

Verantwoording
Inburgeringsexamen Kennis van
de Nederlandse Samenleving

CINOP in samenwerking met LTS en Ordinate
Ten behoeve van de opdrachtgever

Titel	Verantwoording Inburgeringsexamen Kennis van de Nederlandse Samenleving
Projectnummer	10507.01
Auteur	Marli Tijssen (CINOP) / Annet Berntsen CINOP/ John de Jong (LTS) / Matthew Lennig (Ordinate)



Pettelaarpark 1
Postbus 1585
5200 BP 's-Hertogenbosch
Tel: 073-6800800
Fax: 073-6123425
www.cinop.nl

© CINOP 2005

Niets uit deze uitgave mag worden vermenigvuldigd of openbaar gemaakt door middel van druk, fotokopie, op welke andere wijze dan ook, zonder vooraf schriftelijke toestemming van de uitgever.



Inhoud

1	Achtergrond	2
1.1	De opdracht	2
1.2	Het examenprogramma KNS	3
1.3	Toetsafname met inzet van automatische scoring	3
2	Thema's, Toetsconstruct en Examenprogramma....	4
3	De film Naar Nederland	7
3.1	De totstandkoming van de film	7
3.2	De vertalingen	8
3.3	De woordenschat in de film 'Naar Nederland'	8
4	Itemontwikkeling toetsvragen.....	9
4.1	Vooraf	9
4.2	Itemconstructie	9
4.3	Revisie van de items vóór de pretest	10
4.4	De itemselectie	10
4.5	De woordenschat in de examenvragen en antwoorden van het examen KNS ...	11
5	Opzet en uitvoering van de Pretest	13
5.1	Itembank en pretest	13
5.2	Werving van kandidaten	13
5.3	De steekproef	14
5.4	Pretestdesign	16
6	Analyse van de pretestdata.....	17
6.1	Reacties van docenten op de pretest	17
6.2	Ontwikkeling van de taalspecifieke component voor de spraakherkenning.....	17
6.3	Relatie tussen de menselijke oordelen en de spraakherkenner	18
6.4	Eerste analyse van de scores van laagtaalvaardige proefpersonen	19
6.5	Inhoudelijke itemanalyse	21
6.6	Psychometrische itemanalyse en geobserveerde P-waarden	21
6.7	Relatie van de toetsscores en enkele achtergrondgegevens: indicaties voor validiteit	22
7	Cesuurbepaling	26
8	Literatuur.....	27

1 Achtergrond

1.1 De opdracht

In het Hoofdlijnenakkoord voor het kabinet CDA, VVD, D66 van 16 mei 2003 wordt gesteld dat wie zich duurzaam wil vestigen in Nederland actief aan de samenleving moet deelnemen en zich de Nederlandse taal eigen moet maken, zich bewust moet zijn van de Nederlandse waarden, en de normen moet naleven. Iedere nieuwkomer die op vrijwillige basis naar Nederland komt en valt onder de doelgroepen van de Wet inburgering nieuwkomers, moet eerst in eigen land Nederlands op basisniveau leren en enige kennis over de Nederlandse samenleving opdoen als voorwaarde voor toelating. Eenmaal in Nederland aangekomen, moet hij of zij zich dan nog verder verdiepen in de Nederlandse taal en samenleving. Nader af te bakenen groepen oudkomers, in ieder geval zij die onvoldoende Nederlands beheersen en afhankelijk zijn van een uitkering, moeten alsnog een inburgeringsexamen halen. Asielzoekers krijgen pas een definitieve verblijfsstatus na het halen van het examen.

Om aan het gestelde in het Hoofdlijnenakkoord te kunnen voldoen, deed het Ministerie van Justitie in september 2003 een offerteaanvraag uitgaan voor de ontwikkeling van de taaltoets¹. In het voorjaar van 2004 werd besloten dat Kennis van de Nederlandse Samenleving (KNS) eveneens een onderdeel zou zijn van het inburgeringsexamen buitenland. Onderwerpen die in het examen aan de orde zouden moeten komen zijn; geografie en bevolking, geschiedenis, religie en cultuur, wonen en vervoer, wetgeving en democratie, opvoeding en onderwijs, gezondheidszorg, werk en inkomen, de eerste tijd in Nederland en de inburgeringsexamens. Deze onderwerpen zijn door de Commissie Franssen en anderen naar voren gebracht (zie noot hoofdstuk 3).

In overleg met CINOP is bekeken of examinering van dit onderdeel op dezelfde wijze uitgevoerd zou kunnen worden als het taaldeel, namelijk via de telefoon en met gebruikmaking van automatische scoring. In juni 2004 werd aan CINOP gevraagd de ontwikkeling van het examen Kennis van de Nederlandse Samenleving ter hand te nemen. In samenwerking met Ordinate en Language Testing Services (LTS) werd vervolgens een examenprogramma inclusief een toets ontwikkeld. Net zoals bij de taaltoets wordt ook voor de toets Kennis van de Nederlandse Samenleving gebruik gemaakt van automatische scoring en spraaktechnologie zoals ontwikkeld door Ordinate².

Het te ontwikkelen examen moest aan de volgende eisen voldoen:

- Het examen toetst of kandidaten kennis hebben genomen van relevante aspecten van de Nederlandse samenleving.
- Het examen moet uitvoerbaar zijn voor kandidaten op niveau A1- min van mondelinge taalvaardigheid en ook voor analfabeten.
- Het examen moet voldoen aan criteria van validiteit, betrouwbaarheid en zorgvuldigheid.

De ontwikkeling van het examen werd begeleid door een onafhankelijke Resonansgroep waarin ervaringsdeskundigen uit de doelgroep en deskundigen met kennis van de doelgroep en deskundigheid op het gebied van onderzoek, NT2 en alfabetisering vertegenwoordigd waren. Het onderhavige rapport biedt een beschrijving van het ontwikkelingsproces van het examen en zal vergezeld van het advies van de Resonansgroep ter kennis van de Tweede Kamer worden gebracht.

¹ Over de ontwikkeling van de taaltoets is een afzonderlijke rapportage verschenen

² Voor meer informatie over de spraaktechnologie verwijzen wij naar de rapportage over de taaltoets

Het examen Kennis van de Nederlandse Samenleving (hierna: KNS) in het buitenland zal in werking treden op het moment dat het wetsvoorstel in de eerste en tweede Kamer is aangenomen.

1.2 Het examenprogramma KNS

Omdat bij het taalexamen beheersing van het Nederlands op het niveau A1-min als voldoende wordt aangemerkt moet het examen van het onderdeel Kennis van de Nederlandse Samenleving ook uitvoerbaar zijn voor kandidaten die beschikken over een taalvaardigheidsniveau van A1- min in het Nederlands. Om die reden worden de kenniselementen in het examenprogramma door middel van een film aangeboden. De film is onderdeel van een voorbereidingspakket waarin ook de volledige set vragen en antwoorden over de kenniselementen in de film is opgenomen. Tijdens het examen krijgen kandidaten een selectie uit deze vragen ter beantwoording voorgelegd. Door het bekijken van de film en het bestuderen van de bijbehorende vragen en antwoorden doen de kandidaten kennis op over de Nederlandse samenleving en bereiden zich daarbij voor op het examen. Het voorbereidingspakket voor het examen wordt door een uitgeverij in opdracht van het Ministerie van Justitie gedistribueerd in binnen- en buitenland.

1.3 Toetsafname met inzet van automatische scoring

Het examen zal worden afgelegd bij de ambassadeposten in het buitenland. Aspirant Nieuwkomers dienen zelf een aanvraag tot examinering in te dienen bij de betreffende ambassade in het herkomstland. Via de eigen telefoonverbinding van het Ministerie van Buitenlandse Zaken treedt men in verbinding met de computer van Ordinate waarin het testsysteem is opgeslagen. Vervolgens vindt de testafname automatisch plaats. Na afronding van de test is de score binnen enkele minuten beschikbaar. Uitslagen worden per e-mail verzonden naar de betreffende post en naar het ministerie in Den Haag. De ambassades dragen vervolgens zorg voor bekendmaking van de uitslag aan de kandidaten. Voor meer uitgebreide beschrijving van het examen zie examenreglement.

In dit rapport wordt de ontwikkeling en evaluatie van de toets beschreven in de volgende hoofdstukken:

2. Thema's, toetsconstruct en examenprogramma;
3. Film 'Naar Nederland';
4. Itemontwikkeling;
5. Opzet en uitvoering van de pretest;
6. Analyses van de pretestdata en relatie van toetsscores met achtergrondgegevens;
7. Cesurbepaling;
8. Literatuur.

2 Thema's, Toetsconstruct en Examenprogramma

De toets heeft als doel te meten of aspirant nieuwkomers kennis hebben genomen van een set van kenniselementen over de Nederlandse samenleving. Het te toetsen domein is gedefinieerd door :

- (1) eisen van de opdrachtgever ten aanzien de representatie van kennisdomeinen,
- (2) op grond van adviezen van ervaringsdeskundigen in het onderwijs die verantwoordelijk zijn voor inburgeringstrajecten,
- (3) op grond van advies van geschiedkundigen en hoogleraren geschiedenis en de canon van Piet de Rooy,
- (4) door middel van twee informatierondes, een schriftelijke (n=35) en een mondelinge (n=40) .

Ten behoeve van de informatierondes werd op grond van de publicaties Franssen et al. (2004) en Brink et al. (2004) een inventarisatie van mogelijke onderwerpen opgesteld. Deze inventarisatie vormde het bronmateriaal voor vragenlijsten die in de informatierondes werden voorgelegd aan migranten, migrantenorganisaties, beleidsdeskundigen en een groot aantal personen op basis van hun ervaring en deskundigheid en op persoonlijke titel (n=75).

Centraal stonden twee vragen: (1) wat vindt u dat een toekomstige migrant zou *moeten* weten, en (2) wat denkt u dat een toekomstige migrant zou *willen* weten? Bij de inventarisatie van de respons op de informatierondes zijn zeven onderwerpen onderscheiden:

- Nederland: geografie, vervoer en wonen;
- Geschiedenis;
- Staatsinrichting, politiek en wetgeving;
- De Nederlandse taal en het belang van het leren ervan;
- Opvoeding en onderwijs;
- Gezondheidszorg;
- Werk en inkomen.

