

Staatsexamen NT2. Het gewenste niveau.
Deel 4 Cesuurindicatie Spreken en Schrijven Programma I

Cito
Augustus 2005

Henk Kuijper
Aukje Bergsma
Timo Bechger

Inhoudsopgave

1	Inleiding	3
2	Schets van internationale ontwikkelingen rondom het relateren van taalexamens aan het CEFR	4
3	Opzet van het cesuuronderzoek	6
3.1	Het uitgangspunt: het gewenste niveau	6
3.2	Samenstellen van de Raamwerkbeoordelingslijsten	6
3.3	De selectie van de kandidaten	7
3.4	De beoordelingsprocedure	7
4	De relatie tussen Staatsexamen Spreken II en de Raamwerkbeoordeling	8
4.1	Globale beoordeling Spreken	8
4.1.1	Opzet beoordeling	8
4.1.2	Beoordelaarsovereenstemming	8
4.1.3	Relatie tussen de Raamwerkbeoordeling en de STEX-beoordeling	9
4.1.4	Samenvatting beoordelaarsovereenstemming en relatie tussen Raamwerkbeoordeling en STEX-score Spreken	14
4.2	Relatie tussen STEX-scores en Raamwerkbeoordeling: de cesuurkwestie	15
4.2.1	Beheersing B1 volgens het criterium meerderheidsoordeel	16
4.2.2	Afwijkende resultaten Spreken I	22
4.2.3	Samenvatting en aanbevelingen cesuur Spreken I	23
5	De relatie tussen Staatsexamen Schrijven II en de Raamwerkbeoordeling	25
5.1	Globale beoordeling Schrijven	25
5.1.1	Opzet beoordeling	25
5.1.2	Beoordelaarsovereenstemming	25
5.1.3	Relatie tussen de Raamwerkbeoordeling en de STEX-beoordeling	26
5.1.4	Samenvatting beoordelaarsovereenstemming en relatie tussen Raamwerkbeoordeling en STEX-score Schrijven	29
5.2	Relatie tussen STEX-scores en Raamwerkbeoordeling: de cesuurkwestie	29
5.2.1	Beheersing B1 volgens het criterium meerderheidsoordeel	30
5.2.2	Samenvatting en aanbevelingen Schrijven I	34
6	Verantwoording van de gebruikte methode	36
6.1	Inleiding	36
6.2	De ROC-curve analyse	36
6.3	Andere methoden	38

Literatuur

Bijlage 1 Globale raamwerkbeoordeling Spreken

Bijlage 2 Globale raamwerkbeoordeling Schrijven

Bijlage 3 Familiarisatie Spreken

Bijlage 4 Familiarisatie Schrijven

Bijlage 5 Korte impressie kandidaten met grote discrepantie tussen Raamwerk en STEX-beoordeling

1 Inleiding

Begin 2003 heeft het Ministerie van Onderwijs, Cultuur en Wetenschappen het verzoek gedaan om de eindtermen van de opleidingen I en II te evalueren. Citogroep is uitvoerder van dit onderzoek.

Concreet komt het er op neer dat in het onderzoek wordt geïnventariseerd welk taalniveau kandidaten moeten hebben na het behalen van het Staatsexamen NT2, als zij starten met een vervolgopleiding, dan wel instromen op de arbeidsmarkt.

In het onderzoek wordt het taalniveau uitgedrukt in termen van het Raamwerk NT2, dat een vertaling is van het Common European Framework of Reference (Raad van Europa, 2001). Het onderzoek is al uitgevoerd voor Programma II van het Staatsexamen; in het huidige rapport wordt verslag gedaan van het onderzoek naar Programma I.

De resultaten van het onderzoek worden o.a. gebruikt om te komen tot een heroverweging van de cesuren van de verschillende onderdelen van het Staatsexamen NT2.

Het evaluatie-onderzoek bestaat uit twee delen. In het eerste deel van het onderzoek is een behoeftepeiling uitgevoerd bij afnemers van het Staatsexamen Programma I, de MBO-opleidingen. Het doel van de behoeftepeiling was na te gaan welk niveau van het Common European Framework of Reference for Languages (Council of Europe, 2001) door de afnemers gezien wordt als het meest wenselijke niveau voor Staatsexamen NT2, Programma I. Over de resultaten van deze behoeftepeiling is gepubliceerd in Kuijper e.a. (2005).

Het tweede deel van de evaluatie bestaat voor Spreken en Schrijven uit een onderzoek dat gericht is op de vraag bij welke scores van het Staatsexamen men kan aannemen dat een kandidaat op het gewenste niveau Schrijven of Spreken functioneert. Er is onderzocht of er een relatie tussen de Staatsexamenscores en het Raamwerk NT2 bestaat en of deze relatie voldoende is om uitspraken over een eventuele herziening van de cesuur te doen.

In het voorliggende onderzoeksrapport wordt verslag gedaan van het cesuuronderzoek voor de examens Spreken en Schrijven Programma I.

Dit rapport is deel 4 van een reeks publicaties.

Over Programma I en II zijn reeds de volgende publicaties uitgebracht:

- Henk Kuijper, Aukje Bergsma, Timo Bechger (2004). Staatsexamen NT2. Het gewenste niveau.
Deel 1 Behoeftetepeiling Programma II. Citogroep Arnhem, augustus 2004.
- Henk Kuijper, Aukje Bergsma, Timo Bechger (2004). Staatsexamen NT2. Het gewenste niveau.
Deel 2 Cesuurindicatie Spreken en Schrijven Programma II. Citogroep, Arnhem, oktober 2004.
- Henk Kuijper, Aukje Bergsma, Timo Bechger (2004). Staatsexamen NT2. Het gewenste niveau.
Deel 2B Cesuurindicatie Spreken en Schrijven Programma II. Aanvullend onderzoek. Citogroep, Arnhem, december 2004.
- Henk Kuijper, Aukje Bergsma, Timo Bechger (2004). Staatsexamen NT2. Het gewenste niveau.
Deel 3 Behoeftetepeiling Programma I. Citogroep Arnhem, 2005.

Het rapport begint met een korte schets van de internationale ontwikkelingen rond het relateren van een examen aan het CEFR (hoofdstuk 2), vervolgens wordt de onderzoeksopzet gepresenteerd (hoofdstuk 3) en volgen de resultaten voor Spreken (hoofdstuk 4) en Schrijven (hoofdstuk 5). Zowel hoofdstuk 4 als 5 wordt afgesloten met een samenvatting en aanbevelingen. Hoofdstuk 6 bevat een verantwoording van de gebruikte onderzoeksmethode.

2 Schets van internationale ontwikkelingen rondom het relateren van taalexamens aan het CEFR

Sinds het verschijnen van het Common European Framework of Reference (Council of Europe, 2001) worden er in Europa en daarbuiten talloze activiteiten ontplooid rondom het relateren van taalexamens aan het CEFR. Om deze activiteiten te faciliteren en een wildgroei aan linking technieken te voorkomen heeft de Language Policy Division van de Council of Europe een Manual laten opstellen waarin richtlijnen staan die exameninstellingen kunnen volgen om hun taalexamens te relateren aan het CEFR. De pilotversie van deze Manual (Language Policy Division, 2003) is inmiddels verschenen. Deze pilotversie wordt in 2004 en 2005 becommentarieerd en in pilotprojecten getest. Op grond van bevindingen en commentaar zal waarschijnlijk in 2006 of 2007 de definitieve versie van de Manual verschijnen.

De Manual beklemtoont dat zowel inhoudelijke als psychometrische argumenten moeten bijdragen aan de claims die constructeurs van een examen maken met betrekking tot de relatie van hun examens met het CEFR. Volgens de Manual moeten deze claims op drie elkaar aanvullende manieren worden onderbouwd, waarbij met elke stap de claim sterker wordt:

- Specificatie. Bij deze werkwijze wordt een uitgebreide inhoudelijke analyse uitgevoerd van de inhoudelijke aspecten van een examen.
- Standaardisatie. In deze fase wordt het examen vergeleken met benchmark-examens en -opgaven waarvan een bepaald niveau is aangetoond.
- Empirische validatie. Tijdens deze fase worden empirische data aangedragen over de interne validiteit en de relatie met het CEFR.

Inmiddels zijn exameninstellingen – ook buiten Europa – in het proces hun examens te linken aan het CEFR. Ook een associatie als ALTE ontplooit op het ogenblik initiatieven om tot inzichtelijke procedures te komen in deze ingewikkelde problematiek.

Het is nog te vroeg om over resultaten op basis van de toepassing van de Manual uitspraken te doen. In het kader van deze studie – indicaties geven voor de mogelijke nieuwe cesuur van het Staatsexamen Spreken en Schrijven – is het goed kort stil te staan bij twee studies die - ondanks verschillen - qua doelstelling vergelijkbaar zijn met het onderhavige onderzoek.

De eerste studie heeft betrekking op TOEFL, TSE, TWE en TOEIC. In deze studie heeft ETS onderzoek gedaan naar de vraag welke toetsscores gerelateerd zouden zijn aan respectievelijk niveau B1 en C1 van het CEFR. We beperken ons hier tot de Spreek- en Schrijftoetsen (TSE en TWE). Hiertoe werden de descriptoren uit de Illustrative Scales in het CEFR gebruikt om oordelen uit te spreken over het niveau van kandidaten. Vervolgens werden gemiddelde TSE-scores en standaarddeviaties gebruikt om de minimale score te bepalen waarbij men zou kunnen zeggen dat een kandidaat niveau B1 of C1 beheerst (zie Tannenbaum, Wylie, 2004).

Overigens wijst ETS er op zijn website op dat bij de interpretatie van de scores in het licht van het CEF de nodige voorzichtigheid in acht genomen moet worden:

'Institutions, employers, English programs, and learners can interpret the B1 and C1 recommended minimum scores on the TOEFL, TSE, TWE, and TOEIC tests by referring to the global descriptors and the exemplar skills descriptors on page 4.

Although the results help to interpret scores on these ETS tests, it is important to remember the following:

- 1. The CEF is a conceptual framework created by linguistic experts and teaching specialists, but it is not an empirically-based framework (Although the CEF descriptors are inspired by research, they are not empirically tested nor currently supported by performance data.)*
- 2. These results reported above represent the recommendations of informed experts. However, they should be considered **guidelines** because absolute, unequivocal cut-scores do not exist.'* (ETS,2004)

Ook voor PhonePass™ SET10 is een relatie gelegd tussen de overall score op PhonePass en het CEFR (De Jong, Bernstein, z.d.). In dit onderzoek werden de uitingen van 121 kandidaten beoordeeld met een holistisch oordeel over hun positie op de CEFR schaal. In dit onderzoek werd gebruik gemaakt van een holistische beoordelingsschaal die gebaseerd was op de Overall Spoken Interaction Scale (Council of Europe, 2001, pag. 74), maar was herschreven om elementen uit de strategische, pragmatische en linguïstische schalen op te nemen (De Jong, Bernstein, z.d., pag. 3).

Evenals in deze beide onderzoeken is dit Staatsexamenonderzoek erop gericht een relatie te leggen tussen de toetsscores van een bestaand examen en de CEFR-niveaus, door de examenuitingen van kandidaten te beoordelen aan de hand van een beoordelingslijst die ontleend is aan het CEFR. Een belangrijk verschil tussen dit Staatsexamenonderzoek en de beide andere onderzoeken is gelegen in het feit dat in de andere onderzoeken relaties met meerdere niveaus worden gelegd, terwijl we in dit Staatsexamenonderzoek slechts geïnteresseerd zijn in de relatie met één niveau: het door de afnemers gewenste niveau B1. Net als in het ETS- en PhonePass-onderzoek werken wij ook met globale of holistische uitspraken. Daarnaast is in de eerste fase van het cesuuronderzoek voor Programma II tevens een analytische beoordeling toegepast, om een crosscheck uit te voeren op de gevonden resultaten bij de globale beoordeling. Op grond van de resultaten van dit deel van het onderzoek (Kuijper e.a., 2004) is besloten om voor Programma I te kiezen voor uitsluitend een globale beoordelingsmethode.

Methoden van standaardsetting

Het beantwoorden van de vraag naar de relatie tussen de cesuurscores Spreken en Schrijven en het gewenste niveau van kandidaten in termen van het CEFR is een vorm van een standaard setting procedure.

In december 2004 heeft de Language Policy Division van de Raad van Europa een supplement bij de Pilot Manual gepubliceerd (Language Policy Division, 2004).

In de sectie over standaard setting deelt Felianka Kaftandjieva (2004) de talloze methoden van standaardsetting in in twee hoofdcategorieën op grond van de focus van de procedure.

1 Test-gerichte standaardsetting

2 Kandidaat- / prestatiegerichte standaardsetting

In de testgerichte standaardsettingprocedures concentreert de werkwijze zich op het beoordelen van testtaken, terwijl de kandidaatgerichte standaardsetting zich richt op de beoordeling van de prestaties van kandidaten.

In dit onderzoek naar de cesuur van de Staatsexamens Spreken en Schrijven is een vorm van prestatiegerichte standaardsetting toegepast. Dit betekent niet dat de taken in een examen in deze procedure niet van belang zouden zijn. Het is vanzelfsprekend dat de spreek- en schrijftaken geschikt moeten zijn om reacties te ontlokken op de gewenste niveaus. Uit het onderzoek naar de cesuren voor Schrijven en Spreken Programma II (Kuijper e.a., 2004) en de resultaten van het huidige onderzoek voor Programma I blijkt dat de door de Staatsexamens uitgelokte spreek- en schrijfprestaties goed beoordeelbaar zijn op de gewenste niveaus (B2 voor Programma II en B1 voor Programma I). Andere aanwijzingen voor deze positionering worden gevonden in het Referentiekader NT2 (Liemberg en Hulstijn, 1996), waar op grond van eindtermenanalyse en analyses van de voorbeeldexamens Schrijven en Spreken voor Programma I en II respectievelijk worden geplaatst in het NT2-niveaubereik niveau 3 en 4, vergelijkbaar met respectievelijk B1 en B2 van het CEF.

3 Opzet van het cesuuronderzoek

3.1 Het uitgangspunt: het gewenste niveau

Het onderzoek naar het door afnemers gewenste niveau van het Staatsexamen Programma I en naar indicaties voor de consequenties van dit gewenste niveau voor de cesuren van de Spreek- en Schrijfexamens Programma I bestaat uit 3 stappen:

Stap 1: Vaststelling (in Raamwerktermen) van het taalcompetentieprofiel van doelgroep I en II

Stap 2: Beoordelen van schrijf- en spreekproducten aan de hand van de kwalitatieve criteria van het corresponderende Raamwerkniveau.

Stap 3: Beoordelen van dezelfde producten als in stap 2, aan de hand van de beoordelingsmodellen van het staatsexamen.

Over stap 1 is gerapporteerd in Kuijper e.a, 2005. Op grond van de resultaten van het behoefteonderzoek blijkt dat het gewenste niveau voor het Staatsexamen I Spreken en Schrijven niveau B1 is. In Stap 2 en 3 werden allereerst de schrijf- en spreekuitingen van 60 kandidaten beoordeeld op grond van de beheersingsuitspraken in het Raamwerk NT2. Het doel van dit deel van het onderzoek was na te gaan in hoeverre er een relatie bestaat tussen de STEX-scores en het Raamwerk NT2 en of er indicaties gegeven kunnen worden in hoeverre de huidige cesuur overeenkomt met het niveau dat op grond van het behoefteonderzoek (dus B1) als wenselijk wordt gezien.

In deze fase van het onderzoek kunnen de volgende stappen worden onderscheiden:

1. Het vaststellen van een beoordelingslijst die direct aansluit op de niveaubeschrijvingen B1 in het Raamwerk NT2. De constructie van dit raamwerkbeoordelingsmodel wordt beschreven in paragraaf 3.2.
2. Het beoordelen van de spreek- en schrijfuitingen van 60 kandidaten met deze raamwerkbeoordelingslijst. De opzet van deze nieuwe beoordeling van spreek- en schrijfuitingen wordt beschreven in paragraaf 3.3 en 3.4.
3. Het vergelijken van deze resultaten met de scores op grond van de oorspronkelijke beoordeling van de kandidaten in Staatsexamen I. De bespreking van de wijze van vergelijkingsanalyses en de resultaten ervan worden gepresenteerd in hoofdstuk 4 en 5 voor respectievelijke Spreken I en Schrijven I.

3.2 Samenstellen van de raamwerkbeoordelingslijsten

Bij de constructie van het raamwerkbeoordelingsmodel hebben we ons gebaseerd op de relevante beschrijvingen uit het Raamwerk NT2. In de eerste fase van het onderzoekstraject zijn voor Schrijven en Spreken Programma II zowel analytische als globale beoordelingslijsten ontwikkeld. Op grond van de resultaten van dit onderzoek is besloten het onderzoek uit te voeren met uitsluitend globale beoordelingschalen voor de bepaling van het CEFR-niveau (Kuijper e.a. december 2004).

Globale beoordeling

In de globale beoordelingslijsten zijn de algemene beheersingsuitspraken B1 uit het Raamwerk NT2 (Dalderop, Teunisse, 2002) opgenomen van Schrijven (Raamwerk NT2, pag.87) en voor Spreken de algemene beheersingsuitspraken voor Spreken (Raamwerk NT2, pag.72) en Gesprekken voeren (Raamwerk NT2, pag. 53).

Deze Raamwerkuitspraken corresponderen met de formuleringen in de self assessment grid op pagina 26 en 27 van het CEFR (Council of Europe, 2001).

Zowel voor Spreken en Schrijven is een algemene uitspraak toegevoegd waarin de beoordelaar gevraagd werd of de kandidaat gezien het gehele examenwerk B1 beheerst of niet. De globale beoordelingslijsten voor Spreken en Schrijven zijn te vinden in bijlage 1 en 2.

3.3 De selectie van de kandidaten

Voor de beoordeling van het werk van de STEX-I kandidaten zijn in de eerste fase van het cesuuronderzoek 60 kandidaten voor Spreken en Schrijven geselecteerd.

Voor Schrijven I is het examen van januari 2005 gebruikt. De kandidaten zijn geselecteerd uit het scoregebied tussen 18 en 54 ruwe scorepunten. De cesuur van dit examen lag bij 35 ruwe scorepunten. De geselecteerde kandidaten hadden een schaalscore variërend van 454 tot 560.

Voor Spreken I is eveneens het examen van januari 2005 gebruikt. De kandidaten zijn geselecteerd uit het scoregebied tussen 44 en 81 ruwe scorepunten. De cesuur lag bij 59 ruwe scorepunten. De geselecteerde kandidaten hadden een schaalscore variërend van 469 tot 542.

Voor beide programma's geldt dat meer dan 75% van de kandidaten in dit scoregebied zaten.

3.4 De beoordelingsprocedure

Voor zowel Schrijven als Spreken is iedere kandidaat globaal beoordeeld door twee onafhankelijke beoordelaars. De beoordelaars zijn medewerkers van Cito en Bureau ICE die bekend zijn met het Raamwerk NT2 en het Common European Framework of Reference.

Bij deze globale beoordeling moesten de beoordelaars eerst het hele examen van een kandidaat beluisteren of lezen en daarna aangeven of de uitingen al of niet voldeden aan de uitspraken in de beoordelingslijst.

In totaal hebben 10 beoordelaars aan het onderzoek deelgenomen. Bij de verdeling van de kandidaten over de beoordelaars is ervoor gezorgd dat alle mogelijke combinaties voorkwamen. De beoordelaars voerden hun taak onafhankelijk van elkaar uit en wisselden geen informatie over de beoordeelde kandidaten met elkaar uit. De STEX-score van de te beoordelen kandidaten was bij de beoordelaars niet bekend.

