

**De koppeling van de centrale examens leesvaardigheid  
Arabisch, Russisch, Spaans en Turks  
aan het Europees Referentiekader**

Verslag van een onderzoek in opdracht van het  
Ministerie van Onderwijs, Cultuur en Wetenschap

José Noijons, Henk Kuijper, Evelyn Reichard

Januari 2007



Aan deze uitgave hebben meegewerkt:

Fadil Azzarouali  
Timo Bechger  
Tom Duindam  
Anneke de Graaf  
Rahime Gülcü  
Ton Heuvelmans  
Erna Gille  
Çigdem Kinaci  
Henk Kuijper  
José Noijons  
Evelyn Reichard  
Noni Verschoor  
Margriet Welling

© Stichting Cito Instituut voor Toetsontwikkeling Arnhem (2007). Auteursrecht voorbehouden.  
Niets uit dit werk mag zonder voorafgaande schriftelijke toestemming van Stichting Cito Instituut voor Toetsontwikkeling worden openbaar gemaakt en/of verveelvoudigd door middel van druk, fotografie, scanning, computersoftware of andere elektronische verveelvoudiging of openbaarmaking, microfilm, geluidskopie, film- of videokopie of op welke wijze dan ook.

# Inhoud

## Voorwoord 5

## Inleiding 7

<b>1</b>	<b>Koppelingsprocedure .....</b>	<b>9</b>
<b>2</b>	<b>Familiarisatie.....</b>	<b>11</b>
<b>3</b>	<b>Inhoudelijke specificatie .....</b>	<b>13</b>
3.1	Inleiding	13
3.2	Inhoudelijke specificatie van de examenteksten in relatie tot de contextdimensies	14
3.2.1	<i>Tekstdimensies: tekstbron, tekstsoort, onderwerp en domein .....</i>	<i>14</i>
3.2.2	<i>Conclusies en aanbevelingen.....</i>	<i>21</i>
3.3	De inhoudelijke specificatie in relatie tot de schalen voor communicatieve competentie	22
3.3.1	<i>Inleiding .....</i>	<i>22</i>
3.3.2	<i>Werkwijze .....</i>	<i>22</i>
3.3.3	<i>De linguïstische en cognitieve complexiteit van de teksten .....</i>	<i>24</i>
	<i>Mate van abstractie van de teksten.....</i>	<i>24</i>
3.3.4	<i>De complexiteit van de taken .....</i>	<i>34</i>
3.3.5	<i>Conclusies bij de specificatie van teksten en taken .....</i>	<i>44</i>
<b>4</b>	<b>Standaardisatie.....</b>	<b>45</b>
4.1	Inleiding	45
4.2	Het Beoordelingsproces	45
4.3	Data-analyse ter validering van de standaarden	47
4.3.1	<i>Discussie analyse beoordelaarsovereenstemming.....</i>	<i>50</i>
4.4	Het bepalen van de minimumscores voor relevante ERK-niveaus bij ieder examen	51
4.4.1	<i>Discussie bepaling minimumscores voor relevante ERK-niveaus .....</i>	<i>53</i>
<b>5</b>	<b>Samenvatting, conclusies en aanbevelingen .....</b>	<b>54</b>
5.1	Samenvatting	54
5.2	Conclusies	54
5.3	Aanbevelingen	55
<b>6</b>	<b>Referenties .....</b>	<b>56</b>
	<b>Bijlagen.....</b>	<b>58</b>
1	Samenvatting van de examenmodellen Arabisch, Russisch, Spaans en Turks voor vmbo en havo/vwo .....	60
2	De bepaling van minimum scores voor relevante ERK-niveaus .....	61



## Voorwoord

Het Europees Referentiekader (ERK) is de laatste paar jaren uitgegroeid tot een belangrijk raamwerk voor het onderwijs in de moderne vreemde talen binnen Europa. In Nederland vormt het de basis voor een aantal innovatieve ontwikkelingen zoals de ontwikkeling van Europese Taalportfolio's en de Taalprofielen.

In opdracht van het Ministerie van OCW hebben SLO en Cito in 2005-2006 onderzoek uitgevoerd naar de koppeling van examenprogramma's en examens in de moderne vreemde talen aan het ERK. Een dergelijke koppeling aan het ERK sluit aan bij de behoefte:

- (1) om het Nederlandse talenonderwijs meer competentiegericht te laten zijn;
- (2) de exameneisen in Nederland voor docenten en leerlingen transparanter te maken;
- (3) de prestaties van Nederlandse leerlingen internationaal vergelijkbaar te laten zijn.

In januari 2006 is verslag gedaan van een door Cito uitgevoerde koppelingsonderzoek van de centrale examens leesvaardigheid in de moderne vreemde talen Frans, Duits en Engels aan het Europees Referentiekader. Daarbij zijn de procedures gevolgd die zijn beschreven in een handleiding die is gepubliceerd door de Raad van Europa.

Op verzoek van het Ministerie van OCW is nu ook voor de talen Arabisch, Russisch, Spaans en Turks een vergelijkbaar onderzoek met het hierboven genoemde koppelingsonderzoek uitgevoerd. Omdat het aantal kandidaten dat aan deze examens deelneemt aanzienlijk geringer is dan bij Frans, Duits en Engels, zijn de onderzoeksresultaten bescheidener. Desalniettemin bevat het huidige onderzoek voldoende gegevens om een goed beeld te vormen van de positie van de examens tekstbegrip Arabisch, Russisch, Spaans en Turks in relatie tot het ERK. Het stemt daarbij tot voldoening dat de conclusie getrokken mag worden dat deze examens in hoge mate de doelstellingen van het ERK reflecteren.

Erna van Hest

Hoofd afdeling examens vmbo  
Cito



## Inleiding

Eind 2005 gaf het Ministerie van OCW als een vervolg op de werkzaamheden van Cito in 2005 aan Cito de opdracht tot het uitvoeren van een koppelingsproject conform de opzet zoals die ook bij de talen Frans, Duits en Engels was gevolgd:

- Het inhoudelijk analyseren van de examenopgaven leesvaardigheid 2003-2005 voor Arabisch, Spaans en Turks vmbo en havo/vwo aan de hand van de criteria van het *Dutch Grid*;
- Het organiseren en uitvoeren van een standardsetting procedure voor de examens leesvaardigheid Arabisch, Spaans en Turks vmbo en havo/vwo Frans, Duits en Engels vmbo en havo/vwo.