Uiteindelijk is door CINOP in samenwerking met het bureau Odyssee het definitieve script geschreven waarbij rekening werd gehouden met de essentie van wat echt noodzakelijk werd bevonden. Het script werd zodanig vormgegeven dat de informatie toegankelijk en boeiend zou zijn (in de vorm van een verteller) waarbij met betrekking tot het onderwerp geschiedenis is gekozen voor concentratie rondom historische figuren. Vervolgens werd het script door de opdrachtgever geaccordeerd voordat tot filmen werd overgegaan.

Operationalisatie op A1-min

Het begrijpen van informatie over een complexe samenleving als de Nederlandse - zeker voor mensen uit een niet westerse samenleving - is op het taalvaardigheidsniveau A1- min niet mogelijk tenzij er gebruik wordt gemaakt van informatiemateriaal in eigen taal (Franssen, 2004).

Voor KNS is daarom een examenprogramma ontwikkeld dat bestaat uit een film waarin de geïnventariseerde onderwerpen worden behandeld. Naast een versie in het Nederlands is de film beschikbaar in 13 andere talen. Door deze nagesynchroniseerde versies kan ongeveer 95 % van de doelpopulatie de film in de eigen taal zien. De overige 5 % zal gebruik moeten maken van een voor hen bekende andere taal of moet ondersteuning zoeken bij de partner of bij iemand in de eigen omgeving die een van de taalversies wel begrijpt. De kenniselementen uit de film worden tijdens het examen bevraagd door middel van 100 kort-antwoord vragen. De vragen bestaan uit een grafische en een auditieve stimulus. De grafische stimulus is een

foto (in feite een still uit de film) die naar een bepaald kenniselement uit de film verwijst. De auditieve stimulus is de tekst van de kort-antwoord vraag. Er is gekozen voor een toetsvorm naar analogie van het rijexamen. Er is een voorbereidingspakket waarin naast de film een geluiddrager is opgenomen met de volledige set van 100 vragen inclusief de antwoorden. De vragen en antwoorden zijn ingesproken door professionele stemacteurs. Tijdens het examen wordt een selectie van 30 vragen aangeboden. De grafische stimuli staan afgebeeld in een examenboekje en de auditieve stimuli worden ten gehore gebracht via de telefoon. Kandidaten beantwoorden de vragen ook via de telefoon.

Het voorbereidingspakket bestaat uit:

- de film 'Naar Nederland' in 14 talen waaronder het Nederlands;
- opgavenboekje met 100 foto's (stills) uit de film;
- een audiocassette met 100 vragen en antwoorden bij de film;
- een papieren versie van de 100 vragen en antwoorden;
- drie oefentoetsen voor de taaltoets;
- een handleiding voor de kandidaten over de wijze van voorbereiding op zowel de Taal- als de KNS toets.

Wat betreft het taalgebruik in de vragen is rekening gehouden met de beschrijving van het niveau A1-min zoals gehanteerd door de Commissie Franssen en overgenomen voor de ontwikkeling van de taaltoets.

De Commissie Franssen (Franssen, 2004) beschrijft het beoogde beheersingsniveau luisteren en spreken als volgt:

Luisteren

Kan slechts een beperkt aantal vertrouwde woorden en basiszinnen begrijpen die betrekking hebben op de directe, persoonlijke levenssfeer en op de allereerste levensbehoeften; en alleen in direct contact met Nederlandssprekenden die gewend zijn te converseren met anderstaligen, wanneer zij langzaam en duidelijk spreken.

Spreken

Kan zich slechts in zeer beperkte mate uitdrukken, eigenlijk alleen met behulp van losse woorden en standaardformuleringen ('formulaic speech'), op een gering aantal terreinen die verband houden met de directe, persoonlijke levenssfeer.

Het is helder dat de commissie hiermee doelt op een niveau lager dan het niveau A1 op de schaal van Raad van Europa. North (2000) heeft in zijn onderzoek waarop de schaling van het CEF is gebaseerd inderdaad descriptoren voor een niveau onder A1 aangetroffen. Bij de publicatie van CEF is dit echter niet in de schaal opgenomen, omdat het aantal descriptoren te gering werd bevonden. Uitgaande van het advies van de Commissie Franssen is eerder ten behoeve van de ontwikkeling van het inburgeringsexamen Nederlands in het buitenland op grond van het beperkte aantal descriptoren genoemd door North (2000) de volgende omschrijving van het niveau A1-min geformuleerd:

Spreken A1-min

Kan zichzelf verstaanbaar maken met een beperkt aantal zeer frequente losse woorden en standaardformuleringen (formulaic speech). Uitingen bestaan uit 2 of 3 losse woorden en bestaan vaak uit herhalingen van woorden van de gesprekspartner. Van coherentie is niet of nauwelijks sprake. Uitspraak is vaak onverstaanbaar, gesprekspartner moet zeer veel moeite doen om uitingen te begrijpen.

Luisteren A1- min

Begrijpt een beperkt aantal vertrouwde woorden en basiszinnen die betrekking hebben op de directe persoonlijke levenssfeer en ervaringswereld, mits de gesprekspartner zeer zorgvuldig spreekt en veelvuldig gebruik maakt van herhalingen, omschrijvingen en gebaren en zich beperkt tot het gebruik van in de situatie voorspelbare en – voor de toehoorder bekende – losse woorden en basiszinnen

Constructie van de vragen en antwoorden

Bij de constructie van de vragen en antwoorden van het examen KNS is bovengenoemde beschrijving als uitgangspunt genomen. De condities van het examen zijn evenwel zodanig vormgegeven dat beheersing van dit niveau toereikend is om het examen af te leggen. Gezien de beperking van taalvaardigheid is het noodzakelijk vaste, gesloten vragen en antwoorden (formulaic speech) te construeren waarmee kandidaten vooraf kunnen oefenen. Door oefening gaan deze vragen en antwoorden deel uitmaken van de persoonlijke ervaringswereld. Daarnaast noodzaakt het taalvaardigheidsniveau dat een context wordt geboden waardoor het begrijpen van de gesproken taal wordt vergemakkelijkt. Om die reden zijn de vragen en antwoorden ontleend aan de informatie die in de film wordt verstrekt en worden tijdens het examen stimuli verstrekt in de vorm van foto's (stills) uit de film. Rekening houdend met een mogelijk laag taalvaardigheidsniveau van de kandidaten is het woordgebruik in de vragen beperkt en zijn de vragen duidelijk en langzaam ingesproken, opgedeeld in betekenisvolle eenheden en met duidelijk herkenbare woordgrenzen.

3 De film Naar Nederland

De film 'Naar Nederland' vormt de basis voor het examenprogramma KNS.

De film is gemaakt door het productiebedrijf Odyssee Producties, in samenwerking met CINOP en de opdrachtgever (het Ministerie van Justitie).

Vertrekpunt bij de keuze van onderwerpen voor de film was de inventarisatie van het kennisdomein op grond van Franssen (2004), Brink (2004) en de gegevens uit de informatierondes. Conform de zeven onderwerpen die tijdens de inventarisatie waren onderscheiden is de film ingedeeld in zeven afleveringen:

- Nederland: geografie, vervoer en wonen;
- Geschiedenis;
- Staatsinrichting, politiek en wetgeving;
- De Nederlandse taal en het belang van het leren ervan;
- Opvoeding en onderwijs;
- Gezondheidszorg;
- Werk en inkomen.

Op verzoek van de opdrachtgever is een achtste onderwerp toegevoegd om de kandidaten een indruk te geven hoe de afname van het examen op de ambassade in zijn werk gaat. Over dit deel van de film worden geen vragen gesteld in het examen.

- Hoe maak ik het examen op de ambassade?

3.1 De totstandkoming van de film

In het kader van de filmproductie is op diverse punten overleg geweest en overeenstemming bereikt tussen CINOP, productiebedrijf en opdrachtgever van de film i.c. het ministerie van Justitie. Vertrekpunt voor de film was om - vanuit het perspectief van een (niet zo kansrijke) migrant uit alle mogelijke delen van de wereld - een informatief en realistisch beeld van Nederland te geven. Odyssee Producties is een productiebedrijf dat '...op een dynamische manier creatief en compact vormgeeft aan onderwerpen die in eerste instantie complex aandoen'. Een streven daarbij is: 'Er moet een avontuurlijk verfrissend karakter om een project heen hangen, waarin ernst en humor elkaar mogen afwisselen, en waarin een dynamische vorm ervoor zorgt dat de kijker wordt meegenomen, in plaats van dat hij of zij er afstandelijk naar kijkt.' Belangrijke beslissingen binnen het door de opdrachtgever gestelde kader zijn geweest:

- de keuze voor een aansprekende presentatrice/verteller: een vriendelijke gids in een nieuw land;
- de keuze voor fragmenten uit interviews met migranten en enkele Nederlanders; een identificatie- en rustmoment voor de kijker;
- de keuze voor het gebruik van humoristische fragmenten: een man maakt een aantal dingen mee in de rol van nieuwe migrant, die in hun overdrijving een komisch effect hebben op de kijker;
- de keuze voor de delen 'Hoofdpijnen Nederlandse Geschiedenis' en 'Staatinrichting, politiek en wetgeving'. Voorstellen hiertoe zijn door CINOP aan opdrachtgever en aan externe deskundigen voorgelegd voor commentaar en aanvulling.

Op basis van het bovenstaande is door Odyssee Producties het script geschreven.

CINOP, beleidsambtenaren van het ministerie van Justitie en enkele leden die betrokken waren bij de eerdere inventarisatierondes, lazen het script en corrigeerden op correctheid van de inhoud en de toegankelijkheid van het taalgebruik. In een tijdsbestek van 4 weken is de film opgenomen en in een tijdsbestek van 10 weken gemonteerd. Tijdens de montage-

periode zijn er twee momenten geweest waarop CINOP en de opdrachtgever inspraak hebben gehad. Daarop zijn door Odyssee Producties fragmenten verwijderd, veranderd en toegevoegd (zie ook Verduyn, 2005).