Familiarisatie

Hoewel alle beoordelaars goed bekend waren met zowel het CEFR als het Raamwerk NT2 ging er aan de beoordeling een familiarisatiefase vooraf. De beoordelaars kregen de beschikking over de voor de beoordelingstaak relevante schalen en niveaubeschrijvingen. Deze moesten zij vóór het beoordelen doornemen en tijdens het beoordelen raadplegen. De documenten zijn opgenomen als bijlage 3 en 4.

4 De relatie tussen de score op het Staatsexamen Spreken I en de Raamwerkbeoordeling

4.1 Globale beoordeling Spreken

4.1.1 Opzet beoordeling

Bij de globale beoordeling is elke kandidaat beoordeeld door twee onafhankelijke beoordelaars. Zij moesten vijf uitspraken over de spreekvaardigheid van de cursisten beantwoorden met ja of nee. Hun oordelen zijn gebaseerd op het beluisteren van het STEX-I spreekexamen van de kandidaten. De globale raamwerkbeoordelingslijst is opgenomen in bijlage 1.

De eerste vier uitspraken zijn letterlijk ontleend aan de omschrijvingen van het beheersingsniveau B1 Spreken en Gesprekken Voeren van het Raamwerk NT2 (respectievelijk pagina 72 en 53). De laatste uitspraak is de meest algemene beoordeling: een algemene uitspraak in hoeverre een kandidaat volgens de beoordelaar voldoet aan de eisen voor Spreken en Gesprekken Voeren B1. Met deze globale oordelen werden 60 kandidaten beoordeeld door twee onafhankelijke beoordelaars. Eén kandidaat is uit het bestand verwijderd omdat bleek dat deze niet alle opgaven van het Staatsexamen had gemaakt.

Tabel 4.1-1 Globale raamwerkbeoordeling

	Uitspraken Raamwerk NT2	Ja	Nee
S-G1	Kan uitingen op een eenvoudige manier aan elkaar verbinden, zodat hij/zij gebeurtenissen, dromen, verwachtingen en ambities kan beschrijven.		
S-G2	Kan in het kort redenen en verklaringen geven voor meningen en plannen.		
GV-G3	Kan een verhaal vertellen of de plot van een boek of film weergeven en zijn/haar reacties beschrijven.		
GV-G4	Kan onvoorbereid deelnemen aan een gesprek over onderwerpen die vertrouwd zijn, of zijn/haar persoonlijke belangstelling hebben of die betrekking hebben op het dagelijks leven (bijvoorbeeld familie, hobby's, werk, reizen en actuele gebeurtenissen).		
SGV-G5	Beheerst deze kandidaat over het algemeen niveau B1?		

De score van elke beoordelaar kon variëren van 0 tot 5. Indien er tussen de twee beoordelaars van een kandidaat een verschil van 3 of meer scorepunten werd geconstateerd, werd er een derde beoordeling uitgevoerd. Vervolgens werden de twee scores die het dichtst bij elkaar lagen in de verdere analyses meegenomen. Een herbeoordeling werd in 11 gevallen uitgevoerd.

4.1.2 Beoordelaarsovereenstemming

In tabel 4.1-2 wordt voor elk van deze globale oordelen het percentage overeenstemming en de kappa weergegeven.

Tabel 4.1-2 Percentage overeenstemming en kappa voor de globale oordelen van kandidaten

Uitspraak	Percentage overeenstemming	Kappa	Standaardfout kappa
S-G1	91%	.24	0.13
S-G2	88%	.16	0.13
GV-G3	79%	.57	0.14
GV-G4	78%	.34	0.13
SGV-G5	93%	.79	0.13

De percentages overeenstemming zijn hoog en wijzen uit dat er voldoende overeenstemming is tussen de afzonderlijke globale oordelen van verschillende beoordelaars. De kappa's voor uitspraak 1 en 2 zijn betrekkelijk laag. Voor de overige kappas geldt dat ze in termen van Landis en Koch (in

Eggen en Sanders, 1993, p. 449) beschouwd kunnen worden als moderate tot substantial. Bij S-G1, S-G2, en GV-G4 voldeden vrijwel alle kandidaten aan het B1-criterium. De kappa is daardoor niet goed te interpreteren als maat voor overeenstemming. Opvallend is dat de meest algemene uitspraak (SGV-G5) tot een zeer bevredigende overeenstemming leidde. De correlatie tussen de somscores van de eerste en tweede beoordelaar bedraagt .75.

4.1.3 Relatie tussen de Raamwerkbeoordeling en de STEX-beoordeling

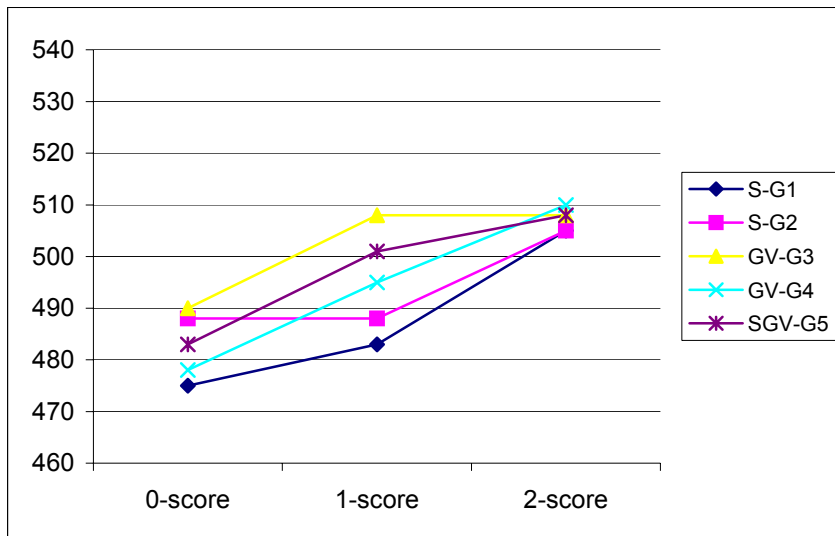
Elke kandidaat is globaal beoordeeld door twee onafhankelijke beoordelaars. De beoordelaars konden voor elke uitspraak aangeven of een kandidaat aan de eisen die in de beoordelingsuitspraak zijn verwoord voldeed of niet.

Voor elke globale uitspraak konden kandidaten de volgende scores krijgen:

- 0 beide beoordelaars zijn van mening dat de kandidaat niet voldoet aan de in de uitspraak gestelde eis
- 1 de ene beoordelaar vindt dat de kandidaat aan de eisen voldoet de andere niet
- 2 beide beoordelaars zijn van mening dat de kandidaat aan de eisen voldoet

Wanneer er een relatie bestaat tussen de score op het Staatsexamen en de Raamwerkbeoordeling kan verondersteld worden dat de groep die een 0 score op de raamwerkbeoordeling de laagste gemiddelde STEX-score heeft. De groep die door beide beoordelaars als B1 wordt gezien zal gemiddeld de hoogste score moeten hebben, terwijl de groep waarover de beoordelaars het niet eens waren (score 1) een score hebben die tussen die van beide andere groepen ligt. Figuur 4.1-1 geeft een weergave van deze relatie voor de 59 kandidaten.

Figuur 4.1-1 Relatie tussen globale oordelen B1 door 2 beoordelaars en de score op Staatsexamen I Spreken



De verwachte relatie wordt voor elk van de 5 globale Raamwerkuitspraken gevonden. Voor de uitspraak GV-G3 blijkt er nauwelijks verschil te zijn tussen de 1- en 2-scoregroep. De correlaties tussen de afzonderlijke raamwerkuitspraken en de STEX-score en tussen de totaalscore van de raamwerkbeoordeling zijn allemaal positief, zoals in tabel 4.1-3 is weergegeven.

Tabel 4.1-3 Correlaties tussen Raamwerkuitspraken B1 en de STEX-schaalscores Spreken I (n=59)

Uitspraak	Pearson productmoment correlatie
S-G1	0,38
S-G2	0,31
GV-G3	0,37
GV-G4	0,55
SGV-G5	0,49
Totaalscore	0,55

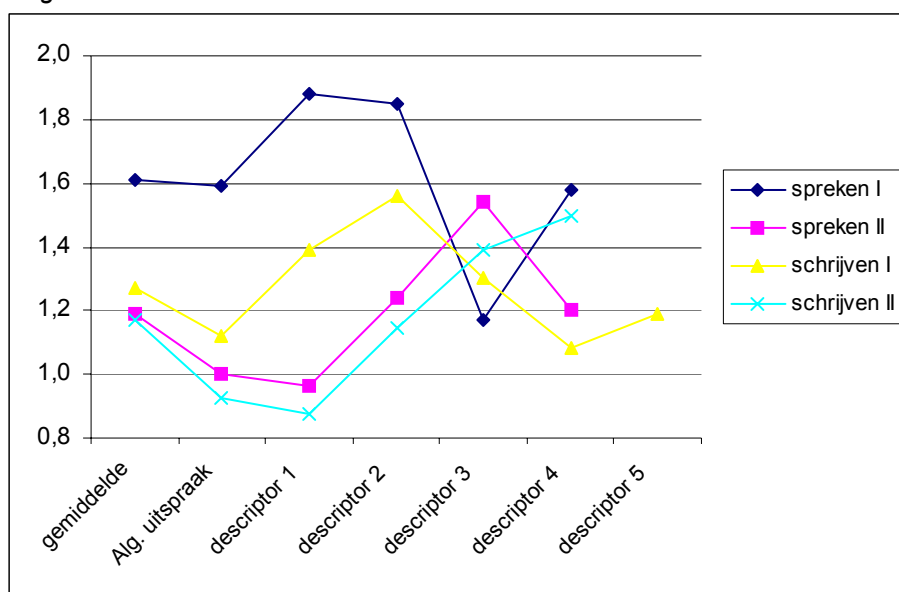
De correlaties tussen Raamwerkuitspraken en de STEX-scores zijn lager dan bij de overige vaardigheden die in dit onderzoek betrokken zijn. De mogelijke verklaring hiervoor is dat de spreiding van de raamwerkscores voor Spreken I aanzienlijk geringer is dan die in de steekproeven voor de overige drie examenonderdelen. Ook de spreiding van de STEX-scores is bij Spreken I iets geringer, terwijl de gemiddelde STEX-scores van de steekproeven niet noemenswaardig verschillen.

Tabel 4.1-2 Gemiddelde Raamwerk- en STEX-scores voor Spreken en Schrijven I en II

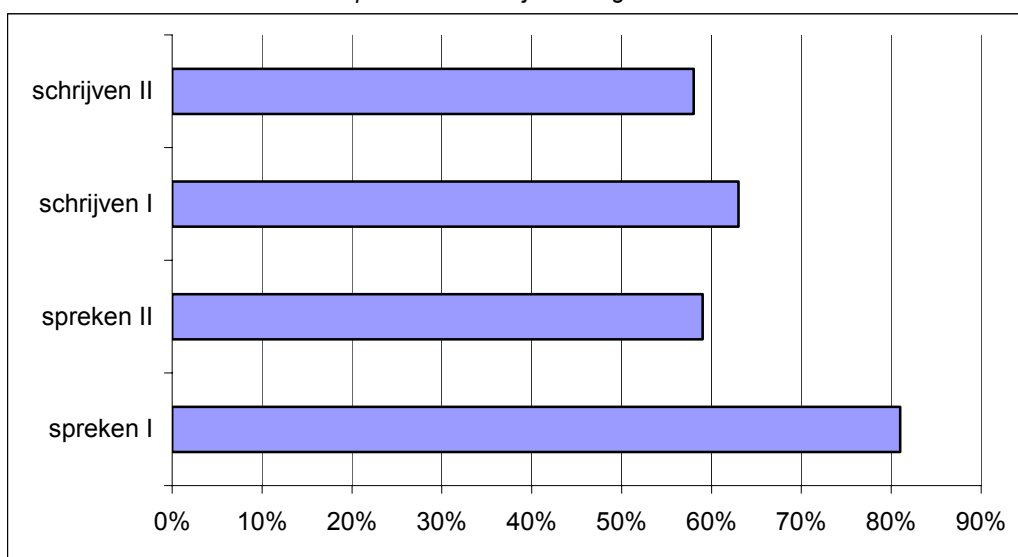
Gemiddelde somscore en standaarddeviaties Raamwerkbeoordeling							
Spreken I		Schrijven I		Spreken II		Schrijven II	
Gemiddelde (max. score 10)	SD	Gemiddelde (max. score 12)	SD	Gemiddelde (max score 10)	SD	Gemiddelde (max score 10)	SD
8,1	2,4	7,6	4,1	5,9	3,6	5,8	3,5
Gemiddelde STEX-score en standaarddeviaties							
Gemiddelde	SD	Gemiddelde	SD	Gemiddelde	SD	Gemiddelde	SD
503,6	20,0	505,6	28,2	506,2	26,0	507,5	28,2

Bovendien blijkt dat de Raamwerkscore voor Spreken I op de afzonderlijke uitspraken (figuur 4.1-2) èn gemeten als percentage van de maximaal te behalen somscore veel hoger (figuur 4.1-3) is dan op de overige onderdelen. Dientengevolge is de proportie van B1-kandidaten aanzienlijk hoger dan bij de overige examenonderdelen (zie ook paragraaf 4.2.1)

Figuur 4.1-2 Gemiddelde scores op de afzonderlijk Raamwerk-criteria voor Schrijven en Spreken Programma I en II



Figuur 4.1-3 Gemiddelde score van kandidaten op de som van de raamwerkscores als proportie van de maximaal te behalen raamwerkscore voor Spreken en Schrijven Programma I en II



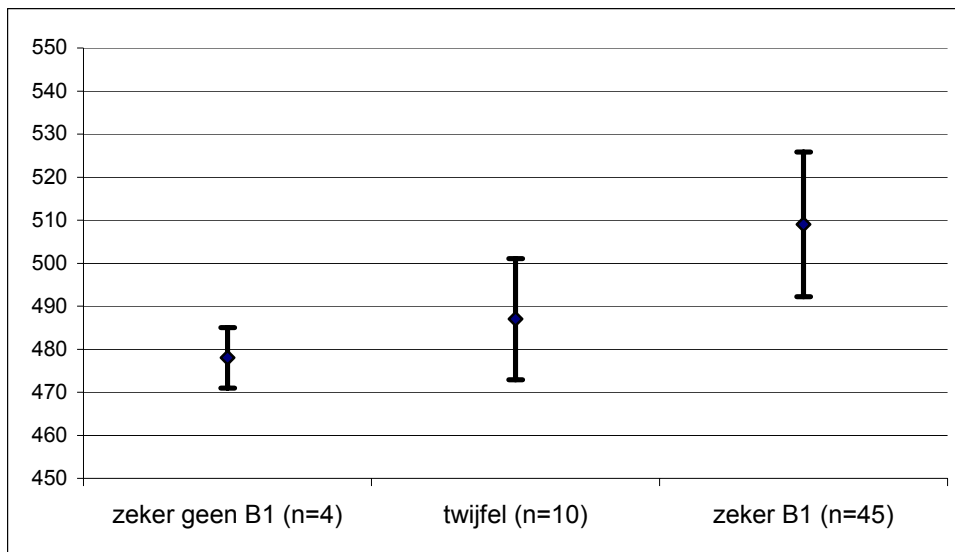
De lagere correlatie tussen de STEX-scores en de Raamwerkscores, betekenen niet dat het overgangsgebied op de STEX-schaal van cursisten die wel of niet B1-sprekers zijn onduidelijker is dan bij de andere vaardigheden (zie tabel 4.2-1).

Op grond van de gevonden samenhang kunnen de raamwerkuitspraken gesommeerd worden. De score op deze gesommeerde beoordeling loopt van 0 (beide beoordelaars geven voor elke uitspraak 0 punten) tot 10 (beide beoordelaars geven voor elke uitspraak 1 punt). In tabel 4.1-5 zijn de kandidaten in drie groepen verdeeld op basis van de scores op de raamwerkbeoordeling. Groep 1 bestaat uit de kandidaten die een lage score hebben (0-3) en dus zeker geen B1-kandidaten zijn. De tweede groep bestaat uit kandidaten die in minimaal 8 van de 10 beoordelingen als B1 zijn aangemerkt. Deze groep kan gezien worden als bestaande uit kandidaten die B1 met grote zekerheid beheersen. De tussenliggende groep is de groep die een score heeft van 4-7. Deze groep kan gezien worden als de groep waar beoordelaars het meest over twijfelden. Vervolgens zijn de STEX-scores van deze drie groepen berekend. Tabel 4.1-5 en figuur 4.1-4 geven de gemiddelde scores en de standaarddeviaties weer van de STEX-scores van deze drie groepen.

Tabel 4.1-5 Gemiddelde STEX-score en standaarddeviaties van kandidaten op grond van hun raamwerkbeoordeling.

Beheerst B1 volgens globale beoordeling	STEX-score		
	N	gemiddeld	SD
Zeker niet	4	478	7,0
Twijfel	10	487	14,1
Zeker wel	45	509	16,8
Alle kandidaten	59	504	20,0

Figuur 4.1-4 Gemiddelde STEX-score en standaarddeviaties van kandidaten op grond van hun raamwerkbeoordeling.



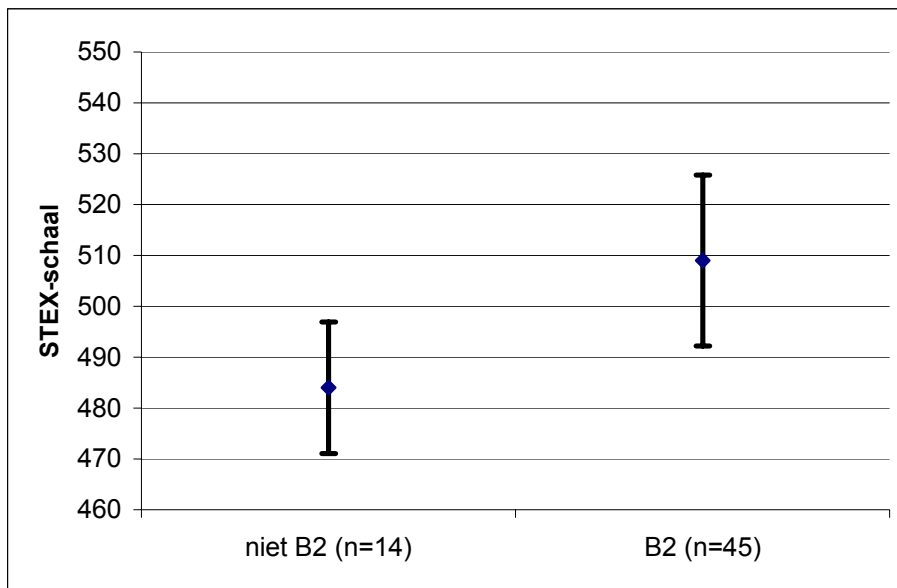
Het verband tussen de Raamwerkbeoordeling en de STEX-score is positief en in de verwachte richting.

Wanneer we de twijfelgroep en de niet-B1-groep samenvoegen kunnen we de kandidaten in twee groepen indelen. De eerste groep is de groep die geacht wordt niveau B1 zeker te beheersen. Deze groep is dezelfde als de 'zeker B1-groep' in tabel 4.1-5 en figuur 4.1-4 (een totaalscore van 8-10). De andere groep is een samenvoeging van kandidaten die volgens de beoordelaars B1 niet beheersen of waarover twijfel bestaat. De gemiddelden en standaarddeviaties van deze twee groepen worden gepresenteerd in tabel 4.1-6 en figuur 4.1-5.

