EXPERTGESPREKKEN & MARKTVERKENNING TOETS GESPROKEN NEDERLANDS

Triarii, Den Haag, 3 oktober 2013
Versie: Eindrapportage
Kenmerk Ministerie van Sociale Zaken en Werkgelegenheid: 210300100.054.005

INHOUDSOPGAVE

(Foto cover page: Robin Utrecht, ANP)

MANAGEMENTSAMENVATTING

De afgelopen maanden is vanuit diverse kanten kritiek opgelaaid op de Toets Gesproken Nederlands (TGN) die afgenomen wordt in het kader van de Wet Inburgering en de Wet Inburgering Buitenland. Bovendien loopt in 2014 het contract af tussen SZW en de huidige leverancier van de TGN. In dit kader is een 20-tal binnen- en buitenlandse experts op het gebied van taalexamens geïnterviewd en bevraagd op hun meningen over de TGN en gerelateerde aspecten, en ideeën over alternatieven voor de TGN.

Uitkomsten TGN-gesprekken

De experts zijn het eens over de volgende punten:

- Examineren van taalvaardigheden is niet eenvoudig;
- Spraaktechnologie is een waardevol hulpmiddel bij taalverwerving;
- De TGN meet extremen in taalbeheersing adequaat.

De experts verschillen van mening over:

- De validiteit van de TGN;
- De validiteit van de GBL (toets Geletterdheid en Begrijpend Lezen);
- Het beoordelen van taalvaardigheid door spraaktechnologie;
- Het examineren van luistervaardigheid A1 in een NVT-omgeving;
- Het examineren van spreekvaardigheid A1 en A2 in één examen;
- Het geïntegreerd examineren van taalvaardigheden.

Validiteit en automatische beoordeling zijn daarbij echte breekpunten; combinatie-examens zijn meer onderhevig aan afwegingen.

De experts vertalen als volgt hun meningen in de verschillende visies op het optimale spreekvaardigheidsexamen:

1. Experts die overtuigd zijn van de TGN willen de TGN handhaven voor het buitenland op A1 niveau en het binnenland op A2 niveau;
2. TGN-kritische commerciële examenontwikkelaars en semi-commerciële examenontwikkelaars / taalopleiders met een meer academische setting willen de TGN vervangen voor zowel buitenland op A1 niveau als binnenland op A2 niveau. De eerste categorie experts wil op beide taalniveaus een vergelijkbaar alternatief; de tweede categorie staat daarentegen daarentegen een veel lichtere vorm van examinering voor op A1 niveau;
3. (Kleine) commerciële taalopleiders, die juist het meest vocaal zijn in hun TGN-kritiek, zijn voorstander van TGN vervangen voor het binnenland op A2 niveau, maar kunnen leven met handhaven voor het buitenland op A1 niveau.

Uitkomsten Marktverkenning

De ideeën ten aanzien van alternatieven voor het huidige examen spreekvaardigheid convergeerden naar vier grote gemene delers, namelijk drie varianten voor examens buitenland op A1 niveau en/of binnenland op A2 niveau, gebaseerd op bekende examenvormen:

1. Verbeterd examen spreekvaardigheid door middel van spraakvergelijking (oftewel automatische examinering en beoordeling) specifiek voor A1 of A2, eventueel voor A2 aangevuld met een gesprekssimulatie;
2. Machinale examinering met menselijke beoordeling achteraf (oftewel automatische examinering en menselijke beoordeling achteraf) op niveaus A1 en A2 aangevuld met een gesprekssimulatie;
3. Geïntegreerd praktijkgericht examen gebaseerd op menselijke examinering in een direct gesprek op A1 en A2 niveau, gevolgd door menselijke beoordeling;

en een fundamenteel nieuwe ontwikkeling, te weten

4. Gameplay met interactieve dialoog: een animatiecomponent geschikt voor automatische examinering.

Conclusies

De globale richting van een nieuwe aanbesteding voor spreekvaardigheidsexamens zal sterk afhankelijk zijn van de weging van betrouwbaarheid / validiteit en de bereidheid tot investeren aan overheidskant:

- Indien de hoogste betrouwbaarheid allesbepalend zal zijn, wordt de keuze in feite beperkt tot een examen spreekvaardigheid door middel van spraakvergelijking;
- Indien de (indruks)validiteit oftewel het maatschappelijk draagvlak doorslaggevend is en er tegelijk een wil is tot verandering van de spreekvaardigheidsexamens voor binnenland en buitenland, is machinale examinering met menselijke beoordeling achteraf het logische alternatief;
- Indien de (indruks)validiteit oftewel het maatschappelijk draagvlak doorslaggevend is en er alleen een wil is tot verandering van spreekvaardigheidsexamens voor binnenland, komen de volgende varianten in principe in aanmerking:
  - Machinale examinering met menselijke beoordeling achteraf;
  - Geïntegreerd praktijkgericht examen.

  De keuze volgt dan waarschijnlijk uit de totale kosten van ontwikkeling en exploitatie.

Voor iedere richting zijn meer aanbieders te verwachten.

Aanbevelingen

Voor het Programma van Eisen bevelen we het volgende aan:

- Specificeer functioneel in plaats van technisch;
- Staar je niet blind op betrouwbaarheid;
- Transparantie over de werking van het examen vergroot de acceptatie;
- Toegankelijkheid in het buitenland waarborgen;
- Geef helder aan op welk ICT platform het examen dient te draaien;
- Let op de integrale kostprijs.

Aanvullend en in het algemeen geven we de overheid in overweging om meer zichtbaarheid te zoeken in het relevante maatschappelijke debat.

1 INLEIDING

1.1 Aanleiding

De afgelopen maanden is vanuit diverse kanten kritiek opgelaaid op de Toets Gesproken Nederlands (TGN) die afgenomen wordt in het kader van de Wet Inburgering en de Wet Inburgering Buitenland. Het ministerie heeft daarom behoefte aan een inventarisatie en analyse van deze kritiek.

Bovendien loopt in 2014 het contract af tussen, inmiddels[1], SZW en de huidige leverancier[2] van de TGN (en tevens de examens Kennis van de Nederlandse Samenleving (KNS), en de toets Geletterdheid en Begrijpend Lezen (GBL) in het buitenland). Het ministerie wil ter voorbereiding van een nieuwe aanbesteding een marktverkenning laten uitvoeren.