Als beoogde producten werden geformuleerd:

- Een inhoudelijke analyse van de examens leesvaardigheid Arabisch, Turks en Spaans vmbo en havo/vwo in termen van het ERK;
- Een classificatie door experts van de examenopgaven Arabisch, Turks en Spaans vmbo en havo/vwo in termen van het ERK met de daarbij horende cut-off scores;
- Een statistische analyse van de gegevens van de standardsetting met o.a. als resultaat de minimumscore die een leerling op een examen Arabisch, Turks en Spaans moet behalen om te kunnen beweren dat hij of zij het door de experts aan dat examen toegekende ERK-niveau beheerst.

Gegeven het feit dat in het onderwijsveld de wens leefde dat een dergelijk onderzoek ook voor Russisch zou worden uitgevoerd, heeft Cito besloten om de tekstbegrip examens Russisch mee te laten lopen in het genoemd onderzoek. Daarbij dient te worden aangetekend dat er voor Russisch slechts op twee niveaus wordt getoetst: havo en vwo.

In 2003 publiceerde de Raad van Europa een concept-handleiding voor het relateren van taalexamens aan het ERK, hierna te noemen: "Handleiding" (Raad van Europa, 2003). In deze Handleiding worden verschillende methodes beschreven om de claim op een koppeling van een examen met het ERK te onderbouwen. In dit rapport worden de activiteiten conform de volgende twee methodes beschreven: (1) koppeling aan ERK op basis van specificatie en (2) koppeling aan ERK op basis van specificatie EN standaardisatie.

Het onderzoek waarover hier wordt gerapporteerd betreft de centrale examens leesvaardigheid omdat de inhoud van deze examens nauwkeurig door CEVO is vastgelegd in het examenprogramma voor de moderne vreemde talen (CEVO, 2006). De concrete inhoud van de *school*examens is formeel de verantwoordelijkheid van de individuele scholen, met dien verstande dat de examens wel gebaseerd dienen te zijn op de handreiking schoolexamens. Over de andere taalvaardigheden naast leesvaardigheid is voor de moderne vreemde talen gerapporteerd in de Handreikingen, gepubliceerd door SLO (2006).



# 1 Koppelingsprocedure

De Handleiding die door de Raad van Europa is gepubliceerd, heeft als doel om belanghebbenden mogelijke procedures aan de hand te doen waarmee men evidentie kan aanvoeren voor de relaties van een examen met het ERK. Het was tot op heden zo dat toetsinstituten ieder op hun eigen wijze relaties legden tussen het ERK en hun toetsen. Sommige van zulke onderzoeken zijn uitvoerig gedocumenteerd, andere minder uitvoerig. De onderlinge vergelijkbaarheid van de onderzoeken is problematisch door de verschillen in gevolgde methodes, waardoor de indruk kan ontstaan dat verschillende instituten verschillende concepties hebben van wat de ERK-niveaus concreet inhouden en hoe deze te koppelen zijn aan niveaus van examens. De Handleiding biedt naar het idee van Cito en SLO een uitgelezen mogelijkheid om deze problemen te voorkomen.

In de Handleiding worden drie methodes van koppeling onderscheiden:

1. Specificatie van de inhoud van examens;
2. Standardisatie van oordelen;
3. Empirische validatie door middel van data-analyse van toetsresultaten.

Bij de eerste methode is de claim op koppeling aan het ERK op basis van alleen specificatie. Deze methode (met varianten) is door een aantal instituten in Europa en daarbuiten gevolgd bij het relateren van hun examens aan het ERK. De tweede methode leidt tot een versterkte claim op een koppeling, namelijk gebaseerd op basis van specificatie EN standardisatie. De derde methode kan leiden tot claims die worden bevestigd door empirische validatie. Al deze methodes zijn in het Cito-onderzoek toegepast, weliswaar met enige beperkingen als het gaat om methode 3. Het aantal deelnemers aan de examens Arabisch, Russisch, Spaans en Turks is zodanig laag dat niet altijd robuuste conclusies kunnen worden getrokken. Ook hebben wij geen externe validatie toegepast: wij hebben onze koppelingsresultaten niet met gegevens van buiten kunnen vergelijken.

Binnen de verschillende methodes worden diverse fases onderscheiden:

- *Familiarisatie*: personen betrokken bij de koppelingsprocedure een gedetailleerde kennis bijbrengen van het doel, de opzet en de niveaus van het ERK (Introductiefase).
- *Specificatie*: nagaan in hoeverre de te onderzoeken examenprogramma's en de centrale examens leesvaardigheid de descriptoren van het ERK dekken.
- *Standardisatie*: via een in de Handleiding omschreven procedure experts (talendocenten, vertegenwoordigers uit politiek en bedrijfsleven) vragen om een selectie van in het onderzoek betrokken examenopgaven te classificeren in termen van ERK (er ERK-niveaus aan te koppelen). Op basis daarvan kan worden bepaald welke score een kandidaat bij een bepaald examen zou moeten behalen om te kunnen zeggen dat hij/zij een bij dat examen relevant ERK-niveau beheerst).
- *Empirische validatie*: psychometrische validering van de gegevens verzameld in de standardisatieprocedure.

In het hiernavolgende schema geven wij aan wat de onderlinge relatie is tussen de verschillende fases. Dit schema is gebaseerd op een schema uit de Handleiding (p. 129). Zoals men kan zien, is het idee dat iedere stap in het koppelingsproces de claim dat een examen is gekoppeld aan het ERK, sterker kan onderbouwen.