Tabel 4.1-6 Gemiddelde STEX-score Spreken I en standaarddeviaties van de kandidaten die B1 wel of niet beheersen volgens 80% criterium

Beheerst B1 volgens globale beoordeling	STEX-score		
	N	gemiddeld	SD
Nee	14	484,5	16,8
Ja	45	508,8	12,9
Alle kandidaten	59	503,6	20,0

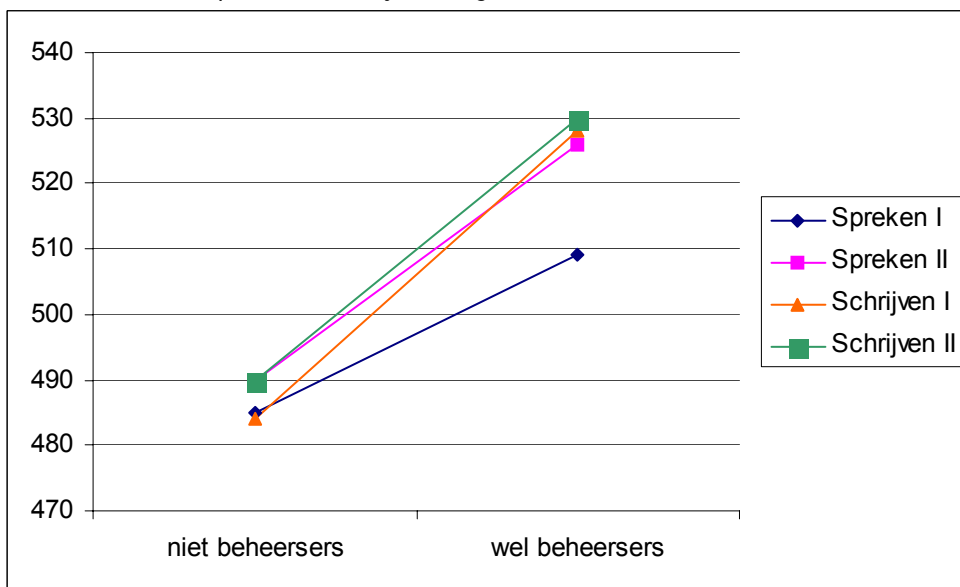
Figuur 4.1-5 Gemiddelde STEX-score Spreken I en standaarddeviaties van de kandidaten die B1 wel of niet beheersen volgens 80% criterium



Volgens deze tweedeling van de kandidaten blijkt de niet-B1-groep gemiddeld 24,3 punten lager te scoren op het Staatsexamen dan de groep die wel op B1 functioneert.

Ook hier wijkt Spreken I af van de overige onderdelen, zoals uit figuur 4.1.6 blijkt. De gemiddelde score van kandidaten die aan het Raamwerk criterium voldoen is aanzienlijk lager dan bij de overige onderdelen. Er is een minder duidelijk verschil tussen beide groepen wat betreft hun STEX-score dan bij de andere onderzochte onderdelen van het staatsexamen. Figuur 4.1-7 geeft een grafische weergave van deze relatie. Het percentage cursisten dat voldoet aan het 80%-raamwerk criterium en dus de desbetreffende niveaus beheerst is bij Spreken I aanzienlijk hoger dan bij de overige onderdelen.

Figuur 4.1-6 Gemiddelde STEX-score van de kandidaten die het raamwerkniveau wel of niet beheersen volgens 80% criterium voor Spreken en Schrijven Programma I en II

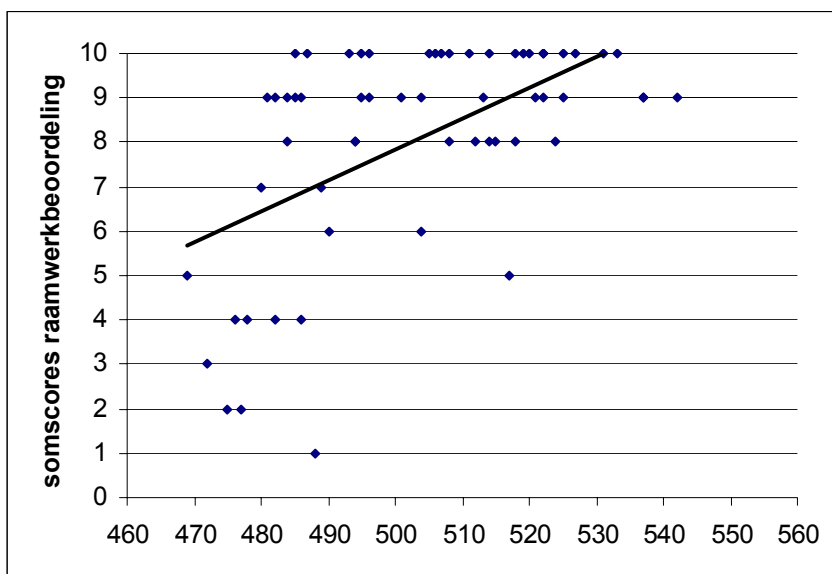


Tabel 4.1-7 Percentage kandidaten dat voldoet aan het Raamwerk criterium (B1 voor programma 1; B2 voor Programma II)

Percentage STEX-kandidaten dat voldoet aan het Raamwerk criterium			
Spreken I	Schrijven I	Spreken II	Schrijven II
77%	49%	46%	44%

De correlatie (Pearson product-moment) tussen de Raamwerkbeoordeling en de STEX-schaalscore bedraagt 0,55.

Figuur 4.1-7 Samenhang tussen de STEX-scores Spreken I en de globale raamwerkbeoordeling Spreken B1.



4.1.4 Samenvatting beoordelaarsovereenstemming en relatie tussen Raamwerkbeoordeling en STEX-score Spreken

- 1 In dit onderzoek zijn 60 kandidaten beoordeeld door twee onafhankelijke beoordelaars die gebruik maakten van een set globale beoordelingsuitspraken uit het Raamwerk NT2 (zie bijlage 1). Indien de somscores van de twee beoordelaars 3 of meer punten van elkaar verschilden, is een derde onafhankelijke beoordelaar ingeschakeld. Voor de verdere analyse zijn de data van de twee beoordelaars die het dichtst bij elkaar lagen, gebruikt. Eén kandidaat is uit het bestand verwijderd omdat bleek dat deze niet alle opgaven van het Staatsexamen had gemaakt.
- 2 De beoordelaarsovereenstemming in de gevolgde procedure voor de globale raamwerkbeoordeling Spreken B1 is voldoende voor verdere analyses met betrekking tot de cesuurkwestie.
- 3 De relatie tussen de STEX-scores en de Raamwerkscores is positief, maar zwakker dan bij de overige examenonderdelen. Steekproeftrekking en de gevolgde beoordelingsprocedure waren hetzelfde als bij de andere vaardigheden en kunnen niet als verklaring gezien worden voor de afwijkende resultaten bij Spreken I. De belangrijkste oorzaak is dat bij Spreken I aanzienlijk meer kandidaten voldoen aan het beheersingscriterium van de Raamwerkbeoordeling en dat ten gevolge daarvan de spreiding in criteriumscore lager is dan bij de andere oordelen, hetgeen de correlatie drukt.

4.2 Relatie tussen STEX-scores en Raamwerkbeoordeling: de cesuurkwestie

Uit paragraaf 4.1 blijkt dat er een positieve relatie bestaat tussen de STEX-scores Spreken I en de globale beoordelingen volgens het Raamwerk NT2. Verder blijkt dat de gemiddelde STEX-score van de B1 kandidaten hoger ligt dan die van de groep kandidaten die B1 niet beheersen. De volgende vraag die nu beantwoord moet worden, is of er aanleiding bestaat voor een mogelijke herziening van de cesuur van het Staatsexamen Spreken I om een goed onderscheid te maken tussen kandidaten die B1 wel of niet beheersen. In het onderzoeksvoorstel wordt dit als volgt verwoord:

“Resultaat Stap 3

Indicatie voor de gewenste cesuur voor de vaardigheden Spreken en Schrijven.”

Voor het beantwoorden van deze vraag wordt in deze paragraaf informatie gepresenteerd die aan deze indicatie kan bijdragen.

De positieve relatie tussen de globale raamwerkbeoordeling en de scores op het Staatsexamen Spreken I geeft aan dat er uitspraken kunnen worden gedaan ter ondersteuning van een eventuele verandering van de cesuurscore. Deze uitspraken zijn echter niet rechtstreeks af te leiden uit de tot nu toe gepresenteerde resultaten. Bij de vaststelling van een cesuur op grond van uitspraken over de beheersing van B1 spelen de volgende aspecten een rol:

- de definiëring van de beheersing van B1
- de consequenties die de hoogte van de cesuur heeft voor het ten onrechte laten zakken of het ten onrechte laten slagen van kandidaten. Afhankelijk van de vraag welke van deze twee consequenties men wil minimaliseren zal de cesuur hoger of lager komen te liggen.

In de volgende paragrafen dragen we de gegevens aan die kunnen bijdragen aan het vaststellen van een indicatie voor een cesuurwijziging.

4.2.1 Beheersing B1 volgens het criterium meerderheidsoordeel

Evenals in het cesuuronderzoek Staatsexamen Programma II beschouwen we een kandidaat als B1 indien deze in 80% van de beoordelingen als B1 beoordeeld wordt. Dit valt samen met een minimale score van 8 van de 10 te behalen punten. Dat betekent dat in minimaal 8 van de 10 beoordelingsuitspraken een beoordelaar uitspreekt dat iemand op B1 functioneert. De resultaten staan in tabel 4.2-1.

Tabel 4.2-1 Beoordeling kandidaten op Staatsexamen Spreken I en de raamwerkbeoordeling met als criterium meerderheidsoordeel van de beoordelaars (80%).

STEX-score	Totaalscore raamwerk	Beheerst B1 met meerderheidsoordeel
542	9	B1
537	9	B1
537	9	B1
533	10	B1
531	10	B1
527	10	B1
525	9	B1
525	10	B1
524	8	B1
522	9	B1
522	10	B1
522	10	B1
521	9	B1
520	10	B1
519	10	B1
518	8	B1
518	10	B1
517	5	
515	8	B1
514	8	B1
514	10	B1
513	9	B1
512	8	B1
511	10	B1
508	8	B1
508	10	B1
507	10	B1
506	10	B1
505	10	B1
504	6	
504	9	B1
501	9	B1
496	9	B1
496	10	B1
495	9	B1
495	10	B1
494	8	B1
494	8	B1
493	10	B1
490	6	
489	7	
488	1	
487	10	B1
486	4	
486	9	B1
485	9	B1
485	10	B1
484	8	B1
484	9	B1
482	4	
482	9	B1
481	9	B1
480	7	
478	4	
477	2	
476	4	
475	2	
472	3	
469	5	

	Geslaagd voor STEX / beheerst niveau B1
	Gezakt voor STEX / beheerst niveau B1 niet

45 kandidaten worden geacht B1 te beheersen; bij slechts 14 kandidaten is dit niet het geval. De proportie B1 kandidaten is aanzienlijk hoger dan bij Schrijven I en dan de proporties B2-kandidaten bij Schrijven II en Spreken II (zie ook tabel 4.1-7) Van de 27 kandidaten die gezakt waren voor het Staatsexamen Spreken I beheersen er 15 B1. In het interval tussen 493 en 481 ligt het overgangsgebied van al of niet B1-beheersing. Van de 14 kandidaten in dit scorebereik worden er 9 beoordeeld als B1 en 5 als niet-B1. Een eventuele nieuwe cesuur zal waarschijnlijk in dit interval liggen. Bij Spreken I is een duidelijk verschil te constateren tussen de beheersingsgrens via de Raamwerkbeoordeling en de cesuur op het Staatsexamen. Anders dan bij Programma II en Schrijven I het geval was, wordt een tamelijke groot aantal kandidaten met een score onder de STEX-cesuur wel als B1 beoordeeld. In bijlage 5 wordt een korte impressie gegeven van het werk van deze kandidaten op grond van een extra inspectie van hun examenproducten. Deze nadere inspectie leidde in de meeste gevallen tot een bevestiging van de eerdere beoordeling.

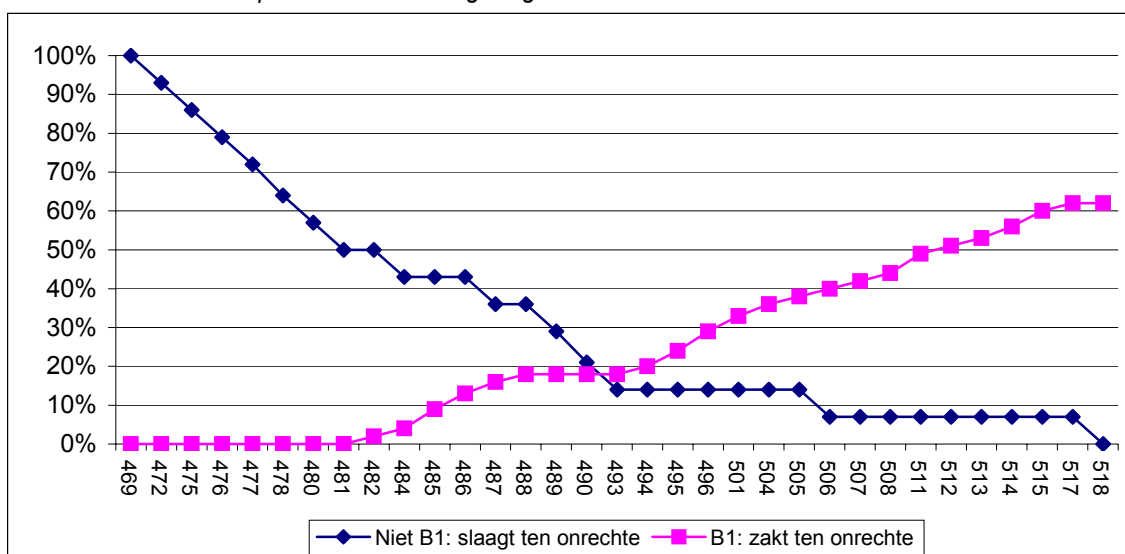
Op grond van tabel 4.2-2 kan dus verwacht worden dat de grens tussen B1 en niet-B1-kandidaten tussen 493 en 481 zal komen te liggen. Voor het vaststellen van de cesuur is echter ook informatie nodig over de proportie kandidaten die bij een gegeven cesuurscore onterecht zakt of onterecht slaagt. Deze twee consequenties moeten tegen elkaar worden afgewogen. Tabel 4.2-3 geeft een overzicht van de gevolgen van de beslissingen voor de cesuurscores tussen 469 en 518.

Tabel 4.2-3 Percentages gezakte en geslaagde kandidaten die B1 al of niet beheersen bij verschillende cesuren (laagste voldoende score) van het Staatsexamen Spreken I. Beheersing B1 gedefinieerd als 80% van maximale score

Cesuur op STEX	Beheerst B1 niet (n=14)		Beheerst B1 (n=45)	
	Zakt terecht	Slaagt ten onrechte	Zakt ten onrechte	Slaagt terecht
469	0%	100%	0%	100%
472	7%	93%	0%	100%
475	14%	86%	0%	100%
476	21%	79%	0%	100%
477	28%	72%	0%	100%
478	36%	64%	0%	100%
480	43%	57%	0%	100%
481	50%	50%	0%	100%
482	50%	50%	2%	98%
484	57%	43%	4%	96%
485	57%	43%	9%	91%
486	57%	43%	13%	87%
487	64%	36%	16%	84%
488	64%	36%	18%	82%
489	71%	29%	18%	82%
490	79%	21%	18%	82%
493	86%	14%	18%	82%
494	86%	14%	20%	80%
495	86%	14%	24%	76%
496	86%	14%	29%	71%
501	86%	14%	33%	67%
504	86%	14%	36%	64%
505	86%	14%	38%	62%
506	93%	7%	40%	60%
507	93%	7%	42%	58%
508	93%	7%	44%	56%
511	93%	7%	49%	51%
512	93%	7%	51%	49%
513	93%	7%	53%	47%
514	93%	7%	56%	44%
515	93%	7%	60%	40%
517	93%	7%	62%	38%
518	100%	0%	62%	38%

In figuur 4.2-1 wordt deze tabel grafisch weergegeven.

Figuur 4.2-1 Percentages gezakte en geslaagde kandidaten die B1 al of niet beheersen bij verschillende cesuren van het Staatsexamen Spreken I. Beheersing B1 gedefinieerd als 80% van maximale score



Figuur 4.2-1 maakt duidelijk dat bij de uiteindelijke vaststelling van de cesuurscore rekening gehouden dient te worden met twee soorten foute beslissingen:

1. Fout negatief. De fout dat men ten onrechte kandidaten laat zakken die B1 wel beheersen.

2. Fout positief. De fout dat men ten onrechte kandidaten laat slagen die B1 niet beheersen.

Zo leidt een cesuurscore van 481 en lager ertoe dat de kans dat een B1-kandidaat ten onrechte zakt nihil is. Daar staat tegenover dat de kans dat kandidaten die B1 nog niet beheersen bij deze scores slagen bij deze cesuurscore zeer hoog zou zijn: 50% of meer. Bij een cesuurscore van 518 maakt men de kans op het laten slagen van niet B1-kandidaten nihil, maar verhoogt men de kans op het onterecht zakken van B1-kandidaten tot 60%. Welke cesuurscore men op basis van deze cijfers kiest, hangt sterk af van de functie van een examen en de ernst van de consequenties van een foute beslissing. Zou figuur 4.2-1 betrekking hebben op bijvoorbeeld een vliegvaardigheidsexamen voor burgerluchtvaartpiloten, dan zal men kiezen voor een cesuur van 518 of hoger. De consequenties van het laten slagen van een incompetent piloot kunnen immers zeer zwaarwegend zijn. Met andere woorden de maatschappelijke consequenties leggen meer gewicht in de schaal dan de persoonlijke consequenties. Het Staatsexamen NT2 geeft recht op een startkwalificatie. Het ligt daarom voor de hand dat het ten onrechte zakken zwaardere consequenties heeft dan het ten onrechte slagen. De persoonlijke consequenties mogen zwaarder wegen. Er zal dus binnen het interval gekozen moeten worden voor een cesuur die het ten onrechte zakken minimaliseert met daarbij een zo klein mogelijke kans op het ten onrechte slagen. We kunnen het optimum van deze keuze vinden door de kans op foute negatieven en foute positieven uit tabel 4.2-3 bij elkaar op te tellen. Een perfect, maar onbestaanbaar examen heeft een grensscore waarbij deze foute keuzes beide 0 en dus hun somscores ook 0 zijn. In de praktijk zal men met een minder scherpe grens tevreden moeten zijn. Tabel 4.2-4 geeft de somwaarden weer onder twee condities:

1 Conditie 1

Onterecht zakken en slagen word even belangrijk gevonden (ongewogen conditie)

Voorbeeld: bij score 501 slaagt 14% ten onrechte en zakt 33% ten onrechte. De optimumscore wordt: $14 + 33 = 47$. Hoe kleiner dit getal hoe minder kans op een foute beslissing.

2 Conditie 2

Men wil voorkomen dat te veel B1-kandidaten ten onrechte zakken. Daarom geeft men ten onrechte zakken een dubbel gewicht.