Opdrachtgever heeft besloten twee versies van de film uit te brengen. Een oorspronkelijke versie waarin de openheid van de Nederlandse samenleving expliciet in beeld is (beelden van een strand met een vrouw met ontbloot bovenlichaam, een popconcert en twee zoenende mannen), en een ‘gekuiste’ versie, waarin deze 3 beelden niet voorkomen. Dit omdat het in bezit hebben van dergelijk beeldmateriaal in een aantal landen tot strafvervolgning zou kunnen leiden. Over deze beelden worden geen vragen gesteld. De film is vervolgens rechtenvrij aan het ministerie van Justitie overgedragen.

3.2 De vertalingen

De film ‘Naar Nederland’ is in 13 talen vertaald en nagesynchroniseerd: in het Engels, Frans, Spaans, Portugees, Russisch, Bahasa Indonesisch, Turks, Kurmanci (Koerdisch zoals in Turkije gesproken) Marokkaans-Arabisch, standaard Arabisch, Berbers (het Riffijns zoals in Noord-Marokko gesproken), Thais en Chinees (Mandarijn). Door de keuze van deze talen kunnen ongeveer 95% van de mensen die zich zullen willen voorbereiden op het examen de informatie in hun eigen taal volgen. Voor ongeveer 5% zal dat niet mogelijk zijn. Zij zullen van een andere taal gebruik moeten maken of van de Nederlandstalige versie, en daarbij de hulp inroepen van de partner of iemand in de eigen omgeving die de taal wel begrijpt.

Vervolgens werd zorggedragen voor zorgvuldige vertalingen en nasynchronisaties, met veel aandacht voor de presentatie. In elke nasynchronisatie is zowel een mannen- als een vrouwenstem te horen. Alle vertalingen zijn door CINOP en moedertaal-sprekers gecontroleerd op juistheid en toegankelijkheid van het woordgebruik.

3.3 De woordenschat in de film ‘Naar Nederland’

De tekst van de film bestaat uit 7085 woorden, 76% van al deze woorden behoort tot de 500 meest frequente woorden in het Corpus Gesproken Nederlands (CGN). Hierdoor kan de film ondersteuning bieden bij de verwerving van het Nederlands, zeker in combinatie met de beelden en een vertaalde versie. De delen 2 (geschiedenis) en 3 (staatsinrichting e.a.) bevatten een groter aantal woorden dan de overige delen. Gezien de onderwerpen is dat niet onverwacht. Op de woorden in het KNS-examen (vragen en antwoorden) komen we terug in paragraaf 4.5.

4 Itemontwikkeling toetsvragen

4.1 Vooraf

Bij het ontwikkelen van de toets Kennis van de Nederlandse Samenleving moest met een aantal bijzondere factoren rekening worden gehouden:

- Het beoogde civiele effect van de Nederlandse toets is groot: kandidaten leggen de toets af in het kader van hun aanvraag voor een Machtiging tot Voorlopig Verblijf in Nederland.
- De toets moet wereldwijd worden afgenomen op Nederlandse posten en ambassades.
- De doelgroep is zeer heterogeen wat betreft achtergrondkenmerken zoals moedertaal, opleidingsniveau en etniciteit. De toets moet bruikbaar zijn voor mensen die niet geletterd zijn.
- De toets moet uitvoerbaar zijn voor kandidaten met een niveau A1 min van taalvaardigheid in het Nederlands.

Daarnaast zijn bij de constructie van de vragen voor het examen Kennis van de Nederlandse Samenleving nog een aantal uitgangspunten in acht genomen. De belangrijkste is dat alle (100) items bekend en beschikbaar zijn in audio en op papier in de vorm van oefenmateriaal. Kandidaten kunnen zich dus vooraf op de toets voorbereiden. De antwoorden moeten gegeven kunnen worden in het Nederlands door mensen met een niveau A1-min wat de Nederlandse taal betreft (zie ook hoofdstuk 2). Met het beantwoorden van de vragen wordt kennis van de Nederlandse Samenleving getoetst in de veronderstelling dat de kandidaten kennis genomen hebben van informatie uit de film 'Naar Nederland' in hun eigen taal.

Tijdens het examen beantwoordt de kandidaat 30 items. Deze 30 vragen vormen één van de 10 subsets/varianten. Ieder subset/variant bevat een andere vaste selectie uit de itembank van 100 vragen. De 30 vragen worden via de telefoon gesteld en worden begeleid door een foto (stills uit de film) in een examenboekje, dat kandidaten bij de afname op de ambassade verstrekt wordt (zie examenreglement). De 30 vragen zijn een selectie uit dezelfde vragen en antwoorden en foto's als waarmee de kandidaat heeft kunnen oefenen.

4.2 Itemconstructie

Bij de itemconstructie zijn deze uitgangspunten als volgt uitgewerkt:

Inhoud van de items

- De items vragen naar kennis over Nederland en sluiten aan bij de kern van de informatie uit de film 'Naar Nederland'.
- Aan het gevraagde in de items moet herkenbaar en duidelijk aandacht geschonken zijn in de film.
- Het item verwijst naar wat er verteld is in de film. In het item wordt geen andere of nieuwe woordenschat gebruikt.
- De items gaan over alle 7 delen van de film.
- De volgorde van de items loopt parallel aan die van de onderwerpen in de film.

Uitvoerbaarheid voor kandidaten met een taalniveau A1-min

- De vraag wordt zo eenvoudig mogelijk gesteld, rekening houdend met een taalniveau van A1-min.
- Bij de uitspraak van de vragen wordt dit criterium gerealiseerd door een langzaam spreektempo, in spreektaal, een duidelijke articulatie, in betekenisvolle eenheden en met inachtneming van woordgrenzen bij de opname van alle mondeling aangeboden vragen.

Bijvoorbeeld: In Nederland / wonen daar / véél mensen / of / weinig mensen?

- Het item moet via één of enkele woorden eenduidig te beantwoorden zijn.
- Het gevraagde veronderstelt dat kandidaten vooraf geoefend hebben met de (100) vragen en antwoorden via het fotoboek en de oefencassette/DVD; via de vragen en antwoorden op papier; via oefening (bij voorkeur aan de telefoon) met hun partner of andere bekenden of een docent in het land van herkomst.
- Het item moet visueel ondersteund zijn door middel van een foto (still uit de film).
- Het item moet betrekking hebben op informatie die in de film aan de orde is geweest.

Conditie voor de spraakherkenner

De antwoorden moet herkenbaar voor de spraakherkenner geproduceerd kunnen worden door een kandidaat met een beheersing van het Nederlands op niveau A1-min.

- Het item moet eenduidig te beantwoorden zijn.
- Het gevraagde mag geen aanleiding geven tot variaties.
- Het gevraagde mag geen aanleiding geven tot grote uitspraakproblemen.

Aard van de items

De items bestaan uit 3 typen:

- (1) Gesloten vragen met een ja/nee antwoord
Voorbeeldvraag: Zijn de kranten, radio en televisie vrij in hun mening?
Antwoord: Ja
- (2) Gesloten vragen met twee antwoordmogelijkheden (of...of vraag waarbij het goede antwoord geselecteerd en gereproduceerd moet worden)
Voorbeeldvraag: Kijk naar de foto. Wie is dit? Willem van Oranje of Prinses Maxima?
Antwoord: Willem van Oranje
- (3) Open vragen met een eenduidig antwoord.
Voorbeeldvraag: Wat zijn de kleuren van de Nederlandse vlag?
Antwoord: Rood, wit, blauw

4.3 Revisie van de items vóór de pretest

Een revisie van de items is door CINOP-medewerkers uitgevoerd door de vragen te screenen op de hierboven genoemde criteria. Ook de leden van de resonansgroep is gevraagd hun commentaar te geven op de relevantie en de formulering van de vragen, voorafgaand aan de pretest. Er is besloten om pas tot itemselectie over te gaan na de pretest.

Voor de pretest zijn 148 items geconstrueerd. Alle items zijn gekoppeld aan een foto uit de film. Alle items alsmede de toets- en iteminstructies werden vervolgens opgenomen in een professionele geluidsstudio onder regie van een van de itemconstructeurs en een geluidstechnicus. De toetsinstructies zijn ingesproken door een mannelijke spreker. De vragen werden vervolgens ingesproken door een vrouwelijke spreker en een (andere) mannelijke spreker heeft de antwoorden ingesproken.

4.4 De itemselectie

De itemselectie op basis van de pretestgegevens heeft op verschillende niveaus plaatsgevonden en uiteindelijk geresulteerd in een itembank van 100 vragen en antwoorden.

Bij die selectie zijn diverse afwegingen gemaakt: op basis van de psychometrische analyse, op basis van inhoudelijke criteria, op basis van uitvoerbaarheid en op basis van de relevantie van de vragen, zoals die door CINOP, door leden van de resonansgroep en door docenten die de pretest uitvoerden aangegeven werden. Het verslag van deze analyse wordt gedaan in hoofdstuk 5.

4.5 De woordenschat in de examenvragen en antwoorden van het examen KNS

De hele tekst van het examen KNS (vragen en antwoorden) bevat in totaal slechts 1571 woorden, waarvan er vele meer malen voorkomen. Bijvoorbeeld, komt het woord “de” 80 maal voor, Nederland 40 keer en Nederlandse 10 keer. Tellen we het aantal verschillende woorden dan komen we uit op 492 woorden (productief en receptief). Ruim de helft hiervan behoort dus tot de meest frequente 500 woorden uit het CGN de andere helft gaat daarboven uit. Veel van deze boven-500 woorden worden in het examenprogramma aangeboden, slechts een zeer beperkt aantal hoeft productief te worden beheerst.

4.5.1 A1-min en de KNS-toets

In de beschrijving van niveau A1-min staat:

Spreken A1-min

*Kan zichzelf verstaanbaar maken met een **beperkt aantal zeer frequente losse woorden en standaardformuleringen (formulaic speech)**. Uitingen bestaan uit 2 of 3 losse woorden en bestaan vaak uit herhalingen van woorden van de gesprekspartner. Van coherentie is niet of nauwelijks sprake. Uitspraak is vaak onverstaanbaar, gesprekspartner moet zeer veel moeite doen om uitingen te begrijpen.*

Luisteren A1- min

*Begrijpt een beperkt aantal **vertrouwde woorden en basiszinnen die betrekking hebben op de directe persoonlijke levenssfeer en ervaringswereld**, mits de gesprekspartner zeer zorgvuldig spreekt en veelvuldig gebruik maakt van herhalingen, omschrijvingen en gebaren en zich beperkt tot het gebruik van in de situatie voorspelbare en – voor de toehoorder bekende – losse woorden en basiszinnen.*

Er is dus sprake van

- a) zeer frequente woorden
en/of
- b) vertrouwde woorden, die betrekking hebben op de directe persoonlijke levenssfeer en ervaringswereld.