Figuur 1-1 Stappenschema ten behoeve van het vergaren van evidentie voor koppeling van toetsen en examens aan het ERK

Specificatie van de inhoud van examens	Standaardisatie van oordelen	Empirische validatie door middel van data-analyse van toetsresultaten
Interne validatie – beschrijving en analyse van: <ul style="list-style-type: none"> <li>• globale inhoud examen</li> <li>• proces van toetsontwikkeling</li> <li>• beoordeling, cijfergeving, resultaten</li> </ul>	Familiarisatie: <ul style="list-style-type: none"> <li>• training met gestandaardiseerde voorbeelden van productieve vaardigheden</li> <li>• training met gestandaardiseerde voorbeelden van receptieve vaardigheden en linguïstische competentie</li> </ul>	Interne validatie: <ul style="list-style-type: none"> <li>• Klassieke toetstheorie</li> <li>• Methoden van kwalitatieve analyse</li> <li>• Generaliseerbaarheidstheorie</li> <li>• Factoranalyse</li> <li>• Itemresponstheorie</li> </ul>
Externe validatie – relateren van: <ul style="list-style-type: none"> <li>• algemene beschrijving van het examen aan een ERK-schaal</li> </ul>	IJking van lokale voorbeelden van taaluitingen	Externe validatie: <ul style="list-style-type: none"> <li>• het correleren aan resultaten op toetsen die al gekalibreerd zijn op het ERK</li> </ul>
<ul style="list-style-type: none"> <li>• beschrijving van getoetste communicatieve activiteiten aan de ERK-schalen</li> </ul>	Standaardbepaling	<ul style="list-style-type: none"> <li>• het correleren van beoordelingen aan ERK-descriptoren</li> </ul>
<ul style="list-style-type: none"> <li>• beschrijving van aspecten van getoetste communicatieve taalcompetentie aan ERK-schalen</li> </ul>	Disseminatie en implementatie	<ul style="list-style-type: none"> <li>• het verankeren van een toets aan een toets die reeds gekalibreerd is op het ERK</li> </ul>
		<ul style="list-style-type: none"> <li>• het direct verankeren van een toets aan de onderliggende ERK-schaalwaarden van de descriptoren.</li> </ul>
↓	↓	↓
<b>Claim op koppeling aan ERK op basis van specificatie</b>	<b>Versterkte claim op koppeling aan ERK op basis van specificatie EN standaardisatie</b>	<b>Bevestiging van claim op koppeling aan ERK op basis van empirische validatie</b>

## 2 Familiarisatie

De eerste fase in het koppelingsproces die in de Handleiding van de Raad van Europa wordt beschreven is die van de familiarisatie. De handleiding geeft aan dat het van groot belang is dat iedere betrokkene bij het koppelingsproces zich vertrouwd maakt met het ERK. Dit kan geschieden door middel van zelfstudie, maar in het koppelingsproject van de examens Arabisch, Russisch, Spaans en Turks dienden de verschillende medewerkers overeen te stemmen in hun interpretatie van categorieën en niveaus van het ERK. De familiarisatie heeft plaats gehad in een sessie waarin de toetsdeskundigen van het Cito hun interpretaties van het ERK met elkaar bespraken en waarin getracht is te komen tot één visie op het ERK.

Alle medewerkers aan het project hebben een academische graad in een of meer van de betrokken vreemde talen en een eerste inventarisatie wees uit dat iedere medewerker in ten minste één vreemde taal functioneerde op minimaal B2-niveau.

Het familiarisatieproces werd uitgevoerd in vijf fasen:

1. Familiarisatie met de algemene en specifieke doelen en de functies van het ERK. Voor dit doel werd het eerste deel van het ERK onder de inhoudelijke medewerkers verspreid ten behoeve van zelfstudie.
2. Discussie naar aanleiding van de vragen die aan het einde van ieder hoofdstuk in het ERK worden gesteld over de relevantie van het betreffende hoofdstuk voor de werksituatie van de gebruiker van het ERK. De discussie werd toegespitst op de relevantie van het ERK bij de ontwikkeling van curricula, syllabi en toetsen.
3. Discussie over de globale descriptorren in het ERK en de niveaus die daarbij horen. De medewerkers maakten een eerste, voorlopige koppeling van Nederlandse onderwijsniveaus aan het ERK. De medewerkers kregen een oefening te doen waarbij zij de globale descriptorren dienden toe te wijzen aan het betreffende ERK-niveau.
4. Het inschatten van het eigen vaardigheidsniveau van de medewerkers in twee talen, met behulp van de matrix voor zelfevaluatie (tabel 2 in het ERK). De eerste taal betrof de taal waarin zij zich door hun studie hadden gekwalificeerd, de tweede taal een taal naar eigen keuze waarin zij eveneens konden functioneren. Over deze inschattingen is uitvoerig gediscussieerd.
5. Het sorteren van meer specifieke descriptorren behorende bij de vaardigheid lezen, de vaardigheid die getoetst wordt in de centrale eindexamens. De descriptorren waren genomen uit Bijlage C1 bij het ERK en ontwikkeld in het kader van het project DIALANG.



## 3 Inhoudelijke specificatie

### 3.1 Inleiding

De tweede fase in het koppelingproces zoals beschreven in de Handleiding en uitgevoerd in het Nederlandse koppelingsonderzoek betreft de specificatiefase. In de Handleiding staat beschreven wat specificatie in het koppelingsproces inhoudt: het beschrijven van de mate waarin een examen de categorieën en de niveaus van het ERK bestrijkt. De Handleiding onderscheidt twee soorten van beschrijving:

1. een beschrijving van het examen in algemene termen;
2. een inhoudsanalyse van het examen in termen van het ERK.

In het hier beschreven koppelingsonderzoek hebben Cito en SLO de eerste activiteit als volgt geïnterpreteerd: er wordt een kwalitatieve analyse gemaakt van het examenprogramma voor de vreemde talen. Deze analyse wordt voor ieder descriptor in het ERK gemaakt in termen van de bestaande, Nederlandse examenspecificaties. Deze analyse is uitgevoerd door SLO en verwerkt in de *Examenprogramma Moderne Vreemde Talen* (CEVO, 2006).