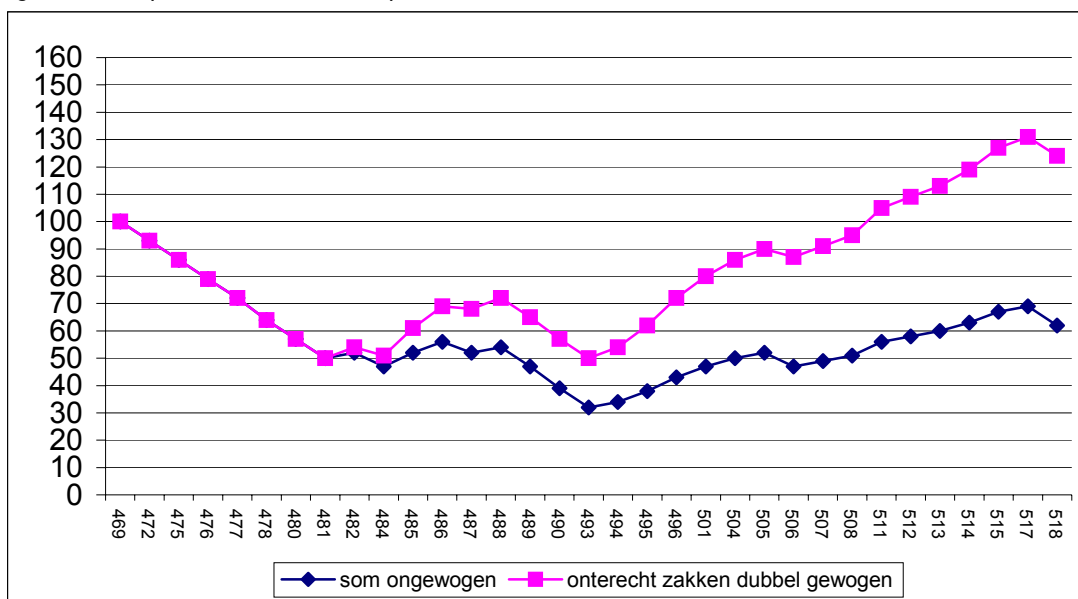
Voorbeeld: bij score 501 slaagt 14% ten onrechte en zakt 33% ten onrechte. De optimumscore als we het onterecht zakken dubbel wegen, wordt nu: $14 + 66 = 80$. Ook hier geldt dat de laagste optimumscore de geringste kans geeft op een foute beslissing.

Figuur 4.2-2 is een grafische weergave van deze tabel.

Tabel 4.2-4 Optimum cesuurscore Spreken I onder 2 condities

	Conditie 1 Ongewogen	Conditie 2 Onterecht zakken telt dubbel
475	86	86
476	79	79
477	72	72
478	64	64
480	57	57
481	50	50
482	52	54
484	47	51
485	52	61
486	56	69
487	52	68
488	54	72
489	47	65
490	39	57
493	32	50
494	34	54
495	38	62
496	43	72
501	47	80
504	50	86
505	52	90
506	47	87
507	49	91
508	51	95
511	56	105
512	58	109
513	60	113
514	63	119
515	67	127
517	69	131
518	62	124

Figuur 4.2-2 Optimum cesuurscore Spreken I onder 2 condities



De optimale cesuurscore in de ongewogen variant ligt op score 493. Wanneer men meer belang hecht aan het voorkomen van het ten onrechte zakken van B1 kandidaten blijft de optimale cesuur op 493, maar heeft de score 481 een even laag optimum. In dit interval komt de cesuur te liggen, wil men een zo goed mogelijk onderscheid maken tussen kandidaten die wel of geen B1 beheersen. Er is echter een groot verschil tussen de bijdrage van ten onrecht slagen en zakken in beide optimumscores, zoals valt af te lezen uit tabel 4.2-3. Bij 481 zakt niemand ten onrechte, maar slaagt tegelijkertijd 50 % ten onrechte. Bij 493 neemt het percentage ten onrechte geslaagde kandidaten af tot 14% en zakt 18% ten onrechte. Voor een uiteindelijke beslissing dient men af te wegen welke verhouding men in dit interval de meest optimale vindt binnen de gevonden schattingen

Het feit dat er zo weinig kandidaten niet als B1 worden beoordeeld (14), maakt dat de procentuele sprongen die gemaakt worden bij de schatting 'onterecht geslaagd' tamelijk groot zijn. Elke kandidaat telt voor 7%. Beoordeling van meer kandidaten in dit interval zal een fijnmaziger schatting kunnen bewerkstelligen. Bij de overige onderzochte examenonderdelen ligt de verhouding tussen wel of niet B1/B2 in de steekproef rond de 50%.

Tabel 4.2-3 Huidige en voorgestelde cesuurscores en de hierbij horende foute beslissingen

Cesuurscore	Niet B1: slaagt ten onrechte	B1: zakt ten onrechte	Optimumscore
481	50%	0%	50
482	50%	2%	54
484	43%	4%	51
485	43%	9%	61
486	43%	13%	69
487	36%	16%	68
488	36%	18%	72
489	29%	18%	65
490	21%	18%	57
493	14%	18%	50

Wanneer de nieuwe cesuur in dit interval gelegd zou worden neemt vanzelfsprekend het percentage kandidaten dat zakt voor het onderdeel Spreken I af. Tabel 4.2-4 geeft het aantal gezakten weer op het Staatsexamen Spreken I van januari 2005. Ook deze gegevens kunnen in de overwegingen worden meegenomen.

Tabel 4.2.4 Percentage gezakte kandidaten in Staatsexamen Spreken I, januari 2005 bij verschillende cesuurscores

Cesuurscore	Percentage gezakte kandidaten
481	18.27
482	19.86
484	20.94
485	21.88
486	23.47
487	24.55
488	25.88
489	27.22
490	28.52
491	30.47
492	31.55
493	31.62
.....
501*	43.03

* De score 500 kwam in het examen januari 2005 niet voor

4.2.2 Afwijkende resultaten Spreken I

Het opvallende verschil tussen enerzijds Spreken I en anderzijds de overige onderzochte onderdelen is dat het overgangsgebied tussen kandidaten die B1 niet en kandidaten die B1 wel beheersen bij Spreken I duidelijk onder de huidige cesuur ligt, terwijl bij de andere vaardigheden het overgangsgebied rondom of iets boven de cesuur ligt. Dit verschil kan een terecht gevolg zijn van de prestaties van de kandidaten, maar ook een ongewenst effect van de toegepaste methodiek. Als dit laatste het geval zou zijn, zouden er geen beslissingen genomen kunnen worden. Daarom worden de onderzoeksopzet en –uitvoering van Spreken I nader onder de loep genomen en vergeleken met de andere vaardigheden.

Steekproeftrekking

Er was geen verschil tussen de wijze van steekproeftrekking bij de verschillende vaardigheden. In alle gevallen is een steekproef getrokken uit de gehele populatie, met uitsluiting van het laagste en hoogste percentiel. De gegevens van de uiteindelijke steekproeven worden weergegeven in tabel 4.2-5. De verschillen tussen Spreken I en de overige onderdelen zijn niet van dien aard dat deze als oorzaak van het verschil beschouwd kunnen worden.

Tabel 4.2-5 Gemiddelden en standaarddeviaties van de steekproeven voor de vier examenonderdelen

	Gemiddelde	Standaarddeviatie
Spreken I	503,6	20,0
Spreken II	506,2	26,0
Schrijven I	505,6	28,2
Schrijven II	507,5	28,2

Beoordelingssystematiek

De gevolgde beoordelingssystematiek is bij alle onderdelen hetzelfde geweest (zie bijlagen 1 - 4). De examenproducten van 60 kandidaten zijn beoordeeld door twee onafhankelijke beoordelaars met een beoordelingslijst die de algemene descriptoren uit het Raamwerk NT2 bevatten voor het bedoelde niveau en de bedoelde vaardigheid. Wanneer we deze descriptoren beschouwen als de juiste omschrijving van de diverse niveaus – en dat moeten we per definitie, omdat het de beschrijvingen zijn uit het Raamwerk en daarmee het CEF – dan is de gekozen operationalisatie voor de raamwerkniveaus voor alle examenonderdelen dezelfde.

Nadat bleek dat de resultaten wezen op het feit dat een B1-cesuur Spreken I onder de huidige cesuur zou komen te liggen, zijn de kandidaten onder de STEX-cesuur nog een extra bekeken. Dit als extra controle. De resultaten van deze controle worden gepresenteerd in bijlage 5. Zij leidden niet tot een heroverweging van de beoordelingen met de raamwerkcriteria.

Beoordelaarsovereenstemming van de raamwerkbeoordeling

De beoordelaarsovereenstemming bij Spreken I is niet afwijkend van die van de andere onderdelen, zoals tabel 4.2-6 laat zien. De oorzaak van de verschillen kan dus niet hier gezocht worden.

Tabel 4.2-6 Beoordelaarsovereenstemming in de Raamwerkbeoordeling voor de vier examenonderdelen

	Proportie overeenstemming	Kappa gemiddeld over de descriptoren
Spreken I	86%	0,42
Spreken II	79%	0,54
Schrijven I	77%	0,48
Schrijven II	75%	0,44

Analysemethoden

De analysemethode en de gevolgde beslisregels zijn bij alle examenonderdelen op dezelfde wijze toegepast.

Conclusie

Uit bovenstaande blijkt dat de gevonden afwijkende resultaten bij Spreken I niet het gevolg zijn van onderzoeksopzet en uitvoering, maar voortkomen uit reële verschillen in de data. Deze wijzen erop dat een aanzienlijk groter percentage van de kandidaten Spreken I het desbetreffende CEF-niveau beheersen en dat de cesuur van Spreken I in relatie tot B1-beheersing volgens het Raamwerk NT2 aan de hoge kant is. Tabel 4.2-5 geeft hiervan een overzicht.

Tabel 4.2-7 Percentage kandidaten voor de vier examenonderdelen dat het betreffende CEF-niveau beheerst en de gemiddelde score en standaarddeviatie op de Raamwerk-somscores

	Percentage kandidaten met niveaubeheersing	Gemiddelde Raamwerk-somscore steekproeven als percentage van de maximumscore	standaarddeviaties raamwerk-somscores
Spreken I	76%	81%	2,4
Spreken II	45%	59%	3,6
Schrijven I	42%	63%	4,1
Schrijven II	40%	58%	3,5

De resultaten geven duidelijk de afwijkende positie aan van Spreken I. Een grotere proportie kandidaten beheerst het gewenste niveau en de spreiding in de raamwerkscore is aanzienlijk lager dan bij de andere onderdelen. Dit verklaart voor een groot deel de lagere correlatie tussen Spreken I en raamwerkscore.

Tenslotte moge worden opgemerkt dat het overgangspatroon van niet- B1- naar wel-B1-kandidaten (zie tabel 4.2.2) niet afwijkt van de overige onderdelen. Het enige verschil is dat deze overgang zich bij Spreken I lager op de STEX-schaal bevindt dan bij de overige examenonderdelen.

4.2.3 Samenvatting en aanbevelingen cesuur Spreken I

Volgens de behoeftepeiling zou het Staatsexamen Spreken I een vaardigheid moeten waarborgen die op niveau B1 van het Raamwerk NT2 ligt (Kuijper e.a., 2005).

In het huidige onderzoek wordt onderzocht of dit consequenties heeft voor de cesuur van Spreken Programma I.

1. In dit onderzoek is dezelfde methodiek toegepast als in het cesuuronderzoek Spreken en Schrijven Programma II (Kuijper e.a., 2004) en Schrijven Programma I.
2. Uit het beoordelingsonderzoek blijkt dat er een positieve relatie bestaat tussen de scores die kandidaten halen op het Staatsexamen Spreken I en de vraag of de respons van een kandidaat beoordeeld wordt als B1. In zijn algemeenheid geldt dat naarmate de STEX-score hoger wordt de kans dat een kandidaat als B1 wordt beoordeeld groter wordt.

3. De gevonden beoordelaarsovereenstemming bij de globale raamwerkbeoordeling Spreken is voldoende om er uitspraken over een cesuurscore op te baseren.
4. Een definitieve beslissing over de cesuur kan niet alleen op statistische gronden worden genomen. Maatschappelijke overwegingen zijn het eerste uitgangspunt, waarna statistische procedures kunnen worden gebruikt om deze uitgangspunten zo goed mogelijk in een cesuurvaststelling om te zetten.
5. Wanneer de cesuur van het Staatsexamen NT2 Spreken Programma I een onderscheid moet maken tussen kandidaten die een spreekvaardigheid hebben die vergelijkbaar is met niveau B1, dan ligt de optimale cesuur in het interval 481-493. Bij de laagste score zou niemand ten onrechte zakken, maar is het percentage dat ten onrechte slaagt aanzienlijk: 50 %. Bij een score van 493 zijn deze percentages meer in evenwicht (14 % slaagt ten onrechte, 18% zakt ten onrechte. Bij het vaststellen van dit cesuurvoorstel moet men in ogenschouw nemen hoe men - in relatie tot de functie van het examen - het ten onrechte slagen of zakken tegen elkaar afweegt. Voor het Staatsexamen NT2 geldt dat het een startkwalificatie verschaft. Dit zou kunnen betekenen dat men het ten onrechte zakken van een kandidaat een slechtere optie vindt dan het ten onrechte slagen. Een cesuurscore van 481 leidt tot een percentage van 18,27 gezakte kandidaten in de examenpopulatie van januari 2005. Een cesuurscore met 493 als eerste voldoende score leidt tot een zakpercentage van 31,62. Bij de huidige cesuur was het zakpercentage 43,03.
6. De resultaten bij Spreken I wijken af van die bij de andere oordelen. Als enige onderdeel wijst het onderzoek op een verlaging van de cesuur. Het verschil tussen Spreken I en de overige onderdelen is niet terug te voeren op verschillen in onderzoeksopzet en –uitvoering.

5 De relatie tussen de score op het Staatsexamen Schrijven I en de Raamwerkbeoordeling

5.1 Globale beoordeling Schrijven

5.1.1 Opzet beoordeling

Bij de globale beoordeling is elke kandidaat beoordeeld door twee onafhankelijke beoordelaars. Zij moesten zes uitspraken over de schrijfvaardigheid van de cursisten beantwoorden met ja of nee. Hun oordelen zijn gebaseerd op de producten van het STEX-I schrijfexamen van de kandidaten. De globale raamwerkbeoordelingslijst is opgenomen in bijlage 2.

De eerste vijf uitspraken zijn letterlijk ontleend aan de omschrijvingen van het beheersingsniveau B1 Schrijven van het Raamwerk NT2 (pagina 87).

De laatste uitspraak is de meest algemene beoordeling: een algemene uitspraak in hoeverre een kandidaat volgens de beoordelaar voldoet aan de eisen voor Schrijven B1.

Met deze globale oordelen werden 60 kandidaten beoordeeld door twee onafhankelijke beoordelaars. Eén kandidaat is uit het bestand verwijderd omdat bleek dat deze niet alle opgaven van het Staatsexamen had gemaakt.

Tabel 5.1-1 Globale raamwerkbeoordelingsuitspraken Schrijven B1

	Uitspraken Raamwerk NT2	Ja	Nee
SC-G1	Kan een eenvoudige, samenhangende tekst schrijven over onderwerpen die vertrouwd of van persoonlijk belang zijn.		
SC-G2	Kan persoonlijke brieven schrijven waarin ervaringen en indrukken beschreven worden.		
SC-G3	Kan notities/berichten schrijven waarin eenvoudige, relevante informatie begrijpelijk overkomt.		
SC-G4	Kan heel korte rapporten schrijven volgens een vastgestelde standaard waarin routinematige feitelijke informatie wordt doorgegeven.		
SC-G5	Kan een korte, eenvoudige zakelijke tekst schrijven.		
SC-G6	Beheerst deze kandidaat over het algemeen niveau B1?		

De somscore van elke beoordelaar kon variëren van 0 tot 6. Indien er tussen de twee beoordelaars van een kandidaat een verschil van 4 of meer scorepunten werd geconstateerd, werd er een derde beoordeling uitgevoerd. Vervolgens werden de twee scores die het dichtst bij elkaar lagen in de verdere analyses meegenomen. Dit was bij zes kandidaten het geval.

5.1.2 Beoordelaarsovereenstemming

In tabel 5.1-2 wordt voor elk van deze globale beoordelingscriteria het percentage overeenstemming en de kappa weergegeven.

Tabel 5.1-2 Percentage overeenstemming en kappa voor de globale oordelen van kandidaten Schrijven I

Uitspraak	Percentage overeenstemming	Kappa	Standaard fout kappa
SC-G1	78%	0,47	0,13
SC-G2	79%	0,41	0,15
SC-G3	72%	0,39	0,13
SC-G4	69%	0,38	0,12
SC-G5	71%	0,39	0,13
SC-G6	93%	0,86	0,13

De percentages overeenstemming en de kappa's wijzen uit dat er voldoende overeenstemming is tussen de afzonderlijke globale oordelen van verschillende beoordelaars. In termen van Landis en Koch (in Eggen en Sanders, 1993, p. 449) kunnen de gevonden kappa's beschouwd worden als fair

tot substantial. Opvallend is dat de meest algemene uitspraak (SC-G6) tot een zeer hoge overeenstemming leidde. De correlatie tussen de somscores van eerste en tweede beoordelaars bedraagt .82. De beoordelaarsovereenstemming was voldoende hoog om de relatie tussen de raamwerkbeoordeling de Staatsexamenscore nader te onderzoeken.

5.1.3 Relatie tussen de Raamwerkbeoordeling en de STEX-beoordeling

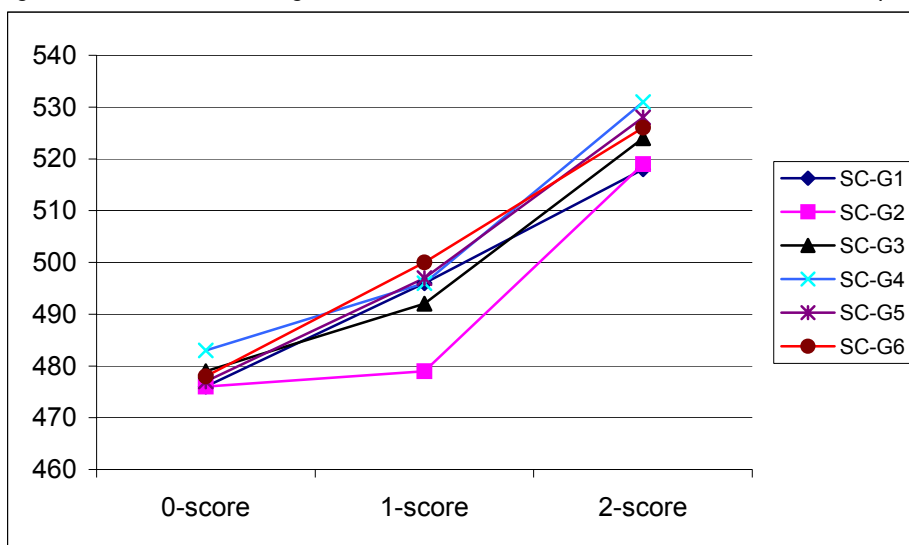
Elke kandidaat is globaal beoordeeld door twee onafhankelijke beoordelaars. De beoordelaars konden voor elke uitspraak aangeven of een kandidaat aan de eisen die in de beoordelingsuitspraak zijn verwoord voldeed of niet.

Voor elke globale uitspraak konden kandidaten de volgende scores krijgen:

- 0 beide beoordelaars zijn van mening dat de kandidaat niet voldoet aan de in de uitspraak gestelde eis
- 1 de ene beoordelaar vindt dat de kandidaat aan de eisen voldoet de andere niet
- 2 beide beoordelaars zijn van mening dat de kandidaat aan de eisen voldoet

Wanneer er een relatie bestaat tussen de score op het Staatsexamen en de Raamwerkbeoordeling kan verondersteld worden dat de groep met een 0-score op een uitspraak van de Raamwerkbeoordeling de laagste gemiddelde STEX-score heeft. De groep die door beide beoordelaars als B1 wordt gezien zal gemiddeld de hoogste score moeten hebben, terwijl de groep waarover de beoordelaars het niet eens waren (score 1) een score hebben die tussen die van beide andere groepen ligt. Figuur 5.1-1 geeft een weergave van deze relatie voor de 59 kandidaten

Figuur 5.1-1 Relatie tussen globale oordelen B1 door 2 beoordelaars en de score op Staatsexamen I Schrijven



De verwachte relatie wordt voor alle globale Raamwerkuitspraken gevonden. De correlaties tussen de afzonderlijke raamwerkuitspraken en de STEX-score en tussen de totaalscore van de raamwerkbeoordeling zijn allemaal positief, zoals in tabel 5.1-3 is weergegeven.