We nemen als uitgangspunt dat om te functioneren op A1-min ongeveer 500 woorden³ voldoende zijn.

Een groot deel van de woorden in de toets KNS voldoen aan voorwaarde a: ze zijn zeer frequent. Een aantal andere woorden voldoet niet aan die voorwaarde. Bijvoorbeeld het woord “grondwet” en het woord “Turkije” behoren niet tot de 500 meest frequente woorden. Deze woorden moeten dus voldoen aan voorwaarde b) om gebruikt te mogen worden in de KNS toets.

Maar wat zijn dat “vertrouwde woorden die betrekking hebben op de directe persoonlijke levenssfeer en ervaringswereld”? Ieders persoonlijke levenssfeer en ervaringswereld verschilt immers van die van andere taalgebruikers. Een timmerman heeft een andere ervaringswereld dan een bakker, en een accountant heeft een andere ervaringswereld dan een violist. Voor ieder van deze mensen behoort dus een andere set van woorden tot de

³ Dit aantal is genoemd door Minister Verdonk

“vertrouwde woorden uit de persoonlijke levenssfeer en ervaringswereld”. Voor de een is een duimstok zeer vertrouwd en voor de ander een bakblik, voor de een behoort balans tot de persoonlijke ervaringswereld en voor de ander strijkstok. Dat lijkt dus een probleem. Maar wij mogen ervan uitgaan, dat kandidaten die deelnemen aan de KNS toets, kennis hebben genomen van de film “Naar Nederland” en van het overige materiaal in het oefenpakket. In die film komt het woord “grondwet” wel 10 keer voor. En ook in het oefenpakket met 100 vragen komen kandidaten die zich voorbereiden op het examen het woord “grondwet” nog een paar maal tegen. We kunnen eigenlijk zeggen dat we door middel van de film en het overige materiaal in het oefenpakket hebben duidelijk gemaakt aan de kandidaten dat zij vertrouwd MOETEN zijn met het woord “grondwet”. Ook kun je zeggen dat we voor alle kandidaten een gemeenschappelijke ervaringswereld hebben gecreëerd. Dus voldoet het woord “grondwet” aan voorwaarde b).

De omvang en de aard van de actieve woordenschat die nodig is om de 100 vragen van het KNS-examen te beantwoorden, is als volgt geanalyseerd:

Er zijn in totaal 120 verschillende woorden nodig om alle 100 vragen correct te beantwoorden (productieve woordenschat). Dit zijn alle woorden, inclusief eigennamen en telwoorden.

Van die 120 verschillende woorden:

- zijn 35 woorden eigennamen, die niet voorkomen bij de 500 meest frequente woorden in het Corpus Gesproken Nederlands;

Maar van die 120 woorden komen er:

- 84 gemiddeld 4,3 ook voor in de vragen;
- 103 gemiddeld 22 maal voor in de tekst van de film Naar Nederland;
- 61 gemiddeld 30 keer voor in de itebank van de taaltoets.

Met ander woorden: er zijn veel minder dan 500 woorden nodig om de KNS toets te kunnen doen, namelijk er zijn maar 120 woorden nodig. Niet alle 120 woorden behoren tot de 500 meest frequente woorden. Zij voldoen dus niet allemaal aan voorwaarde a). Maar als zij niet aan voorwaarde a) voldoen, voldoen ze wel aan voorwaarde b): ze worden vaak in de film en/of in andere delen van het oefenmateriaal gebruikt en behoren daarmee tot de – weliswaar opgelegde – ervaringswereld van de kandidaten.

5 Opzet en uitvoering van de Pretest

5.1 Itembank en pretest

Na de hiervoor in hoofdstuk 4 geschetste procedures bevatte de itembank ten behoeve van de pretest in totaal 148 vragen en antwoorden over de film die aan de noodzakelijke criteria voldeden. Bij voorbaat was duidelijk dat slechts korte tijd beschikbaar was om pretestkandidaten voldoende kennis te laten nemen van de film en de vragen. Daarom werd besloten bij de voorbereiding op de pretest de set van “te leren vragen” te beperken tot 30 vragen per kandidaat. Er werden 10 toetsboekjes ontwikkeld, elk met 30 foto’s (stills) uit de film met de daarbij horende vragen en antwoorden, verspreid over de verschillende onderwerpen uit de film. De toetsboekjes werden zodanig verspreid over de deelnemende instellingen dat elk item voldoende responsen zou opleveren .

5.2 Werving van kandidaten

Bij het samenstellen van de steekproef ten behoeve van het pretesten van de items moest met een aantal punten rekening worden gehouden.

In de eerste plaats was vanaf de start van het project duidelijk dat de pretest in principe in het buitenland zou moeten plaatsvinden om een steekproef te kunnen samenstellen die 100% representatief was voor de doelgroep van de toets: aspirant nieuwkomers die het inburgeringsexamen buitenland afleggen als verplicht onderdeel van de procedure voor het aanvragen van een Machtiging tot Voorlopig Verblijf in Nederland. Gezien het korte tijdspad van ontwikkeling bleek dat niet haalbaar.

De pretest zou daarom in Nederland moeten plaatsvinden. Om toch de pretestsituatie zoveel mogelijk vergelijkbaar te maken met de authentieke toetsituatie, is voor de pretest met name gezocht naar kandidaten die pas kort in Nederland waren en die nog over weinig taalvaardigheid in het Nederlands beschikten. Om voldoende aantallen kandidaten te bereiken zijn, net als bij de taaltoets, opnieuw Regionale Opleidingen Centra verspreid over het land, benaderd.

Een bijkomend knelpunt bij het pretesten was dat het hele project onder hoge tijdsdruk ontwikkeld moest worden. De film was weliswaar al beschikbaar maar nog niet alle versies in de verschillende talen. Daardoor konden sommige pretest kandidaten de film slechts in het Nederlands bekijken.

In de winter van 2004 zijn door CINOP in totaal 11 ROC’s benaderd met het verzoek mee te werken aan de pretest. De betrokken ROC’s werden persoonlijk geïnformeerd over de achtergrond en inhoud van het project, de kenmerken en doelen van de te ontwikkelen toets en de opzet van de pretest.

Bovendien werden de tegenprestaties beschreven die CINOP zou leveren:

- elke pretestkandidaat zou per afgelegde pretest een bedrag van 10 Euro ontvangen;
- elke coördinator zou een tegoedbon ontvangen ter waarde van 50 euro, daarnaast zou de instelling te zijner tijd een gratis exemplaar ontvangen van het examenpakket.

De meewerkende docenten hebben aan de hand van door CINOP verstrekte materialen, hun cursisten geïnformeerd over de doelen en achtergronden van de toets waarvoor hun medewerking werd gevraagd. CINOP heeft elke school die bereid was aan het project deel te nemen persoonlijk een set pretestmaterialen bezorgd waarbij werd uitgegaan van een maximale respons. Bij de bezorging van de set pretestmaterialen werd aan de deelnemende

scholen gevraagd een contract te tekenen voor geheimhouding en retourgarantie van de pretestmaterialen.

Voor elke NT2-docent waren instructie en materialen beschikbaar ten behoeve van de afname van de pretest in elk van zijn/haar lesgroepen, waarbij de omvang van een lesgroep werd geschat op 20 deelnemers. Coördinatoren wisten van te voren in welke talen versies van de film beschikbaar waren en konden bij de selectie van de pretestdeelnemers met dit gegeven rekening houden.

Het is duidelijk dat meer pretestmaterialen werden verspreid dan er pretestkandidaten bereikt zouden worden: kandidaten konden weigeren, de omvang van lesgroepen in het volwassenenonderwijs loopt zeer uiteen (zeker in alfabetiseringsgroepen is het aantal deelnemers veel geringer) en ook het aantal aanwezige cursisten kan wisselen per dag.

In het kader van het streven naar een steekproef van vooral laag taalvaardigen en analfabeten is in elk ROC nadrukkelijk gevraagd om met name ook deelnemers aan alfabetiseringscursussen te benaderen voor deelname aan de pretest.

5.3 De steekproef

Van alle deelnemers werden ook een aantal achtergrondgegevens gevraagd met betrekking tot geslacht, opleidingsachtergrond, taalvaardigheidsniveau, nationaliteit, jaren van verblijf in Nederland, aantal uren NT2, wel of niet gealfabetiseerd, recent behaalde scores op andere NT2 toetsen en de inschatting van het CEF-niveau door de docent voor de vaardigheden spreken en luisteren.

Helaas vulden niet alle docenten voor iedere deelnemer achtergrondgegevens in. De omvang van de achtergronddata verschilt per instelling, docent en kandidaat. Van de gegevens die in principe voor alle kandidaten konden worden geleverd zijn de meeste gegevens beschikbaar voor land van herkomst (n=755) en de minste voor het door de docent ingeschatte CEF-niveau voor taalvaardigheid (n=437). In totaal hebben iets meer dan 1000 deelnemers deelgenomen aan de pretest KNS.

Enkele achtergrondgegevens van de kandidaten (N = 1000)

Het totaal aantal deelnemende proefpersonen bedroeg ongeveer 1000 cursisten. De aantallen waarover gegevens zijn ingevuld verschillen echter per variabele. Daarom vermelden we steeds de totalen waarover gegevens beschikbaar zijn. Het exacte aantal “missing” is niet bekend maar bedraagt ongeveer 1000 min de opgegeven n.

Geslacht

In de pretest hebben meer vrouwen dan mannen deelgenomen. Er is niet doelbewust op geslacht geselecteerd maar er nemen nu eenmaal meer vrouwen dan mannen deel aan onderwijs in Regionale Opleidingscentra. Uit de beschikbare gegevens blijkt dat 228 mannen versus 527 vrouwen aan de pretest hebben deelgenomen (n=755).