1.2 Vraagstelling

De opdracht aan Triarii is door het ministerie van Sociale Zaken en Werkgelegenheid als volgt geformuleerd:

1. Het interviewen van experts op het gebied van taalexamens (in het bijzonder spreek- en luistervaardigheid, meer in het bijzonder de TGN), (in het vervolg "TGN-gesprekken")
    - Het betreft ca. 20 experts van diverse soorten instellingen in Nederland, eventueel aangevuld met enkele buitenlandse experts. Opdrachtgever levert een conceptlijst aan.
    - Opdrachtgever levert een concept topiclijst aan. De gesprekken moeten minimaal ingaan op:
        - De ervaringen van de experts
        - De meningen van de experts
        - Opvattingen over alternatieven
        - De wenselijkheid van aparte spreekvaardigheidsexamen voor A1 en A2
        - De wenselijkheid van een examen luistervaardigheid in het buitenland

[1] Via achtereenvolgens de ministeries van Binnenlandse Zaken, Justitie en VROM

[2] Een consortium van Ordinate / Pearson en CINOP

2. Het uitvoeren van een marktverkenning op het gebied van NT2 examens spreekvaardigheid en op het gebied van NVT examen spreek-, lees- en luistervaardigheid, in het vervolg de "Marktverkenning".
    - Welke vormen zijn er om de examens af te nemen? Opdrachtnemer dient dit ruim te interpreteren. Opdrachtgever levert een concept topic lijst aan.
    - Welke nieuwe mogelijkheden zijn er ontstaan sinds 2003 in Nederland, Europa en de rest van de wereld? Opdrachtnemer dient dit ruim te interpreteren. Opdrachtgever levert een concept topic lijst aan.
    - Wat is de betrouwbaarheid en validiteit van de verschillende vormen van examinering?
    - Hoe lang moet elk van de examens duren? Uitgaande van een hoge betrouwbaarheid (0,90)?
    - Is het mogelijk dat de examens op 1 november 2014 operationeel zijn?
    - Welke partijen zijn bereid in te schrijven op een aanbesteding?
    - Welke aspecten moeten zeker vermeld worden in het programma van eisen voor aanbesteding van de examens?

Het gewenste resultaat is een rapportage met een analyse van de gesprekken met experts en conclusies en aanbevelingen. De marktverkenning zal input zijn voor het programma van eisen voor de aanbesteding van de examens spreek-, lees- en luistervaardigheid.

1.3 Aanpak

Voor de werkstromen TGN-gesprekken en Marktverkenning zijn de meningen verzameld en geïnventariseerd van een groep binnen- en buitenlandse experts op het gebied van taalverwerving, taalexaminering, (taal)psychometrie en spraaktechnologie. De samenstelling van de expertpopulatie is primair bepaald door het opdrachtgevende Ministerie van Sociale Zaken en Werkgelegenheid; in overleg met opdrachtgever hebben wij enige vervangers voor niet tijdig beschikbare experts en alle buitenlandse experts aangedragen. De uiteindelijke lijst van geraadpleegde experts is opgenomen als bijlage.

Op grond van onze visie dat alle experts ook potentiële marktpartijen / deelnemers aan een consortium kunnen zijn, hebben wij beide werkstromen in ieder interview geadresseerd.

Voor de werkstroom Marktverkenning is aanvullend een systematiseringsslag uitgevoerd door middel van schriftelijke enquêtes. Om de responsedrempel zo laag mogelijk te leggen, zijn de enquêtes door ons vooringevuld voor de vier

geïdentificeerde grootste gemene delers uit de interviewronde. Deze enquêtes zijn vervolgens weer voorgelegd aan de inhoudelijke deskundige experts (in de praktijk alle eerder geraadpleegden behalve de individuele taalopleiders).

Uit de aard van een verkenning is deze rapportage beperkt tot kwalitatieve beschrijvingen van de belangrijkste onderscheidende aspecten. Het kader van de Marktverkenning, en niet in de laatste plaats de korte doorlooptijd in een vakantieperiode, stond bovendien geen diepgaande technisch-inhoudelijke studie van huidige systeem en mogelijke alternatieven toe.

Deze opdracht is uitgevoerd in de periode 3 juli tot 23 september 2013 door Hans Modder, Ron Overgoor en Gert-Jan van der Panne.

1.4 Leeswijzer

Waar de werkstromen TGN-gesprekken en Marktverkenning gezamenlijk zijn geadresseerd tijdens de uitvoering, worden zij voor de meest directe aansluiting met de vraagstelling separaat gerapporteerd.

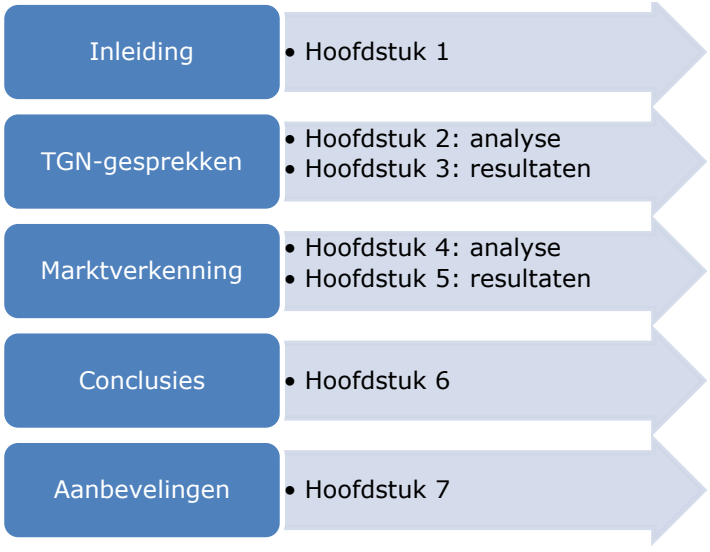
Hoofdstukken 2 en 3 behandelen respectievelijk onze analyse en de onderliggende resultaten van de TGN-gesprekken.

Hoofdstukken 4 en 5 doen aansluitend datzelfde voor de Marktverkenning.

Vervolgens behandelt hoofdstuk 6 de conclusies.

Tenslotte volgen in hoofdstuk 7 de aanbevelingen voor het Programma van eisen.

Hieronder is de indeling schematisch weergegeven:



2 ANALYSE TGN-GESPREKKEN

2.1 Meningen kwantitatief niet noodzakelijk representatief

Getalsmatige representativiteit is onduidelijk

Het is onduidelijk hoe de meningen ten aanzien van de TGN en gerelateerde aspecten verdeeld zijn in het veld. Ook in dit geval zoeken de tegenstanders van de huidige situatie vaker de publiciteit en zijn ze vocaler in het debat dan de voorstanders. Feit is dat de TGN-critici in de populatie van geraadpleegde experts in de meerderheid zijn.

Zuiverheid niet absoluut

De meningen van experts zijn vaak een combinatie van "eerste en tweedehands": experts geven eigen meningen maar herhalen ook meningen van anderen die goed bij de eigen opvattingen passen en/of in het algemeen vertrouwd worden. In ieder geval blijkt lang niet iedere expert zelf de TGN-(proef)examen gedaan te hebben.

Al dan niet expliciet toegegeven kunnen de gegeven meningen ingegeven worden door achterliggende belangen op zakelijk, wetenschappelijk en/of ethisch gebied.

Daarnaast lijken per expert de meningen op de verschillende deelonderwerpen niet altijd even ontwikkeld onder andere door (meer) zijdelingse betrokkenheid of gedateerde / vervaagde ervaring.

Tenslotte blijken niet alle meningen volledig gebaseerd op feiten: sommige experts zijn onbekend met het feit dat de TGN ook voor A2-NT2 wordt gebruikt of blijken de toets GBL met LLS te verwarren.

