

Staatsexamen NT2. Het gewenste niveau.
Deel 2B Cesuurindicatie Spreken en Schrijven Programma II

Aanvullend onderzoek

December 2004

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	9
4.1	Globale beoordeling Spreken	9
4.1.1	Opzet beoordeling	9
4.1.2	Beoordelaarsovereenstemming	9
4.1.3	Relatie tussen de Raamwerkbeoordeling en de STEX-beoordeling	10
4.1.4	Samenvatting beoordelaarsovereenstemming en relatie tussen raamwerkbeoordeling en STEX-score Spreken	12
4.2	Relatie tussen STEX-scores en Raamwerkbeoordeling: de cesuurkwestie	13
4.2.1	Beheersing B2 volgens volledige overeenstemming: strengste norm	13
4.2.2	Beheersing B2 volgens het criterium meerderheidsoordeel	15
4.2.3	Samenvatting en aanbevelingen cesuur Spreken	19
5	De relatie tussen Staatsexamen Schrijven II en de raamwerkbeoordeling	21
5.1	Globale beoordeling Schrijven	21
5.1.1	Opzet beoordeling	21
5.1.2	Beoordelaarsovereenstemming	21
5.1.3	Relatie tussen de Raamwerkbeoordeling en de STEX-beoordeling	22
5.1.4	Samenvatting beoordelaarsovereenstemming en relatie tussen raamwerkbeoordeling en STEX-score Schrijven	25
5.2	Relatie tussen STEX-scores en Raamwerkbeoordeling: de cesuurkwestie	25
5.2.1	Beheersing B2 volgens volledige overeenstemming: Strengste norm	26
5.2.2	Beheersing B2 volgens het criterium meerderheidsoordeel	27
5.2.3	Samenvatting en aanbevelingen Schrijven	31

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 programma II, als zij starten met een vervolgopleiding op hbo- of universitair niveau, dan wel instromen in een midden- of hogere kaderfunctie 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 wordt in eerste instantie uitgevoerd voor programma II van het Staatsexamen; programma I volgt daarna.

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 II, de universiteiten en hogescholen. 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 II. Over de resultaten van deze behoeftepeiling is gepubliceerd in Kuijper e.a., augustus 2004).

Het tweede deel van de evaluatie bestond voor Spreken en Schrijven uit een onderzoek dat gericht was op de vraag bij welke scores van het Staatsexamen men kan aannemen dat een kandidaat op het gewenste niveau Schrijven of Spreken functioneert. Dit deel van het onderzoek bestond uit twee fasen. Over de eerste fase is gerapporteerd in Kuijper e.a., oktober 2004. In deze fase van het onderzoek zijn 20 kandidaten beoordeeld met zowel een analytische als globale beoordelingslijst, gebaseerd op het Raamwerk NT2 (Dalderop en Teunisse, 2002). Er is onderzocht of er een relatie tussen de Staatsexamenscores en het Raamwerk NT2 bestaat en of deze relatie voldoende was om uitspraken over een eventuele herziening van de cesuur te doen. Op grond van het rapport is in overleg met de Staatsexamencommissie besloten het onderzoek uit te breiden naar 60 kandidaten met het doel een meer gefundeerde beslissing te kunnen nemen dan mogelijk was op grond van de eerste analyses op 20 kandidaten.

In het voorliggende onderzoeksrapport wordt verslag gedaan van dit aanvullende onderzoek. Dit rapport maakt deel uit van een serie: in deel 1 wordt verslag gedaan van de behoeftepeiling die onder afnemers van het Staatsexamen heeft plaatsgevonden. In oktober 2004 is het eerste verslag van het onderzoek naar de cesuurindicatie gepubliceerd (Kuijper e.a., oktober 2004).

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.

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 vorig jaar 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 –niet alleen in 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 toetscores 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 B2. Analoot aan het ETS- en PhonePass-onderzoek werken wij ook met globale of holistische uitspraken. Daarnaast is in de eerste fase van het cesuuronderzoek tevens een analytische beoordeling toegepast, om een crosscheck uit te voeren op de gevonden resultaten bij de globale beoordeling. In dit aanvullende onderzoek is op grond van de bevindingen in fase 1 alleen gewerkt met een globale of holistische beoordeling.

3 Opzet van het cesuuronderzoek

3.1 Het uitgangspunt: het gewenste niveau

Het onderzoek naar het door afnemers gewenste niveau van het Staatsexamen programma II en naar indicaties voor de consequenties van dit gewenste niveau voor de cesuren van de Spreek- en Schrijfexamens programma 2 bestaat uit 4 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, augustus 2004. Op grond van de resultaten van het behoefteonderzoek blijkt dat het gewenste niveau voor het Staatsexamen II Spreken en Schrijven niveau B2 is. In Stap 2 en 3 werden allereerst de schrijf- en spreekuitingen van 20 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 B2) 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 in het Raamwerk NT2. De constructie van dit raamwerkbeoordelingsmodel wordt beschreven in deze paragraaf.
2. Het beoordelen van de spreek- en schrijfuitingen van kandidaten met deze Raamwerkbeoordelingslijst. De opzet van deze nieuwe beoordeling van spreek- en schrijfuitingen wordt beschreven in paragraaf 3.2, 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 II. De bespreking van de wijze van vergelijkingsanalyses en de resultaten ervan worden gepresenteerd in hoofdstuk 4 en 5.

In Kuijper e.a., oktober 2004 is verslag gedaan van de eerste fase van dit deel van het onderzoek. Op grond van deze voorlopige resultaten is besloten een aanvullend onderzoek te doen waarin de beoordelingsprocedure is herhaald op een groep van 40 andere kandidaten om zo op basis van steekproeven van 60 kandidaten per vaardigheid te komen tot een meer gefundeerd advies over de hoogte van de cesuur.

3.2 Samenstellen van de raamwerkbeoordelingslijsten

Bij de constructie van het raamwerkbeoordelingsmodel hebben we ons gebaseerd op de relevante beschrijvingen uit het Raamwerk NT2. Er zijn voor Schrijven en Spreken zowel analytische als globale beoordelingslijsten ontwikkeld.

Globale beoordeling

In de globale beoordelingslijsten zijn de algemene beheersingsuitspraken uit het Raamwerk NT2 (Dalderop, Teunisse, 2002) opgenomen van Schrijven (Raamwerk NT2, pag.91) en voor Spreken de algemene beheersingsuitspraken voor Spreken (Raamwerk NT2, pag.75) en Gesprekken voeren (Raamwerk NT2, pag. 58).

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 B2 beheerst of niet. De globale beoordelingslijsten voor Spreken en Schrijven zijn te vinden in bijlage 1 en 2.

Analytische beoordeling

Behalve een globale beoordelingslijst is er in de eerste fase tevens een analytische beoordelingslijst ontwikkeld om zonedig aanvullende informatie te vergaren op de resultaten van de globale beoordeling. Deze analytische beoordelingsprocedure bestond uit een vaste set van beoordelingsuitspraken die letterlijk ontleend zijn aan de 'tekstkenmerken productief' in het Raamwerk NT2 voor Gesprekken Voeren en Spreken en voor Schrijven.

Met deze set beoordelingsaspecten werd elke kandidaat beoordeeld door één beoordelaar. De analytische beoordelingslijst voor Spreken, gesprekken voeren staat in Bijlage 3. De analytische beoordelingslijst Schrijven is te vinden in bijlage 4.

In dit aanvullende onderzoek is besloten alleen de globale beoordelingssystematiek te hanteren, omdat beide beoordelingssystematieken tot vergelijkbare resultaten leidden en de globale beoordeling veel minder arbeidsintensief is dan de analytische beoordeling.

3.3 De selectie van de kandidaten

Voor de beoordeling van het werk van de STEX-II kandidaten zijn in de eerste fase van het cesuuronderzoek 20 kandidaten voor Spreken en Schrijven geselecteerd.

Voor Schrijven II is het examen van juli 2003 gebruikt. De kandidaten zijn geselecteerd uit het scoregebied tussen 20 en 39 ruwe scorepunten. De cesuur van dit examen lag bij 27 ruwe scorepunten. De geselecteerde kandidaten hadden een schaalscore variërend van 460 tot 561.

Voor Spreken II is het examen van oktober 2003 gebruikt. De kandidaten zijn geselecteerd uit het scoregebied tussen 55 en 74 ruwe scorepunten. De cesuur lag bij 59 ruwe scorepunten. De geselecteerde kandidaten hadden een schaalscore variërend van 479 tot 549.

Voor beide programma's geldt dat meer dan 50% van de kandidaten in dit scoregebied zaten. Er is bewust gekozen voor een grotere selectie boven de huidige cesuur dan eronder, vanwege de verwachting dat het gewenste niveau van de examens hoger zou moeten zijn dan het huidige niveau.

In het aanvullende onderzoek is voor beide vaardigheden een nadere selectie van 40 extra kandidaten gemaakt volgens de hierboven beschreven selectiecriteria. Dit leidde tot een analysegroep van 60 kandidaten voor Spreken II en eveneens 60 kandidaten voor Schrijven II. Voor deze aanvullende selectie is voor zowel Schrijven II als Spreken II gebruik gemaakt van het examen van juli 2004. Er kon geen gebruik meer gemaakt worden van de examens uit de eerste fase, omdat deze inmiddels vernietigd waren. Doordat de verschillende versies van het Staatsexamen vergelijkbaar zijn levert dit geen bezwaren op.

De schaalscores van de uiteindelijk geselecteerde 60 kandidaten Spreken varieerden van 445 tot 568. Voor Schrijven lagen de scores tussen 451 en 561.

3.4 De beoordelingsprocedure in het aanvullende onderzoek

In het eerste deel van het cesuuronderzoek is voor zowel Schrijven als Spreken iedere kandidaat globaal beoordeeld door twee onafhankelijke beoordelaars. Daarnaast is elke kandidaat door een derde beoordelaar beoordeeld met het analytische beoordelingsmodel.

De beoordelaars waren Cito-medewerkers die bekend waren met het Raamwerk NT2 en het Common European Framework of Reference.

Voor Spreken zijn van 20 kandidaten de 7 middellange en de lange opdracht globaal beoordeeld door twee onafhankelijke beoordelaars. Iedere kandidaat is tevens door een derde beoordelaar beoordeeld met een analytisch beoordelingsmodel.

Voor Schrijven is een zelfde procedure toegepast. De beoordeling had betrekking op de twee middellange en de lange opdracht in het examen. De resultaten van dit deel van het onderzoek zijn gepubliceerd in Kuijper e.a., oktober 2004.

Het aanvullende onderzoek

In dit aanvullende onderzoek is alleen de globale beoordeling toegepast op 40 extra kandidaten per vaardigheid..

De beoordelaarsgroep is uitgebreid met medewerkers van bureau ICE.