Herkomstland

Wat betreft land van herkomst waren de pretestdeelnemers afkomstig uit 96 verschillende landen (n=755).

De meest frequente landen waren:

Marokko	156
Turkije	122
Irak	59
Afghanistan	42
Thailand	25
Indonesië	23
China	19
Saoudi	
Arabie	17
Somalie	17
Polen	16
Iran	15
Angola	12
Sri Lanka	12

Verblijfsduur in Nederland

Ongeveer 295 van de 722 kandidaten van wie gegevens beschikbaar waren, verbleven minder dan twee jaar in Nederland . Daarvan waren er 91 minder dan 1 jaar in Nederland. De overigen waren verdeeld over een groot aantal verschillende jaren van verblijf.

Taalvaardigheidsniveau

Hieronder in tabel 2 het taalvaardigheidsniveau conform het CEF zoals geschat door docenten (n=437). 63% van de pretestkandidaten waarover gegevens beschikbaar zijn heeft een taalvaardigheidsniveau van A1 of lager.

Tabel 2: Ingeschat niveau (n=437)

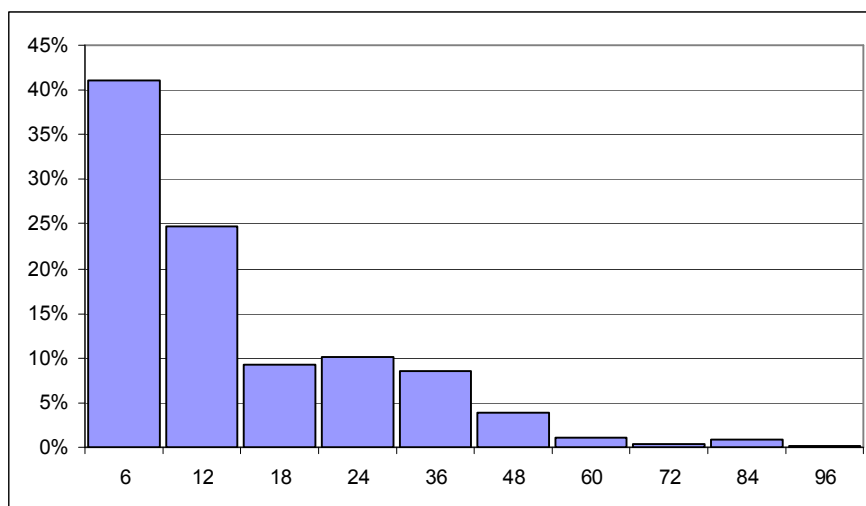
Niveau	Aantal	Percentage	Cumulatief
Onder A1-	17	4%	4%
A1-min	74	17%	21%
A1	182	42%	63%
A2	137	31%	94%
B1	26	6%	100%
B2	1	0%	100%
C1	0	0%	100%
C2	0	0%	100%

Analfabeten

Het was uiteraard van belang om aan de pretest zoveel mogelijk analfabeten te laten deelnemen. Het korte tijdsbestek voor de afname van de pretest leverde aanzienlijke knelpunten op met betrekking tot de voorbereiding met deze doelgroep. Uiteindelijk blijkt uit de beschikbare informatie dat 72 analfabeten aan de pretest hebben deelgenomen versus 600 geletterden (n=672).

Onderwijs in Nederlandse taal (NT2)

Circa 40 % van de deelnemers had 6 maanden of minder NT2 onderwijs in Nederland genoten, 25 % 6 tot 12 maanden en de resterende 35 % waren verdeeld over 1,5 tot 5 jaar onderwijs (n= 621).



Figuur 1: Aantal maanden NT2 onderwijs.

Opleidingsniveau

Circa 25 % (181) van de pretest deelnemers (n=715) had niet meer dan 6 jaar onderwijs genoten in eigen land, circa 50 % (358) had tussen de 6 en 12 jaar onderwijs genoten en circa 25 % (176) had meer dan 12 jaar onderwijs gevolgd.

5.4 Pretestdesign

Opzet dataverzameling

Voor de pretest KNS zijn alleen allochtone kandidaten benaderd. Alle deelnemers aan de pretest kregen een subset van 30 vragen uit de totale set van 148 vragen. Deze 30 vragen en antwoorden, inclusief de bijbehorende foto's (stills) uit de film, waren aan de kandidaten ter voorbereiding beschikbaar gesteld. Het was de bedoeling dat docenten met de pretest deelnemers naar de film in het Nederlands zouden kijken, dat de deelnemers vervolgens de film meerdere malen in de eigen taal zouden bekijken (zelfstandig of in groepjes) en dat de deelnemers zelfstandig of in de groep met de vragen en antwoorden zouden oefenen. Vervolgens zou op een bepaald moment de toetsafname telefonisch plaatsvinden. Uitgaande van minimaal 100 responsen per item zouden er gelet op het design 500 deelnemers nodig zijn. Omdat echter uit ervaring bekend is dat toezeggingen van deelname meestal niet worden gerealiseerd, is bij de werving uitgegaan van minimaal 750 toegezegde deelnames. Uiteindelijk bleken meer dan 1000 kandidaten aan de pretest te hebben deelgenomen waardoor het minimale aantal responsen ruimschoots werd gerealiseerd.

Doelstelling dataverzameling

De verzameling van data tijdens de pretest diende vier afzonderlijke doelen:

1. reactie van deelnemende docenten op aard en inhoud van de vragen en antwoorden;
2. ontwikkeling van de taalspecifieke component voor de spraakherkenning;
3. toets- en itemanalyse;
4. indicaties voor validiteit van de toetsscores.

6 Analyse van de pretestdata

6.1 Reacties van docenten op de pretest

Op 11 ROC's hebben ruim 25 docenten het mogelijk gemaakt dat er in 54 groepen NT2-cursisten de film 'Naar Nederland' is bekeken. Er is geoefend met de antwoorden op de 30 vragen en de antwoorden zijn ingebeld. Via een evaluatieformulier is docenten gevraagd hun indruk te geven wat pretest-kandidaten vonden van de film, de voorbereiding op de toets en op de toetsvragen zelf.

Over de pretest merken docenten het volgende op:

De film 'Naar Nederland' is serieus bekeken en er is in de meeste groepen meerdere malen geoefend met de 30 vragen. In slechts één op de drie groepen is de film ook in de eigen taal bekeken. De toetsprocedure (het inbellen) heeft niet tot grote problemen geleid. Pretest-kandidaten waren enerzijds serieus en soms zenuwachtig voor het doen van de pretest, maar er is wisselend op de vragen gereageerd. Sommigen vonden de vragen te gemakkelijk, maar leuk om te doen. Anderen vonden het moeilijk of niet alle vragen relevant. De indruk is dat cursisten redelijk zelfstandig en met enthousiasme de pretest hebben voorbereid en uitgevoerd. Docenten die met name aan laag opgeleide en analfabete cursisten les geven, waren het meest kritisch over de toets. Hun opmerkingen betreffen het hoge receptieve niveau van het Nederlands en de aard van de vragen. In hun observaties zagen zij dat dit ertoe leidde dat kandidaten antwoorden uit het hoofd leren zonder de vragen te begrijpen. Een aantal pretest-items is door meerdere docenten aangemerkt als minder relevant. De reacties van de docenten zijn gebruikt als één van de selectiecriteria bij het samenstellen van de definitieve itembank.

6.2 Ontwikkeling van de taalspecifieke component voor de spraakherkenning

De verzameling van de pretestgegevens heeft als eerste doel de taalspecifieke component voor de spraakherkenning te ontwikkelen. De verzamelde spraakdata voor de Nederlandse toets omvatte 25.800 uitingen van proefpersonen afkomstig uit alle werelddelen en uit 96 verschillende landen.

Het doelwit van spraakherkenning in de gegeven context is de predictie van antwoorden van kandidaten in de perceptie van interactanten met deze kandidaten. Uiteindelijk vormen daarom menselijke oordelen de norm waarvoor moet worden geoptimaliseerd.

De maat voor correctheid wordt machinaal berekend met behulp van de spraakherkenner die op grond van de door de pretest verkregen data, specifiek is getraind op de antwoorden van de kandidaten.

In het systeem van Ordinate is specifieke software ontwikkeld om snel en doeltreffend data te verzamelen en daarmee modellen te bouwen die het mogelijk maken continue spraak zelfs met de vaak zeer afwijkende uitspraakvarianten van kandidaten te herkennen. Hiermee is de correctheidsmaat onafhankelijk gemaakt van de kwaliteit van de spraak. Voor de KNS toets is uitspraak slechts in zoverre van belang dat menselijke beoordeling- gelet op het feit dat de goede en foute antwoorden bekend zijn - moet kunnen herkennen of de kandidaat het goede antwoord heeft gegeven. De spraakherkenner is op alle uitspraakvarianten van de goede antwoorden die uit de pretest kwamen, getraind.

Voor de correctheidsmaat waarop de score is gebaseerd zijn verschillende menselijke oordelen verzameld. Allereerst zijn de responsen van proefpersonen getranscribeerd door getrainde transcribenten (moedertaalsprekers van het Nederlands) werkend met specifiek daartoe ontwikkelde software. Wat de transcribenten tonen te kunnen verstaan (doordat zij het opschrijven) moet uiteindelijk ook zijn wat door middel van spraakherkenning wordt verstaan.

De transcriptie diende nog een tweede doel: het achterhalen van mogelijke antwoord alternatieven. Bij de ontwikkeling van de items zijn door de itemconstructeurs verschillende alternatieve goede en foute antwoorden gesuggereerd en in de itembank opgenomen. Na de verzameling van de responsen in de pretest werd nagegaan in hoeverre antwoorden van de kandidaten overeenstemmen met de veronderstelde antwoorden en werden alternatieve en/of aanvullende goede antwoorden die door kandidaten waren geleverd, alsnog aan de itembank toegevoegd.

De volledige set van 25 800 antwoorden is handmatig op basis van de transcripties gescoord als goed (1) of fout (0). Deze scoring dient voor een initiële schatting van de moeilijkheidsgraad van de items, inzicht in de soort van fouten die door kandidaten wordt gemaakt en een grotere set van goede antwoorden die de herkenner moet leren te herkennen als “goed”. Voor iedere respons werd nagegaan of door proefpersonen gegeven antwoorden die afweken van het antwoordmodel, mogelijk toch als goed antwoord konden worden geïnterpreteerd.