2.2 Hardheid en vertaling van meningen varieert

Indruksvaliditeit[3] en automatische beoordeling breekpunten

De experts blijven fundamenteel van mening verschillen over het al dan niet valide zijn van de TGN. Men is of overtuigd van de validiteit, met verwijzing naar de uitgevoerde onderzoeken, of van mening dat de indruksvaliditeit gewoonweg te laag blijft, ongeacht wat de rapporten zeggen. Hierbij worden met name onvoldoende aansluiting bij CEF en kennelijke bias ten nadele van Oost-Aziaten en Spaans/Portugeestaligen aangehaald.

Eenzelfde situatie bestaat rondom het gebruik van op spraaktechnologie gebaseerde automatische beoordeling. Voorstanders wijzen op de hoge betrouwbaarheid en efficiëntie, waar tegengenstanders simpelweg niet kunnen aanvaarden dat een

[3] Vertaling van het begrip "face validity" die in onze optiek het beste de lading dekt

complexe uiting als taal voldoende te vangen en zuiver te beoordelen is met een (star) computersysteem.

Combinatie-examens meer onderhevig aan afwegingen

Vanuit betrouwbaarheidsoverwegingen hebben vooral psychometrici een voorkeur voor zo specifiek mogelijke examens, die één enkele taalvaardigheid voor één niveau pogen te beoordelen. Taalopleiders redeneren vooral vanuit de validiteit en zijn daarom wel voor geïntegreerde examens op één niveau, maar minder op verschillende niveaus omdat te hoge niveaus niet wenselijk geacht worden voor lageropgeleiden. Desalniettemin is er breed begrip voor het uiteindelijk toepassen van combinatie-examens in het algemeen vanuit efficiëntie- en kostenoverwegingen.

Meningen worden verschillend vertaald naar het optimale spreekvaardigheidsexamen

Drie opties zijn geïdentificeerd voor het optimale spreekvaardigheidsexamen, te weten:

1. TGN handhaven voor A1-NVT en A2-NT2;
2. TGN vervangen voor zowel A1-NVT als A2-NT2;
3. TGN handhaven voor A1-NVT, vervangen voor A2-NT2.

Experts die overtuigd zijn van de TGN prefereren zonder uitzondering de eerste optie oftewel handhaving van de huidige situatie (waarbij wel ook gewezen wordt op het verbeterpotentieel, zie sectie 4.1).

Experts die niet overtuigd zijn van de TGN willen verandering, maar niet allemaal in dezelfde vorm:

- Wellicht vooral ingegeven door de zakelijke belangen, geven de commerciële taal-examenontwikkelaars onder hen de voorkeur aan de tweede optie, oftewel een totale verandering van het examen spreekvaardigheid. Zij benadrukken tevens dat dit goed uitvoerbaar is met handhaving van de huidige eisen voor taalvaardigheid op A1 en A2 niveau.
- Semi-commerciële examenontwikkelaars / taalopleiders met een meer academische setting zijn eveneens voor de tweede optie met totale verandering, maar neigen om validiteitsredenen naar een veel lichtere vorm van examinering op A1 niveau.
- Ondanks dat ze het meest vocaal zijn in hun kritiek op de TGN, blijken juist de (kleine) commerciële taalopleiders een voorstander van de laatste optie namelijk alleen verandering van de A2-NT2 situatie. Verwijzend naar de moeilijkheid om een betere combinatie van examen spreekvaardigheid en logistiek te ontwikkelen voor de A1-NVT situatie dat aan alle eisen voldoet en niet veel duurder zal zijn, accepteren ze met tegenzin handhaving indien de TGN maar vervangen wordt voor A2-NT2. Het valt niet uit te sluiten dat men de A1-NVT acceptatie ook ziet als een vorm van wisselgeld om maar iets

gedaan te krijgen voor de A2-NT2 situatie waar duidelijk hun pijn zit waar zij niet alleen inhoudelijke, maar ook emotionele argumenten voor opvoeren.

2.3 Meningen kwalitatief wel representatief

Ondanks dat de meningen in kwantitatieve zin waarschijnlijk niet representatief zijn voor het veld, hebben wij de indruk dat zij wel een goed beeld geven van de variatie in opvattingen:

- De populatie experts mag divers genoeg geacht worden om het merendeel aan argumenten af te dekken en het totaal convergeerde steeds naar het hier gerapporteerde;
- De globale uitkomst vertaald in visies op het optimale spreekvaardigheidsexamen – twee extreme vormen en één meest logische compromis – dekt het spectrum.

3 RESULTATEN TGN GESPREKKKEN

Voor de werkstroom TGN-gesprekken zijn de experts gevraagd naar hun meningen ten aanzien van:

1. Taalexamens voor inburgeringsdoeleinden in het algemeen;
2. De Toets Gesproken Nederlands in het bijzonder, en;
3. Aspecten gerelateerd aan de huidige TGN implementatie door Nederland.

In de praktijk bleken de geraadpleegde buitenlandse experts weinig specifieke kennis te hebben over de Nederlandse aanpak en beperkte de interviews zich tot het eerste onderdeel. De resultaten hiervan zijn in een aparte paragraaf aan het eind opgenomen, voornamelijk ter kleuring.

Bij de binnenlandse experts daarentegen vormden de onderdelen 2 en 3 juist de hoofdmoot. Dit hoofdstuk begint met een korte beschrijving van de enkele punten waar men het over eens is. Vervolgens wordt de meeste plaats ingeruimd voor die aspecten waarop de meningen uiteenlopen. Hier is gekozen voor een rapportage in tabelvorm waarbij per punt de voor- en tegenargumenten direct tegenover elkaar geplaatst zijn.

Dit hoofdstuk is gebaseerd op onze interpretatie van de expliciete meningen zoals geuit in de interviews en eventuele impliciete inter- en extrapolaties daarvan (respondenten laten men name overeenkomende aspecten onbesproken en/of laten conclusies onbenoemd); nergens zijn meningen verdraaid of doelbewust weggelaten wanneer niet passend in onze weergave.

3.1 Gedeelde meningen

3.1.1 Examineren van taalvaardigheden is niet eenvoudig

Taalbeheersing is het resultaat van een nauw samenspel van zaken als woordenschat, grammaticakennis, begrip, uitspraak, etc. Dit complex van factoren zorgt ervoor dat het examineren van specifieke taalvaardigheden een specialistische aangelegenheid is.

3.1.2 Spraaktechnologie is een waardevol hulpmiddel bij taalverwerving

Spraaktechnologie is slechts zo goed als de training van de gebruikte automatische spraakherkenner en derhalve intrinsiek beperkt omdat nooit alle aspecten van spraak vooraf in te schatten zijn. Voor (uit)spraaktraining wegen eventuele onvolkomenheden niet op tegen de voordelen van immer beschikbare, geduldige en relatief goedkope computer ondersteuning.

3.1.3 De TGN meet extremen in taalbeheersing adequaat

De TGN is voldoende onderscheidend voor het inschatten van extremen in kennis van en vaardigheid met de Nederlandse taal:

- Degenen zonder enige kennis en vaardigheden halen niveau A1 niet;
- Degenen met meer dan voldoende kennis en vaardigheden halen zeker niveau A2.