In dit hoofdstuk wordt verslag gedaan van de inhoudsanalyse van de centrale examens leesvaardigheid in termen van het ERK. Deze inhoudelijke analyse van de examens bestaat uit een beschrijving van de examens aan de hand van de communicatieve activiteitenschalen en de schalen voor communicatieve competentie van het ERK.

Bij de specificatie van de examenteksten en –opgaven zijn de volgende stappen te onderscheiden.

- Een beschrijving van de teksten en opgaven aan de hand van de vier globale descriptoren en de daarbij horende gedetailleerde descriptoren van Taalprofielen (Liemberg & Meijer, 2004).
- Deze analyse is voor de talen in dit rapport - Arabisch, Russisch, Spaans en Turks - niet uitgevoerd. De reden hiervoor is dat uit de analyse voor Frans, Duits en Engels (Noijons & Kuijper, 2006) is gebleken dat het overgrote deel van de opgaven betrekking heeft op de schaal 'lezen om informatie op te doen'. Een globale analyse van de examens Arabisch, Russisch, Spaans en Turks wees uit dat dit ook voor deze examens geldt. Het uitvoeren van een nadere analyse werd daarom niet meer nodig geacht.
- Een beschrijving van teksten en opgaven met een op het ERK gebaseerd beschrijvingsmodel, het *Dutch Grid*, zoals deze ook is uitgevoerd bij de examens Frans, Duits en Engels. Voor een beschrijving van het *Dutch Grid* zie paragraaf 3.3.2.
  - Allereerst zijn de teksten beschreven aan de hand van de volgende contextvariabelen uit het ERK: tekstbron, tekstsoort, communicatieve thema's en domein. Dit komt in paragraaf 3.2 aan de orde..
  - Vervolgens is er met behulp van de *Dutch Grid* een analyse van de linguïstische en cognitieve complexiteit van de examenteksten en -opgaven uitgevoerd. In dit deel van de specificatie ligt de nadruk op de schalen voor communicatieve competentie. Dit deel van de specificatiefase wordt beschreven in paragraaf 3.3.

## 3.2 Inhoudelijke specificatie van de examenteksten in relatie tot de contextdimensies

### 3.2.1 Tekstdimensies: tekstbron, tekstsoort, onderwerp en domein

Het ERK bevat een aantal inhoudelijke dimensies om teksten te beschrijven. Deze dimensies zijn opgenomen in het *Dutch Grid*, een beschrijvingsmodel voor lees- en luisteropgaven. Voor een korte beschrijving van het *Dutch Grid* verwijzen we naar paragraaf 3.3.2. In het *Dutch Grid* worden de volgende tekstdimensies gebruikt: tekstbron, tekstsoort en communicatiethema's en domein. Tekstbron, communicatiethema's en domein zijn direct aan het ERK ontleend (resp. p. 49, 51-52 en 45) en kunnen worden beschouwd als generieke dimensies over alle niveaus heen. Tekstsoort is een indeling afkomstig uit DIALANG en is ontwikkeld als een generalisatie van de minder systematische beschrijving van deze dimensie in het ERK. Deze dimensies kunnen gebruikt worden om de variatie van teksten en onderwerpen in examens te beschrijven en te reguleren, afhankelijk van het doel en de functie van de examens. In deze paragraaf worden de examens leesvaardigheid Arabisch, Russisch, Spaans en Turks aan de hand van deze dimensies beschreven. Het aantal teksten dat is geanalyseerd wordt weergegeven in tabel 3.3.

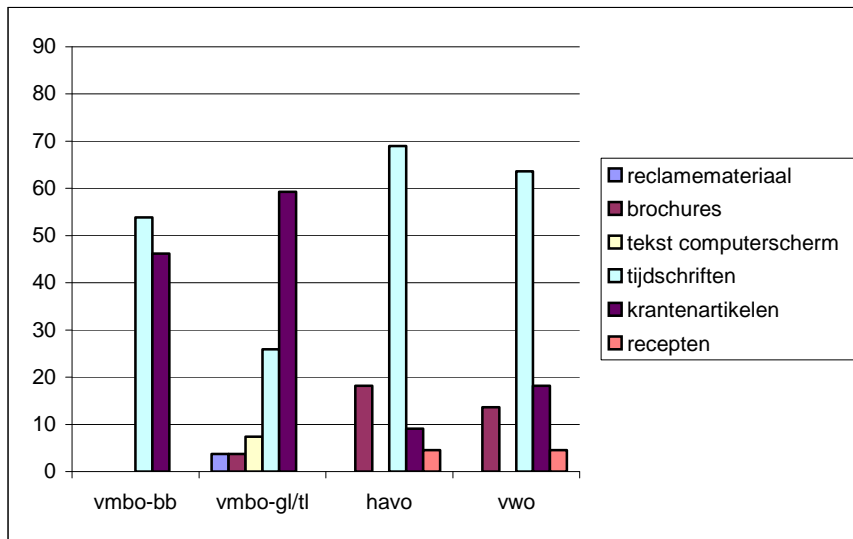
#### **Tekstbron**

Het ERK vermeldt de volgende tekstbronnen.