Bijvoorbeeld: in de transcripties werd het antwoord Willem van Oranje in allerlei varianten gegeven b.v. willem, wilhelm van oranje, william van oranje . Het antwoord ‘willem’ alléén werd fout gerekend, de beide andere antwoorden werden goed gerekend.

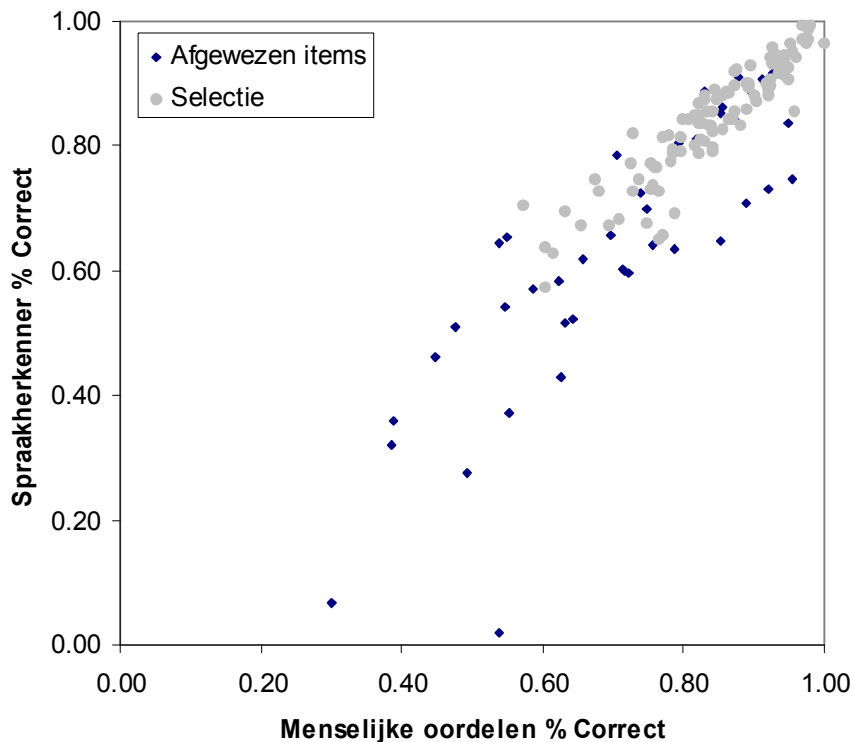
6.3 Relatie tussen de menselijke oordelen en de spraakherkenner

Overeenkomst tussen de score gebaseerd op de handmatige transcriptie en de score als gegenereerd door de machine is het ultieme criterium voor de kwaliteit van het machinale oordeel.

Nog een voorbeeld: bij de vraag ‘wat is typisch Nederlands in het verkeer ‘ bij een foto met allemaal fietsen (still uit de film waarbij verteld wordt in de film dat de fiets iets typisch Nederlands is in het verkeer) is het correcte antwoord uiteraard ‘fietsen’. Op basis van de transcripties is fiets, veel fietsen, fiezen, fiesen, fies, vies, een fiets/fies,vies eveneens correct bevonden . Uitspraak doet niet ter zake, zolang het antwoord herkenbaar is voor een moedertaalspreker die weet wat hij mag verwachten in de gegeven context. Over alle items werden meer dan 150 transcripties door de transcribenten en door verschillende itemconstructeurs beoordeeld en alle aanvaardbare varianten werden in de spraakcomputer ingevoerd.

De figuur hieronder laat zien wat de relatie is tussen de menselijke oordelen en de spraakherkenner en toont ook het effect van de itemsselectie.

De correlatie tussen beide maten bedraagt 0.894 op de ongeselecteerde set opgaven. Op grond van de tweede selectie van 107 opgaven waarbij te moeilijke en niet goed functionerende opgaven zijn verwijderd, neemt die overeenstemming toe tot 0.948.



Alle 148 opgaven zijn gepretest, verdeeld over tien overlappende subsets van 30 items ieder. De antwoorden van de kandidaten zijn dichotoom (goed/fout) gescoord door de spraakherkenner. De betrouwbaarheid van de 10 toetsen varieert van 0.81 tot 0.94 en bedraagt gemiddeld (RMS) 0.92. De gemiddelde p-waarde van de opgaven is 0.78 en de correlaties tussen de opgaven zijn behoudens enkele uitzonderingen hoog. (N.B.: het programma drukt 99.999 af wanneer geen geldige waarde kan worden berekend.)

6.4 Eerste analyse van de scores van laagtaalvaardige proefpersonen

In een eerste analyse is met name gekeken naar de scores van proefpersonen met een zeer laag taalvaardigheidsniveau of proefpersonen die analfabeet waren. Op dat lage taalvaardigheidsniveau moet het examen immers uitvoerbaar zijn. Het ging hierbij om 6000 getranscribeerde responsen van kandidaten op de laagste niveaus. Het betreft de antwoorden van de 17 kandidaten lager dan A1-min, 73 kandidaten op A1-min en 173 kandidaten op A1 niveau. Van deze kandidaten waren op het moment van analyse 17 tot 28 responsen getranscribeerd. In het navolgende worden enkele gegevens op grond van deze transcripties gepresenteerd.

Tabel 3 hieronder toont een overzicht van de proporties goede antwoorden per opgave. Hieruit valt te lezen dat 24% van de opgaven door 90% of meer van de proefpersonen zodanig beantwoord is dat de transcriptie van hun spraaksignaal overeenkwam met het goede antwoord. Bij ruim de helft van alle opgaven konden de antwoorden van 80% of meer van de proefpersonen als goed worden herkend en 92% van de opgaven werd door meer dan de helft van de proefpersonen correct beantwoord.

Tabel 3: Proportie goede antwoorden per opgave/menselijke beoordeling

<i>p-waarde</i>	<i>CumFreq. %</i>	<i>CumFreq.Aantal</i>
≥ 0.90	24%	35
≥ 0.80	51%	75
≥ 0.70	73%	108
≥ 0.60	86%	127
≥ 0.50	92%	136
≥ 0.40	97%	144
≥ 0.30	100%	148
≥ 0.20	100%	148
≥ 0.10	100%	148
≥ 0.00	100%	148

Gebaseerd op 263 proefpersonen niveau ≤ A1

Tabel 4 presenteert resultaten voor de proefpersonen op de drie laagste niveaus. In de eerste kolom het percentage goede antwoorden; in kolommen 2 tot en met 4 de cumulatieve frequenties voor de drie onderscheiden niveaus en in de kolommen 5 tot en met 7 de absolute aantallen proefpersonen. Uit de tabel blijkt dat circa een kwart van de kandidaten op de niveaus A1-min en daaronder 90% of meer van de opgaven herkenbaar goed heeft beantwoord. Bijna de helft van de proefpersonen op A1-min heeft 80% of meer van de vragen goed beantwoord. Het is dus goed mogelijk om onverlet het lage taalniveau een zeer hoge score op de verzameling KNS opgaven te behalen.

Tabel 4: Percentage goede antwoorden voor proefpersonen op de drie laagste niveaus/menselijke beoordeling

<i>% goede antwoorden</i>	<i>CumFreq. %</i>			<i>CumFreq.Aantal</i>		
	<i><A1-</i>	<i>A1-</i>	<i>A1</i>	<i><A1-</i>	<i>A1-</i>	<i>A1</i>
≥ 90%	24%	27%	54%	4	20	93
≥ 80%	29%	47%	66%	5	34	115
≥ 70%	41%	67%	75%	7	49	129
≥ 60%	41%	74%	84%	7	54	145
≥ 50%	47%	78%	91%	8	57	157
≥ 40%	76%	89%	92%	13	65	160
≥ 30%	88%	96%	95%	15	70	164
≥ 20%	94%	96%	97%	16	70	167
≥ 10%	100%	99%	97%	17	72	168
≥ 0%	100%	100%	100%	17	73	173

6.5 Inhoudelijke itemanalyse

Ook inhoudelijk zijn twee nadere inspecties uitgevoerd: naar de “moeilijke” opgaven dat wil zeggen opgaven waarop veel proefpersonen niet het juiste antwoord hebben gegeven en personen met een lage score.

Uit beluistering van kandidaten op A1-min niveau en van kandidaten die daar zelfs nog onder zitten, blijkt dat kandidaten die moeilijk verstaanbaar zijn en/of een zwaar accent hebben, toch correct zijn getranscribeerd. Uit de transcripties is ook gebleken dat de antwoorden op een aantal opgaven woorden bevatten die voor deze kandidaten moeilijk uitspreekbaar zijn. Dergelijke opgaven zijn dan ook in de definitieve versie zoveel mogelijk vermeden.

Bij minder dan 1% van alle responsen hebben de transcribenten aangegeven dat zij de respons van de betreffende kandidaat niet hebben kunnen verstaan. Deze onverstaanbare antwoorden zijn verdeeld over 44 kandidaten, de meesten met één of twee onverstaanbare responsen, slechts bij één proefpersoon werden zoveel als vier antwoorden als onverstaanbaar gemarkeerd. Om als onverstaanbaar te worden gemarkeerd moet het antwoord bij meer malen luisteren en met kennis van de mogelijke alternatieve antwoorden, niet verstaanbaar zijn. Deze laatste proefpersoon blijkt echter niet geheel onverstaanbaar, antwoorden als *Amsterdam, c w i, de gemeente, in de landbouw, katholiek* komen er goed verstaanbaar uit.

Opgaven waarbij veel proefpersonen voor het foute antwoord gaven zijn bijvoorbeeld opgaven waarbij de huidige feiten afwijken van hetgeen onder toekomstige wetgeving zal gelden. Bijvoorbeeld bij “Wie betaalt de taal cursus, de school of uzelf?” antwoordt een belangrijk deel van de proefpersonen met “de school” en ook antwoorden als “gemeente” en “sociale dienst” komen voor. Ook bestaat bij deze kandidaten kennelijk onzekerheid over de vraag of je in Nederland een uitkering kunt krijgen (gewenst antwoord “Nee”).