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 II 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-II 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 B2 Spreken en gesprekken Voeren van het Raamwerk NT2 (respectievelijk pagina 75 en 58). 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 B2. Met deze globale oordelen werden 60 kandidaten beoordeeld door twee onafhankelijke beoordelaars.

Tabel 4.1-1 Globale raamwerkbeoordeling

	Uitspraken Raamwerk NT2	Ja	Nee
S-G1	Kan duidelijke, gedetailleerde beschrijvingen presenteren over een breed scala van onderwerpen die betrekking hebben op mijn interessegebied.		
S-G2	Kan een standpunt over een actueel onderwerp verklaren en de voordelen en nadelen van diverse opties uiteenzetten.		
GV-G3	Kan zodanig deelnemen aan een vloeiend gesprek, dat normale uitwisseling met moedertaalsprekers redelijk mogelijk is.		
GV-G4	Kan binnen een vertrouwde context actief deelnemen aan een discussie en hierin standpunten uitleggen en ondersteunen.		
SGV-G5	Beheerst deze kandidaat over het algemeen niveau B2?		

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 12 gevallen uitgevoerd. Één kandidaat is na de beoordeling uit de analyse verwijderd. Deze kandidaat was inhoudelijk adequaat, maar had stotterproblemen, waardoor de raamwerkbeoordeling ten onrechte geen B2 kon zijn.

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	81%	.63	0.10
S-G2	75%	.47	0.12
GV-G3	78%	.38	0.15
GV-G4	75%	.47	0.11
SGV-G5	88%	.76	0.08

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 moderate tot substantial). 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 .84.

De beoordelaarsovereenstemming voor Spreken blijkt aanzienlijk hoger te zijn dan in de eerste fase van het cesuuronderzoek (Kuijper e.a., oktober 2004). De meest waarschijnlijke verklaring hiervoor is het feit dat er in deze tweede fase een derde beoordelaar is ingeschakeld in gevallen waarbij de beoordelaars 3 of meer punten van elkaar verschilden.

De beoordelaarsovereenstemming is voldoende hoog om de relatie tussen de raamwerkbeoordeling de Staatsexamenscore nader te onderzoeken.

4.1.3 Relatie tussen de Raamwerkbeoordeling en de STEX-beoordeling

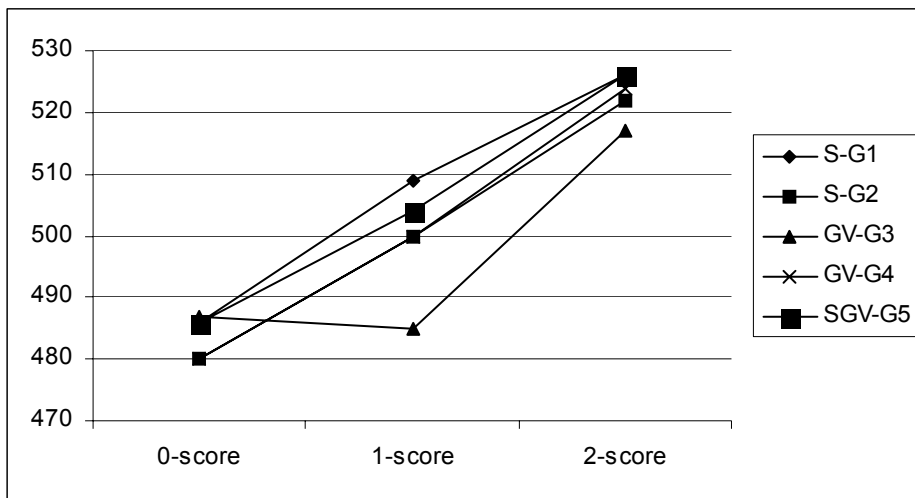
Elke kandidaat is globaal beoordeeld door twee onafhankelijke beoordelaars. De beoordelaars konden voor elke uitspraak aangeven of een kandidaat aan de eisen die in de beoordelingsuitspraak zijn verwoord voldeed of niet.

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 B2 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 B2 door 2 beoordelaars en de score op Staatsexamen II Spreken



De verwachte relatie wordt voor elk van de 5 globale Raamwerkuitspraken gevonden. Alleen voor de uitspraak GV-G3 blijkt er nauwelijks verschil te zijn tussen de 0- en 1-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 en de STEX-schaalscores Spreken (n=59)

Uitspraak	Pearson productmoment correlatie
S-G1	0,70
S-G2	0,68
GV-G3	0,51
GV-G4	0,71
SGV-G5	0,73
Totaalscore	0,79

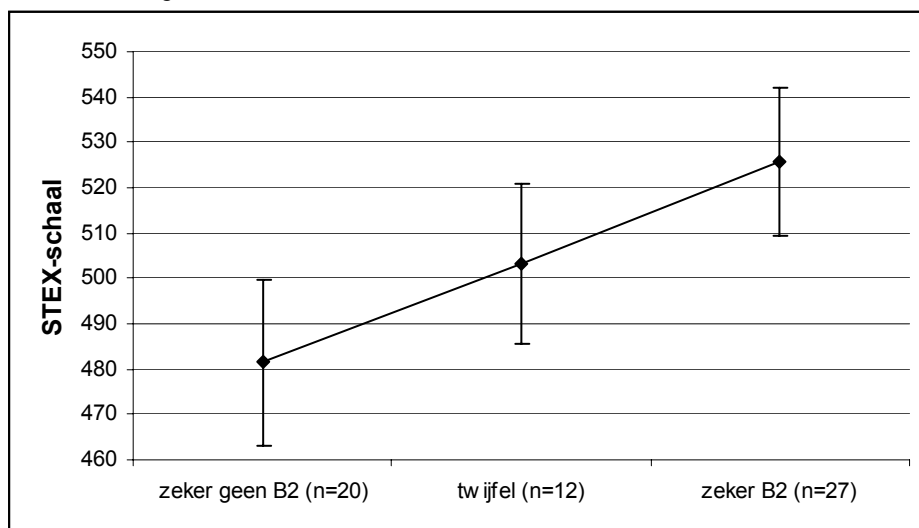
De correlaties tussen raamwerkuitspraken en de STEX-scores zijn zeer bevredigend, zij het met voorbehoud voor uitspraak 3, die een wat matiger positief verband heeft.

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 10 (beide beoordelaars geven voor elke uitspraak 1 punt). In tabel 4.1-4 zijn de kandidaten in 3 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 B2 kandidaten zijn. De tweede groep bestaat uit kandidaten die in minimaal 8 van de 10 beoordelingen als B2 zijn aangemerkt. Deze groep kan gezien worden als bestaande uit kandidaten die B2 met grote zekerheid beheersen. De tusseliggende 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 3 groepen berekend. Tabel 4.1-4 en figuur 4.1-2 geven de gemiddelde scores en de standaarddeviaties weer van de STEX-scores van deze drie groepen.

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

Beheerst B2 volgens globale beoordeling	STEX-score		
	N	gemiddeld	SD
Zeker niet	20	481,5	18,2
twijfel	12	503,3	17,5
Zeker wel	27	525,7	16,5
Alle kandidaten	59	506,2	26,0

Figuur 4.1-2 Gemiddelde STEX-score en standaarddeviaties van kandidaten op grond van hun raamwerkbeoordeling.



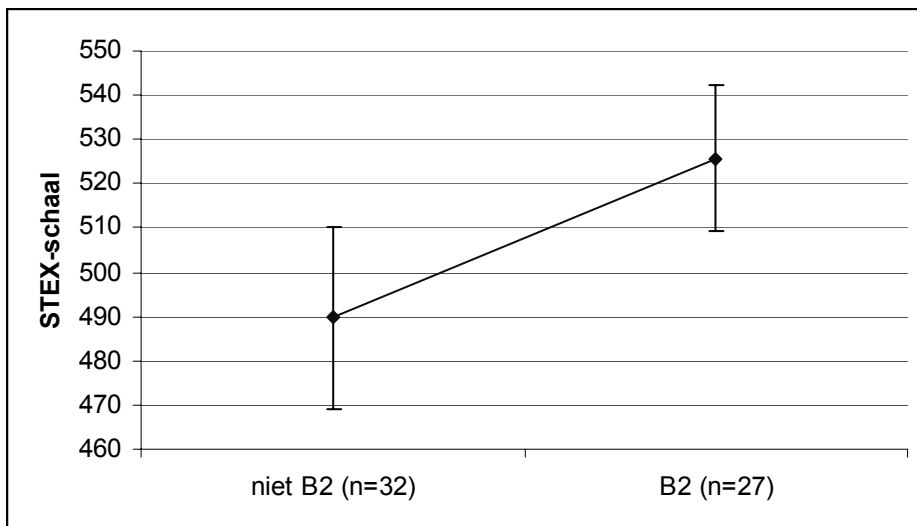
Het verband tussen de Raamwerkbeoordeling en de STEX-score is duidelijk en in de verwachte richting.

Wanneer we de twijfelgroep en de niet-B2-groep samenvoegen kunnen we de kandidaten in twee groepen indelen. De eerste groep is de groep die geacht wordt niveau B2 zeker te beheersen. Deze groep is dezelfde als de 'zeker B2-groep' in tabel 4.1-4 en figuur 4.1-2 (een totaalscore van 8-10). De andere groep is een samenvoeging van kandidaten die volgens de beoordelaars B2 niet beheersen of waarover twijfel bestaat. De gemiddelden en standaarddeviaties van deze twee groepen worden gepresenteerd in tabel 4.1-5 en figuur 4.1-3.