Tabel 5.1-3 Correlaties tussen Raamwerkuitspraken en de STEX-schaalscores Schrijven I (n=59)

Uitspraak	Pearson productmoment correlatie
S-G1	0,60
S-G2	0,64
S-G3	0,67
S-G4	0,72
S-G5	0,74
S-G6	0,80
Somscore	0,84

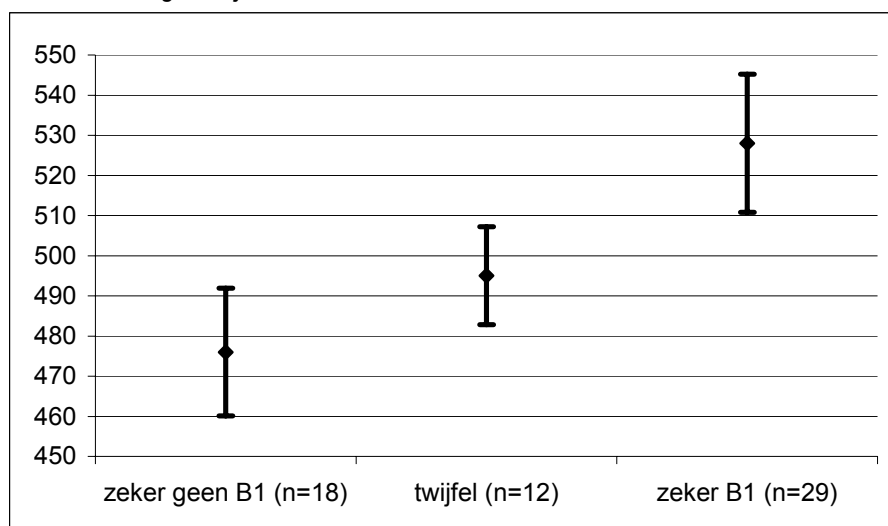
De correlaties tussen raamwerkuitspraken en de STEX-scores zijn zeer bevredigend.

Op grond van de gevonden samenhang kunnen de uitspraken gesommeerd worden. De score op deze gesommeerde beoordeling loopt van 0 (beide beoordelaars geven voor elke uitspraak 0 punten) tot 12 (beide beoordelaars geven voor elke uitspraak 1 punt). In tabel 5.1-4 zijn de kandidaten in drie groepen verdeeld op basis van de scores op de raamwerkbeoordeling. Groep 1 bestaat uit de kandidaten die een lage score hebben (0-4) en dus zeker geen B1 kandidaten zijn. De tweede groep bestaat uit kandidaten die in minimaal 10 van de 12 beoordelingen als B1 zijn aangemerkt. Deze groep kan gezien worden als bestaande uit kandidaten die B1 met grote zekerheid beheersen. De tussenliggende groep is de groep die een score heeft van 5-9. Deze groep kan gezien worden als de groep waar beoordelaars het meest over twijfelden. Vervolgens zijn de STEX-scores van deze drie groepen berekend. Tabel 5.1-4 en figuur 5.1-2 geven de gemiddelde scores en de standaarddeviaties weer van de STEX-scores van deze drie groepen.

Tabel 5.1-4 Gemiddelde STEX-score en standaarddeviaties van kandidaten op grond van hun Raamwerkbeoordeling Schrijven.

Beheerst B1 volgens globale beoordeling	STEX-score		
	N	gemiddeld	SD
Zeker niet	18	476,0	15,9
Twijfel	12	495,3	12,2
Zeker wel	29	528,2	17,2
Alle kandidaten	59	505,6	28,2

Figuur 5.1-2 Gemiddelde STEX-score Schrijven I en standaarddeviaties van kandidaten op grond van hun Raamwerkbeoordeling Schrijven.



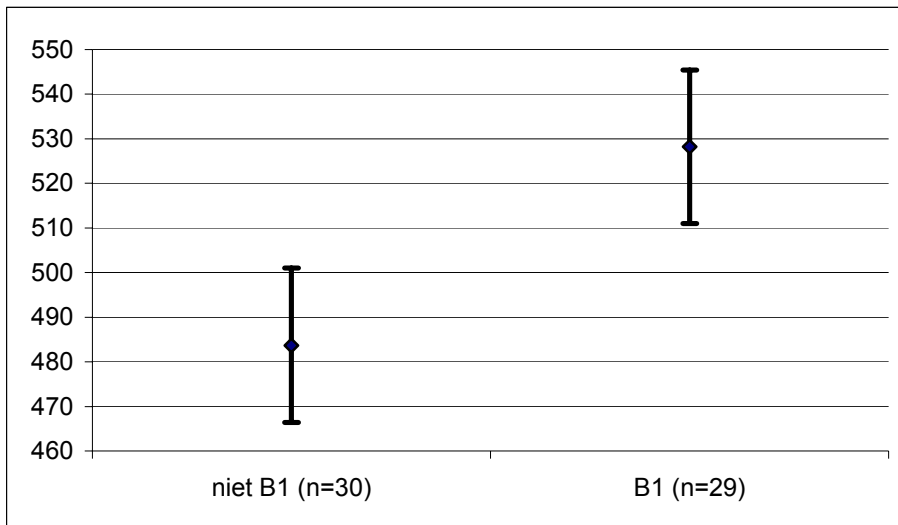
Het verband tussen de Raamwerkbeoordeling en de STEX-score is duidelijk en in de verwachte richting.

Wanneer we de twijfelgroep en de niet-B1-groep samenvoegen kunnen we de kandidaten in twee groepen indelen. De eerste groep is de groep die geacht wordt niveau B1 zeker te beheersen. Deze groep is dezelfde als de 'zeker B1-groep' in tabel 5.1-4 en figuur 5.1-2 (een totaalscore van 10-12). De andere groep is een samenvoeging van kandidaten die volgens de beoordelaars B1 niet beheersen of waarover twijfel bestaat. De gemiddelden en standaarddeviaties van deze twee groepen worden gepresenteerd in tabel 5.1-5 en figuur 5.1-3.

Tabel 5.1-5 Gemiddelde STEX-score Schrijven I en standaarddeviaties van de kandidaten die B1 wel of niet beheersen volgens 80% criterium

Beheerst B1 volgens globale beoordeling	STEX-score		
	N	gemiddeld	SD
Nee	30	483,7	17,3
Ja	29	528,2	17,2
Alle kandidaten	59	505,6	28,2

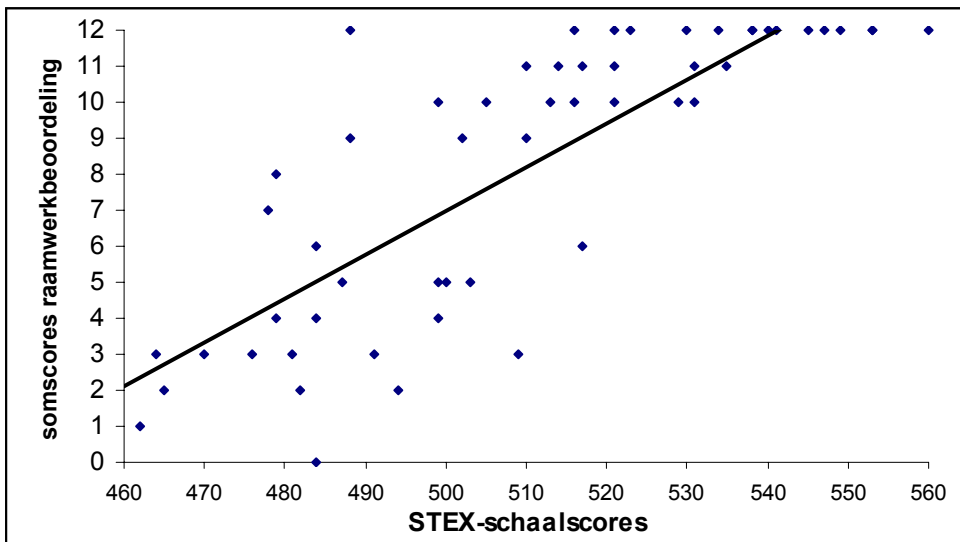
Figuur 5.1-3 Gemiddelde STEX-score Schrijven I en standaarddeviaties van de kandidaten die B1 wel of niet beheersen volgens 80% criterium



Volgens deze tweedeling van de kandidaten blijkt de niet-B1-groep gemiddeld 44,5 punten lager te scoren op het Staatsexamen dan de groep die wel op B1 functioneert. Er is dus een duidelijk verschil tussen beide groepen wat betreft hun STEX-score.

De correlatie (Pearson product-moment correlatie) tussen de somscore van de raamwerkbeoordeling en de STEX-score bedraagt 0,84. Figuur 5.1-4 laat geeft een grafische weergave van deze relatie.

Figuur 5.1-4 Samenhang tussen de STEX-scores Schrijven I en de globale raamwerkbeoordeling Schrijven B1.



5.1.4 Samenvatting beoordelaarsovereenstemming en relatie tussen raamwerkbeoordeling en Staatsexamen-score Schrijven

- 1 In dit onderzoek zijn 60 kandidaten beoordeeld door twee onafhankelijke beoordelaars die gebruik maakten van een set globale beoordelingsuitspraken uit het Raamwerk NT2 (zie bijlage 2). Indien de somscores van de twee beoordelaars 3 of meer punten van elkaar verschilden, is een derde onafhankelijke beoordelaar ingeschakeld. Voor de verdere analyse zijn de data van de twee beoordelaars die het dichtst bij elkaar lagen, gebruikt. Eén kandidaat is uit het bestand verwijderd omdat bleek dat deze niet alle opgaven van het Staatsexamen had gemaakt.
- 2 De beoordelaarsovereenstemming in de gevolgde procedure voor de globale raamwerkbeoordeling Schrijven is goed en voldoende voor verdere analyses met betrekking tot de cesuurkwestie. Er bestaat een sterke positieve samenhang (.84) tussen de globale raamwerkbeoordeling en STEX-score. Kandidaten die volgens het globale oordeel functioneren op B1 hebben gemiddeld een hogere STEX-score dan kandidaten die B1 nog niet beheersen.

5.2 Relatie tussen STEX-scores en Raamwerkbeoordeling: de cesuurkwestie

Uit paragraaf 5.1 blijkt dat er een sterke relatie gevonden is tussen de STEX-scores en de globale beoordelingen volgens het Raamwerk NT2. Verder blijkt dat de gemiddelde STEX-score van de B1 kandidaten hoger ligt dan die van de groep kandidaten die B1 niet beheersen. De volgende vraag die nu beantwoord moet worden is of er aanleiding is voor een mogelijke herziening van de cesuur van het Staatsexamen Schrijven I om een goed onderscheid te maken tussen kandidaten die B1 wel of niet beheersen. In het onderzoeksvoorstel wordt dit als volgt verwoord:

Resultaat Stap 3

Indicatie voor de gewenste cesuur voor de vaardigheden Spreken en Schrijven.”

Voor het beantwoorden van deze vraag wordt in deze paragraaf informatie gepresenteerd die aan deze indicatie kan bijdragen.

De positieve relatie tussen de globale raamwerkbeoordeling en de scores op de Staatsexamen Schrijven I geeft aan dat er uitspraken kunnen worden gedaan ter ondersteuning van een eventuele verandering van de cesuurscore. Deze uitspraken zijn echter niet rechtstreeks af te leiden uit de tot nu toe gepresenteerde resultaten. Bij de vaststelling van een cesuur op grond van uitspraken over de beheersing van B1 spelen de volgende aspecten een rol:

- de definiëring van de beheersing van B1
- de consequenties die de hoogte van de cesuur heeft voor het ten onrechte laten zakken of het ten onrechte laten slagen van kandidaten. Afhankelijk van de vraag welke van deze twee consequenties men wil minimaliseren zal de cesuur hoger of lager komen te liggen.

In de volgende paragrafen dragen we de gegevens aan die kunnen bijdragen aan het vaststellen van een indicatie voor een cesuurwijziging.

5.2.1 Beheersing B1 volgens het criterium meerderheidsoordeel

Evenals in het cesuuronderzoek Staatsexamen Programma II beschouwen we een kandidaat als B1 indien deze in 80% van de beoordelingen als B1 beoordeeld wordt. Dit valt samen met een minimale score van 9,6 van de 12 te behalen punten. Dat betekent dat in minimaal 10 van de 12 beoordelingsuitspraken een beoordelaar uitspreekt dat iemand op B1 functioneert. De resultaten staan in tabel 5.2-1.

Tabel 5.2-1 Beoordeling kandidaten op Staatsexamen Schrijven I en de Raamwerkbeoordeling met als criterium meerderheidsoordeel van de beoordelaars (80%).

STEX-score	Totaalscore raamwerk	Beheersing B1 met meerderheidsoordeel
560	12	
553	12	
553	12	
549	12	
547	12	
545	12	
541	12	
540	12	
538	12	
538	12	
535	11	
534	12	
531	10	
531	11	
530	12	
529	10	
523	12	
521	10	
521	11	
521	12	
517	6	
517	11	
516	10	
516	12	
514	11	
513	10	
510	9	
510	11	
509	3	
505	10	
503	5	
502	9	
500	5	
499	4	
499	5	
499	10	
496	5	
494	2	
491	3	
488	9	
488	12	
487	5	
484	0	
484	4	
484	6	
482	2	
481	3	
479	4	
479	8	
478	7	

STEX-score	Totaalscore raamwerk	Beheersing B1 met meerderheidsoordeel
476	3	
470	3	
465	2	
464	3	
462	1	
459	2	
458	4	
457	0	
454	1	

	Geslaagd voor STEX / beheerst niveau B1
	Gezakt voor STEX / beheerst niveau B1 niet

Uit tabel 5.2-1 blijkt dat 29 kandidaten geacht worden B1 te beheersen en dat dit bij 30 kandidaten niet het geval is. Van de 26 kandidaten die gezakt waren voor het Staatsexamen beheersen er slechts twee B1. Een van deze kandidaten is met 499 maar net gezakt. Van de 26 kandidaten met een STEX-score 513 of hoger worden er 25 als B1 beoordeeld. In het interval tussen 500 en 513 bestaat de geringste samenhang tussen de STEX-score en de Raamwerkscore. Van de 7 kandidaten in dit scorebereik worden er 2 beoordeeld als B1. Van de 5 kandidaten die niet als B1 worden gezien, bevinden er zich 2 (met een score van 9) net op de grens van B1. Een eventuele nieuwe cesuur zal waarschijnlijk in dit interval liggen. Er zijn enkele kandidaten met een aanmerkelijk verschil tussen STEX- en Raamwerkbeoordeling, terwijl de Raamwerkbeoordelaars het wel met elkaar eens waren. In bijlage 5 wordt een korte impressie gegeven van het werk van deze kandidaten.

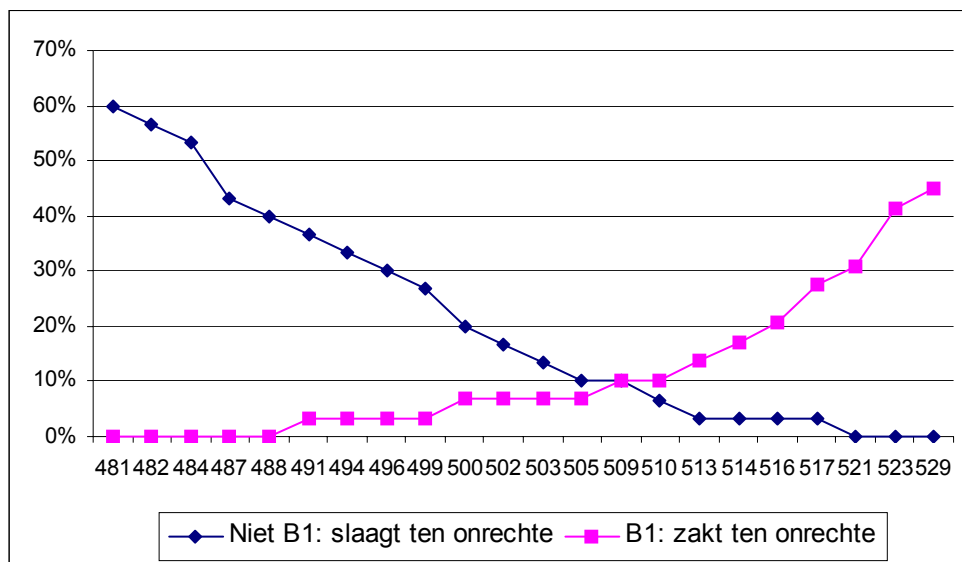
Voor het vaststellen van de cesuur is echter ook informatie nodig over de proportie kandidaten die bij een gegeven cesuurscore onterecht zakt of onterecht slaagt. Deze twee consequenties moeten tegen elkaar worden afgewogen. Tabel 5.2-2 geeft een overzicht van de gevolgen van de beslissingen voor de cesuurscores tussen 481 en 529.

Tabel 5.2-2 Percentages gezakte en geslaagde kandidaten die B1 al of niet beheersen bij verschillende cesuren (laagste voldoende score) van het Staatsexamen Schrijven I. Beheersing B1 gedefinieerd als 80% van maximale score

Cesuur op STEX	Beheerst B1 niet (n=30)		Beheerst B1 (n=29)	
	Zakt terecht	Slaagt ten onrechte	Zakt ten onrechte	Slaagt terecht
481	40,0%	60,0%	0%	100%
482	43,3%	56,7%	0%	100%
484	46,7%	53,3%	0%	100%
487	56,7%	43,3%	0%	100%
488	60,0%	40,0%	0%	100%
491	63,3%	36,7%	3,4%	96,6%
494	66,7%	33,3%	3,4%	96,6%
496	70,0%	30,0%	3,4%	96,6%
499	73,3%	26,7%	3,4%	96,6%
500	80,0%	20,0%	6,9%	93,1%
502	83,3%	16,7%	6,9%	93,1%
503	86,7%	13,3%	6,9%	93,1%
505	90,0%	10,0%	6,9%	93,1%
509	90,0%	10,0%	10,3%	89,7%
510	93,3%	6,7%	10,3%	89,7%
513	96,7%	3,3%	13,8%	86,2%
514	96,7%	3,3%	17,2%	82,8%
516	96,7%	3,3%	20,7%	79,3%
517	96,7%	3,3%	27,6%	72,4%
521	100%	0%	31,0%	69,0%
523	100%	0%	41,3%	58,7%
529	100%	0%	44,8%	55,2%

In figuur 5.2-1 wordt deze tabel grafisch weergegeven.