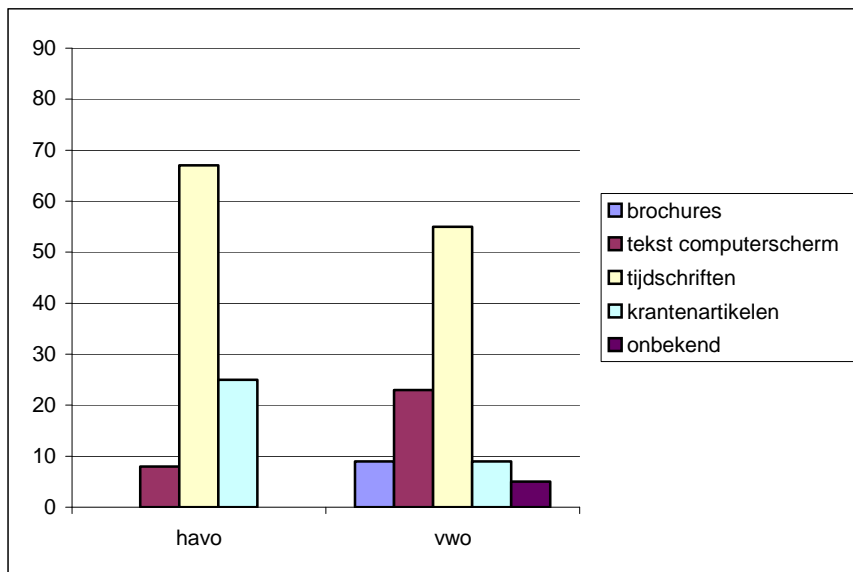
<b>Persoonlijk</b>	<b>Publiek</b>
Teletekst	Aankondigingen en berichten
Garantief formulier	Labels
Recept	Pamfletten, graffiti
Instructiemateriaal	Tickets, dienstregelingen
Roman	Regelgeving
Tijdschrift	Programma's
Krant	Contracten
Junk mail	Menu's
Brochures	Gewijde teksten
Persoonlijke brief	
<b>Beroepsmatig</b>	<b>Onderwijs en opleiding</b>
Zakenbrief	Authentieke teksten
Rapport, memorandum	Schoolboeken
Veiligheidsinstructies	Naslagwerken
Instructiemateriaal	Tekst op schoolbord
Regels	Tekst op sheet
Reclamemateriaal	Tekst op computerscherm
Etiketten e.d.	Videotekst
Beroepsomschrijving	Oefentekst
Richtingaanwijzingen	Oefenmateriaal
Visitekaartjes	Artikelen uit tijdschriften
	Uittreksels
	Woordenboeken

Van alle examenteksten Arabisch, Russisch, Spaans en Turks is aangegeven bij welke van deze tekstbronnen ze horen. Figuur 3-1 t/m 3-4 geven weer welke tekstbronnen ten grondslag liggen aan de teksten voor de centrale examens leesvaardigheid van deze talen.

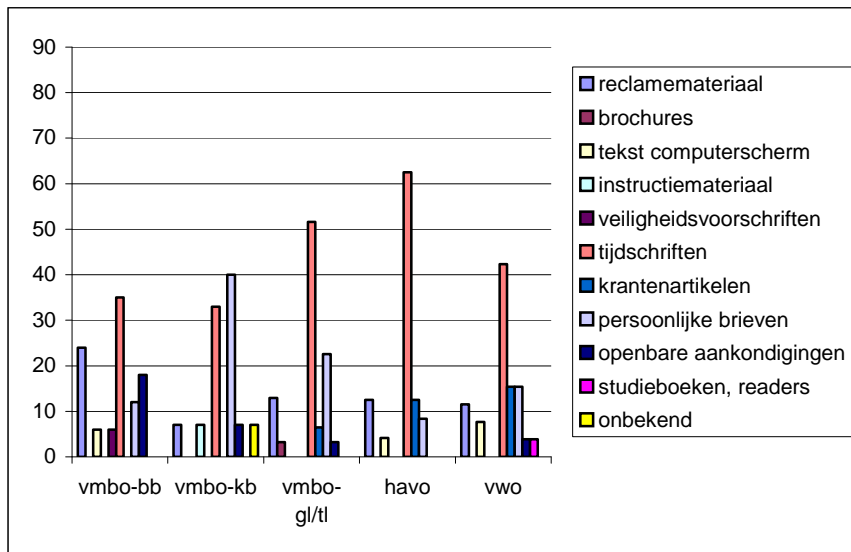
Figuur 3-1 Tekstbron van de teksten in de centrale examens leesvaardigheid Arabisch als percentage



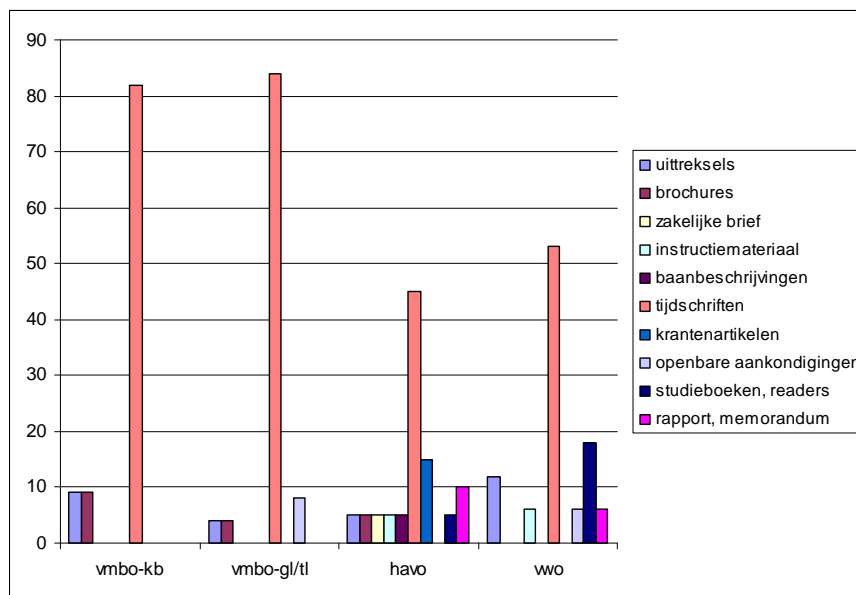
Figuur 3-2 Tekstbron van de teksten in de centrale examens leesvaardigheid Russisch als percentage



Figuur 3-3 Tekstbron van de teksten in de centrale examens leesvaardigheid Spaans als percentage



Figuur 3-4 Tekstbron van de teksten in de centrale examens leesvaardigheid Turks als percentage