Tabel 4.1-5 Gemiddelde STEX-score Spreken II en standaarddeviaties van de kandidaten die B2 wel of niet beheersen volgens 80% criterium

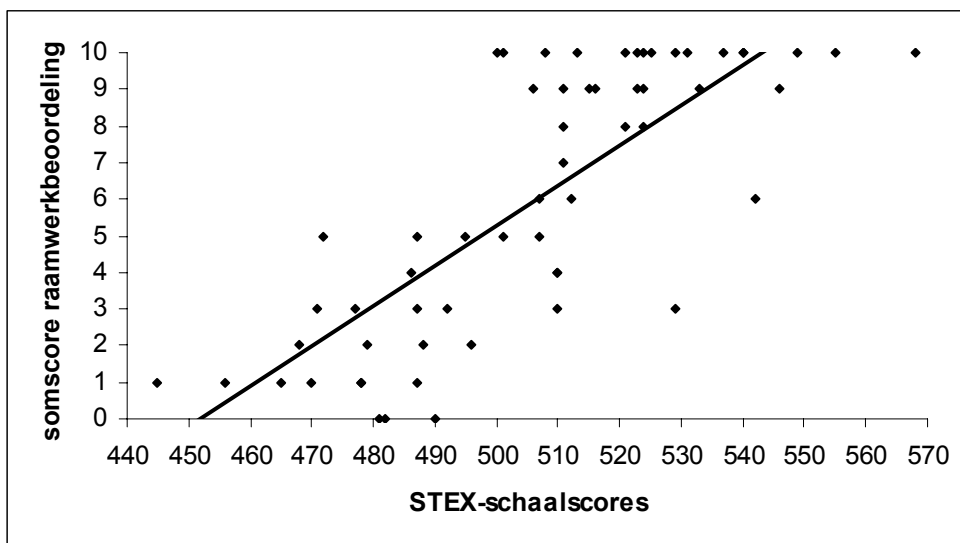
Beheerst B2 volgens globale beoordeling	STEX-score		
	N	gemiddeld	SD
Nee	32	489,7	20,7
Ja	27	525,7	16,5
Alle kandidaten	59	506,2	26,0

Figuur 4.1-3 Gemiddelde STEX-score Spreken II en standaarddeviaties van de kandidaten die B2 wel of niet beheersen volgens 75% criterium



Volgens deze tweedeling van de kandidaten blijkt de niet-B2-groep gemiddeld 36 punten lager te scoren op het Staatsexamen dan de groep die wel op B2 functioneert. Er is dus een duidelijk verschil tussen beide groepen wat betreft hun STEX-score. Er is dus een substantiële samenhang tussen de scores van beide beoordelingsbenaderingen. Figuur 4.1-4 laat geeft een grafische weergave van deze relatie. De correlatie (Pearson product-moment) tussen de raamwerkbeoordeling en de STEX-schaalscore bedraagt 0,79.

Figuur 4.1-4 Samenhang tussen de scores van kandidaten op het Staatsexamen en op de globale raamwerkbeoordeling Spreken



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

1. In dit aanvullende onderzoek zijn naast de 20 kandidaten uit de eerste fase van het cesuuronderzoek 40 extra STEX-kandidaten beoordeeld door twee onafhankelijke beoordelaars. Het totaal aantal respondenten in het onderzoek bedraagt daarmee 60 kandidaten. De uitingen van al deze 60 kandidaten zijn 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. Één kandidaat is uit het bestand verwijderd.

- De beoordelaarsovereenstemming in de gevolgde procedure voor de globale raamwerkbeoordeling Spreken is goed en voldoende voor verdere analyses met betrekking tot de cesuur kwestie. Het inschakelen van een derde beoordeling heeft aan deze goede overeenstemming bijgedragen. Kandidaten die volgens het globale oordeel functioneren op B2 hebben gemiddeld een hogere STEX-score dan kandidaten die B2 nog niet beheersen. Dit gegeven wordt ondersteund door het feit dat de correlatie tussen STEX-score en somscore op de raamwerkbeoordeling met .79 substantieel is.

In de volgende paragraaf wordt ingegaan op de hoogte van de cesuur van het Staatsexamen Spreken II in relatie tot het beheersingsniveau B2.

4.2 Relatie tussen STEX-scores en Raamwerkbeoordeling: de cesuurkwestie

Uit paragraaf 4.1 blijkt dat er een bevredigende relatie gevonden is tussen de STEX-scores en de globale beoordelingen volgens het Raamwerk. Verder blijkt dat de gemiddelde STEX-score van de B2 kandidaten hoger ligt dan de van de groep kandidaten die B2 niet beheersen. De volgende vraag die nu beantwoord moet worden is of deze gegevens aanleiding geven voor een mogelijke herziening van de cesuur van de Staatsexamen Spreken II.

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 Spreken II geeft aan dat er tot zekere hoogte 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 B2 spelen de volgende aspecten een rol:

- de definiëring van de beheersing van B2
- 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. Voor twee verschillende definities van de beheersing van B2 worden de consequenties van verschillende cesuurscores voor het slagen en zakken van kandidaten beschreven. De eerste definitie is een strenge, bij de tweede definitie worden minder hoge eisen gesteld aan de overeenstemming van beoordelaars.

4.2.1 Beheersing B2 volgens volledige overeenstemming: strengste norm

Tabel 4.2-1 geeft een totaaloverzicht van de globale raamwerkscores van alle 59 kandidaten. In kolom 1 staan de scores op het Staatsexamen Spreken II. Kolom 2 geeft de totaalscore over de 5 beoordelingsuitspraken. Uit de laatste kolom kan worden afgelezen of een kandidaat niveau B2 beheerst volgens het criterium volledige overeenstemming. De gearceerde cellen geven aan dat een kandidaat geslaagd is voor het Staatsexamen Spreken II en/of B2 beheerst volgens het criterium volledige overeenstemming.

Tabel 4.2-1 Beoordeling kandidaten op Staatsexamen Spreken II en de raamwerkbeoordeling met als criterium volledige overeenstemming van de beoordelaars.

STEX-score	Totaalscore raamwerk	Beheerst B2 met volledige overeenstemming
568	10	B2
555	10	B2
549	10	B2
546	9	
542	6	

STEX-score	Totaalscore raamwerk	Beheerst B2 met volledige overeenstemming
540	10	B2
540	10	B2
537	10	B2
533	9	
531	10	B2
529	10	B2
529	3	
525	10	B2
524	9	
524	8	
524	10	B2
523	9	
523	10	B2
521	10	B2
521	8	
516	9	
515	9	
513	10	B2
512	6	
511	7	
511	9	
511	8	
510	4	
510	3	
510	4	
508	10	B2
507	6	
507	5	
506	9	
501	5	
501	10	B2
500	10	B2
496	2	
495	5	
492	3	
490	0	
488	2	
487	1	
487	5	
487	3	
486	4	
482	0	
481	0	
479	2	
478	1	
478	1	
477	3	
472	5	
471	3	
470	1	
468	2	
465	1	
456	1	
445	1	

	Geslaagd voor STEX / beheerst niveau B2
	Gezakt voor STEX / beheerst niveau B2 niet

Zestien kandidaten beheersen B2 volgens het strenge criterium van volledige overeenstemming. Van kandidaten die gezakt zijn voor het Staatsexamen beheerst er niemand B2 volgens de hier gehanteerde norm van volledige overeenstemming. Er geen duidelijke overgang in de STEX-score te vinden waarbij kandidaten een zeer grote kans hebben volgens de raamwerkbeoordeling als B2 te worden gezien

Er kunnen volgens dit strenge criterium van volledige overeenstemming geen duidelijke conclusies getrokken worden over een eventuele herziening van de cesuur.

4.2.2 Beheersing B2 volgens het criterium meerderheidsoordeel

In paragraaf 4.2.1 hebben we een zeer strenge norm gehanteerd om te bepalen of iemand B2 beheerst. Het is de vraag of de eis dat beide beoordelaars het op alle criteria eens moeten zijn over de beheersing van B2 niet een te zware is. Uit figuur 4.1-1 blijkt voor vier van de vijf beheersingsuitspraken dat kandidaten over wie de beoordelaars het niet eens zijn een STEX-score hebben die ligt tussen die van de kandidaten die volgens beide beoordelaars B2 beheersen en van de kandidaten die B2 niet beheersen. We zouden een verschil van mening tussen beoordelaars, behalve als een gebrek aan overeenstemming, wellicht ook kunnen beschouwen als een uitvloeisel van het feit dat een kandidaat net op de grens van B2 verkeert, waardoor de ene beoordelaar zo'n kandidaat als B2 en de ander als niet B2 beoordeelt. Een score 1 zou in deze redenering een beheersing vertegenwoordigen die tussen 0 en 2 ligt.

Op grond van deze overwegingen kunnen we het 100% beheersingscriterium vervangen door een meerderheidsoordeel. We beschouwen een kandidaat als B2 indien deze in 80% van de beoordelingen als B2 beoordeeld wordt. Dit valt samen met een minimale score van 8 van de 10 te behalen punten. Dat betekent dat in 8 van de 10 beoordelingen een beoordelaar uitspreekt dat iemand op B2 functioneert. De resultaten staan in tabel 4.2-2.

Tabel 4.2.2 Beoordeling kandidaten op Staatsexamen en globale raamwerkbeoordeling met als criterium meerderheidsoordeel van de beoordelaars (80%).

STEX-score	Totaalscore raamwerk	Beheerst B2 met meerderheidsoordeel
568	10	B2
555	10	B2
549	10	B2
546	9	B2
542	6	
540	10	B2
540	10	B2
537	10	B2
533	9	B2
531	10	B2
529	10	B2
529	3	
525	10	B2
524	9	B2
524	8	B2
524	10	B2
523	9	B2
523	10	B2
521	10	B2
521	8	B2
516	9	B2
515	9	B2
513	10	B2
512	6	
511	7	
511	9	B2
511	8	B2
510	4	
510	3	
510	4	
508	10	B2
507	6	
507	5	
506	9	B2
501	5	
501	10	B2
500	10	B2
496	2	
495	5	
492	3	
490	0	

STEX-score	Totaalscore raamwerk	Beheerst B2 met meerderheidsoordeel
488	2	
487	1	
487	5	
487	3	
486	4	
482	0	
481	0	
479	2	
478	1	
478	1	
477	3	
472	5	
471	3	
470	1	
468	2	
465	1	
456	1	
445	1	

	Geslaagd voor STEX / beheerst niveau B2
	Gezakt voor STEX / beheerst niveau B2 niet

Uit tabel 4.2-2 blijkt dat het aantal kandidaten dat geacht wordt B2 te beheersen is toegenomen van 16 tot 27. Van de 22 kandidaten die gezakt waren voor het Staatsexamen beheerst niemand B2. Van de 23 kandidaten met een STEX-score 513 of hoger worden er 21 als B2 beoordeeld. Van de 36 kandidaten met een STEX-score onder 513 worden er slechts 6 als B2 beoordeeld. 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.

In het interval tussen 500 en 513 bestaat de geringste samenhang tussen de STEX-score. Van de 14 kandidaten in dit scorebereik worden er 6 beoordeeld als B2 en 8 als niet-B2. Een eventuele nieuwe cesuur zal in dit interval liggen.