Figuur 5.2-1 Percentages gezakte en geslaagde kandidaten die B1 al of niet beheersen bij verschillende cesuren van het Staatsexamen Schrijven I. Beheersing B1 gedefinieerd als 80% van maximale score



Figuur 5.2-1 maakt duidelijk dat bij de uiteindelijke vaststelling van de cesuurscore rekening gehouden dient te worden met twee soorten foute beslissingen:

1. Fout negatief. De fout dat men ten onrechte kandidaten laat zakken die B1 wel beheersen.
2. Fout positief. De fout dat men ten onrechte kandidaten laat slagen die B1 nog niet beheersen. Zo leidt een cesuurscore van 481 ertoe dat de kans dat een B1-kandidaat ten onrechte zakt nihil is. Daar staat tegenover dat de kans dat kandidaten die B1 nog niet beheersen bij deze scores slagen bij deze cesuurscore zeer hoog zou zijn: 60%. Bij een cesuurscore van 529 maakt men de kans op het laten slagen van niet B1-kandidaten nihil, maar verhoogt men de kans op het onterecht zakken van B1-kandidaten tot 45%.

Welke cesuurscore men op basis van deze cijfers kiest, hangt sterk af van de functie van een examen en de ernst van de consequenties van een foute beslissing. Zou figuur 5.2-1 betrekking hebben op bijvoorbeeld een vliegvaardigheidsexamen voor burgerluchtvaartpiloten, dan zal men kiezen voor een cesuur van 529 of hoger. De consequenties van het laten slagen van een incompetent piloot kunnen immers zeer zwaarwegend zijn. Met andere woorden de maatschappelijke consequenties leggen meer gewicht in de schaal dan de persoonlijke consequenties. Het Staatsexamen NT2 geeft recht op een startkwalificatie. Het ligt daarom voor de hand dat het ten onrechte zakken zwaardere consequenties heeft dan het ten onrechte slagen. De persoonlijke consequenties mogen zwaarder wegen. Er zal dus binnen het interval gekozen moeten worden voor een cesuur die het ten onrechte zakken minimaliseert met daarbij een zo klein mogelijke kans op het ten onrechte slagen. We kunnen het optimum van deze keuze vinden door de kans op foute negatieven en foute positieven uit tabel 5.2-2 bij elkaar op te tellen. Een perfect, maar onbestaanbaar examen heeft een grensscore waarbij deze foute keuzes beide 0 en dus hun somscores ook 0 zijn. In de praktijk zal men met een minder scherpe grens tevreden moeten zijn.

Tabel 5.2-3 geeft de somwaarden weer onder twee condities:

1 Conditie 1

Onterecht zakken en slagen word even belangrijk gevonden (ongewogen conditie)

Voorbeeld: bij score 509 slaagt 10% ten onrechte en zakt 10,3% ten onrechte. De optimumscore wordt: $10 + 10,3 = 20,3$. Hoe kleiner dit getal hoe minder kans op een foute beslissing.

2 Conditie 2

Men wil voorkomen dat te veel B1-kandidaten ten onrechte zakken. Daarom geeft men ten onrechte zakken een dubbel gewicht.

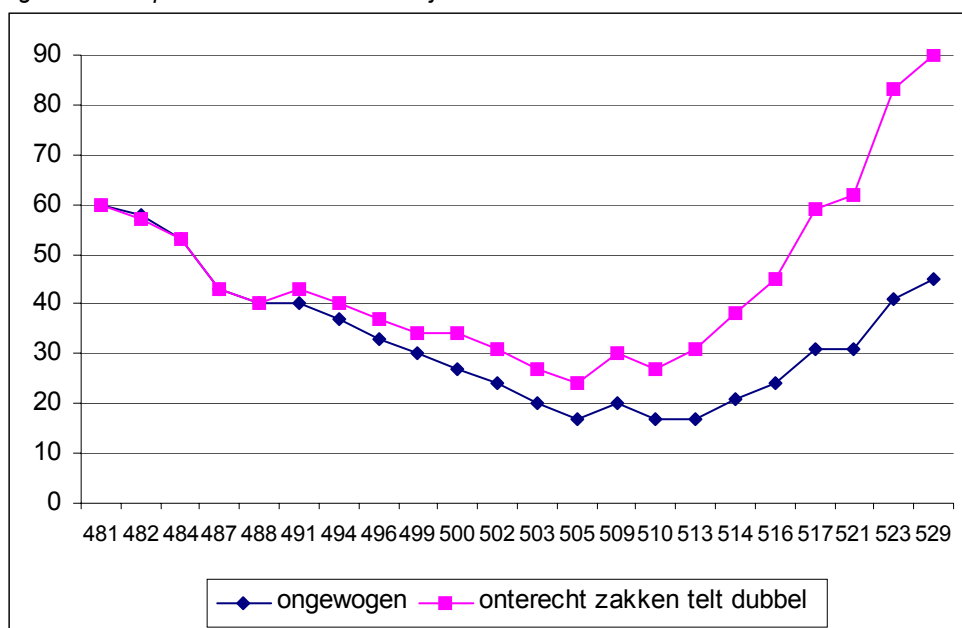
Voorbeeld: bij score 509 slaagt 10% ten onrechte en zakt 10,3% ten onrechte. De optimumscore als we het onrecht zakken dubbel wegen, wordt nu: $10 + 20,6 = 30,6$. Ook hier geldt dat de laagste optimumscore de geringste kans geeft op een foute beslissing.

Figuur 5.2-2 is een grafische weergave van deze tabel.

Tabel 5.2-3 Optimum cesuurscore Schrijven I onder 2 condities

	Conditie 1 Ongewogen	Conditie 2 Onterecht zakken telt dubbel
481	60	60
482	57	57
484	53	53
487	43	43
488	40	40
491	40	43
494	37	39
496	33	36
499	30	33
500	27	34
502	24	31
503	20	27
505	17	24
509	20	30
510	17	27
513	17	31
514	21	37
516	24	45
517	31	59
521	31	62
523	41	82
529	45	90

Figuur 5.2-2 Optimum cesuurscore Schrijven I onder 2 condities



De optimale cesuurscore in de ongewogen variant ligt tussen 505 en 513. Wanneer men meer belang hecht aan het voorkomen van het ten onrechte zakken van B1 kandidaten ligt de optimale cesuur op 505. Hieronder worden de proporties ten onrechte geslaagde en gezakte kandidaten voor dit cesuurvoorstel en voor de huidige cesuur weergegeven.

Tabel 5.2-4 Huidige en voorgestelde cesuurscores en de hierbij horende foute beslissingen

Cesuurscore	Niet B1: slaagt ten onrechte	B1: zakt ten onrechte	Optimumscore
500	20%	6,9%	34
505	10%	6,9%	24

Wanneer de nieuwe cesuur bij 505 gelegd zou worden neemt vanzelfsprekend het percentage kandidaten dat zakt voor het onderdeel Schrijven I toe. Tabel 5.2-5 geeft het aantal gezakten weer op het Staatsexamen Schrijven I van januari 2005.

Tabel 5.2.5 Percentage gezakte kandidaten in Staatsexamen Schrijven I, januari 2005 bij verschillende cesuurscores

Cesuurscore	Percentage gezakte kandidaten
500	49,18%
501	50,37%
502	50,52%
503	52,09%
504	53,35%
505	53,65%

5.2.2 Samenvatting en aanbevelingen cesuur Schrijven I

Volgens de behoeftepeiling zou het Staatsexamen Schrijven I een vaardigheid moeten waarborgen die op niveau B1 van het Raamwerk NT2 ligt (Kuijper e.a., 2005).

- 1 In dit onderzoek is dezelfde methodiek toegepast als in het cesuuronderzoek Spreken en Schrijven Programma II (Kuijper e.a., 2004).

- 2 Uit het beoordelingsonderzoek blijkt dat er een sterke positieve relatie bestaat tussen de scores die kandidaten halen op het Staatsexamen Schrijven I en de vraag of de respons van een kandidaat beoordeeld wordt als B1. In zijn algemeenheid geldt dat naarmate de STEX-score hoger wordt de kans dat een kandidaat als B1 wordt beoordeeld groter wordt.
- 3 De gevonden beoordelaarsovereenstemming bij de globale raamwerkbeoordeling Schrijven is hoog en voldoende om er uitspraken over een cesuurscore op te baseren.
- 4 Een definitieve beslissing over de cesuur kan niet alleen op statistische gronden worden genomen. Maatschappelijke overwegingen zijn het eerste uitgangspunt, waarna statistische procedures kunnen worden gebruikt om deze uitgangspunten zo goed mogelijk in een cesuurvaststelling om te zetten.
- 5 Wanneer de cesuur van het Staatsexamen NT2 Schrijven Programma I een onderscheid moet maken tussen kandidaten die een schrijfvaardigheid hebben die vergelijkbaar is met niveau B1, dan ligt de optimale cesuur bij en 505. Bij het vaststellen van dit cesuurvoorstel moet men in ogenschouw nemen hoe men - in relatie tot de functie van het examen - het ten onrechte slagen of zakken tegen elkaar afweegt. Voor het Staatsexamen NT2 geldt dat het een startkwalificatie verschaft. Dit zou kunnen betekenen dat men het ten onrechte zakken van een kandidaat een slechtere optie vindt dan het ten onrechte slagen. Een cesuurscore van 505 als eerste voldoende score ligt dan voor de hand. Het aantal gezakte kandidaten neemt dan toe van 49,18% tot 53,65%.

6 Verantwoording van de gebruikte methode

6.1 Inleiding

In het huidige onderzoek is gekozen voor een methode van cesuurbepaling die gebaseerd is op het principe dat de classificatie van kandidaten zo goed mogelijk moet zijn. Het Staatsexamen is hierbij opgevat als een efficiënte maar feilbare methode om te bepalen of kandidaten minimaal een bepaald CEF-niveau hebben.

Een voorwaarde voor het gebruik van de methode is dat het CEF-niveau foutloos kan worden bepaald. Evenals bij de medische diagnostiek valt te betwijfelen of dit het geval is.

Om het CEF-niveau zo goed mogelijk te bepalen lieten we het niveau van elke kandidaat door twee onafhankelijke beoordelaars beoordelen. De functie van het tweede oordeel was om kandidaten waarvan het CEF-niveau niet met voldoende grote zekerheid kon worden vastgesteld uit te sluiten.

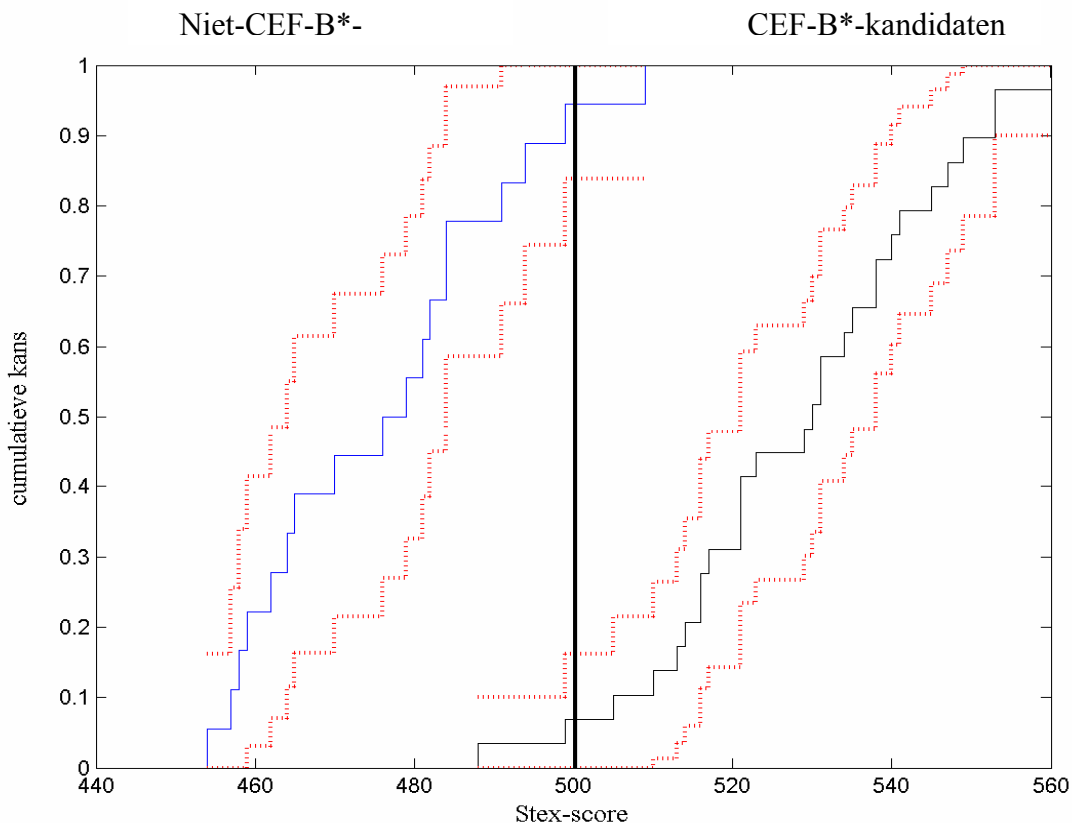
Concreet werden alleen kandidaten waarover de oordelen in meerderheid overeenkwamen gebruikt voor cesuurbepaling (zie Maris en Bechger, 2004).

We hebben verschillende mogelijkheden om een meerderheidsoordeel te bepalen. Omdat beoordelaars meerdere aspecten hebben beoordeeld kunnen we beoordelingen over aspecten sommeren. Dat is in het huidige onderzoek gedaan. We zouden ook een globaal oordeel als uitgangspunt kunnen nemen.

In het huidige onderzoek zijn verschillende CEF-niveaus bepaald voor verschillende vaardigheden, Spreken en Schrijven, en verschillende examens, Programma I en II. De overeenstemming is afhankelijk van de vaardigheid die werd beoordeeld, het CEF-niveau en het examen. Het kan betekenen dat men voor sommige vaardigheden meer kandidaten moet beoordelen dan voor anderen om voldoende kandidaten over te houden voor de in dit onderzoek toegepaste ROC-curve analyse. Deze analyse wordt in 6.2 toegelicht.

6.2 De ROC-curve analyse

De methode staat bekend onder de naam ROC-curve analyse. Ze wordt vooral in de medische diagnostiek en de mathematische psychologie toegepast en lijkt niet eerder te zijn gebruikt als methode voor cesuurbepaling. Vandaar dat we in deze paragraaf kort de methode en de voor- en nadelen van ROC-curve analyse willen bespreken.



Hierboven wordt geïllustreerd hoe de methode werkt. Het plaatje geeft de geschatte verdeling van de staatsexamenscore in twee groepen. Links van het midden staat de verdeling van diegenen die niet aan het criterium voldoen (niet-CEF- kandidaten). Rechts van het midden die van de kandidaten waarvan is vastgesteld dat ze wel aan het criterium voldoen (CEF-kandidaten). De vetgedrukte verticale lijn geeft aan waar op de horizontale as de huidige cesuur op het staatsexamen ligt.

Ten eerste zien we in dit voorbeeld, dat bij een cesuur van 500, een procent of 10 van de niet-CEF-kandidaten (CEF-) toch slaagt voor het staatsexamen (EX+). We noteren dit als volgt:

$$P(\text{EX+} \mid \text{CEF-}, C = 500) = 0.10 \quad \text{“Onterecht Slagen”}$$

waarbij $C = 500$ betekent: “Bij een cesuur van 500”. We zien ook dat

$$P(\text{EX-} \mid \text{CEF+}, C = 500) = 0.08. \quad \text{“Onterecht Zakken”}$$

Gegeven dat het staatsexamen dient om CEF-kandidaten (CEF+) te onderscheiden van niet-CEF-kandidaten zijn dit, respectievelijk, de kans op onterecht slagen en de kans op onterecht zakken.

Ten tweede valt op dat de verdeling van de CEF-kandidaten rechts ligt van die van de niet-CEF-kandidaten. Dit betekent dat de kans dat een CEF-kandidaat slaagt voor het examen voor alle mogelijke cesuren hoger is dan die van een niet-CEF-kandidaat. We noteren dit als volgt: voor alle grenzen x geldt dat

$$P(\text{Score} > x \mid \text{CEF+}, C = x) > P(\text{Score} > x \mid \text{CEF-}, C = x).$$

Dit is een belangrijke eigenschap waaraan het Staatsexamen in alle gevallen voldeed. Als deze eigenschap, die bekend staat als stochastische ordening, niet zou gelden dan zouden er cesuren zijn, waarvoor het staatsexamen negatief samenhangt met het CEF-niveau.

Tenslotte merken we op dat de stippellijnen rond de curven betrouwbaarheidsintervallen zijn. Deze

betrouwbaarheidsintervallen geven aan dat er een grote onzekerheidsmarge moet worden aangenomen rond de geschatte percentages. Als het aantal personen in de steekproef toeneemt worden de onzekerheidsintervallen kleiner. De geschatte verdelingen stabiliseren en worden steeds minder rafelig van uiterlijk.

Zowel de kans op onterecht zakken als ook de kans op onterecht slagen zijn afhankelijk van de cesuur. Bij verschuiving van de cesuur veranderen deze kansen. Als de cesuur wordt verhoogd wordt de kans op ten onrechte slagen kleiner en die op ten onrechte zakken groter. Bij verlaging van de cesuur gebeurt het omgekeerde. ROC-curve analyse bestaat hieruit dat we de cesuur kiezen waarbij fouten zo klein mogelijk zijn. In bovenstaande illustratie lijkt de huidige cesuur hier goed in te slagen, wanneer we de kans op ten onrechte zakken en ten onrechte slagen even zwaar wegen.

Vaak wordt er gekozen de gewogen som van de kansen te minimaliseren. Dat wil zeggen: kies x zodanig dat

$$GW_1 \times P(\text{EX-} \mid \text{CEF+}, C = x) + GW_2 \times P(\text{EX+} \mid \text{CEF-}, C = x)$$

zo klein mogelijk is. De gewichten GW_1 en GW_2 worden zo gekozen dat het relatieve belang van de ene fout ten opzichte van de andere overeenkomt met de mening van de betrokkenen. ROC-curve is in deze zin dus geen objectieve methode. Er is altijd sprake van (inter)subjectieve oordelen over de beste weging van de kansen.

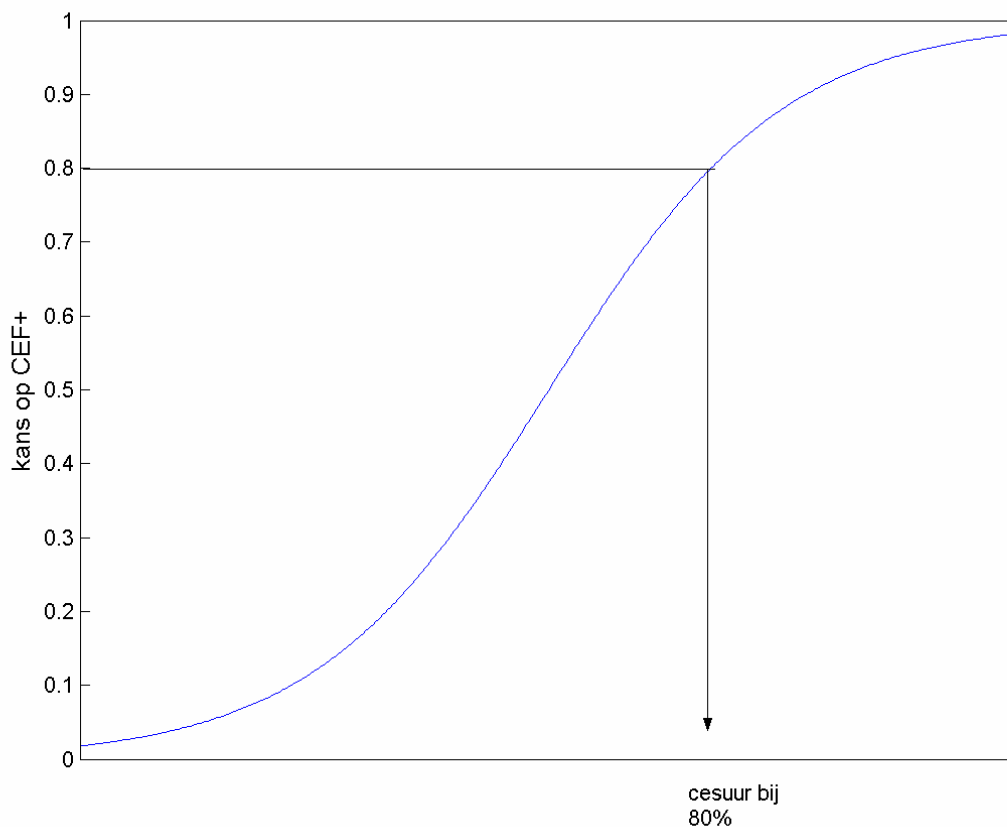
De verdelingen zijn essentieel voor het vaststellen van de cesuur. Dat betekent dat de methode populatieafhankelijk is. Als de populatie verandert, dan dient de cesuur opnieuw geëvalueerd te worden.