3.2 Meningen over de validiteit van de TGN

Validiteit heeft betrekking op de vraag of een examen daadwerkelijk meet wat hij zou moeten meten. De TGN wordt momenteel gebruikt in het buitenland en het binnenland om vast te stellen of een kandidaat tenminste spreekvaardigheid op respectievelijk niveau A1 of A2 bezit.

De TGN is onvoldoende valide	**De TGN is voldoende valide**

De TGN sluit onvoldoende aan bij de CEF-schaal:
▪ Geen (volledige) luister- en spreekvaardigheid:
  ▫ "Napapegaaien" is geen spreken
  ▫ (Te) korte stimuli zonder context
  ▫ Geen gesprekspartner die de vraag kan herhalen
▪ Geen gespreksvaardigheid
	CEF is een raamwerk, geen wet. TGN examineert weliswaar indirecter, maar:
	▪ Wel degelijk representatief voor luister- en spreekvaardigheid
	▪ Gebaseerd op echt taalgebruik (geen gekunsteld taalgebruik bedoeld om multiple choice vragen over te kunnen stellen)
	▪ Spreekvaardigheid weer gerelateerd aan gespreksvaardigheid

Oost-Aziaten, Spaans-/Portugees-taligen scoren ondergemiddeld, wat een bias suggereert.	Oost-Aziaten, Spaans-/Portugees-taligen hebben intrinsiek meer moeite met uitspraak Nederlandse taal.
Terugslageffect op taalonderwijs is negatief en bedreigt de validiteit van het examen.	Ieder examen heeft een terugslageffect, maar dat hoeft de taalverwerving niet in de weg te staan.
Items liggen inmiddels op straat.	Ondanks dat de items op straat liggen verwerft de kandidaat een woordenschat en spreekervaring.
Examen geeft geen voorspelling van het haalbare taalniveau van de kandidaat.	Geen enkel examen kan het haalbare taalniveau voorspellen.
Onnatuurlijke setting van de examinering werkt stress-verhogend.	Onnatuurlijke setting geldt in zekere mate voor ieder examen. Bovendien is

	een telefoon wereldwijd een bekend fenomeen.
De kandidaat krijgt geen feedback op onderdelen, geen menselijke herbeoordeling.	De feedback op onderdelen en menselijke herbeoordeling is per 1 juli 2013 ingevoerd.
De TGN is niet nauwkeurig genoeg om niveau A2 scherp in te schatten.	De TGN omvat het hele CEF-spectrum, waarbij niveau A2 net zo scherp wordt ingeschat als alle andere niveaus.
De TGN op NVT-A1 sluit onvoldoende aan bij belevingswereld kandidaat.	-

3.3 Meningen over de validiteit van de toets GBL

De toets GBL wordt momenteel gebruikt in het buitenland om vast te stellen of een kandidaat tenminste leesvaardigheden op niveau A1 bezit. Omdat bij de toets GBL afname plaatsvindt via hetzelfde systeem als de TGN is er overlap in de voor- en tegenargumenten. In deze paragraaf zijn alleen de aanvullende meningen opgenomen.

Toets GBL is onvoldoende valide	**Toets GBL is voldoende valide**
Benadeelt analfabeten en laagopgeleiden.	Benadeling van analfabeten en laagopgeleiden zal altijd het geval zijn bij een eis op gebied van leesvaardigheid.
Onnodig indirecte examinering door beoordeling op basis van spraaktechnologie.	Misschien omslachtig door gebruikmaking van het TGN-systeem, maar toch een goed examen.

3.4 Meningen over beoordelen door spraaktechnologie

De TGN is gebaseerd op beoordeling van de taalvaardigheid door spraaktechnologie. Waar er breed gedragen waardering is voor het gebruik van spraaktechnologie in taalvaardigheidstrainingen, lopen de meningen uiteen ten aanzien van inzet voor automatische beoordeling.

Beoordeling door spraaktechnologie: nee	**Beoordeling door spraaktechnologie: ja**
Taalniveau beoordelen is complex door grote variëteit in input. Spraaktechnologie kan dit niet aan.	Taalniveau beoordelen door goed getrainde spraaktechnologie kan, mits met vangnet op basis van menselijke (her)beoordeling.

Menselijke beoordeling is met strakke protocollen, "geijkte" beoordelaars en segmentatie voldoende betrouwbaar.	Menselijke beoordeling blijft minder betrouwbaar dan beoordeling op basis van spraaktechnologie.
Directe examinering via praten met een mens ligt dichter bij de praktijk.	-
In het TGN geval is de feitelijke beoordeling een "black box".	TGN beoordeling onafhankelijk getoetst en binnen specificaties.

3.5 Meningen over de mogelijkheid tot examineren luistervaardigheid A1-NVT

Het taalniveau A1 binnen het CEF is het allerlaagste taalniveau. De vraag is of het mogelijk en wenselijk is luistervaardigheid te eisen op dit minimaal niveau in een omgeving waar geen Nederlands gesproken wordt.

Luistervaardigheidsexamen A1-NVT: nee	**Luistervaardigheidsexamen A1-NVT: ja**
A1 niveau is in het algemeen moeilijk te examineren gezien de beperkte taalvaardigheid.	Eisen aan luistervaardigheid en examineren daarvan kan en levert een bijdrage aan de taalontwikkeling.

3.6 Meningen over het examineren van spreekvaardigheid A1 en A2 in één examen

De TGN beoordeelt meer taalniveaus in één en hetzelfde examen. De vraag is of juist het tegelijk examineren van de laagste taalniveaus mogelijk en wenselijk is.

A1 en A2 in één examen: nee	**A1 en A2 in één examen: ja**
A1 en A2 spreekvaardigheid apart examineren: A1 kandidaat niet confronteren met uitingen van een te hoog niveau. Een examen dat voor een specifiek niveau ontwikkeld is, is meer valide.	A1 en A2 spreekvaardigheid in hetzelfde examen methodologisch geen probleem. Bovendien is het efficiënter.

3.7 Meningen over het geïntegreerd examineren van taalvaardigheden

Een geïntegreerd examen examineert meerdere taalvaardigheden tegelijk, in tegenstelling tot specifieke examens die gericht zijn op afzonderlijke taalvaardigheden. Geïntegreerd examineren wordt over het algemeen gezien als meer valide omdat het dichter bij de realiteit staat, waar een apart examen per vaardigheid doorgaans een hogere betrouwbaarheid kan garanderen.

Geïntegreerd examineren: nee	Geïntegreerd examineren: ja
Luister- en spreekvaardigheid apart is meer valide per vaardigheid. Anders is luistervaardigheid bepalend voor spreekvaardigheid.	Luister- en spreekvaardigheid in hetzelfde examen ligt dichter bij de praktijk/ functioneel taalgebruik en efficiënter.
-	Niet geëxamineerde vaardigheden correleren positief met wel geëxamineerde in bijvoorbeeld Cloze-examens.