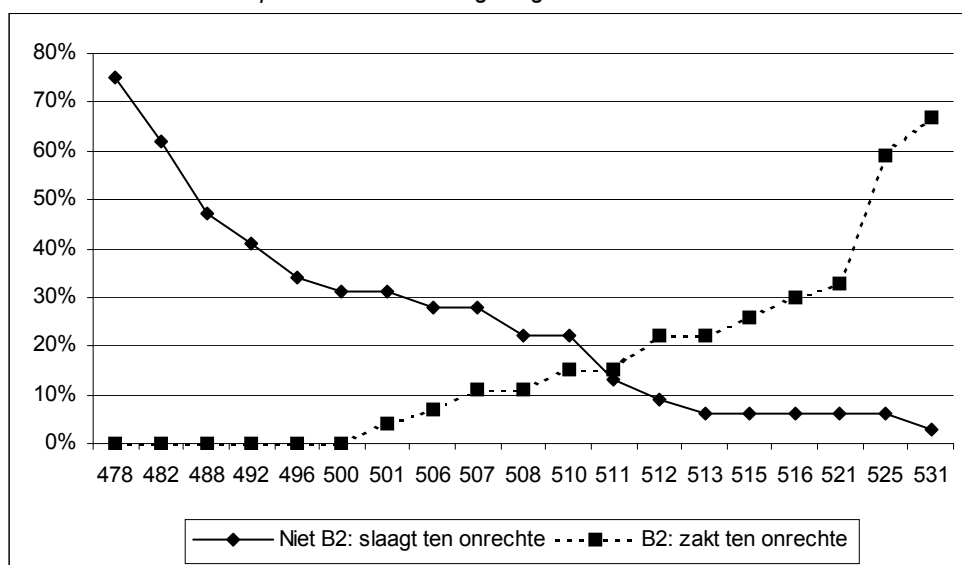
Voor het vaststellen van de cesuur is echter ook informatie nodig over de proportie kandidaten die bij een gegeven cesuurscore onterecht zakt of onterecht slaagt. Deze twee consequenties moeten tegen elkaar worden afgewogen. Tabel 4.2-3 geeft een overzicht van de gevolgen van de beslissingen voor de cesuurscores van 500 tot en met 516.

Tabel 4.2-3 Percentages gezakte en geslaagde kandidaten die B2 al of niet beheersen bij verschillende cesuren van het Staatsexamen Spreken II. Beheersing B2 gedefinieerd als meerderheidsoordeel: 80% van maximale score

Cesuur op STEX	Beheerst B2 niet (n=32)		Beheerst B2 (n=27)	
	Zakt terecht	Slaagt ten onrechte	Zakt ten onrechte	Slaagt terecht
531	97%	3%	67%	33%
525	94%	6%	59%	41%
521	94%	6%	33%	67%
516	94%	6%	30%	70%
515	94%	6%	26%	74%
513	94%	6%	22%	78%
512	81%	9%	22%	78%
511	87%	13%	15%	85%
510	78%	22%	15%	85%
508	78%	22%	11%	89%
507	72%	28%	11%	89%
506	72%	28%	7%	93%
501	69%	31%	4%	96%
500	69%	31%	0%	100%
496	66%	34%	0%	100%
492	59%	41%	0%	100%
488	53%	47%	0%	100%
482	38%	62%	0%	100%
478	25%	75%	0%	100%

In figuur 4.2-1 wordt deze tabel grafisch weergegeven.

Figuur 4.2-1 Percentages gezakte en geslaagde kandidaten die B2 al of niet beheersen bij verschillende cesuren van het Staatsexamen Spreken II. Beheersing B2 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 B2 wel beheersen.
 2. Fout positief. De fout dat men ten onrechte kandidaten laat slagen die B2 nog niet beheersen.
- Zo leidt een cesuurscore van 500 ertoe dat de kans dat een B2-kandidaat ten onrechte zakt nihil is. Daar staat tegenover dat van de kandidaten die B2 nog niet beheersen 31% zal slagen. Bij een cesuurscore van 516 maakt men de kans op het laten slagen van niet B2-kandidaten klein, maar verhoogt men de kans op het onterecht zakken tot 30%.

Welk cesuurscore men 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 516

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 4.2-4 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 minder 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 516 slaagt 6% ten onrechte en zakt 30% ten onrechte. De optimumscore wordt: $6 + 30 = 36$. Hoe kleiner dit getal hoe minder kans op een foute beslissing.

2. Conditie 2

Men wil voorkomen dat te veel B2-kandidaten ten onrechte zakken. Daarom geeft men ten onrechte zakken een dubbel gewicht.

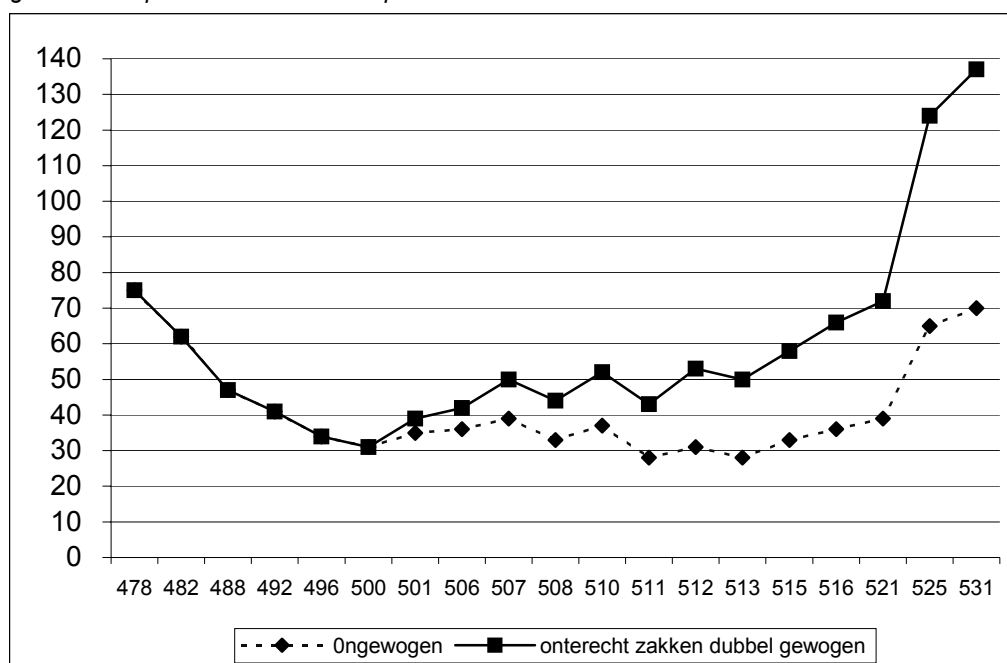
Voorbeeld: score bij score 515 slaagt 6% ten onrechte en zakt 30% ten onrechte. De optimumscore als we het onterecht zakken dubbel wegen, wordt nu: $6 + 60 = 66$. Ook hier geldt dat de laagste optimumscore de geringste kans geeft op een foute beslissing.

Figuur 4.2-2 is een grafische weergave van deze tabel. De optimale cesuurscore voor beide condities is vet gedrukt.

Tabel 4.2-4 Optimum cesuurscore Spreken II onder 2 condities

Ceuurscore	Ongewogen	Onterecht zakken telt dubbel
478	75	75
482	62	62
488	47	47
492	41	41
496	34	34
500	31	31
501	35	39
506	35	42
507	39	50
508	33	44
510	37	52
511	28	43
512	31	53
513	28	50
515	32	58
516	36	66
521	39	72
525	56	106

figuur 4.2-2 Optimum cesuurscore Spreken II onder 2 condities



De optimale cesuurscore in de ongewogen variant ligt op 511 of 513.

Wanneer men meer belang hecht aan het voorkomen van het ten onrechte zakken van B2 kandidaten komt een cesuur van 500 het meest in aanmerking. De percentages geslaagde en gezakte kandidaten bij deze scores staan in tabel 4.2-4. Gezien de functie van het Staatsexamen ligt het handhaven van de cesuur op 500 het meest voor de hand.

4.2.3 Samenvatting en aanbevelingen cesuur Spreken

Volgens de behoeftepeiling zou het Staatsexamen Spreken II een vaardigheid moeten waarborgen die op niveau B2 van het Raamwerk NT2 ligt (Kuijper e.a., augustus 2004).

1. Uit het beoordelingsonderzoek blijkt dat er een substantiële positieve relatie bestaat tussen de scores die kandidaten halen op het Staatsexamen Spreken II en de vraag of de respons van een kandidaat beoordeeld wordt als B2. In zijn algemeenheid geldt dat naarmate de STEX-score hoger wordt de kans dat een kandidaat als B2 wordt beoordeeld groter wordt.
2. De gevonden beoordelaarsovereenstemming bij de globale raamwerkbeoordeling Spreken is voldoende om er uitspraken over een cesuurscore op te baseren.
3. De methodiek die in dit onderzoek gehanteerd is – globale beoordeling van de schrijfproducten van kandidaten – kan in zijn algemeenheid als bevredigend worden gezien.
 - a. De globale en analytische beoordelingsmethode leiden grosso modo tot vergelijkbare resultaten (Kuijper e.a, augustus 2004). Gezien de grote tijdsinvestering die nodig is voor de analytische methode ligt het gebruik van een globale beoordeling voor het doel van dit onderzoek voor de hand.
 - b. De beoordelaarsovereenstemming van de globale raamwerkbeoordeling is goed. Het inschakelen van een derde beoordeling heeft tot een hogere overeenstemming geleid dan in de eerste fase van dit onderzoek.
 - c. De relatie tussen raamwerkbeoordeling en de STEX-score is substantieel.
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 II een onderscheid moet maken tussen kandidaten die een spreekvaardigheid hebben die vergelijkbaar is met niveau B2, dan zal deze cesuur liggen in het interval tussen de schaalscores 500-513. Bij het vaststellen van de cesuur 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. Indien men dit als uitgangspunt neemt ligt een cesuur van 500 voor de hand. Hierbij kan in aanmerking te worden genomen dat 40% van de bij deze score geslaagde niet-B2-kandidaten een score heeft van minstens 60% van de maximale raamwerksscore en dus een vaardigheid niet ver onder B2 bezitten. Indien men het voorkomen van het ten onrechte zakken van kandidaten minder belangrijk vindt, is 511 de optimale cesuurscore.

5 De relatie tussen de score op het Staatsexamen Schrijven II 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 vijf uitspraken over de schrijfvaardigheid van de cursisten beantwoorden met ja of nee. Hun oordelen zijn gebaseerd op de productenvan het STEX-II schrijfexamen van de kandidaten. De globale raamwerkbeoordelingslijst is opgenomen in bijlage 2.

De eerste vier uitspraken zijn letterlijk ontleend aan de omschrijvingen van het beheersingsniveau B2 Schrijven van het Raamwerk NT2 (pagina 91).

De laatste uitspraak is de meest algemene beoordeling: een algemene uitspraak in hoeverre een kandidaat volgens de beoordelaar voldoet aan de eisen voor Schrijven B2.

Met deze globale oordelen werden 60 kandidaten beoordeeld door twee onafhankelijke beoordelaars.