Een tweede type moeilijke opgaven bij deze kandidaten vormen de opgaven over geschiedenis, verwachte antwoorden als “tweede wereldoorlog” en “gouden eeuw” worden sterk verhaspeld of totaal vermeden met stilte of “Ik weet het niet”.

Proefpersonen met zeer lage scores geven overwegend duidelijk foute antwoorden of antwoorden met een helder “Ik weet het niet” (zij hebben kennelijk de stof niet geleerd), hun lage score komt niet door gebrekkige verstaanbaarheid.

6.6 Psychometrische itemanalyse en geobserveerde P-waarden

Uitgangspunt voor een eerste selectie was dat de gemiddelde p-waarde berekend op de transcripties boven de 0.7 zou moeten liggen. Vervolgens is gekeken naar de overlap tussen de verzameling geselecteerd op inhoudelijk gronden en de selectie gebaseerd op de p-waarden berekend op de transcripties. Items die niet op beide gronden waren afgewezen of goedgekeurd werden voorlopig als reserve aangemerkt.

Op het moment dat de data van de automatische scoring beschikbaar kwamen zijn toets- en itemanalyse uitgevoerd met het computerprogramma OPLM (Verhelst, Glas en Verstralen, 1991. Hieruit bleek dat 17 van de oorspronkelijke 148 items niet voldeden aan de criteria voor psychometrische kwaliteit. Deze criteria betroffen de moeilijkheidsgraad en de mate van samenhang van ieder individueel item met de rest van de verzameling. Met de resterende 131 items kon redelijke passing aan het meetmodel worden gevonden ($R1c = 442.677$; $df = 379$; $p = .0125$)

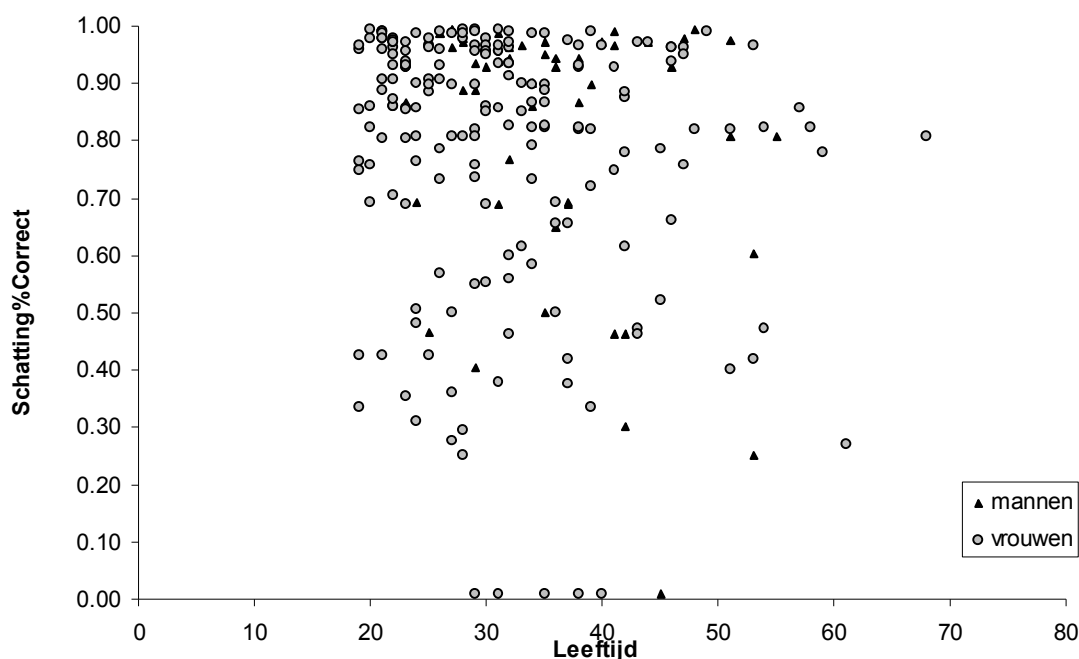
Combinatie van de bevindingen van de analyses op basis van de inhoudelijke criteria, de transcripties en de psychometrische criteria leverde een verzameling van 97 items op waarbij volledige overlap bestond tussen de criteria. Vervolgens is met drie items uit de reserve categorie de verzameling aangevuld tot 100. Voor deze verzameling kon een nog licht verbeterde passing aan het meetmodel worden gevonden ($R^2 = 239.644$; $df = 201$; $p = .0309$)

6.7 Relatie van de toetsscores en enkele achtergrondgegevens: indicaties voor validiteit

In voorgaande secties van dit rapport zijn gegevens gepresenteerd over de samenstelling van de steekproef en de resultaten van de pretest. Op grond van deze beide soorten gegevens kunnen nu ook de toetsresultaten in verband worden gebracht met enkele van de achtergrondgegevens over deelnemers aan de pretesten.

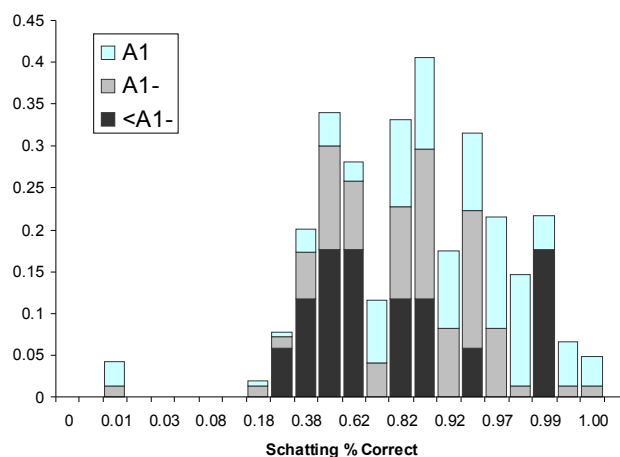
Resultaten naar leeftijd

De relatie met **leeftijd voor mannen en voor vrouwen** wordt in de volgende figuur in beeld gebracht. Het aantal mannen is beduidend minder dan het aantal deelnemende vrouwen. Personen van uiteenlopende leeftijden blijken onafhankelijk van hun leeftijd zowel lage als hoge scores te kunnen behalen. Ook valt op dat één enkele man en een zeer gering aantal vrouwen een minimale score hebben behaald, zij hebben in het geheel geen goed antwoord gegeven.



Resultaten naar taalvaardigheid

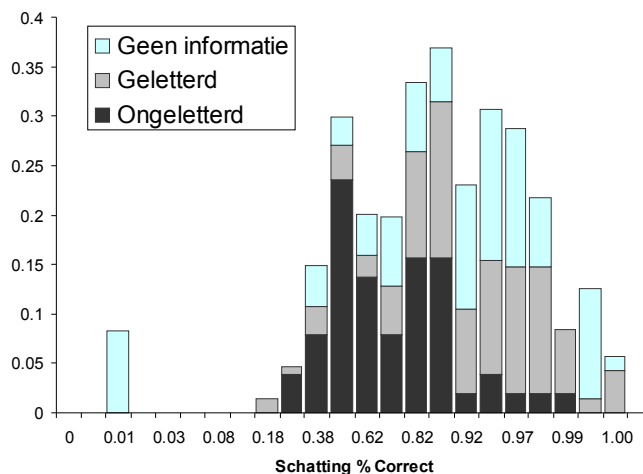
Taalvaardigheid mag - als tenminste het niveau A1-min wordt beheerst – geen belemmering vormen. Onderstaande figuur toont aan dat ook personen met een door hun docent ingeschatte niveau onder A1-min, nog kans maken meer dan 80% correct te scoren.



Toetsscore in relatie tot geletterdheid

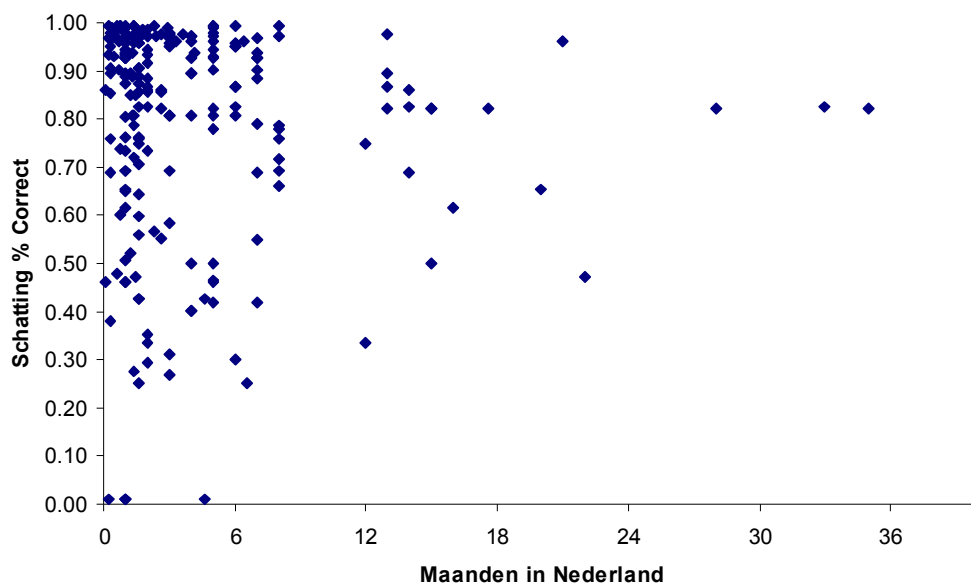
Gelet op de eerst beoogde doelgroep voor de toets (mensen die nog woonachtig zijn in het land van herkomst) is besloten dat geletterdheid geen rol mocht spelen bij de examinering KNS.

Hieronder is de scoreverdeling weergegeven voor geletterden en niet-geletterden. Uit deze grafiek blijkt dat niet-geletterden weliswaar gemiddeld een lagere score behalen dan geletterden, dat echter niet-geletterden wel degelijk kans maken een hoge toetsscore te verwerven.