Ten opzichte van andere methoden geldt dat bij ROC-curve analyse de cesuur kan veranderen maar de voorspellingsfouten constant kunnen blijven, terwijl bij andere methoden de cesuur constant blijft maar de kwaliteit van de voorspelling fluctueert over de tijd.

ROC-curve analyse is gebaseerd op geschatte verdelingen. In bovenstaande figuur hebben we de geschatte verdelingen getekend. Wat geschat wordt zijn de verdelingen in de populatie van examenkandidaten. Daarom is het van belang dat de personen die bij het onderzoek worden betrokken op basis van toeval uit de populatie van examenkandidaten werden getrokken. Omdat het om schattingen gaat moeten we, bij het bepalen van de cesuur, rekening houden met steekproefvariaties. Op het oog zou, in het voorbeeld, de cesuur tussen 490 en 510 kunnen liggen zonder dat de fouten significant groter of kleiner zouden worden. Deze onzekerheid kan worden teruggebracht door meer personen bij het onderzoek te betrekken. Merk op dat, in het huidige onderzoek de cesuur alleen is aangepast indien zowel de ROC-curve analyse als ook de behoeftepeiling hier aanleiding toe gaf. Dat biedt een bescherming tegen het risico dat de cesuur bepaald zou zijn door louter toeval.

6.3 Andere methoden

De voordelen van ROC-curve analyses zijn dat de methode relatief goedkoop is en naast een cesuur veel informatie oplevert over de kwaliteit van het examen met betrekking tot de voorspellingen van het examen voor het CEF-niveau. Er kunnen ook andere methoden gebruikt worden. Bijvoorbeeld dat het oordeel over het CEF niveau van een kandidaat wordt meegekalibreerd met de responsen op het Staatsexamen. De wijze waarop de gegevens gekalibreerd kunnen worden staat beschreven in Maris en Bechger (2004).



Men gaat bij deze methode eveneens uit van een steekproef van kandidaten die het staatsexamen hebben gemaakt en waarvan tevens een oordeel over hun CEF-niveau beschikbaar is. Het oordeel wordt opgenomen als item en samen met de responsen op de vragen op het staatsexamen gekalibreerd. De eerste voorwaarde is dat het oordeel als item past bij het staatsexamenscores.

Met behulp van de "item response functie (IRF)" van het oordeel bepalen we nu de cesuur. In bovenstaande figuur is te zien hoe dat gaat. We kiezen een minimale kans om CEF+ te zijn, bijvoorbeeld 80%. Dan lopen we via de IRF terug naar de bijbehorende vaardigheid. Dat wordt dan de cesuur.

Voordelen:

- a) De cesuur wordt bepaald voor alle populaties waarbij het examen dezelfde vaardigheid meet.
- b) We kunnen bepalen of het CEF-niveau betrekking heeft op dezelfde vaardigheid die door het examen wordt gemeten.

Een nadeel van deze methode is dat ze veel grotere aantallen observaties vereist dan ROC-curve analyse.

De methode is onafhankelijk van de populatie maar garandeert evenmin dat de kans op ten onrechte zakken of slagen constant blijven over de tijd. Een voordeel ten opzichte van de ROC-curve analyse is dat we kunnen nagaan of het CEF-niveau en de staatsexamenscore betrekking hebben op dezelfde vaardigheid.

Literatuur

- Council of Europe (2001).** *Common European Framework of Reference for Languages: learning, teaching, assessment.* Cambridge University press, Cambridge.
- Dalderop, Kaatje, Liemberg, Ellie en Teunisse, Franke (2002).** *Raamwerk NT2.* BVE-Raad, De Bilt.
- De Jong, John, Bernstein, Jared (z.d.).** *Relating PhonePass™ overall scores to the Council of Europe Framework level descriptors.* Language Testing Services, Velp, Ordinate Corporation, Menlo Park CA.
- Eggen, T.J.H.M., Sanders, P.F., (red). (1993).** *Psychometrie in de praktijk.* Cito, Arnhem.
- ETS (2004).** *Mapping TOEFL, TSE, TWE, AND TOEIC on the Common European Framework. Executive Summary.* www.ets.org/ell
- Kaftandjieva, Felianka (2004).** *Standard setting.* In: Reference Supplement to the preliminary pilot version of the Manual for Relating Language examinations to the CEF, Section B. Language Policy Division, Strasbourg
- Kuijper, H., Bergsma, A., Bechger, T. (2004).** *Staatsexamen NT2. Het gewenste niveau Deel 1 Behoeftetepeiling Programma II.* Augustus 2004, Citogroep, Arnhem.
- Kuijper, H., Bergsma, A., Bechger, T. (2004).** *Staatsexamen NT2. het gewenste niveau. Deel 2 cesuurindicatie Spreken en Schrijven Programma II.* Oktober 2004, Citogroep, Arnhem.
- Kuijper H., Bergsma, A., Bechger, T. (2004).** *Staatsexamen NT2. Het gewenste niveau. Deel 2B Cesuurindicatie Spreken en Schrijven Programma II. Aanvullend onderzoek.* Citogroep, Arnhem, december 2004.
- Kuijper, H., Bergsma, A., Bechger, T. (2005).** *Staatsexamen NT2. Het gewenste niveau Deel 3 Behoeftetepeiling Programma I.* Mei 2005, Citogroep, Arnhem.
- Liemberg, E., Hulstijn, J. (1996).** *Referentiekader NT2, Concept 2.* PROVE, Amersfoort, november 1996.
- Language Policy Division (2003).** *Relating Language Examinations to the Common European Framework of Reference for Languages: learning, Teaching, Assessment (CEF). Manual Preliminary Pilot Version.* Language Policy Division, Strasbourg. http://www.coe.int/T/E/Cultural_Co-operation/education/Languages
- Language Policy Division (2004).** *Reference Supplement to the preliminary pilot version of the Manual for Relating Language examinations to the CEF.* Language Policy Division, Strasbourg. http://www.coe.int/T/E/Cultural_Co-operation/education/Languages
- Maris, G., en Bechger, T.M. (2004).** *Het scoren van open vragen: theorie en praktijk.* POK Memorandum 2004-2, Cito: Arnhem.
- Tannenbaum, R.J., Wylie, E.C. (2003).** *Mapping test scores onto the Common European Framework: Setting standards of language proficiency on the Test of English as a Foreign Language (TOEFL), The Test of Spoken English (TSE), The test of Written English (TWE), and the test of English for International Communication (TOEIC).* Princeton, NJ: Educational Testing Service.

Bijlage 1 Globale raamwerkbeoordeling Spreken I

Kandidaatnummer:

--	--	--	--	--	--

Beoordelaar:

--

Globaal oordeel

Algemene beheersingsuitspraken Spreken B1

In te vullen voor het examen als geheel

U heeft het gehele examen van deze kandidaat beluisterd.

Wilt u op grond van uw algemene indruk aangeven of deze kandidaat naar uw mening in het dagelijks leven in staat is de volgende taalhandelingen op niveau B1 uit te voeren?

Baseer uw oordeel op het document: Familiarisatie Spreken B1

Algemene beheersingsuitspraken Spreken B1

	Uitspraken Raamwerk NT2 (aangepast)	Ja	Nee
S-G1	Kan uitingen op een eenvoudige manier aan elkaar verbinden, zodat hij/zij gebeurtenissen, dromen, verwachtingen en ambities kan beschrijven.		
S-G2	Kan in het kort redenen en verklaringen geven voor meningen en plannen.		
S-G3	Kan een verhaal vertellen of de plot van een boek of film weergeven en zijn/haar reacties beschrijven.		
GV-G4	Kan onvoorbereid deelnemen aan een gesprek over onderwerpen die vertrouwd zijn, of zijn/haar persoonlijke belangstelling hebben of die betrekking hebben op het dagelijks leven (bijvoorbeeld familie, hobby's, werk, reizen en actuele gebeurtenissen).		
SGV-G5	Beheerst deze kandidaat over het algemeen niveau B1?		

Bijlage 2 Globale raamwerkbeoordeling Schrijven I

Kandidaatnummer:

--	--	--	--	--	--

Beoordelaar:

--

Globaal oordeel Algemene beheersingsuitspraken Schrijven B1

In te vullen voor het examen als geheel

U heeft het gehele examen van deze kandidaat bekeken.

Wilt u op grond van uw algemene indruk aangeven of deze kandidaat naar uw mening in het dagelijks leven

in staat is de volgende taalhandelingen op niveau B1 uit te voeren?

Baseer uw oordeel op het document: Familiarisatie Schrijven B1

	Uitspraken Raamwerk NT2 (aangepast)	Ja	Nee
SC-G1	Kan een eenvoudige, samenhangende tekst schrijven over onderwerpen die vertrouwd of van persoonlijk belang zijn.		
SC-G2	Kan persoonlijke brieven schrijven waarin ervaringen en indrukken beschreven worden.		
SC-G3	Kan notities/berichten schrijven waarin eenvoudige, relevante informatie begrijpelijk overkomt.		
SC-G4	Kan heel korte rapporten schrijven volgens een vastgestelde standaard waarin routinematige feitelijke informatie wordt doorgegeven.		
SC-G5	Kan een korte, eenvoudige zakelijke tekst schrijven.		
SC-G6	Beheerst deze kandidaat over het algemeen niveau B1?		

Bijlage 3 Familiarisatie Spreken

Familiarisatie Spreken en Gesprekken Voeren B1

Globale schalen

Zelfevaluatiematrix

Subvaardigheden en descriptoren Gesprekken Voeren B1

Subvaardigheden en descriptoren Spreken B1

De globale schalen

De gemeenschappelijke referentieniveaus: de globale schalen

Bij alle onderverdelingen die mogelijk en ook vaak wenselijk zijn is het belangrijk om vast te houden aan gemeenschappelijke referentiepunten, opdat in verschillende landen en in verschillende onderwijssectoren helder blijft wat iemand kan die een taal beheerst op bijvoorbeeld niveau B1. Daartoe biedt het CEF de zogenaamde 'common reference levels', de gemeenschappelijke referentieniveaus. Die worden eerst gepresenteerd in een overzichtsvorm, vanuit de gedachte dat op die manier de systematiek duidelijk te maken is aan niet-specialisten: leerlingen, ouders, vakdocenten, werkgevers.

We geven hier een vertaling van de globale schalen uit het Common European Framework (CUP editie, pg 24. Vertaling van John de Jong en Francis Staatsen).

De niveaus zijn *cumulatief*, dat wil zeggen dat beheersing op een bepaald niveau inhoudt dat alles wat beschreven is op de onderliggende niveaus ook beheerst wordt.

Lees nauwkeurig de beschrijving van B1 voordat u aan het beoordelen begint.

Basisgebruiker	
A1	Kan vertrouwde dagelijkse uitdrukkingen en basiszinnen gericht op de bevrediging van concrete behoeften begrijpen en gebruiken. Kan zichzelf aan anderen voorstellen en kan vragen stellen en beantwoorden over persoonlijke gegevens zoals waar hij/zij woont, mensen die hij/zij kent en dingen die hij/zij bezit. Kan op een simpele wijze reageren, aangenomen dat de andere persoon langzaam en duidelijk praat en bereid is om te helpen.
A2	Kan zinnen en regelmatig voorkomende uitdrukkingen begrijpen die verband hebben met zaken van direct belang (bijvoorbeeld persoonsgegevens, familie, winkelen, plaatselijke geografie, werk). Kan communiceren in simpele en alledaagse taken die een eenvoudige en directe uitwisseling over vertrouwde en alledaagse kwesties vereisen. Kan in eenvoudige bewoordingen aspecten van de eigen achtergrond, de onmiddellijke omgeving en kwesties op het gebied van diverse behoeften beschrijven.

Onafhankelijk gebruiker	
B1	Kan de belangrijkste punten begrijpen uit duidelijke standaardteksten over vertrouwde zaken die regelmatig voorkomen op het werk, op school en in de vrije tijd. Kan zich redden in de meeste situaties die kunnen optreden tijdens het reizen in gebieden waar de betreffende taal wordt gesproken. Kan een eenvoudige lopende tekst produceren over onderwerpen die vertrouwd of die van persoonlijk belang zijn. Kan een beschrijving geven van ervaringen en gebeurtenissen, dromen, verwachtingen en ambities en kan kort redenen en verklaringen geven voor meningen en plannen.
B2	Kan de hoofdgedachte van een ingewikkelde tekst begrijpen, zowel over concrete als over abstracte onderwerpen, met inbegrip van technische besprekingen in het eigen vakgebied. Kan zo vloeiend en spontaan reageren dat een normale uitwisseling met moedertaalsprekers mogelijk is zonder dat dit voor een van de partijen inspanning met zich meebrengt. Kan duidelijke, gedetailleerde tekst produceren over een breed scala van onderwerpen; kan een standpunt over een actuele kwestie uiteenzetten en daarbij ingaan op de voor- en nadelen van diverse opties.

Vaardig gebruiker	
C1	Kan een uitgebreid scala van veeleisende, lange teksten begrijpen en de impliciete betekenis herkennen. Kan zichzelf vloeiend en spontaan uitdrukken zonder daarvoor aantoonbaar naar uitdrukkingen te moeten zoeken. Kan flexibel en effectief met taal omgaan ten behoeve van sociale, academische en beroepsmatige doeleinden. Kan een duidelijke, goed gestructureerde en gedetailleerde tekst over complexe onderwerpen produceren en daarbij gebruikmaken van organisatorische structuren en verbindingswoorden.
C2	Kan vrijwel alles wat hij hoort of leest gemakkelijk begrijpen. Hij kan informatie die afkomstig is van verschillende gesproken en geschreven bronnen samenvatten, argumenten reconstrueren en hiervan samenhangend verslag doen. Hij kan zichzelf spontaan, vloeiend en precies uitdrukken en kan hierbij fijne nuances in betekenis, zelfs in complexere situaties onderscheiden.

Zelfevaluatie matrix Spreken en Gesprekken Voeren (Raamwerk NT2, pag. 12)

	A1	A2	B1	B2	C1	C2
Spreken	Ik kan deelnemen aan een eenvoudig gesprek, wanneer de gesprekspartner bereid is om zaken in een langzamer spreektempo te herhalen of opnieuw te formuleren en mij helpt bij het formuleren van wat ik probeer te zeggen. Ik kan eenvoudige vragen stellen en beantwoorden die een directe behoefte of zeer vertrouwde onderwerpen betreffen.	Ik kan communiceren over eenvoudige en alledaagse taken die een eenvoudige en directe uitwisseling van informatie over vertrouwde onderwerpen en activiteiten betreffen. Ik kan zeer korte sociale gesprekken aan, alhoewel ik gewoonlijk niet voldoende begrijp om het gesprek zelfstandig gaande te houden.	Ik kan de meeste situaties aan die zich kunnen voordoen tijdens een reis in een gebied waar de betreffende taal wordt gesproken. Ik kan onvoorbereid deelnemen aan een gesprek over onderwerpen die vertrouwd zijn, of mijn persoonlijke belangstelling hebben of die betrekking hebben op het dagelijks leven (bijvoorbeeld familie, hobby's, werk, reizen en actuele gebeurtenissen).	Ik kan zodanig deelnemen aan een vloeiend en spontaan gesprek, dat normale uitwisseling met moedertaalsprekers redelijk mogelijk is. Ik kan binnen een vertrouwde context actief deelnemen aan een discussie en hierin mijn standpunten uitleggen en ondersteunen.	Ik kan mezelf vloeiend en spontaan uitdrukken zonder merkbaar naar uitdrukkingen te hoeven zoeken. Ik kan de taal flexibel en effectief gebruiken voor sociale en professionele doeleinden. Ik kan ideeën en meningen met precisie formuleren en mijn bijdrage vaardig aan die van andere sprekers relateren.	Ik kan zonder moeite deelnemen aan welk gesprek of discussie dan ook en ben zeer vertrouwd met idiomatische uitdrukkingen en spreektaal. Ik kan mezelf vloeiend uitdrukken en de fijnere betekenisnuances precies weergeven. Als ik een probleem tegenkom, kan ik mezelf hernemen en mijn betoog zo herstructureren dat andere mensen het nauwelijks merken.
Gesprekken voeren	Ik kan eenvoudige uitdrukkingen en zinnen gebruiken om mijn woonomgeving en de mensen die ik ken, te beschrijven.	Ik kan een reeks uitdrukkingen en zinnen gebruiken om in eenvoudige bewoordingen mijn familie en andere mensen, leefomstandigheden, mijn opleiding en mijn huidige of meest recente baan te beschrijven.	Ik kan uitingen op een simpele manier aan elkaar verbinden, zodat ik ervaringen en gebeurtenissen, mijn dromen, verwachtingen en ambities kan beschrijven. Ik kan in het kort redenen en verklaringen geven voor mijn meningen en plannen. Ik kan een verhaal vertellen, of de plot van een boek of film weergeven en mijn reacties beschrijven	Ik kan duidelijke, gedetailleerde beschrijvingen presenteren over een breed scala van onderwerpen die betrekking hebben op mijn interessegebied. Ik kan een standpunt over een actueel onderwerp verklaren en de voordelen en nadelen van diverse opties uiteenzetten.	Ik kan duidelijke, gedetailleerde beschrijvingen geven over complexe onderwerpen en daarbij sub-thema's integreren, specifieke standpunten ontwikkelen en het geheel afronden met een passende conclusie.	Ik kan een duidelijke, goedlopende beschrijving of redenering presenteren in een stijl die past bij de context en in een doeltreffende logische structuur, zodat de toehoorder in staat is de belangrijke punten op te merken en te onthouden.

Vaardigheid: Gesprekken voeren

Niveau: B1

Beheersingsniveau:

Kan onvoorbereid deelnemen aan een gesprek over onderwerpen die vertrouwd zijn, of zijn/haar persoonlijke belangstelling hebben of die betrekking hebben op het dagelijks leven (bijvoorbeeld familie, hobby's, werk, reizen en actuele gebeurtenissen).