Tabel 5.1-1 Globale raamwerkbeoordelingsuitspraken Schrijven B2

	Uitspraken Raamwerk NT2	Ja	Nee
SC-G1	Kan een duidelijke, gedetailleerde tekst schrijven.		
SC-G2	Kan een opstel of verslag schrijven.		
SC-G3	Kan informatie doorgeven of redenen aanvoeren ter ondersteuning vóór of tégen een specifiek standpunt.		
SC-G4	Kan brieven schrijven waarin het persoonlijk belang van ervaringen en indrukken aangegeven wordt.		
SC-G5	Beheerst deze kandidaat over het algemeen niveau B2?		

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.

Vier van de 60 kandidaten bleken geen volledig examen gemaakt te hebben. Dit heeft hun STEX-score nadelig beïnvloed. Daarom zijn deze kandidaten van de analyse uitgesloten.

5.1.2 Beoordelaarsovereenstemming

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

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

Uitspraak	Percentage overeenstemming	Kappa	Standaard fout kappa
SC-G1	0,74	0,47	0,12
SC-G2	0,70	0,39	0,12
SC-G3	0,73	0,36	0,14
SC-G4	0,68	0,19	0,15
SC-G5	0,90	0,79	0,08

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, met alleen voorbehoud voor SC-G4). Opvallend is dat de meest algemene uitspraak (SC-G5) tot een zeer bevredigende overeenstemming leidde. De correlatie tussen de somscores van eerste en tweede beoordelaars bedraagt .75. De beoordelaarsovereenstemming voor Schrijven blijkt aanzienlijk hoger te zijn dan in de eerste fase van het cesuuronderzoek (Kuijper e.a., oktober 2004). De meest waarschijnlijke verklaring hiervoor is het feit dat er in deze tweede fase een derde

beoordelaar is ingeschakeld in gevallen waarbij de beoordelaars 3 of meer punten van elkaar verschilden.

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

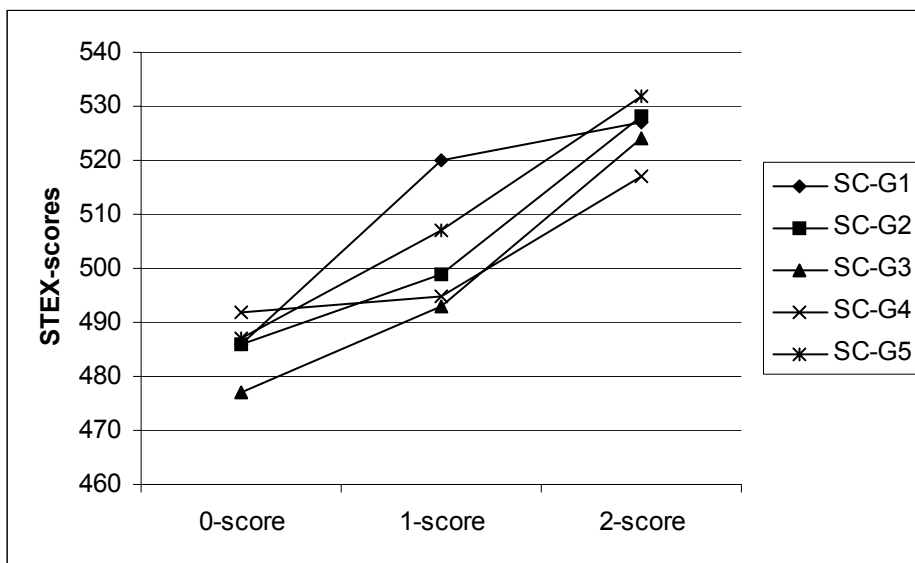
Elke kandidaat is globaal beoordeeld door twee onafhankelijke beoordelaars. De beoordelaars konden voor elke uitspraak aangeven of een kandidaat aan de eisen die in de beoordelingsuitspraak zijn verwoord voldeed of niet.

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 B2 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 56 kandidaten

Figuur 5.1-1 Relatie tussen globale oordelen B2 door 2 beoordelaars en de score op Staatsexamen II 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 (n=56)

Uitspraak	Pearson productmoment correlatie
S-G1	0,63
S-G2	0,71
S-G3	0,71
S-G4	0,44
S-G5	0,75
Totaalscore	0,74

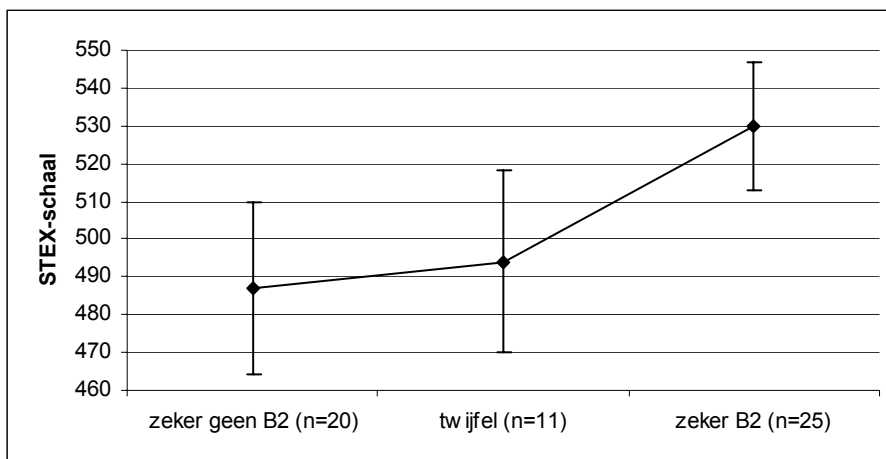
De correlaties tussen raamwerkuitspraken en de STEX-scores zijn zeer bevredigend, met uitzondering van uitspraak 4, die een matiger positief verband heeft.

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 10 (beide beoordelaars geven voor elke uitspraak 1 punt). In tabel 5.1-4 zijn de kandidaten in 3 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 B2 kandidaten zijn. De tweede groep bestaat uit kandidaten die in minimaal 8 van de 10 beoordelingen als B2 zijn aangemerkt. Deze groep kan gezien worden als bestaande uit kandidaten die B2 met grote zekerheid beheersen. De tusseliggende 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 3 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 B2 volgens globale beoordeling	STEX-score		
	N	gemiddeld	SD
Zeker niet	20	487,0	20,9
twijfel	11	494,1	23,8
Zeker wel	25	529,8	17,0
Alle kandidaten	57	507,5	28,2

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



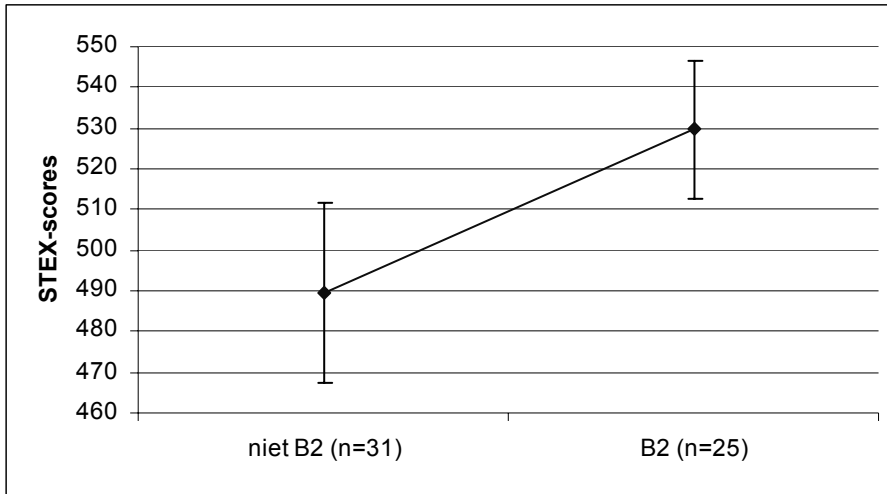
Het verband tussen de Raamwerkbeoordeling en de STEX-score is duidelijk en in de verwachte richting.

Wanneer we de twijfelgroep en de niet-B2-groep samenvoegen kunnen we de kandidaten in twee groepen indelen. De eerste groep is de groep die geacht wordt niveau B2 zeker te beheersen. Deze groep is dezelfde als de 'zeker B2-groep' in tabel 5.1-4 en figuur 5.1-2 (een totaalscore van 8-10). De andere groep is een samenvoeging van kandidaten die volgens de beoordelaars B2 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 II en standaarddeviaties van de kandidaten die B2 wel of niet beheersen volgens 75% criterium

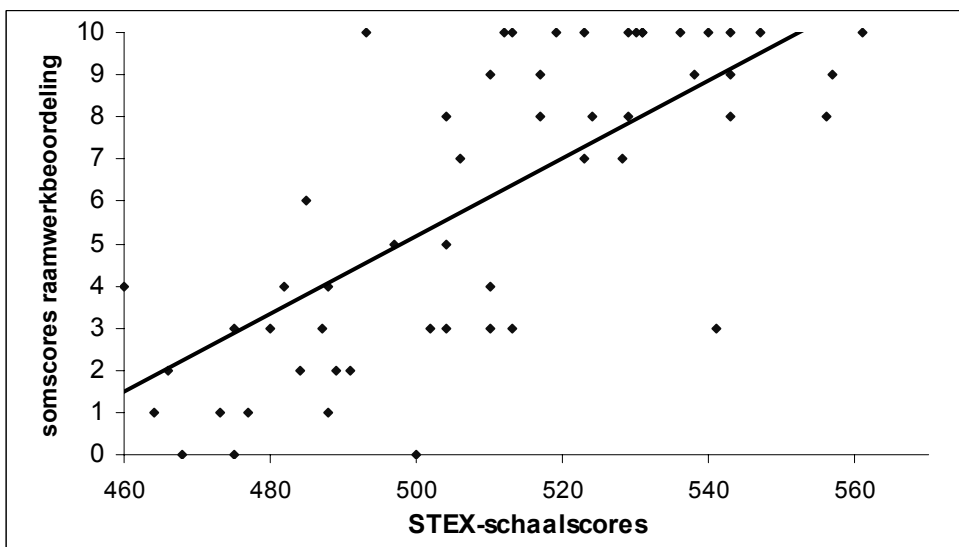
Beheerst B2 volgens globale beoordeling	STEX-score		
	N	gemiddeld	SD
Nee	31	489,5	21,9
Ja	25	529,8	17,0
Alle kandidaten	56	507,5	28,2

Figuur 5.1-3 Gemiddelde STEX-score Schrijven II en standaarddeviaties van de kandidaten die B2 wel of niet beheersen volgens 75% criterium



Volgens deze tweedeling van de kandidaten blijkt de niet-B2-groep gemiddeld 40,3 punt lager te scoren op het Staatsexamen dan de groep die wel op B2 functioneert. Er is dus een duidelijk verschil tussen beide groepen wat betreft hun STEX-score. Er is dus een substantiële samenhang tussen de scores van beide beoordelingsbenaderingen. Figuur 5.1-4 laat geeft een grafische weergave van deze relatie. De correlatie (Pearson product-moment correlatie) tussen de raamwerkbeoordeling en de STEX-schaalscore bedraagt 0,74.