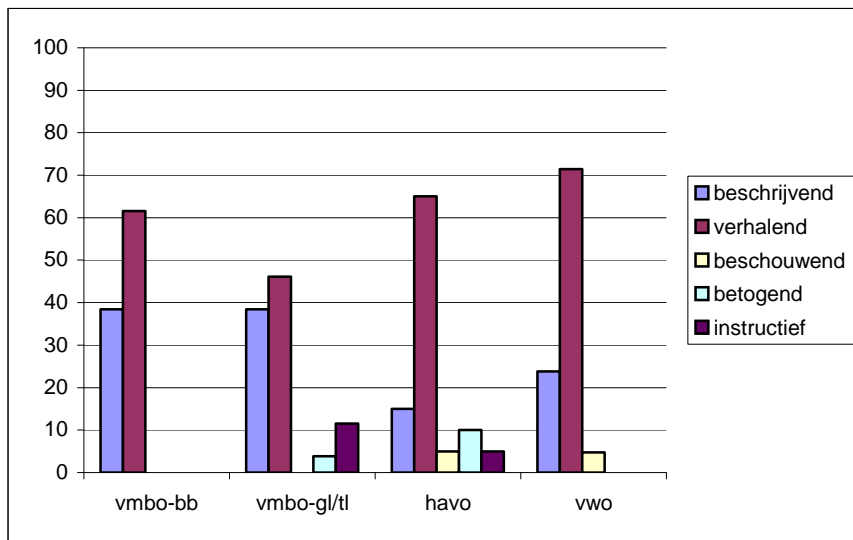
Van de in het ERK genoemde tekstbronnen komen er tussen 4 (Russisch) en 10 tekstsoorten (Turks) voor. In alle examens ligt de nadruk op actuele teksten uit kranten en tijdschriften.

### Tekstsoort

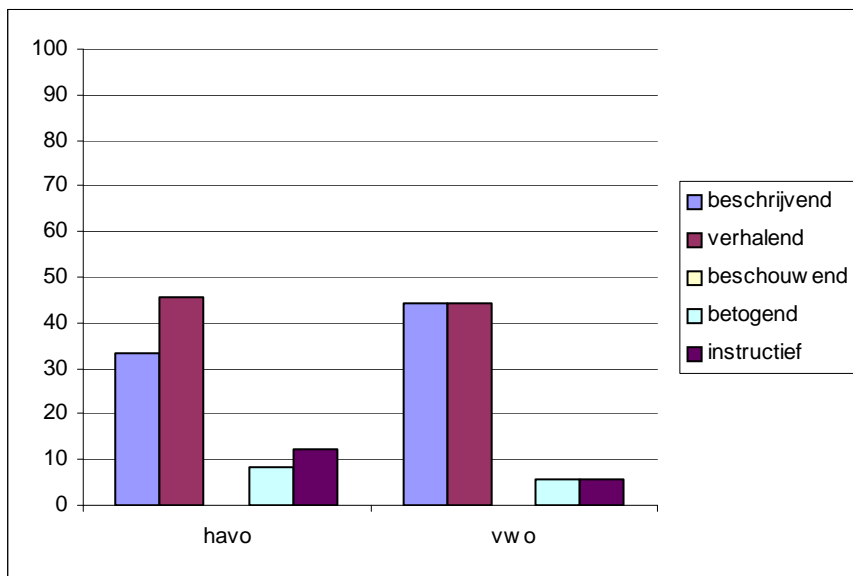
In de leesscalen van het ERK treffen we vanaf de lagere naar hogere niveaus een verschuiving aan van meer feitelijke tekstsoorten naar meer beschouwende en argumentatieve tekstsoorten. Tijdens de specificatiefase zijn de examenteksten ingedeeld naar tekstsoort om na te gaan of deze tendens zich ook in de centrale examens leesvaardigheid voordoet.

Figuren 3-5 t/m en 3-8 geven het resultaat van deze indeling weer voor de examenteksten Arabisch, Russisch, Turks en Spaans.

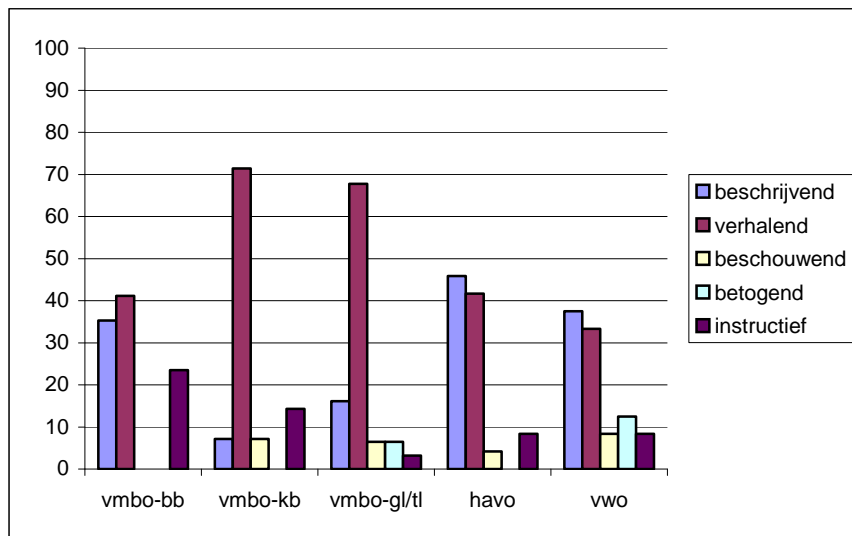
Figuur 3-5 Tekstsoorten per schoolsoort centrale examens leesvaardigheid Arabisch in percentages



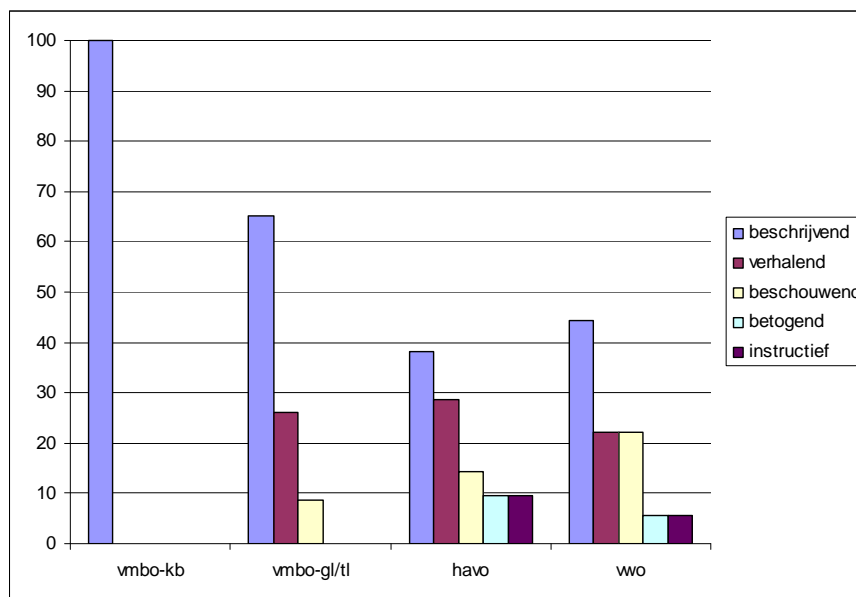
Figuur 3-6 Tekstsoorten per schoolsoort centrale examens leesvaardigheid Russisch in percentages