Subvaardigheden en Descriptoren	Voorbeelden bij de descriptoren	Tekstkenmerken receptief	Tekstkenmerken productief
<p>1. Informele gesprekken Kan gevoelens uitdrukken en erop reageren. Kan uitleggen waarom iets een probleem is. Kan persoonlijke standpunten en meningen geven en zoeken in gesprekken over onderwerpen die binnen de eigen interessegebied liggen, en commentaar geven op standpunten van anderen.</p> <p>2. Bijeenkomsten en vergaderingen Kan een standpunt duidelijk overbrengen, maar heeft moeite mee te doen in een debat. Kan deelnemen aan routinematige formele discussies over vertrouwde onderwerpen, waarbij duidelijk gesproken wordt in de standaardtaal en waarbij sprake is van uitwisseling van feitelijke informatie, het geven van instructies of het bespreken van praktische problemen.</p> <p>3. Zaken regelen Kan bij het bespreken van hoe verder te gaan met een klus of mogelijke oplossingen voor iets, zeggen wat hij/zij ervan vindt en daarbij kort een reden of een uitleg geven Kan minder routinematige situaties aan in winkels, postkantoren, banken, bijvoorbeeld bij het terugbrengen van een aankoop of bij klachten over producten en/of diensten.</p> <p>4. Informatie uitwisselen Kan eenvoudige feitelijke informatie achterhalen en doorgeven. Kan om gedetailleerde aanwijzingen vragen en ze opvolgen. Kan meer gedetailleerde informatie verkrijgen. Kan in beperkte mate initiatieven nemen in een vraaggesprek, bijvoorbeeld over een nieuw onderwerp beginnen.</p>	<p>De voorbeelden zijn opgenomen op de volgende pagina, uitgesplitst naar de domeinen werk, opleiding en dagelijks leven.</p>	<ul style="list-style-type: none"> ▪ onderwerp Het onderwerp is vertrouwd, valt binnen de persoonlijke belangstelling of heeft betrekking op het dagelijks leven, eigen werk of opleiding. ▪ woordgebruik en zinsbouw Het woordgebruik is toereikend voor alledaagse zaken. Elementaire structuren en veel herhalingen. ▪ tempo en articulatie Woorden worden duidelijk uitgesproken in een vertrouwd accent. ▪ hulp De taalgebruiker wordt direct aangesproken. Soms moeten bepaalde woorden en uitdrukkingen herhaald worden. Ongebruikelijke idiomatische uitdrukkingen worden vermeden en er wordt duidelijk gearticuleerd. 	<ul style="list-style-type: none"> ▪ onderwerp Het onderwerp is vertrouwd, valt binnen de persoonlijke belangstelling of heeft betrekking op het dagelijks leven, eigen werk of opleiding. ▪ woordenschat en woordgebruik De woordenschat is toereikend om overalldagse onderwerpen, zoals familie, hobby's, werk, reizen en actuele gebeurtenissen te spreken eventueel door middel van omschrijvingen. ▪ grammaticale correctheid Redelijk accuraat gebruik van frequente routines en patronen die horen bij voorspelbare situaties. ▪ interactie Kan een eenvoudig face-to-face gesprek over bekende onderwerpen waarvoor persoonlijke interesse bestaat beginnen, voeren en afsluiten. Kan gedeeltelijk herhalen wat iemand gezegd heeft om te bevestigen dat men elkaar begrepen heeft. ▪ vloeiendheid Is goed te volgen, alhoewel pauzes voor grammaticale en lexicale planning en herstel van fouten, vooral in langere stukken vrij geproduceerde tekst, veel voorkomen. ▪ coherentie Series van kortere, eenvoudige afzonderlijke elementen zijn verbonden tot een samenhangende lineaire reeks van punten. ▪ uitspraak De uitspraak is duidelijk verstaanbaar, alhoewel gekleurd door een accent en af en toe een verkeerd uitgesproken woord.

Vaardigheid: Spreken

Niveau: B1

Beheersingsniveau:

Kan uitingen op een eenvoudige manier aan elkaar verbinden, zodat hij/zij ervaringen, gebeurtenissen, dromen, verwachtingen en ambities kan beschrijven. Kan in het kort redenen en verklaringen geven voor meningen en plannen. Kan een verhaal vertellen of de plot van een boek of film weergeven en zijn/haar reacties beschrijven.

Subvaardigheden en descriptoren	Voorbeelden	Kenmerken van de taakuitvoering
<p>1. Monologen Kan zonder omhaal een beschrijving geven van verscheidene vertrouwde zaken binnen zijn/of haar interessegebied. Kan gedetailleerd verslag doen van ervaringen, en meningen en reacties beschrijven. Kan details vertellen van onverwachte gebeurtenissen, zoals bijvoorbeeld een ongeluk. Kan echte of verzonnen gebeurtenissen beschrijven. Kan een verhaal vertellen. Kan vertellen over zijn/haar dromen, verwachtingen en toekomstwensen.</p> <p>2. Een publiek toespreken Kan een eenvoudige voorbereide presentatie geven over een onderwerp binnen zijn/haar werkveld, die ondanks afwijkend accent en intonatie over het algemeen duidelijk genoeg is om zonder moeilijkheden te volgen, en waarin de hoofdpunten redelijk gedetailleerd geformuleerd zijn. Kan hierop volgende vragen beantwoorden, moet eventueel wel om herhaling vragen als het spreektempo hoog lag.</p>	<p>De voorbeelden zijn opgenomen op de volgende pagina, uitgesplitst naar de domeinen werk, opleiding en dagelijks leven.</p>	<ul style="list-style-type: none">▪ onderwerp Het onderwerp is vertrouwd, valt binnen de persoonlijke belangstelling of heeft betrekking op het dagelijks leven, eigen werk of opleiding.▪ woordgebruik en woordenschat De woordenschat is toereikend om eventueel met behulp van omschrijvingen vertrouwde onderwerpen, zoals familie, hobby's, werk, reizen en actuele gebeurtenissen te beschrijven.▪ grammaticale correctheid Redelijk accuraat gebruik van frequente 'routines' en patronen die horen bij over het algemeen voorspelbare situaties.▪ vloeiendheid De spreker is goed te volgen, alhoewel pauzes voor grammaticale en lexicale planning en herstel van fouten, vooral in langere stukken vrij geproduceerde tekst, veel voorkomen.▪ coherentie Series van kortere, eenvoudige afzonderlijke elementen zijn verbonden tot een samenhangende lineaire reeks van punten.▪ uitspraak De uitspraak is duidelijk verstaanbaar, zelfs met een accent en af en toe een verkeerd uitgesproken woord.

Bijlage 4 Familiarisatie Schrijven

Familiarisatie Schrijven B1

Globale schalen
Zelfevaluatiematrix
Subvaardigheden en descriptoren Schrijven B1

De globale schalen

De gemeenschappelijke referentieniveaus: de globale schalen

Bij alle onderverdelingen die mogelijk en ook vaak wenselijk zijn is het belangrijk om vast te houden aan gemeenschappelijke referentiepunten, opdat in verschillende landen en in verschillende onderwijssectoren helder blijft wat iemand kan die een taal beheerst op bijvoorbeeld niveau B1. Daartoe biedt het CEF de zogenaamde 'common reference levels', de gemeenschappelijke referentieniveaus. Die worden eerst gepresenteerd in een overzichtsvorm, vanuit de gedachte dat op die manier de systematiek duidelijk te maken is aan niet-specialisten: leerlingen, ouders, vakdocenten, werkgevers.

We geven hier een vertaling van de globale schalen uit het Common European Framework (CUP editie, pg 24. Vertaling van John de Jong en Francis Staatsen).

De niveaus zijn *cumulatief*, dat wil zeggen dat beheersing op een bepaald niveau inhoudt dat alles wat beschreven is op de onderliggende niveaus ook beheerst wordt.

Lees nauwkeurig de beschrijving van B1 voordat u aan het beoordelen begint.

Basisgebruiker	
A1	Kan vertrouwde dagelijkse uitdrukkingen en basiszinnen gericht op de bevrediging van concrete behoeften begrijpen en gebruiken. Kan zichzelf aan anderen voorstellen en kan vragen stellen en beantwoorden over persoonlijke gegevens zoals waar hij/zij woont, mensen die hij/zij kent en dingen die hij/zij bezit. Kan op een simpele wijze reageren, aangenomen dat de andere persoon langzaam en duidelijk praat en bereid is om te helpen.
A2	Kan zinnen en regelmatig voorkomende uitdrukkingen begrijpen die verband hebben met zaken van direct belang (bijvoorbeeld persoonsgegevens, familie, winkelen, plaatselijke geografie, werk). Kan communiceren in simpele en alledaagse taken die een eenvoudige en directe uitwisseling over vertrouwde en alledaagse kwesties vereisen. Kan in eenvoudige bewoordingen aspecten van de eigen achtergrond, de onmiddellijke omgeving en kwesties op het gebied van diverse behoeften beschrijven.

Onafhankelijk gebruiker	
B1	Kan de belangrijkste punten begrijpen uit duidelijke standaardteksten over vertrouwde zaken die regelmatig voorkomen op het werk, op school en in de vrije tijd. Kan zich redden in de meeste situaties die kunnen optreden tijdens het reizen in gebieden waar de betreffende taal wordt gesproken. Kan een eenvoudige lopende tekst produceren over onderwerpen die vertrouwd of die van persoonlijk belang zijn. Kan een beschrijving geven van ervaringen en gebeurtenissen, dromen, verwachtingen en ambities en kan kort redenen en verklaringen geven voor meningen en plannen.
B2	Kan de hoofdgedachte van een ingewikkelde tekst begrijpen, zowel over concrete als over abstracte onderwerpen, met inbegrip van technische besprekingen in het eigen vakgebied. Kan zo vloeiend en spontaan reageren dat een normale uitwisseling met moedertaalsprekers mogelijk is zonder dat dit voor een van de partijen inspanning met zich meebrengt. Kan duidelijke, gedetailleerde tekst produceren over een breed scala van onderwerpen; kan een standpunt over een actuele kwestie uiteenzetten en daarbij ingaan op de voor- en nadelen van diverse opties.

Vaardig gebruiker	
C1	Kan een uitgebreid scala van veeleisende, lange teksten begrijpen en de impliciete betekenis herkennen. Kan zichzelf vloeiend en spontaan uitdrukken zonder daarvoor aantoonbaar naar uitdrukkingen te moeten zoeken. Kan flexibel en effectief met taal omgaan ten behoeve van sociale, academische en beroepsmatige doeleinden. Kan een duidelijke, goed gestructureerde en gedetailleerde tekst over complexe onderwerpen produceren en daarbij gebruikmaken van organisatorische structuren en verbindingswoorden.
C2	Kan vrijwel alles wat hij hoort of leest gemakkelijk begrijpen. Hij kan informatie die afkomstig is van verschillende gesproken en geschreven bronnen samenvatten, argumenten reconstrueren en hiervan samenhangend verslag doen. Hij kan zichzelf spontaan, vloeiend en precies uitdrukken en kan hierbij fijne nuances in betekenis, zelfs in complexere situaties onderscheiden.

Zelfevaluatie matrix Schrijven (Raamwerk NT2, pag. 12)

	A1	A2	B1	B2	C1	C2
Schrijven	Ik kan een korte, eenvoudige ansichtkaart schrijven, bijvoorbeeld voor het zenden van vakantiegroeten. Ik kan op formulieren persoonlijke details invullen, bijvoorbeeld mijn naam, nationaliteit en adres noteren op een hotel-inschrijvingsformulier	Ik kan korte, eenvoudige notities en boodschappen opschrijven. Ik kan een zeer eenvoudige persoonlijke brief schrijven, bijvoorbeeld om iemand voor iets te bedanken.	Ik kan eenvoudige samenhangende teksten schrijven over onderwerpen die vertrouwd of van persoonlijk belang zijn. Ik kan persoonlijke brieven schrijven waarin ik mijn ervaringen en indrukken beschrijf.	Ik kan een duidelijke, gedetailleerde tekst schrijven over een breed scala van onderwerpen die betrekking hebben op mijn interesses. Ik kan een opstel of verslag schrijven, informatie doorgeven of redenen aanvoeren ter ondersteuning vóór of tégen een specifiek standpunt. Ik kan brieven schrijven waarin ik het persoonlijk belang van gebeurtenissen en ervaringen aangeef.	Ik kan me in duidelijke, goed gestructureerde tekst uitdrukken en daarbij redelijk uitgebreid standpunten uiteenzetten. Ik kan in een brief, een opstel of een verslag schrijven over complexe onderwerpen en daarbij de voor mij belangrijke punten benadrukken. Ik kan schrijven in een stijl die is aangepast aan de lezer die ik in gedachten heb.	Ik kan een duidelijke en vloeiend lopende tekst in een gepaste stijl schrijven. Ik kan complexe brieven, verslagen of artikelen schrijven waarin ik een zaak weergeef in een doeltreffende, logische structuur, zodat de lezer de belangrijke punten kan opmerken en onthouden. Ik kan samenvattingen van en kritieken op professionele of literaire werken schrijven.

Vaardigheid: Schrijven

Niveau: B1

Beheersingsniveau:

Kan eenvoudige, samenhangende tekst schrijven over onderwerpen die vertrouwd of van persoonlijk belang zijn. Kan persoonlijke brieven schrijven waarin ervaringen en indrukken beschreven worden.

Subvaardigheden en Descriptoren	Voorbeelden bij de descriptoren	Tekstkenmerken productief
<p>1. Correspondentie Kan vrij gedetailleerde persoonlijke brieven schrijven over ervaringen, gevoelens en gebeurtenissen. Kan een korte, eenvoudige zakelijke brief schrijven.</p> <p>2. Aantekeningen, berichten, formulieren Kan notities/berichten schrijven waarin eenvoudige informatie van onmiddellijke relevantie voor vrienden, mensen van diensten, docenten en anderen die in zijn/haar dagelijks leven een rol spelen overgebracht wordt waarbij de belangrijke punten begrijpelijk overkomen.</p> <p>3. Verslagen en rapporten Kan heel korte rapporten schrijven volgens een vastgestelde standaard, waarin routinematige, feitelijke informatie doorgegeven wordt en redenen voor het ondernemen van bepaalde acties aangegeven worden.</p> <p>4. Vrij schrijven Kan eenvoudige, gedetailleerde beschrijvingen maken over een aantal bekende onderwerpen binnen het eigen interessegebied. Kan verslag doen van ervaringen, en daarbij gevoelens en reacties beschrijven in eenvoudige lopende tekst. Kan een gebeurtenis, een recent uitstapje - waar gebeurd of verzonnen - beschrijven.</p>	<p>De voorbeelden zijn opgenomen op de volgende pagina, uitgesplitst naar de domeinen werk, opleiding en dagelijks leven.</p>	<ul style="list-style-type: none">▪ onderwerp Het onderwerp is vertrouwd, valt binnen de persoonlijke belangstelling of heeft betrekking op het dagelijks leven, eigen werk of opleiding.▪ woordenschat en woordgebruik De woordenschat is toereikend om eventueel met behulp van omschrijvingen over vertrouwde onderwerpen, zoals familie, hobby's, werk, reizen en actuele gebeurtenissen te schrijven.▪ grammaticale correctheid Redelijk correct gebruik van frequente 'routines' en patronen die horen bij voorspelbare situaties.▪ spelling en interpunctie Lopende tekst die over het algemeen in het geheel begrijpelijk is. Spelling, interpunctie en lay-out zijn in de meeste gevallen accuraat genoeg om te volgen.▪ coherentie Een serie van kortere, eenvoudige afzonderlijke elementen is verbonden tot een samenhangende lineaire reeks van punten.

Bijlage 5 Korte impressie kandidaten met grote discrepantie tussen raamwerk- en STEX-beoordeling

SCHRIJVEN		
somscore Raamwerk	STEX-score	Opmerkingen n.a.v. nadere inspectie “rare gevallen”
12	488	De examenopdrachten zijn wisselend gemaakt; sommige erg goed andere minder en één slecht. Dit kan geleid hebben tot een lage STEX-score, terwijl de inferentie over het B1 niveau –gezien zijn globale karakter- toch en terecht leidt tot de beslissing: kandidaat is absoluut B1.
3	509	Inhoudelijk voldoet deze kandidaat wel aan de STEX-opdrachten. Deze kandidaat maakt echter veel grammaticale fouten. Het vocabulair laat te wensen over, en spelling en interpunctie zijn verre van correct. STEX-score lijkt te hoog.
6	517	Inhoudelijk voldoet deze kandidaat aan de STEX-opdrachten, hoewel de score wat aan de hoge kant lijkt te zijn. Woordgebruik en grammatica zijn aan de povere kant. De raamwerk beoordeling lijkt echter aan de lage kant. Zou ook als net B1 kunnen worden gezien. Hier lijkt sprake van een effect van enerzijds een wat geflatteerde STEX-score en anderzijds een strenge raamwerkbeoordeling door één van de beoordelaars.
SPREKEN		
somscore Raamwerk	STEX-score	Opmerkingen n.a.v. nadere inspectie “rare gevallen”
9	482	Geen enkele twijfel over niveau B1: STEX-score wellicht te laag? herbeoordeling STEX leidt tot vergelijkbare score: 487
9	484	Realiseert vereiste taalhandeling niet altijd, praat tamelijk binnensmonds; sommige opgaven heel goed, andere heel slecht
8	484	Veel fouten op detailniveau: vervoeging, soms onverstaanbaar. daardoor weinig STEX-punten, maar globaal wel B1
9	485	Zeker B1, spreekt heel vlot, soms niet goed verstaanbaar, maar o.h.a. goed
10	485	Praat onzeker en binnensmonds, komt niet overtuigend over, maar op zich wel goed w.b. grammatica en vocabulaire: zeker B1
9	486	Praat heel helder, grammaticaal niet sterk, vocabulaire schiet soms wat te kort, soms iets te onvolledig om STEX-punten te scoren, maar globaal wel goed : B1
10	487	Veel kleine foutjes (lidw, ww, inversie) die worden gemaskeerd door vlot spreken. Soms moeite om juiste woorden te vinden. Globaal zeker B1.
10	493	Heldere uitspraak, grammaticaal goed, bij deel 2 veel aarzelender dan in deel 1; daardoor waarschijnlijk. punten misgelopen op STEX
8	494	Beheerst globaal wel B1, maar voert niet altijd taalhandeling uit waardoor STEX-score laag uitvalt
8	494	Heel aarzelend (nervus), kost moeite om het goed te verstaan. Grammaticaal. en vocabulair goed. Wel B1.
10	495	Spreekt duidelijk en vlot. Wel B1. Terechte scores. Bij sommige opdrachten wat kort of onvolledig.
9	495	Grensgeval voor STEX, dus terechte score. Globaal wel B1, maar fouten in gram, inversie, moet soms woorden zoeken
10	496	Heel aarzelend, klinkt erg nerveus, moeite om iets langer te spreken (middellange opdrachten): wel B1
9	496	Stexscore waarschijnlijk onderschat, globaal wel B1
6	504	Gram. niet sterk, qua vocabulair ook niet, maar vlot. STEX grensgeval, maar van beoordelaars voordeel van de twijfel gekregen. Een lagere stexscore zou niet vreemd zijn geweest: er zijn kandidaten met lagere STEX-score in deze tabel die beter spreken!
5	517	STEX-score misschien iets te hoog, maar terecht geslaagd. Meer punten mogelijk voor B1. Bushalte-opdracht niet sterk, dan ineens weer heel verrassend correct: “als ik u was zou ik de Mondriaanschool kiezen want....”

Algemene opmerkingen m.b.t. de beoordelingen Spreken 1

Bij Spreken I is een duidelijk verschil constateren tussen de beheersingsgrens via de raamwerkbeoordeling en de cesuur op het Staatsexamen. Kandidaten die vlot spreken, kunnen een hoge raamwerkbeoordeling krijgen, terwijl ze slecht scoren als je ze beoordeelt met het STEX-model. Dit leidt ertoe dat – anders dan bij Programma II het geval was, - kandidaten met een score onder de STEX-cesuur wel als B1 worden beoordeeld.

© Cito Arnhem 2005. Auteursrechten voorbehouden.

Niets uit dit werk mag zonder voorafgaande schriftelijke toestemming van Cito worden openbaar gemaakt en/of verveelvoudigd door middel van druk, fotokopie/reprografie, scanning, computersoftware of andere elektronische verveelvoudiging of openbaarmaking, microfilm, geluidskopie, film- of videokopie of op welke wijze dan ook.