Figuur 5.1-4 Samenhang tussen de scores van kandidaten op het Staatsexamen en op de globale raamwerkbeoordeling Schrijven



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

- 1 In dit aanvullende onderzoek zijn naast de 20 kandidaten uit de eerste fase van het cesuuronderzoek 40 extra STEX-kandidaten beoordeeld door twee onafhankelijke beoordelaars. Het totaal aantal respondenten in het onderzoek bedraagt daarmee 60 kandidaten. De uitingen van al deze 60 kandidaten zijn 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. Vier kandidaten zijn uit het bestand verwijderd omdat bleek dat zij niet alle opgaven van het Staatsexamen hadden gemaakt.
- 2 De beoordelaarsovereenstemming in de gevolgde procedure voor de globale raamwerkbeoordeling Schrijven is goed en voldoende voor verdere analyses met betrekking tot de cesuur kwestie. Het inschakelen van een derde beoordeling heeft aan deze goede overeenstemming bijgedragen. Er bestaat een substantiële positieve samenhang tussen de globale raamwerkbeoordeling en STEX-score. Kandidaten die volgens het globale oordeel functioneren op B2 hebben gemiddeld een hogere STEX-score dan kandidaten die B2 nog niet beheersen. Dit gegeven wordt ondersteund door het feit dat de correlatie tussen STEX-score en somscore op de raamwerkbeoordeling met .74 substantieel is.

5.2 Relatie tussen STEX-scores en Raamwerkbeoordeling: de cesuurkwestie

Uit paragraaf 5.1 blijkt dat er een bevredigende relatie gevonden is tussen de STEX-scores en de globale beoordelingen volgens het raamwerk. Verder blijkt dat de gemiddelde STEX-score van de B2 kandidaten hoger ligt dan de van de groep kandidaten die B2 niet beheersen. De volgende vraag die nu beantwoord moet worden is of deze gegevens aanleiding geven voor een mogelijke herziening van de cesuur van de Staatsexamen Schrijven II.

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 II 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 B2 spelen de volgende aspecten een rol:

- de definiëring van de beheersing van B2
- 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. Voor twee verschillende definities van de beheersing van B2 worden de consequenties van verschillende cesuurscores voor het slagen en zakken van kandidaten beschreven. De eerste definitie is een strenge, bij de tweede definitie worden minder hoge eisen gesteld aan de overeenstemming van beoordelaars.

5.2.1 Behersing B2 volgens volledige overeenstemming: strengste norm

Tabel 5.2-1 geeft een totaaloverzicht van alle globale raamwerkscores van alle 56 kandidaten. In kolom 1 staan de scores op het Staatsexamen Schrijven II. Kolom 2 geeft de totaalscore over de 5 beoordelingsuitspraken. Uit de laatste kolom kan worden afgelezen of een kandidaat niveau B2 beheerst volgens het criterium volledige overeenstemming. De gearceerde cellen geven aan dat een kandidaat geslaagd is voor het Staatsexamen Schrijven II en/of B2 beheerst volgens het criterium volledige overeenstemming.

Tabel 5.2-1 Beoordeling kandidaten op Staatsexamen Schrijven II en de raamwerkbeoordeling met als criterium volledige overeenstemming van de beoordelaars.

STEX-score	Totaalscore raamwerk	Beheerst B2 met volledige overeenstemming
561	10	B2
557	9	
556	8	
547	10	B2
543	10	B2
543	8	
543	9	
543	2	
540	10	B2
538	9	
536	10	B2
531	10	B2
531	10	B2
530	10	B2
529	10	B2
529	8	
528	7	
524	8	
523	10	B2
523	7	
519	10	B2
517	8	
517	9	
513	3	
513	10	B2
512	10	B2
510	9	
510	3	
510	4	
506	7	
504	3	
504	5	
504	8	
502	3	
500	0	
497	5	
493	10	B2
491	2	
489	2	
488	4	
488	1	
487	3	
485	6	
484	2	
482	4	
480	3	
477	1	
475	3	
475	0	
473	1	
468	0	
466	2	
464	1	
460	4	
452	4	
451	3	

Veertien kandidaten beheersen B2 volgens het strenge criterium van volledige overeenstemming. Van kandidaten die gezakt zijn voor het Staatsexamen beheerst er slechts één B2 volgens de hier gehanteerde norm van volledige overeenstemming. Er geen duidelijke overgang in de STEX-score te vinden waarbij kandidaten een zeer grote kans hebben volgens de raamwerkbeoordeling als B2 te worden gezien

Er kunnen volgens dit strenge criterium van volledige overeenstemming geen duidelijke conclusies getrokken worden over een eventuele herziening van de cesuur.

5.2.2 Beheersing B2 volgens het criterium meerderheidsoordeel

In paragraaf 5.2.1 hebben we een zeer strenge norm gehanteerd om te bepalen of iemand B2 beheerst. Het is de vraag of de eis dat beide beoordelaars het op alle criteria eens moeten zijn over de beheersing van B2 niet een te zware is. Uit figuur 5.1-1 blijkt voor elk van de vijf beheersingsuitspraken dat kandidaten over wie de beoordelaars het niet eens zijn een STEX-score hebben die ligt tussen die van de kandidaten die volgens beide beoordelaars B2 beheersen en van de kandidaten die B2 niet beheersen. We zouden een verschil van mening tussen beoordelaars, behalve als gebrek aan overeenstemming, wellicht ook kunnen beschouwen een uitvloeisel van het feit dat een kandidaat net op de grens van B2 verkeert, waardoor de ene beoordelaar zo'n kandidaat als B2 en de ander al niet B2 beoordeelt. Een score 1 zou in deze redenering een beheersing vertegenwoordigen die tussen 0 en 2 ligt.

Op grond van deze overwegingen vervangen we het 100% beheersingscriterium vervangen door een meerderheidsoordeel. We beschouwen een kandidaat als B2 indien deze in 80% van de beoordelingen als B2 beoordeeld wordt. Dit valt samen met een minimale score van 8 van de 10 te behalen punten. Dat betekent dat in 8 van de 10 beoordelingen een beoordelaar uitspreekt dat iemand op B2 functioneert. De resultaten staan in tabel 5.2-2.

Tabel 5.2-2 Beoordeling kandidaten op Staatsexamen Schrijven en de raamwerkbeoordeling met als criterium meerderheidsoordeel van de beoordelaars (80%).

STEX-score	Totaalscore raamwerk	Beheersing B2 met meerderheidsoordeel
561	10	B2
557	9	B2
556	8	B2
547	10	B2
543	10	B2
543	8	B2
543	9	B2
543	2	
540	10	B2
538	9	B2
536	10	B2
531	10	B2
531	10	B2
530	10	B2
529	10	B2
529	8	B2
528	7	
524	8	B2
523	10	B2
523	7	
519	10	B2
517	8	B2
517	9	B2
513	3	
513	10	B2
512	10	B2
510	9	B2
510	3	
510	4	
506	7	
504	3	
504	5	
504	8	
502	3	

STEX-score	Totaalscore raamwerk	Beheersing B2 met meerderheidsoordeel
500	0	
497	5	
493	10	
491	2	
489	2	
488	4	
488	1	
487	3	
485	6	
484	2	
482	4	
480	3	
477	1	
475	3	
475	0	
473	1	
468	0	
466	2	
464	1	
460	4	
452	4	
451	3	

	Geslaagd voor STEX / beheerst niveau B2
	Gezakt voor STEX / beheerst niveau B2 niet

Uit tabel 5.2-2 blijkt dat het aantal kandidaten dat geacht wordt B2 te beheersen is toegenomen van 14 tot 25. Van de 21 kandidaten die gezakt waren voor het Staatsexamen beheerst er slechts één B2. Van de 26 kandidaten met een STEX -score 512 of hoger worden er 22 als B2 beoordeeld. Van de 30 kandidaten met een STEX-score onder 512 worden er slechts 3 als B2 beoordeeld. In het interval tussen 512 en 528 bestaat de geringste samenhang tussen de STEX-score. Van de 10 kandidaten in dit scorebereik worden er 7 beoordeeld als B2. Van de 3 kandidaten die niet als B2 worden gezien, bevinden er zich 2 (met een score van 7) net op de grens van B2. 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.

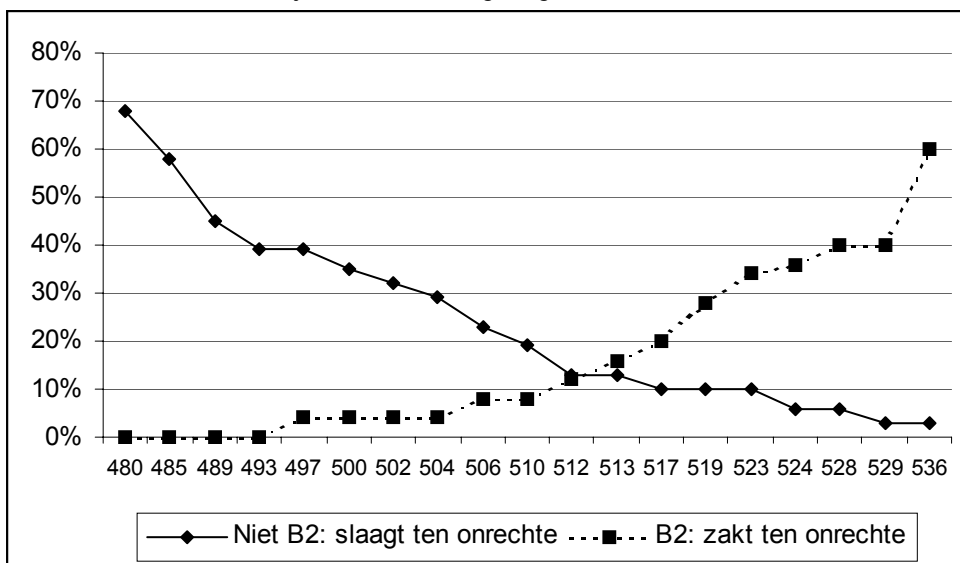
Voor het vaststellen van de cesuur is echter ook informatie nodig over de proportie kandidaten die bij een gegeven cesuurscore onterecht zakt of onterecht slaagt. Deze twee consequenties moeten tegen elkaar worden afgewogen. Tabel 5.2-3 geeft een overzicht van de gevolgen van de beslissingen voor de cesuurscores tussen 500 en 528.