3.8 Visies op het optimale spreekvaardigheidsexamen

Gevraagd naar hun mening over het optimale spreekvaardigheidsexamen, komen de experts met de volgende visies:

1. TGN handhaven voor het examineren van niveau A1 in het buitenland en A2 in het binnenland;
2. TGN vervangen voor zowel het examineren van niveau A1 in het buitenland en A2 in het binnenland;
3. TGN handhaven voor het examineren van niveau A1 in het buitenland maar vervangen voor het examineren van niveau A2 in het binnenland.

Onderstaande tabel vat dit schematisch samen:

	A1-NVT	A2-NT2
1. TGN handhaven op A1-NVT en A2-NT2	TGN	TGN
2. Nieuw examen A1-NVT en A2-NT2	Nieuw examen	Nieuw examen
3. TGN alleen op A1-NVT handhaven	TGN	Nieuw examen

3.9 Buitenlandse kijk op taalexamens voor inburgeringsdoeleinden

In België heerst op het gebied van opleiden en examens een andere filosofie dan in Nederland. Waar in Nederland de nadruk op incidenteel en centraal examineren ligt, besteedt men in België meer aandacht aan continu examineren tijdens de taalopleiding. Voor integratie geldt in België dan ook een verplichte cursus zonder examen.

Op Europees niveau speelde de ALTE[4] Special Interest Group "Language Assessment for Migration & Integration[5]" tot 2009 een grote rol. Deze groep is sindsdien omgevormd tot een Expertteam onder de Raad van Europa. In beide hoedanigheiden worden periodiek EU-brede enquêtes op het gebied van examenbeleid en uitvoering uitgevoerd en gerapporteerd. Verder heeft deze groep algemene examenrichtlijnen voor beleidsmakers ontwikkeld[6]. Deze richtlijnen benadrukken het belang van continu monitoren ten behoeve van de kwaliteit, de mogelijke inbreuk op mensenrechten en het beschikbaar maken van voldoende middelen (tijd, geld) om een goed, eerlijk examen te ontwikkelen. Over de TGN heeft deze groep geen formele mening.

[4] Association of Language Testers Europe

[5] http://www.alte.org/projects/language_assessment_for_migration_and_integration_lami

[6] Language tests for social cohesion and citizenship - an outline for policymakers, ALTE (2008)

4 ANALYSE MARKTVERKENNING

Dit hoofdstuk is beperkt tot de varianten voor examens A1-NVT en/of A2-NT2 gebaseerd op bekende examenvormen:

1. Verbeterd examen spreekvaardigheid door middel van spraakvergelijking;
2. Machinale examinering met menselijke beoordeling achteraf;
3. Geïntegreerd praktijkgericht examen.

Doordat Gameplay een component is en nog niet ontwikkeld voor examentoepassing zijn eventuele uitspraken te speculatief.

4.1 Beschrijvingen dekken afname- en beoordelingsspectrum

De alternatieve examenvarianten verschillen met name op de feitelijke examinering (door mens of machine) en examenbeoordeling (door mens of machine). Zoals geïllustreerd met onderstaande figuur dekken ze tezamen alle logische combinaties.

		Examenbeoordeling door	
		Machine	**Mens**
Examinering door	**Machine**	Verbeterd examen spreekvaardigheid dmv spraakvergelijking	Machinale examinering met menselijke beoordeling achteraf
	Mens	-	Geïntegreerd praktijkgericht examen

Alle varianten kunnen bewijsvoering overleggen die suggereert dat er reële examens te ontwikkelen zijn zowel voor A1-NVT als A2-NT2.

4.2 Spraakvergelijking gebaseerde variant enige optie bij betrouwbaarheidsgrens 0,9 of hoger

Examens gebaseerd op menselijke beoordeling (en eventueel afname) kunnen met veel maatregelen een hoge beoordelingsbetrouwbaarheid halen in de buurt van de 0,8 die COTAN[7] voldoende acht voor "high stakes" examens[8]. Echter, vasthouden aan waarden van 0,9 of meer houdt impliciet een keuze in voor een volledig automatische examensysteem.

[7] Commissie Testaangelegenheden Nederland van het Nederlands Instituut voor psychologen (NIP)

[8] Arne Evers, Wouter Lucassen, Rob Meijer, Klaas Sijtsma (2010), "COTAN Beoordelingssysteem voor de kwaliteit van tests", p33.

4.3 Geïntegreerd praktijkgericht examen winnaar op indruksvaliditeit

Het is onwaarschijnlijk dat een aanscherping van meetniveaus bij en/of toevoeging van gesprekssimulatie aan een op spraakvergelijking gebaseerd examen de indruksvaliditeit zodanig zal beïnvloeden dat de kritiek wezenlijk minder wordt. Alleen veel meer transparantie en actieve communicatie daarover zou naar verwachting enig verschil kunnen maken.

Met name de menselijk examinering in combinatie met het aanzienlijke aandeel laagopgeleiden van de inburgeraars, maken dat een geïntegreerd praktijkgericht examen variant de hoogste indruksvaliditeit zal krijgen. Machinale examinering met menselijke beoordeling achteraf zal daar wel bij in de buurt kunnen komen.

4.4 Buitenland-logistiek lijkt onverenigbaar met geïntegreerd praktijkgericht examen

De aantallen kandidaten noch de binnenland-logistiek worden als beperkende factor gezien. De buitenland-logistiek is dat wel en daarmee de maatstaf voor de systeemcompatibiliteit van de alternatieve examenvarianten. Alle varianten claimen een examenvorm te kunnen ontwikkelen die uitvoerbaar is met de huidige, telefoon-gebaseerde buitenlandlogistiek. Desalniettemin is het duidelijk dat de menselijke examinering van het geïntegreerd praktijkgericht examen op gespannen voet staat met de 24/7 eis en alleen gegarandeerd kan worden met een organisatie die naar verwachting niet goedkoop zal zijn.

4.5 Fraudegevoeligheid / veiligheid lastig; Geheimhouding onmogelijk

Fraudegevoeligheid / veiligheid van een examen is een meer universeel probleem onafhankelijk van de examenvariant. Het punt en de voortdurende toename van (technische) mogelijkheden op dit gebied worden erkend, maar vooralsnog komt men hier niet met suggesties om fraude in te perken of de veiligheid te optimaliseren.

Geheimhouding lijkt een illusie en op zijn best alleen tijdelijk bereikbaar onder de continue druk van nieuwsgierige kandidaten en taalopleiders. Examenontwikkelaars zien snel verversen van de items, dat wil zeggen sneller dan ze bekend worden, als de voornaamste strategie. Voor een geïntegreerd praktijkgericht examen wordt geheimhouding vanwege het positieve terugslageffect als minder noodzakelijk gezien.

4.6 Operationeel per 1 november 2014 haalbaar voor examen

Voor alle alternatieve examenvarianten wordt aangegeven dat de doeldatum van 1 november 2014 operationeel inpasbaar moet zijn bij een aanvang van de ontwikkeling per begin 2014. Of dit ook voor het systeem als geheel geld is onduidelijk want geen van de (potentiële) examenontwikkelaars beschrijft het benodigde platform en de integratie daarmee (zie sectie 7.1 Aanbevelingen PvE).