Figuur 3-7 Tekstsoorten per schoolsoort centrale examens leesvaardigheid Spaans in percentages



Figuur 3-8 Tekstsoorten per schoolsoort centrale examens leesvaardigheid Turks in percentages



Het blijkt dat examens vooral gekenmerkt worden door de meer feitelijke tekstsoorten. In het havo en het vwo neemt het aandeel van de beschouwende en betogende teksten iets toe, hoewel ook hier de meer feitelijke tekstsoorten de overhand hebben. Dit is in lijn met de descriptors in het ERK. Uit de standaardbepalingsprocedure (zie paragraaf 4.3) blijkt dat de vwo-examens voor de vier talen tussen B1 en B2 liggen. Het relatief grote aandeel van de meer feitelijke tekstsoorten op vwo en havo is in overeenstemming met de niveaubeschrijvingen van het ERK.

### Communicatiethema's

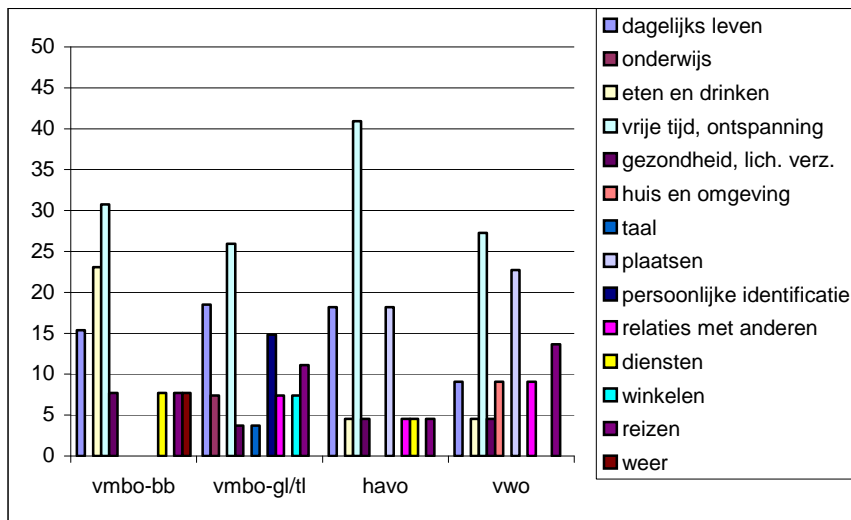
Het ERK (pag.51) bevat de volgende lijst van communicatiethema's.

1 Persoonlijke identificatie	5 Reizen	9 Winkelen
2 Huiselijke en directe omgeving	6 Relaties met anderen	10 Eten en drinken
3 Dagelijks leven	7 Gezondheid en lichaamsverzorging	11 Diensten
4 Vrije tijd en ontspanning	8 Onderwijs en opleiding	12 Plaatsen
		13 Taal
		14 Weer

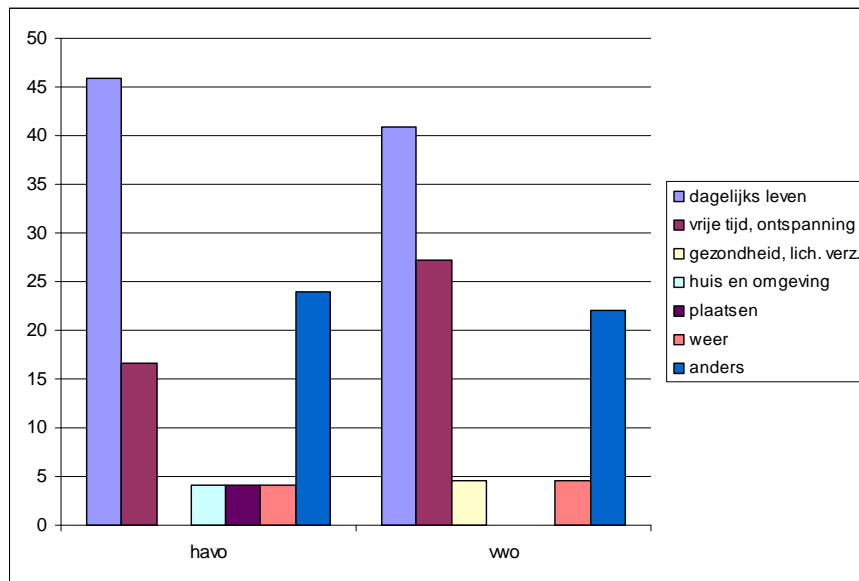
Evenals bij het ERK-overzicht voor tekstbron pretendeert het ERK niet dat deze lijst volledig en definitief is. Evenmin zijn de categorieën wederzijds uitsluitend of voorschrijvend. Er is er bij de beschrijving voor gekozen een tekst toe te wijzen aan de categorie waar deze tekst het meest thuishoort.

Figuren 3-9 t/m 3-12 geven een overzicht van de mate waarin deze communicatiethema's in de centrale examens leesvaardigheid Arabisch, Russisch, Turks en Spaans voorkomen.