Tabel 5.2-3 Percentages gezakte en geslaagde kandidaten die B2 al of niet beheersen bij verschillende cesuren van het Staatsexamen Schrijven II. Beheersing B2 gedefinieerd als 80% van maximale score

Cesuur op STEX	Beheerst B2 niet (n=31)		Beheerst B2 (n=25)	
	Zakt terecht	Slaagt ten onrechte	Zakt ten onrechte	Slaagt terecht
480	32%	68%	0%	100%
485	42%	58%	0%	100%
489	55%	45%	0%	100%
493	61%	39%	0%	100%
497	61%	39%	4%	96%
500	65%	35%	4%	96%
502	68%	32%	4%	96%
504	71%	29%	4%	96%
506	77%	23%	8%	92%
510	81%	19%	8%	92%
512	87%	13%	12%	88%
513	87%	13%	16%	84%
517	90%	10%	20%	80%
519	90%	10%	28%	72%
523	90%	10%	32%	68%
524	94%	06%	36%	64%
528	94%	6%	40%	60%
529	97%	3%	40%	60%
536	97%	3%	60%	40%

In figuur 5.2-1 wordt deze tabel grafisch weergegeven.

Figuur 5.2-1 Percentages gezakte en geslaagde kandidaten die B2 al of niet beheersen bij verschillende cesuren van het Staatsexamen Schrijven II. Beheersing B2 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 B2 wel beheersen.
 2. Fout positief. De fout dat men ten onrechte kandidaten laat slagen die B2 nog niet beheersen.
- Zo leiden een cesuurscores van 500 tot 504 ertoe dat de kans dat een B2-kandidaat ten onrechte zakt klein is. Daar staat tegenover dat de kans dat kandidaten die B2 nog niet beheersen bij deze scores slagen afneemt van 35% tot 29%. Bij een cesuurscore van 528 maakt men de kans op het laten slagen van niet B2-kandidaten klein, maar verhoogt men de kans op het onterecht zakken tot 40%. Welk cesuurscore men 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 528 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 5.2-3 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 minder tevreden moeten zijn.

Tabel 5.2-4 geeft de somwaarden weer onder twee condities:

1 conditie 1

Onterecht zakken en slagen word even belangrijk gevonden (ongewogen conditie)
 Voorbeeld: bij score 519 slaagt 10% ten onrechte en zakt 28% ten onrechte. De optimumscore wordt: $10 + 28 = 38$. Hoe kleiner dit getal hoe minder kans op een foute beslissing.

2 Conditie 2

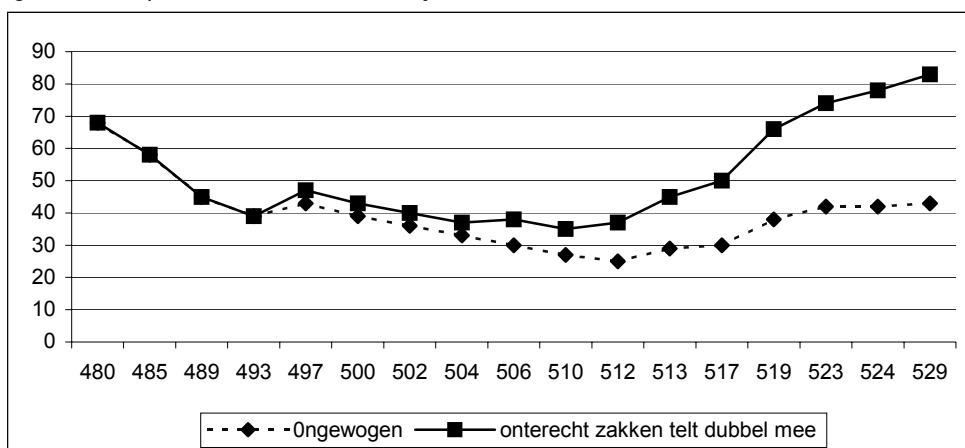
Men wil voorkomen dat te veel B2-kandidaten ten onrechte zakken. Daarom geeft men ten onrechte zakken een dubbel gewicht.
 Voorbeeld: bij score 519 slaagt 10% ten onrechte en zakt 28% ten onrechte. De optimumscore als we het onterecht zakken dubbel wegen, wordt nu: $10 + 56 = 66$. Ook hier geldt dat de laagste optimumscore de geringste kans geeft op een foute beslissing.

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

Tabel 5.2-4 Optimum cesuurscore Schrijven II onder 2 condities

	Ongewogen	Onterecht zakken telt dubbel
480	68	68
485	58	58
489	45	45
493	39	39
497	43	47
500	39	43
502	36	40
504	33	37
506	30	38
510	27	35
512	25	37
513	29	45
517	30	50
519	38	66
523	42	74
524	42	78
529	43	83

Figuur 5.2-2 Optimum cesuurscore Schrijven II onder 2 condities



De optimale cesuurscore in de ongewogen variant ligt op 512. Wanneer men meer belang hecht aan het voorkomen van het ten onrechte zakken van B2 kandidaten komt een cesuur van 510 het meest in aanmerking. Maar ook 504 en 512 liggen hier dicht bij. Een cesuurscore van 504 of 510 ligt voor de hand. Bij de keuze voor één van deze twee alternatieven zal de doorslag moeten geven welk percentage ten onrechte gezakten men bij deze twee scores het meest aanvaardbaar vindt. Een compromis zou gevonden kunnen worden door een score tussen 504 en 510 te kiezen. Hieronder worden de proporties ten onrechte geslaagde en gezakte kandidaten voor deze scores vermeld.

Cesuurscore	Niet B2: slaagt ten onrechte	B2: zakt ten onrechte	Optimum-score
504	29%	4%	37
510	19%	8%	35
512	13%	12%	37

5.2.3 Samenvatting en aanbevelingen cesuur Schrijven

Volgens de behoeftepeiling zou het Staatsexamen Schrijven II een vaardigheid moeten waarborgen die op niveau B2 van het Raamwerk NT2 ligt (Kuijper e.a., augustus 2004).

1. Uit het beoordelingsonderzoek blijkt dat er een substantiële positieve relatie bestaat tussen de scores die kandidaten halen op het Staatsexamen Schrijven II en de vraag of de respons van een kandidaat beoordeeld wordt als B2. In zijn algemeenheid geldt dat naarmate de STEX-score hoger wordt de kans dat een kandidaat als B2 wordt beoordeeld groter wordt.
2. De gevonden beoordelaarsovereenstemming bij de globale raamwerkbeoordeling Schrijven is voldoende om er uitspraken over een cesuurscore op te baseren.
3. De methodiek die in dit onderzoek gehanteerd is – globale beoordeling van de schrijfproducten van kandidaten – kan in zijn algemeenheid als bevredigend worden gezien.
 - a. De globale en analytische beoordelingsmethode leiden grosso modo tot vergelijkbare resultaten (Kuijper e.a., augustus 2004). Gezien de grote tijdsinvestering die nodig is voor de analytische methode ligt het gebruik van een globale beoordeling voor het doel van dit onderzoek voor de hand.
 - b. De beoordelaarsovereenstemming van de globale raamwerkbeoordeling is goed. Het inschakelen van een derde beoordeling heeft tot een hogere overeenstemming geleid dan in de eerste fase van dit onderzoek.
 - c. De relatie tussen raamwerkbeoordeling en de STEX-score is substantieel.
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 II een onderscheid moet maken tussen kandidaten die een spreekvaardigheid hebben die vergelijkbaar is met

niveau B2, dan zal deze cesuur liggen in het interval tussen de schaalscores 504-512. Bij het vaststellen van de cesuur 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 504 of 510 ligt dan voor de hand. Bij de keuze voor één van deze twee alternatieven zal de doorslag moeten geven welk percentage ten onrechte gezakten men bij deze twee scores het meest aanvaardbaar vindt. Vindt men 4% acceptabel dan zal men moeten kiezen voor 504, waarbij het percentage ten onrechte geslaagden 29% bedraagt. Vindt men dit percentage te hoog dan moet men kiezen voor een cesuur van 510. Het percentage ten onrechte geslaagden daalt dan tot 19%, terwijl 8% van de kandidaten met niveau B2 zal zakken. Een compromis zou gevonden kunnen worden door een score tussen 504 en 510 te kiezen, vanuit de veronderstelling dat de lage B2 scores tussen 504 en 510 een artefact van de steekproef zijn.

Literatuur

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.

Council of Europe (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment.* Cambridge University press, Cambridge.

Dalderop, Kaatje 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

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

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

Beoordelaar:

Globaal oordeel Algemene beheersingsuitspraken Spreken B2

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 B2 uit te voeren?

Baseer uw oordeel op het document: Familiarisatie Spreken B2

Algemene beheersingsuitspraken Spreken B2

	Uitspraken Raamwerk NT2 (aangepast)	Ja	Nee
S-G1	Kan duidelijke, gedetailleerde beschrijvingen presenteren over een breed scala van onderwerpen die betrekking hebben op mijn interessegebied.		
S-G2	Kan een standpunt over een actueel onderwerp verklaren en de voordelen en nadelen van diverse opties uiteenzetten.		
GV-G3	Kan zodanig deelnemen aan een vloeiend gesprek, dat normale uitwisseling met moedertaalsprekers redelijk mogelijk is.		
GV-G4	Kan binnen een vertrouwde context actief deelnemen aan een discussie en hierin standpunten uitleggen en ondersteunen.		
SGV-G5	Beheerst deze kandidaat over het algemeen niveau B2?		

Bijlage 2 Globale raamwerkbeoordeling Schrijven

Beoordelingslijst Schrijven II B2

Kandidaatnummer:

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

Beoordelaar:

--

Globaal oordeel Algemene beheersingsuitspraken Schrijven B2

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 B2 uit te voeren?
Baseer uw oordeel op het document: Familiarisatie Schrijven B2

	Uitspraken Raamwerk NT2 (aangepast)	Ja	Nee
SC-G1	Kan een duidelijke, gedetailleerde tekst schrijven.		
SC-G2	Kan een opstel of verslag schrijven.		
SC-G3	Kan informatie doorgeven of redenen aanvoeren ter ondersteuning vóór of tégen een specifiek standpunt.		
SC-G4	Kan brieven schrijven waarin het persoonlijk belang van ervaringen en indrukken aangegeven wordt.		
SC-G5	Beheerst deze kandidaat over het algemeen niveau B2?		

Bijlage 3 Familiarisatie Spreken

Familiarisatie Spreken en Gesprekken Voeren B2

Globale schalen

Zelf evaluatie matrix

Subvaardigheden en descriptoren Gesprekken Voeren B2

Subvaardigheden en descriptoren Spreken B2

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 CEFR 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 B2 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: B2

Beheersingsniveau:
 Kan zodanig deelnemen aan een vloeiend gesprek, dat normale uitwisseling met moedertaalsprekers redelijk mogelijk is. Kan binnen een vertrouwde context actief deelnemen aan een discussie en hierin standpunten uitleggen en ondersteunen.