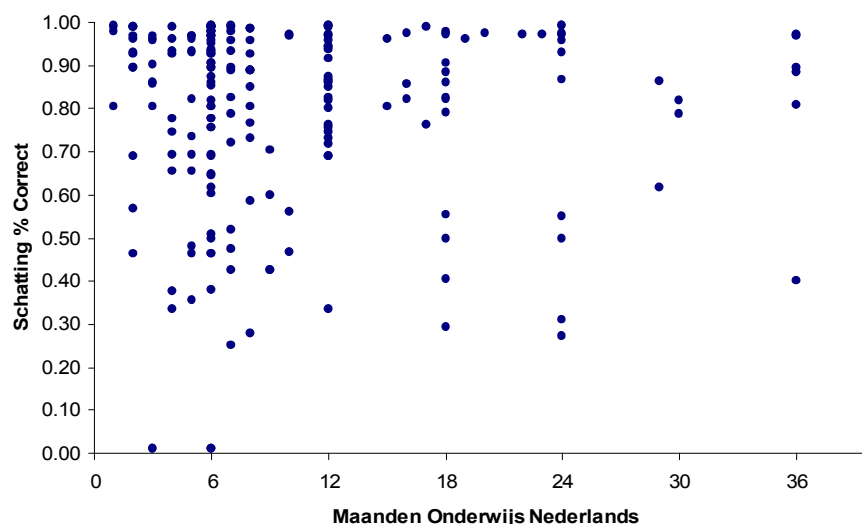
Toetsscore in relatie tot jaren van verblijf in Nederland

Aangezien de toets bestemd is voor mensen die nog niet in Nederland zijn, is er bij de pretest naar gestreefd personen met een zo kort mogelijke **verblijfsduur in Nederland** te selecteren. Onderstaande figuur toont dat er inderdaad een sterke concentratie is van deelnemers die korter dan een half jaar in Nederland waren op het moment van de toetsafname. Onder deze groep komen zowel lage als hoge tot zeer hoge scores voor.



Toetsscore in relatie tot onderwijs in de Nederlandse taal

De mate waarin potentiële kandidaten de kans hebben gehad lessen **Nederlandse taal** te volgen mag ook weinig of geen invloed hebben op de kansen voor dit examengedeelte te slagen. Onderstaande figuur toont de relatie tussen aantal maanden les in de Nederlandse taal en de toetsscore. Geen lessen in het Nederlands was onder de steekproefpopulatie niet mogelijk aangezien de kandidaten via onderwijsinstellingen zijn benaderd.

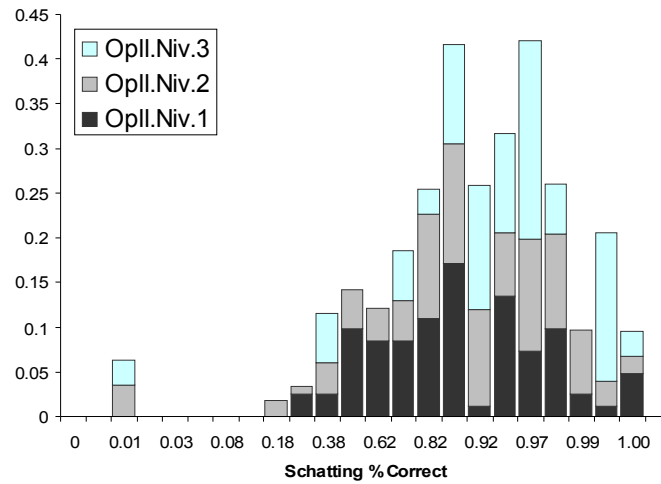


Toetsscore in relatie tot genoten opleiding

Ook **opleidingsniveau** mag geen hinderpaal vormen. Opleiding is ingedeeld in drie klassen: 6 jaar of minder, 6 tot 12 jaar en meer dan 12 jaar. Onderstaande figuur toont dat men ook met uitsluitend een lagere school opleiding een hoge score op de toets kan halen.

Conclusie: de gegevens uit de pretest tonen aan dat achtergrondvariabelen zoals opleiding, verblijfsduur in Nederland, leeftijd en geslacht niet van invloed zijn op de toetsscore. Ook een laag taalvaardigheidsniveau vormt op zich geen belemmering om een hoge score op de

toets te halen. Ofschoon het voor analfabeten uit de pretest mogelijk blijkt om voldoende en zelfs hoge scores te behalen (bij een cesuur van 0.8 blijkt nog 40 % de toets te halen), blijken geletterden gemiddeld duidelijk beter te presteren dan analfabeten. Analfabeten zullen naar verwachting meer inspanning moeten leveren dan geletterden.



7 Cesuurbepaling

Vanwege de schaling van de itembank met een passend IRT model kan voor iedere kandidaat, op grond van de afname van een gedeelte van de verzameling opgaven, de kans op goede antwoorden op ieder andere verzameling uit de itembank – inclusief de totale set worden geschat. Dit biedt het voordeel dat op grond van ieder van de tien parallel versies die voor het examen worden samengesteld de beheersing van alle opgaven in de bank kan worden geschat.

Bij de bepaling van de grensscore moeten een aantal kanttekeningen worden gemaakt.

De schattingen van item- en persoonsparameters zijn gebaseerd op de afname van sets van opgaven in pretest condities. Het is bekend dat kandidaten in dergelijke condities eerder onder hun eigen niveau presteren dan in examencondities. In de echte examensituatie zullen kandidaten ook gemotiveerder zijn om de toets af te leggen en in het herkomstland zullen kandidaten meer kunnen oefenen met de band in de eigen taal en in het Nederlands en met de vragen en antwoorden. Zij kunnen zoveel tijd voor de voorbereiding nemen als zij maar wensen hetgeen in de pretest niet het geval was.

Ook de leeftijd van de pretestdeelnemers verschilt met die van de kandidaten aan het examen; naar verwachting zullen de laatstgenoemde jonger zijn aangezien het gaat om gezinsvorming en gezinshereniging

Een ander verschil tussen pretest en authentieke examensituatie is dat tijdens de pretest slechts 30 items ter voorbereiding moesten worden geleerd en dat zullen er in de authentieke examensituatie 100 zijn.

Maar de pretest bevatte items van een uiteenlopende kwaliteit, ca 30% van de opgaven zijn uitgesloten van opname in het definitieve examen. Het voorkomen van een relatief hoog percentage minder geslaagde items in een toets kan de resultaten op de andere opgaven beïnvloeden.

En als laatste dient te worden opgemerkt dat alle voorliggende analyses gebaseerd zijn op een cesuur van 0.8. Dit is behoorlijk hoog voor de inspanningsverplichting die van aspirant nieuwkomers wordt gevraagd.

Al met al zijn er behoorlijk wat verschillen tussen de condities tijdens de pretest en die van de authentieke toetsituatie. Dit gegeven bemoeilijkt het vaststellen van de grens waarop de kandidaat is gezakt of geslaagd (cesuur) .

Gelet op het feit dat a)de pretestsituatie verschilt van de authentieke situatie b)dat de voorbereiding tijdens de pretest verschilt van de authentieke situatie en c) dat analfabeten een geringere kans hebben om een voldoende score te behalen dan geletterden, is enige voorzichtigheid in dit stadium geboden. Wij stellen daarom voor de cesuur in ieder geval niet hoger te leggen dan 0.70 en deze bij de wettelijk voorgestelde evaluatie te betrekken.

8 Literatuur

Referenties

- Baddeley, A. (2000) The episodic buffer: a new component of working memory? *Trends in Cognitive Science*, 4 (11), 417-423.
- Baddely, A. (1986) *Working memory*. Oxford: Clarendon Press.
- Bernstein, J. and H. Franco, (1996), Speech recognition by computer, in Lass, N.J. (ed.) *Principles of Experimental Phonetics*, Mosby, pp. 408-434.
- Bernstein, J., De Jong, J.H.A.L., Pisoni, D., and Townshend, B. (2000). "Two Experiments on Automatic Scoring of Spoken Language Proficiency." In: P. Delcloque(ed.), Proceedings of InSTIL2000 (Integrating Speech Technology in Learning, University of Abertay Dundee, Scotland, August, 2000, pp. 57-61.
- Brink, M e.a. (2004) *Evaluatie eindtermen maatschappijoriëntatie*, Amsterdam, Regioplan
- Byrne, W., Khudanpur, S., Knodt, E. and J. Bernstein.(1997) Is automatic speech recognition ready for non-native speech? a data collection effort and initial experiments in modeling conversational Hispanic English. In *ESCA-ITR Workshop on speech technology in language learning*.
- Carroll, J.B. (1986) Second Language. In: R.F. Dillon and R.J. Sternberg (Eds) *Cognition and Instruction*. Orlando, FL: Academic Press.
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Cremelie, N. (2000) Heuristische zoekstrategie en automatische aanmaak van lexica voor continue spraakherkenning. Diss. Gent, UGent.
- Franssen, J. et al. (2004) *Inburgering getoetst Advies over het niveau van het inburgeringsexamen in het buitenland*.
- Gibson, E., 1991. A Computational Theory of Human Linguistic Processing: Memory Limitations and Processing Breakdown. PhD Thesis, Carnegie Mellon University, Pittsburgh, PA.
- Miller, G.A., Isard, S., 1964. Free recall of self-embedded English sentences. *Information and Control* 7, 292–303.
- North, B. (2000) *The Development of a Common Framework Scale of language Proficiency*. New York, NY: Peter Lang.
- Poelmans, P. (2003) *Developing Second Language Listening Comprehension: Effects of training lower-order skills versus higher-order strategy*. Diss. Amsterdam UvA.
- Schneider, W. and R.M. Shiffrin (1977) Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84, 1-66.
- Verduyn, M (2005) Een artikel over de totstandkoming van de film Naar Nederland is gepubliceerd in Justitie Magazine, Uitgave van het Ministerie van Justitie, 4^e jaargang nummer 2, maart 2005 'The making of... Naar Nederland' p 8-10.
- Verhelst, N.D., Glas, C.A.W., and H.H.F.M. Verstralen (1991) OPLM: A Computer Program and Manual. Arnhem: CITO

Wright, B.D. & G.N. Masters (1982) *Rating Scale Analysis: Rasch Measurement*. Chicago: MESA Press.

Wright, B.D. & M.H. Stone (1979) *Best Test Design: Rasch Measurement*. Chicago: Mesa Press.

Wylie, E. (2002). An overview of the International Second Language Proficiency Ratings (ISLPR). Centre for Applied Linguistics and Languages (CALL), Griffith University.
<http://www.gu.edu.au/centre/call/ISLPROverview.pdf> (retrieved 23/12/2004).