kandidaten het examen afleggen? Hoe kan een kandidaat die problemen ondervindt met de locatie waar het onderzoek plaatsvindt hier zijn/haar recht halen?

Technische problemen zijn bijvoorbeeld storingen aan de gebruikte apparatuur of aan de verbindingen waardoor de potentiële nieuwkomer wordt verhinderd het basisexamen af te leggen of daarbij ernstig wordt belemmerd. De examens worden in een aparte ruimte afgelegd. Er wordt op toegezien dat de omstandigheden in deze ruimte qua temperatuur en dergelijke aanvaardbaar zijn, en dat er tijdens het examen geen verstoringen kunnen optreden. Op een beperkt aantal posten waar nog geen aparte spreekkamers beschikbaar zijn, zijn speciale maatregelen getroffen om de examens ongestoord te kunnen laten verlopen. Omdat examens alleen op afspraak kunnen worden afgelegd, heeft de post de regie hiervan volledig zelf in de hand. Mocht de kandidaat problemen ondervinden met de examenruimte dat dient hij of zij dat direct en ter plaatse

aan te geven bij de aanwezige medewerker van het ministerie van Buitenlandse Zaken. Deze zal dat als een klacht in de zin van de Algemene wet bestuursrecht opvatten en de kandidaat naar bevind van zaken opnieuw het basisexamen kunnen laten afleggen.

72

Is de minister voornemens een klachtenregeling en bezwaar- en beroepsprocedure in te stellen – gezien de mogelijkheid tot een kosteloos herexamen – bij technische problemen? Welke procedure dienen kandidaten bij problemen te volgen? Hoe worden kandidaten over een dergelijke regeling of procedure geïnformeerd? Waarover, wanneer en op welke manier worden kandidaten in staat gesteld om een klacht in te dienen?

Nee, ik ben niet voornemens een klachtenregeling en een bezwaar- en beroepsprocedure in te stellen daar de Algemene wet bestuursrecht hier reeds in voorziet. Artikel 8:4 onder e, van de Awb, sluit bezwaar en beroep uit, voor zover het gaat om de inhoudelijke beoordeling van de het examenresultaat – het «kennen en kunnen» van de geëxamineerde. Dit kennen en kunnen houdt tevens in dat er geen bezwaar kan worden ingediend tegen de vaststelling van de opgaven, de beoordelingsnormen en nadere regels voor de examinering. Wel is het mogelijk om tegen de examenomstandigheden bezwaar te maken. Daarnaast is het mogelijk om een klacht in te dienen bij de betreffende ambassade. Een kandidaat zal reeds voordat het examen begint worden gewezen op de mogelijkheid om direct melding te maken van onregelmatigheden, zoals storend omgevingsgeluid, ruis in de telefoonlijn, een koptelefoon die niet werkt, e.d. Een ambassademedewerker kan direct besluiten een kandidaat het examen kosteloos opnieuw af te laten leggen. Daarbij heeft het Ministerie van Buitenlandse Zaken er zoveel mogelijk aan gedaan om de examenomstandigheden optimaal te laten zijn. Zo zijn er geluidsdichte ruimtes geïnstalleerd, wordt er aan klimaatbeheersing gedaan enzovoorts. Bij de bekendmaking van de examenuitslag wordt de kandidaat gewezen op de mogelijkheid bezwaar te maken of beroep in te stellen. Tevens is het mogelijk om gebruik te maken van de reguliere bezwaar- en beroepsmogelijkheid bij afwijzing van de mvv-aanvraag.

73

Met een nauwkeurige monitor wordt het uitvoeringsproces op een fors aantal relevante aspecten gevolgd: wie is hier verantwoordelijk voor? Zal hier een «onafhankelijke» partij voor ingeschakeld worden, of wordt één van de partijen ingeschakeld die ook al bij de totstandkoming betrokken is geweest?

In mijn brief aan uw Kamer van 11 november 2005 inzake de wetenschappelijke rapportage examen buitenland heb ik aangegeven het van belang te vinden dat de prestaties van het examensysteem voortdurend nauwlettend worden gevolgd en beoordeeld. Dat zal dan ook gebeuren met behulp van een nauwkeurige monitor waarmee het uitvoeringsproces op een aantal relevante aspecten gevolgd kan worden. Deze monitor zal kwantitatieve informatie opleveren op basis van de examengegevens. Ik heb er voor gekozen deze monitor te laten uitvoeren door INDIAC, het Informatie- en analysecentrum van de IND, omdat zij reeds ruime ervaring heeft met onderzoek binnen de vreemdelingenketen en bovendien reeds de migratiegegevens verzamelen. INDIAC heeft niet eerder een rol gehad in de totstandkoming van de Wet inburgering in het Buitenland.

74

Tijdens het onderzoek zal een herbeoordeling plaatsvinden van het oordeel van de computer. Daarbij worden alle uitslagen die zich in een bepaalde marge rond de A1min norm bevinden betrokken. Waarom

kiezen ze ervoor alleen die uitslagen te kiezen die binnen een bepaalde marge vallen? Welke marge zal dit zijn? Waarom is er niet voor gekozen alle uitslagen te gebruiken? Ook iemand die buiten deze marge valt kan namelijk nadeel hebben ondervonden van het oordeel van de computer.

TNO adviseert in haar rapport om de herbeoordeling te laten uitvoeren voor *alle* examenkandidaten (en dus niet alleen van diegenen, wier uitslag zich in een bepaalde marge rond de A1-min norm bevindt) om de juistheid van de huidige instelling van de zaken slaaggrens te bepalen. Zoals ook in het antwoord op vraag 60 is beschreven, zullen de eerste kandidaten, ongeacht hun uitslag worden herbeoordeeld. In de praktijk betekent dit dat de kandidaten die kort na de inwerkingtreding van de Wet het examen afleggen en volgens de computer zouden zakken, als gevolg van een herbeoordeling door een panel van examinatoren alsnog een positieve uitslag kunnen krijgen.

75

Ook staat in dezelfde alinea dat op basis van de gegevens die tijdens de uitvoering van het onderzoek worden verzameld, de zak/slaaggrens zo nauwkeurig mogelijk zal worden bepaald. Deze zak/slaaggrens moet immers al vast staan voor er überhaupt tot examinering wordt overgegaan. Waarom kan deze grens nog worden veranderd, terwijl de examinering al lang van toepassing is?

De zak/slaaggrens is nu ingesteld op basis van de onderzoeken die door CINOP zijn uitgevoerd. Bij een zak/slaag-toets gaat het om twee typen fouten: onterecht slagen en onterecht zakken. De instelling van de zak/slaag-grens bepaalt de verhouding tussen de twee typen fouten. Als de zak/slaag-grens lager wordt ingesteld (de toets wordt makkelijker) zullen er minder kandidaten onterecht zakken, maar meer kandidaten onterecht slagen. Andersom, als de zak/slaaggrens hoger wordt ingesteld (de toets wordt moeilijker), zullen minder kandidaten onterecht slagen, maar meer onterecht zakken. Bij het optimaal instellen van de zak/slaag-grens gaat het dus om het afwegen van deze twee typen fouten. Hoe belangrijk het ène of het andere type fout is, is een beleidsmatige beslissing, maar is ook afhankelijk van de te verwachten verdeling van taalvaardigheidsniveaus. Pas na het verzamelen van praktijkdata kan de optimale zak/slaaggrens bepaald worden.