Subvaardigheden en Descriptoren	Voorbeelden bij de descriptoren	Tekstkenmerken receptief	Tekstkenmerken productief
<p>1. Informele gesprekken Kan deelnemen aan uitvoerige gesprekken over de meeste algemene onderwerpen op een duidelijk actieve manier, zelfs in een rumoerige omgeving. Kan mate van emotie uitdrukken en de persoonlijke betekenis van gebeurtenissen en ervaringen naar voren brengen.</p> <p>2. Bijeenkomsten en vergaderingen Kan actief participeren in routinematige en niet-routinematige formele discussies. Kan zijn/haar mening naar voren brengen, verantwoorden en handhaven, alternatieve voorstellen op waarde toetsen en hypothesen opstellen en erop reageren.</p> <p>3. Zaken regelen Kan de voortgang van het werk op weg helpen door anderen uit te nodigen om mee te doen, om te zeggen wat zij denken etc. Kan een vraagstuk of probleem helder schetsen, door te speculeren over oorzaken en consequenties, en voor- en nadelen van verschillende wijzen van aanpak af te wegen.</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 De onderwerpen kunnen algemeen, wetenschappelijk of beroepsmatig van aard zijn of te maken hebben met vrijetijdsbesteding. ▪ woordgebruik en zinsbouw Er worden complexe argumentaties gebruikt. Idiomatiche uitdrukkingen worden niet altijd begrepen. Zinsbouw kan complex zijn. ▪ tempo en articulatie Het spreektempo is het gemiddelde spreektempo van een moedertaalspreker. Er wordt standaardtaal gesproken. ▪ hulp Gesprekspartners praten op een normaal tempo in de standaardtaal. 	<ul style="list-style-type: none"> ▪ onderwerp De onderwerpen kunnen algemeen, wetenschappelijk of beroepsmatig van aard zijn of te maken hebben met vrijetijdsbesteding. ▪ woordenschat en woordgebruik Taal is toereikend om duidelijke beschrijvingen te geven, en meningen te verkondigen over de meeste algemene onderwerpen. Een enkel maal komt een minder gebruikelijk of minder passend woord voor. Het is echter meestal duidelijk wat wordt bedoeld... ▪ grammaticale correctheid De grammatica is redelijk correct. Eventuele kleine afwijkingen doen geen afbreuk aan de begrijpelijkheid en worden vaak hersteld. Enige complexe zinsvormen komen voor. ▪ interactie De aanvang van een gesprek, het beurt nemen en het beëindigen van een gesprek gebeurt op over het algemeen op gepaste wijze, soms op een minder elegante manier. Het taalgebruik bevat aanmoedigingen en bevestigingen van

<p>Kan een probleem uitleggen en met de desbetreffende instantie of personen onderhandelen over een oplossing voor een onenigheid.</p> <p>4. Informatie uitwisselen Kan betrouwbaar gedetailleerde informatie doorgeven. Kan een heldere, gedetailleerde beschrijving geven van hoe een procedure uitgevoerd moet worden. Kan informatie en argumenten uit een aantal bronnen bijeenvoegen en daarover rapporteren. Kan initiatief nemen in een vraaggesprek, kan ideeën ontwikkelen en ze uitbreiden met een beetje hulp of stimulans van de gesprekspartner.</p>			<p>de gesprekspartner die dienen om het gesprek gaande te houden..</p> <ul style="list-style-type: none"> ▪ vloeiendheid <i>Het tempo is over het algemeen en voor langere gedeelten vrij vast, alhoewel ook aarzelingen voorkomen waar de spreker zoekt naar patronen en uitdrukkingen. Er zijn nog maar weinig opvallend lange pauzes.</i> ▪ coherentie Helderheid en coherentie wordt bewerkstelligd door middel van een beperkt aantal verbindingswoorden, hoewel dat in een lange bijdrage niet altijd even goed lukt. ▪ uitspraak Duidelijke, natuurlijke uitspraak en intonatie.
---	--	--	--

Vaardigheid: Spreken
Niveau: B2

Beheersingsniveau:
Kan duidelijke, gedetailleerde beschrijvingen presenteren over een breed scala van onderwerpen die betrekking hebben op zijn/haar interessegebied. Kan een standpunt over een actueel onderwerp verklaren en de voordelen en nadelen van diverse opties uiteenzetten.

Subvaardigheden en Descriptoren	Voorbeelden bij de descriptoren	Tekstkenmerken productief
<p>1. Monologen Kan over een groot aantal onderwerpen binnen zijn/haar interessegebied een argumentatie systematisch opbouwen en op adequate wijze de belangrijke punten en relevante details markeren. Kan een reeks van beredeneerde argumenten opbouwen. Kan een standpunt uitleggen en daarbij de voor- en nadelen van verschillende opties presenteren</p> <p>2. Een publiek toespreken Kan een duidelijke, voorbereide presentatie geven waarbij hij/zij argumenten kan noemen voor of tegen een bepaald standpunt of voor- en nadelen van diverse opties. Kan omgaan met een aantal daaropvolgende vragen met een mate van vloeiendheid en spontaniteit die noch voor hem/haarzelf, noch voor het publiek ongemak oplevert. Kan aankondigen over de meeste algemene onderwerpen zo helder, vlot en spontaan brengen dat toehoorders het zonder inspanning kunnen volgen.</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 De onderwerpen kunnen algemeen, wetenschappelijk of beroepsmatig van aard zijn of te maken hebben met vrijetijdsbesteding. ▪ woordgebruik en woordenschat De woordenschat is toereikend om duidelijke beschrijvingen te geven, en meningen te verkondigen over de meeste algemene onderwerpen. Het valt niet op dat de spreker naar woorden moet zoeken. ▪ grammaticale correctheid Vertoont een vrij uitgebreide beheersing van de grammatica en maakt gebruik van complexe zinsvormen. Er zijn geen fouten die de begrijpelijkheid in de weg staan. De meeste fouten worden door de spreker zelf hersteld. ▪ vloeiendheid <i>Stukken tekst worden in een vrij vast tempo geproduceerd, alhoewel de spreker soms aarzelt wanneer hij/zij zoekt naar patronen en uitdrukkingen. Er zijn nog maar weinig opvallend lange pauzes.</i> ▪ coherentie De spreker maakt gebruik van een beperkt aantal middelen voor tekstcohesie om zijn/haar uitingen te verbinden tot een heldere, coherente tekst, hoewel dat in een lange bijdrage niet altijd even goed lukt. ▪ uitspraak Duidelijke, natuurlijke uitspraak en intonatie.

Bijlage 4 Familiarisatie Schrijven

Familiarisatie Schrijven B2

Globale schalen
Zelf evaluatie matrix
Subvaardigheden en descriptoren Schrijven B2

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 B2 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 tekst 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: B2

Beheersingsniveau:
Kan een duidelijke, gedetailleerde tekst schrijven over een breed scala van onderwerpen die de persoonlijke interesse van de schrijver hebben.
Kan een opstel of verslag schrijven, informatie doorgeven of redenen aanvoeren ter ondersteuning vóór of tégen een specifiek standpunt. Kan brieven schrijven waarin het persoonlijk belang van ervaringen en indrukken aangegeven wordt.

Subvaardigheden en Descriptoren	Voorbeelden bij de descriptoren	Tekstkenmerken productief
<p>1. Correspondentie Kan brieven schrijven met een bepaalde mate van emotie en kan de persoonlijke betekenis van gebeurtenissen en ervaringen onderstrepen en het nieuws en standpunten van de correspondent becommentariëren. Kan adequate zakelijke brieven schrijven.</p> <p>2. Aantekeningen, berichten, formulieren Als op B1</p> <p>3. Verslagen en rapporten Kan een essay of rapport schrijven waarbij een argument wordt uitgewerkt, en daarbij redenen aangeven voor of tegen een bepaalde mening en de voordelen en nadelen van verschillende keuzes uitleggen. Kan een geheel maken van informatie en argumenten uit een aantal bronnen.</p> <p>4. Vrij schrijven Kan heldere, gedetailleerde beschrijvingen geven over een scala aan onderwerpen gerelateerd aan zijn/haar interessegebied.</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 De onderwerpen kunnen algemeen, wetenschappelijk of beroepsmatig van aard zijn of te maken hebben met vrijetijdsbesteding. ▪ woordenschat en woordgebruik Voldoende taal om duidelijke beschrijvingen te geven, en meningen te verkondigen over de meeste algemene onderwerpen. ▪ grammaticale correctheid Vertoont een vrij uitgebreide beheersing van de grammatica. Er zijn geen fouten die de begrijpelijkheid in de weg staan. ▪ spelling en interpunctie Duidelijke, begrijpelijke, lopende tekst waarbij tekstconventies zoals indeling in paragrafen en standaard lay-out zijn toegepast. Spelling en interpunctie zijn redelijk accuraat maar kunnen invloeden van de moedertaal bevatten. ▪ coherentie Een gelimiteerd aantal middelen voor tekstcohesie is aangewend om de uitingen van de schrijver te verbinden tot een heldere, coherente tekst, hoewel dat in een lange tekst niet altijd even goed is gelukt.

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

SPREKEN					
Somscore beoordelaar1	Somscore beoordelaar 2	verschil	Somscore totaal	STEX-score	opmerkingen
5	5	0	10	500	Spreekt heel vlot, klets lekker door, grammatica niet perfect, maar dit wordt gemaskeerd door vloeiendheid en tempo. Woordenschat goed
5	4	1	9	506	Compenseert grammaticale fouten en gebrek aan woordenschat met heel vlot spreken. Globaal dus wel B2, maar in details wel fouten. Dat verklaart de terechte STEX-score. 9 punten raamwerk-beoordeling is aan de hoge kant.
5	5	0	10	508	Soms wat aarzelend, wel goed verstaanbaar. Grammatica voldoende. STEX-score terecht, 10 punten op raamwerk aan de hoge kant.
1	2	1	3	529	Weinig vloeiend, erg slechte uitspraak, grammaticaal onvoldoende. Niet vloeiend.
4	2	2	6	542	Hoewel inhoudelijk wel goed, is vloeiendheid, verre van soepel, woordgebruik grammatica beperkt. Hortend dus géén B2.
SCHRIJVEN					
5	5	0	10	493	Geeft korte beschrijvingen, die volgens beide beoordelaars voldoen aan B2 wat betreft tekstkenmerken productief. Inhoud voldoet waarschijnlijk niet aan de STEX-beoordeling.
2	0	0	2	543	Taalhandelingen en inhoud zijn goed. Dit verklaart de hoge STEX-score. Lage score raamwerk omdat de grammatica niet voldoet aan de b2 eisen. Veel fouten

© Citogroep Arnhem 2004. Auteursrechten voorbehouden.

Niets uit dit werk mag zonder voorafgaande schriftelijke toestemming van de Citogroep 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.