4.7 Voor iedere vorm meer aanbiedingen te verwachten

Met de bekende commerciële examenaanbieders hebben voor iedere examenvariant verscheidene andere partijen interesse getoond in een toekomstige aanbesteding. Er zijn dan ook voor iedere examenvariant meer aanbiedende consortia denkbaar en te verwachten.

5 RESULTATEN MARKTVERKENNING

Voor de werkstroom Marktverkenning zijn de experts eerst in een interview gevraagd naar hun meningen ten aanzien van eventuele alternatieven voor het huidige examen spreekvaardigheid. In de praktijk bleek het lastig de vele dimensies van het interviewprotocol voor dit onderdeel in een gesprek volledig te adresseren. Daarom is aanvullend een systematiseringsslag uitgevoerd door middel van door ons ontwikkelde schriftelijke enquêtes. Alle hierop betrekking hebbende vragen uit het interviewprotocol waarop in redelijkheid een betekenisvol antwoord verwacht mag worden, zijn hierin verwerkt.[9]

Een eerste analyse van de interviewresultaten suggereerde vier grote gemene delers in de ideeën, namelijk drie varianten voor examens A1-NVT en/of A2-NT2 gebaseerd op bekende examenvormen:

1. Verbeterd examen spreekvaardigheid door middel van spraakvergelijking;
2. Machinale examinering met menselijke beoordeling achteraf;
3. Geïntegreerd praktijkgericht examen;

en een fundamenteel nieuwe ontwikkeling, te weten:

4. Gameplay met interactieve dialoog.

Om de responsedrempel zo laag mogelijk te leggen zijn deze vier vormen door ons in concept uitgewerkt in de enquête-formats en weer voorgelegd aan de experts. De vier grote gemene delers zijn daarbij (h)erkend en overeind gebleven en, ondanks een uitdrukkelijke uitnodiging, zijn er geen nieuwe meer toegevoegd.

Dit hoofdstuk volgt bovenstaande opdeling en is gebaseerd op een integratie van alle verkregen resultaten.

5.1 Verbeterd examen spreekvaardigheid door middel van spraakvergelijking

Beschrijving en onderbouwing: Een examen spreekvaardigheid door middel van spraakvergelijking is op een aantal punten te verbeteren ten opzichte van de huidige situatie:

- Aparte examens voor niveau A1 en niveau A2. Dit maakt het specifieke examen passender in de verschillende settings;
- Op niveau A2 kan een gesprek gesimuleerd worden via een korte reeks afhankelijke items, waarbij de computer de vervolgvraag laat afhangen van

[9] Alleen de punten transparantie, toegankelijkheid en terugkoppeling naar BZ en of DUO zijn niet opgenomen omdat de teneur van de interviews suggereerde dat alles inpasbaar was, mits pragmatisch gespecificeerd. Wij komen terug op deze aspecten in sectie 6.1 Aanbevelingen voor het Programma van Eisen.

het antwoord dat de kandidaat geeft. Hier is (voor andere talen) reeds enige ervaring mee.

Betrouwbaarheid: Onveranderd ten opzichte van huidige systeem, dat wil zeggen dat de betrouwbaarheid van dit alternatief hoger is dan 0,9. De verwachting is dat een apart examen per niveau minder tijd kost dan de huidige TGN. Het is onduidelijk hoeveel tijd een gesimuleerd gesprek zal kosten.

Validiteit: Minstens zo valide als het huidige systeem. Met de twee voorgestelde verbeteringen verbetert de validiteit.

Logistiek: De logistiek (examenafname en –beoordeling door computer) blijft onveranderd. De veranderingen zullen alleen in de software zitten. Het resultaat van het examen is direct beschikbaar.

Fraudegevoeligheid en veiligheid / Geheimhouding: Beide aspecten blijven ongewijzigd ten opzichte van het huidige systeem.

Ontwikkeling: Het wordt haalbaar geacht een verbeterd examen spreekvaardigheid door middel van spraakvergelijking per 1 november 2014 operationeel te krijgen. Het pretesten van nieuwe items en het organiseren van de juiste pretest populaties zijn het meest tijdrovend bij de ontwikkeling van nieuwe examens.

5.2 Machinale examinering met menselijke beoordeling achteraf

Beschrijving en onderbouwing: Machinale examinering met menselijke beoordeling achteraf wordt momenteel toegepast bij het Staatsexamen. Het Staatsexamen Nederlands als Tweede Taal (NT2) wordt gebruikt om het taalniveau van niet-Nederlandstaligen te examineren op het niveau B1 en B2. Taalvaardigheden worden apart geëxamineerd op aangewezen examenlokaties. Machinale examinering met menselijke beoordeling achteraf voor A1 en A2 is gebaseerd op de opzet van het Staatsexamen, maar aangepast aan woordenschat en belevingswereld van niveau A1 en A2 met toevoeging van een gesimuleerd gesprek.

- Voor niveau A1: vervanging huidige examen spreek- en luistervaardigheid en examen leesvaardigheid op basis van twee examens:
  - Examen Luisteren en Lezen op basis van multiple choice;
  - Examen Spreken en Gesprekken op basis van zinnen produceren en reactie op gesprekken.
- Voor niveau A2: een analoog examen Spreken en Gesprekken, aangepast aan woordenschat en belevingswereld A2.

Betrouwbaarheid: Op basis van de ervaringen met het Staatsexamen en de beoordelingssystematiek moet een betrouwbaarheid van ruim 0,8 haalbaar zijn. De verwachting is dat ieder examen zo'n 20 à 30 minuten zou moeten duren.

Validiteit: Het Staatsexamen maakt al gebruik van authentieke situaties. Vergelijkbaar materiaal in de toepasselijke context/belevingswereld is al ontwikkeld voor het Elektronisch Praktijk Examen op niveaus A1 en A2.

Logistiek: De logistiek van het Staatsexamen kan grotendeels gekopieerd worden: examinering via de telefoon (A1 en A2) of computer (A2), beoordeling op afstand door goed opgeleide en geïnstrueerde beoordelaars. Dit gebeurt gesegmenteerd en gerandomiseerd. Het resultaat van het examen kan binnen enkele weken beschikbaar zijn.

Fraudegevoeligheid en veiligheid / Geheimhouding: Menselijke beoordeling is fraudegevoeliger dan automatische beoordeling.

- A1 (buitenland): Procedures voor veiligheid en geheimhouding zullen in grote mate vergelijkbaar zijn aan die van de huidige TGN;
- A2 (binnenland): Procedures voor veiligheid en geheimhouding zullen in grote mate vergelijkbaar zijn aan die van de huidige Staatsexamen.

De items zullen regelmatig ververst moeten worden.

Ontwikkeling: Het wordt haalbaar geacht machinale examinering met menselijke beoordeling achteraf per 1 november 2014 operationeel te krijgen. Het pretesten van nieuwe items is het meest tijdrovend. De ontwikkeling kan versneld worden door het beschikbaar stellen van de technische infrastructuur voor de nieuwe examens.

5.3 Geïntegreerd praktijkgericht examen

Beschrijving en onderbouwing: In een geïntegreerd praktijkgericht examen worden meerdere taalvaardigheden tegelijk geëxamineerd in een gesprek. De persoon die het examen afneemt (de examinator) is niet noodzakelijkerwijs de persoon die het examen beoordeeld. Het gesprek gaat over een situatie waar de kandidaat mee bekend is. Het huidige CNaVT is hier een voorbeeld van.

Betrouwbaarheid: Examinering door middel van een gesprek met een mens is minder betrouwbaar dan automatische examinering. Examenbeoordeling door een mens is eveneens minder betrouwbaar dan door een computer. Dit is te ondervangen door de volgende maatregelen:

- Goed opleiden en instrueren van examinatoren;
- Verschillende onderdelen door verschillende beoordelaars laten beoordelen (segmenteren);
- Opnames van het examen at random naar beoordelaars sturen (randomiseren);
- Twee beoordelaars een beoordeling uit laten voeren, bij twijfel een derde beoordelaar inschakelen;
- Antwoordmodellen en beoordelingscriteria goed vastleggen.

Met dergelijke maatregelen is een betrouwbaarheid van ongeveer 0,8 haalbaar (betrouwbaarheid van het CNaVT schommelt tussen de 0,75 en 0,9). Het examen zal waarschijnlijk ruim 20 minuten duren.

Validiteit: De validiteit van dit examen is hoog omdat er is namelijk sprake is van een gesprek met een gesprekspartner die de vraag kan herhalen (conform CEF-beschrijving A1 en A2).

Logistiek: De examinering en examenbeoordeling gebeuren beide door een mens. De examinator hoeft niet noodzakelijkerwijs het examen te beoordelen. Een opname van het gesprek kan naar de beoordelaar(s) gestuurd worden. Het resultaat van het examen kan binnen enkele weken beschikbaar zijn, afhankelijk van de ingezette middelen.

- Op niveau A1, in het buitenland, zal het gesprek over de telefoon of via Skype plaatsvinden. Om deze variant over de hele wereld dagelijks aan te kunnen bieden, zullen examinatoren nachtdiensten moeten draaien;
- Op niveau A2, in Nederland, kunnen directe gesprekken plaatsvinden.

Fraudegevoeligheid en veiligheid / Geheimhouding: Geheimhouding is bij een geïntegreerd praktijkgericht examen niet van groot belang. Indien kandidaten zich op dit examen voorbereiden heeft dit een positief effect op het taalniveau. Hoe persoonlijker het examen ingevuld kan worden, hoe minder de noodzaak voor geheimhouding.

Examinatoren zullen goed getraind moeten worden en het protocol moet zodanig vormgegeven worden, dat de kandidaat niet kan volstaan met uit het hoofd geleerde antwoorden.

Ontwikkeling: Het wordt haalbaar geacht het geïntegreerd praktijkgericht examen per 1 november 2014 operationeel te krijgen. Dit zal zeker het geval zijn indien de ontwikkeling gebaseerd wordt op bestaande ervaringen met deze examenvorm.

5.4 Gameplay met interactieve dialoog

Beschrijving en onderbouwing: Via gameplay met interactieve dialoog wordt op een (visueel) aantrekkelijke wijze het examen afgenomen. Het examen kan vervolgens automatisch of door mensen beoordeeld worden. Gameplay is een veel toegepaste manier om taalvaardigheden te trainen. Examinering via gameplay staat nog in de kinderschoenen.

Betrouwbaarheid: De betrouwbaarheid is afhankelijk van de wijze waarop deze variant toegepast wordt. Gameplay zal niet meer zijn dan een component van een groter geheel.

Validiteit: De validiteit is afhankelijk van de wijze waarop deze variant toegepast wordt. Indien de gameplay erin slaagt op realistische wijze praktijksituaties na te bootsen, kan dit de validiteit bevorderen.

Logistiek: Examinering gebeurt via de computer, de beoordeling kan automatisch of menselijk plaatsvinden.

- Op niveau A1 is examinering via de telefoon denkbaar, maar dit lijkt weinig zinvol. Directe afname door de computer en beeldscherm is passender. Dit betekent dat er computers met voldoende rekenkracht op de posten gezet moeten worden;
- Op niveau A2 zal uitgezocht moeten worden of de huidige computers op de examenlokaties geschikt zijn voor de gameplay.

Fraudegevoeligheid en veiligheid / Geheimhouding: Gezien de interactieve dialoog is er veel variatie in het examen, wat het het examen goed houdbaar maakt.

Ontwikkeling: Systemen voor taalverwerving op basis van gameplay zijn beschikbaar. Echter, taalexaminering op basis van gameplay is nog geen bewezen technologie. Het kost tijd dit te ontwikkelen. Het lijkt dan ook onwaarschijnlijk dat dit examen op 1 november 2014 operationeel kan zijn.

De ontwikkeling kan versneld worden indien TGN geluidsbestanden beschikbaar worden gesteld aan onderzoekers (geluidsbestanden gekoppeld aan de triggers, oftewel de vragen. Bij voorkeur ook gekoppeld aan persoonlijke gegevens kandidaat: geboorteland, geslacht, leeftijd, etc).

6 Conclusies

De globale richting van een nieuwe aanbesteding voor spreekvaardigheidsexamens zal sterk afhankelijk zijn van de weging van betrouwbaarheid / validiteit en de bereidheid tot investeren aan overheidskant:

6.1 Bij hoogste betrouwbaarheid is een spraakvergelijking gebaseerde variant enige keuze

Indien de hoogste betrouwbaarheid allesbepalend zal zijn, wordt de keuze in feite beperkt tot een examen spreekvaardigheid door middel van spraakvergelijking.

6.2 Validiteit en verandering A1-NVT en A2-NT2 leidt tot machinale examinering met menselijke beoordeling achteraf

Indien de (indruks)validiteit oftewel het maatschappelijk draagvlak doorslaggevend is en er tegelijk een wil is tot verandering van de spreekvaardigheidsexamens voor binnenland en buitenland, is machinale examinering met menselijke beoordeling achteraf het logische alternatief.

6.3 Validiteit en alleen verandering A2-NT2 dan keuze tussen machinale examinering met menselijke beoordeling achteraf en geïntegreerd praktijkgericht examen bepaald door integrale kostprijs

Indien de (indruks)validiteit oftewel het maatschappelijk draagvlak doorslaggevend is en er alleen een wil is tot verandering van spreekvaardigheidsexamens voor binnenland, komen de volgende varianten in principe in aanmerking:

- Machinale examinering met menselijke beoordeling achteraf;
- Geïntegreerd praktijkgericht examen.

De keuze volgt dan waarschijnlijk uit de totale kosten van ontwikkeling en exploitatie. Hoe deze keuze uitvalt is niet op voorhand te zeggen.

7 AANBEVELINGEN VOOR HET PROGRAMMA VAN EISEN

7.1 Aanbevelingen voor het programma van eisen

7.1.1 Specificeer functioneel in plaats van technisch

Technisch specificeren is alleen productief indien de opdrachtgever precies weet welk product of welke dienst hij aan wil schaffen. Indien de overheid technisch zou specificeren, neemt zij in feite de rol van de experts over.

Het zoveel mogelijk functioneel specificeren van de eisen aan het examen zal creatieve uitwerkingen bij indienende partijen prikkelen en hen uitdagen. Ook kunnen partijen reeds delen of elementen van een gewijzigd of nieuw spreekvaardigheidsexamen "op de plank hebben liggen" wat de ontwikkeltijd zou kunnen verkorten. Dat betekent overigens niet dat er geen randvoorwaarden gesteld moeten worden waaraan een examen zou moeten voldoen.

7.1.2 Staar je niet blind op betrouwbaarheid

Een hoge betrouwbaarheid is belangrijk bij een examen, zeker bij zogenaamde High Stake examens. Met het vooraf definiëren van een betrouwbaarheid van 0,90 worden echter alternatieven voor het huidige spraakvergelijking gebaseerde examen uitgesloten. Ook andere High Stake examens zoals de Staatsexamens NT2 halen deze betrouwbaarheid niet. Daarnaast gaat een hoge betrouwbaarheid ten koste van de validiteit en zeker de indruksvaliditeit. Laat partijen die aanbiedingen doen helder maken hoe hoog de betrouwbaarheid wordt. Gebruik de betrouwbaarheid als één van selectiecriteria en niet als een knock-out criterium.

7.1.3 Transparantie over de werking van het examen vergroot de acceptatie

Een belangrijk deel van de kritiek op de TGN spitst zich toe op de werking van het examen. Voor veel deskundigen is het feit dat het examen en de wijze van beoordeling door de machine een "black box" zijn een bron van ergernis en zelfs achterdocht. Dit leidt tot een laag acceptatieniveau van de TGN bij wetenschappers, taaldocenten en referenten. Om deze weerstand tegen een toekomstig examen te ondervangen, zou transparantie over de werking van het examen en de wijze van beoordeling, inclusief scoring, een goed selectiecriterium kunnen zijn in de gunning van een aanbesteding voor een nieuw examen.

7.1.4 Toegankelijkheid in het buitenland waarborgen

Door de afname van het aantal posten in het buitenland, neemt voor een aantal inburgeraars de reistijd voor het afleggen van een examen toe. Om het aantal posten waar het examen afgenomen wordt niet nog verder af te laten nemen, en daarmee de reistijd nog verder te vergroten, dient het nieuwe examen aan de ICT

eisen van de post met de laagste kwaliteit ICT infrastructuur (telefoonlijnen, Internetverbinding, bandbreedte, bedrijfszekerheid, etc.) te voldoen.

7.1.5 Geef helder aan op welk ICT platform het examen dient te draaien

Momenteel draait het inburgeringsexamen in Nederland op het platform van SCI van Andriessen. Andriessen is een subcontractor van DUO, de uitvoerder van de inburgeringsexamens in Nederland. Dit contract loopt nog tenminste een jaar door. Daarnaast wordt momenteel door DUO gewerkt aan de aanbesteding van een nieuw platform met een bredere functionaliteit. Tegelijkertijd vinden er ontwikkelingen plaats rond het systeem Facet van het CvE waar in de toekomst alle door het CvE begeleide examens (inclusief Staatsexamens NT2) op gaan draaien. Dit zou ook een optie kunnen zijn voor (delen) van het toekomstig inburgeringsexamen in Nederland.

De keuze van het ICT-platform hangt ook ten nauwste samen met de wijze waarop de terugkoppeling van het resultaat naar Buitenlandse Zaken, inclusief de betreffende post, en DUO geïmplementeerd kan worden. Het verdient in ieder geval aanbeveling vooraf met beide eindgebruikers te overleggen over wat hun wensen en mogelijkheden naar aanleiding van hun leerervaringen met de TGN om daar vervolgens een pragmatische vertaling van te maken voor het PvE.

7.1.6 Let op de integrale kostprijs

De ontwikkelingskosten van een nieuw examen zullen naar verwachting aanzienlijk zijn maar toch slechts een deel van de totale kosten. Additionele kosten kunnen onder andere zijn: kosten van implementatie op ICT systemen, kosten voor aanvullende onderzoeken (zoals in het verleden) om betrouwbaarheid aan te vullen of nader te onderbouwen, exploitatiekosten, kosten voor verversing van de items, pretestkosten van nieuwe items, etc. Probeer in de aanbesteding ook zicht te krijgen op deze kostenposten en welke acties de aanbiedende partijen zullen ondernemen om deze transparant te maken, zodanig dat ze een rol in de keuze van de nieuw examen kunnen spelen.

7.2 Algemene aanbevelingen

Bij de uitvoering van deze opdracht is het ons opgevallen dat er zowel aan de kant van het veld als aan die van de overheid een groot wij-zij gevoel leeft. Niet de illusie hebbende dat alle kritiek daarmee weg te nemen is, denken wij dat de scherpte uit het debat te halen zal zijn met een meer outgoing en communicatieve houding van de overheid. Wij geven u dan ook in overweging om als overheids-vertegenwoordigers meer zichtbaar te worden in de verschillende professionele fora die zich met de problematiek van inburgeringsexamens bezig houden en met regelmaat wat tekst en uitleg te geven over de verschillende keuzes die gemaakt zijn of overwogen (moeten) worden. Niets leidt sneller tot onbegrip en polarisatie dan afstand en afwezige communicatie of indirecte communicatie via media of rapporten.

BIJLAGE: LIJST VAN GERAADPLEEGDE EXPERTS

Voornaam	Achternaam	Organisatie
Ad	Appel	Ad Appel Taaltrainingen
Bart	Bossers	VU
Annelies	Braams	Nedles
Annemarie	de Groot	CINOP
John	de Jong	Pearson Knowledge Technologies / VU Amsterdam
Hanneke	de Weger	Taalunie
Jan Erik	Grezel	Onafhankelijk taaldocent
Elwine	Halewijn	ITTA UvA
Jan	Hulstijn	UvA
Judith	Janssen	CITO
Martine	Jetter	CITO
Anne	Kerkhoff	Fontys
Judith	Kessens	TNO
Jeanne	Kurvers	Tilburg University / ROC Tilburg
Jo Fond	Lam	CINOP
Martin	Nuttel	ALTE
Mirna	Pit	Bureau ICE
Anne	Poppema	DUO
Inge	Rogiers	Huis van het Nederlands (BE)
Helmer	Strik	Radboud Universiteit Nijmegen, CLST
Franke	Teunisse	Bureau ICE
Yolanda	Thurkow	Ministerie van Buitenlandse Zaken
Jeanine	Treep	CITO
Piet	van Avermaet	ALTE / Universiteit van Gent (BE)
Ineke	van de Craats	Gepensioneerd (voormalig Radboud Universiteit Nijmegen)
Marja	van den Dungen	CINOP
Teun	van Iperen	Flevotaal
Bernard	Veldkamp	Universiteit Twente
Simon	Verhallen	CvE
Bert	Werkman	DUO