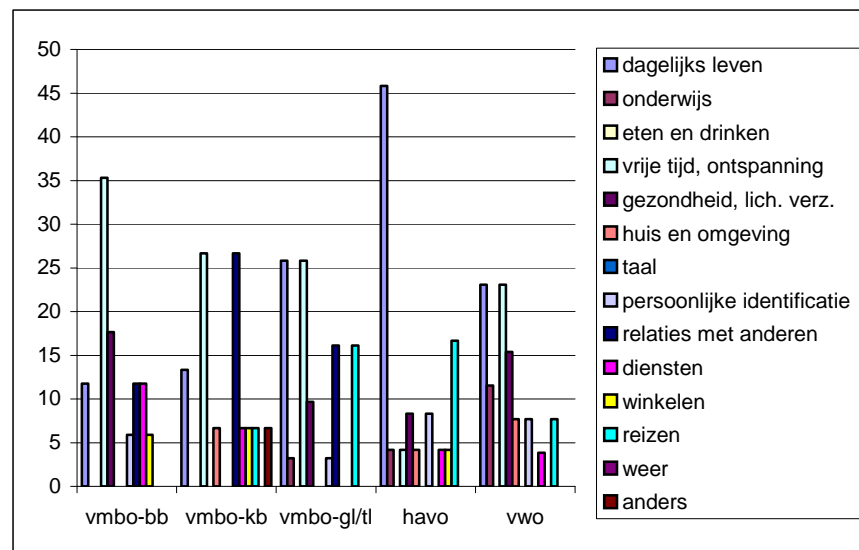
Figuur 3-9 Communicatiethema's in de centrale examens leesvaardigheid Arabisch per schoolsoort in percentages



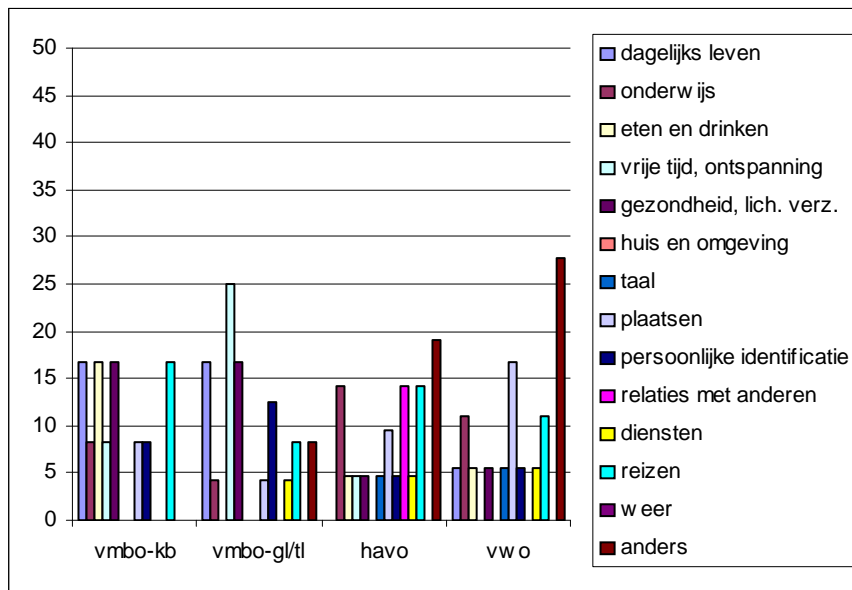
Figuur 3-10 Communicatiethema's in de centrale examens leesvaardigheid Russisch per schoolsoort in percentages



Figuur 3-11 Communicatiethema's in de centrale examens leesvaardigheid Spaans per schoolsoort in percentages



Figuur 3-12 Communicatiethema's in de centrale examens leesvaardigheid Turks per schoolsoort in percentages



Het blijkt dat vrijwel alle communicatiethema's uit het ERK in de centrale examens leesvaardigheid voorkomen, zij het dat niet in elk examen alle thema's vertegenwoordigd zijn. Met name bij Turks komt de categorie 'anders' relatief veel voor. Hier werden de volgende thema's genoemd als verkeer, politiek, dieren, sport, handel, migratie, literatuur, computeren en literatuur.

### Domeinen

Het ERK (p. 45) bevat vier domeinen:

- persoonlijk
- publiek
- werk
- onderwijs

Uit de indeling van de teksten in deze domeinen blijkt dat het overgrote deel van de examens betrekking heeft op het persoonlijke en het publieke domein.

### 3.2.2 Conclusies en aanbevelingen

In paragraaf 3.2.1 is verslag gedaan van de specificatie van de opgaven in de centrale examens leesvaardigheid Arabisch, Russisch, Spaans en Turks.

- Er is geen uitvoerige analyse gedaan van de verdeling van de opgaven over de vier globale descriptor en de daarbij behorende specifieke descriptor. Een globale beschouwing van de opgaven in de examens Arabisch, Russisch, Spaans en Turks bevestigden echter het beeld van de analyses die eerder uitgevoerd waren voor de examens Frans, Duits en Engels (Noijons & Kuijper, 2006): de opgaven gericht op lezen om informatie op te doen hebben het grootste aandeel. Dit heeft te maken met de voorschriften in de examenmodellen, die juist aan dit soort lezen het grootste gewicht toekennen.
- Anderzijds is het huidige examenprogramma en het daaruit afgeleide examenmodel (zie bijlage 1) in feite heel ruim wat de mogelijkheden voor teksten en taken betreft. De weidsheid van het uitgangsmateriaal genoemd in de descriptor in Taalprofielen komt overeen met de grote variatie aan tekstmateriaal dat in principe voor de eindexamens gebruikt kan worden. Deze variatie wordt bevestigd door het feit dat de examenteksten wat betreft tekstsoort, tekstbron en communicatiethema's een ruim scala vertegenwoordigen van wat in het ERK beschreven wordt.
- Juist deze mogelijke grote variatie maakt de noodzaak van een uitgebreide revisie van de huidige examens niet erg voor de handliggend.

- Hoewel niet alle vier de globale descriptoren in de examens in dezelfde mate lijken voor te komen, kan men veronderstellen dat de vaardigheden die beschreven worden in de gedetailleerde descriptoren in een redelijke afspiegeling van het ERK in de examens voor te komen. Dit heeft te maken met het feit dat de leesvaardigheden die onder de globale descriptoren beschreven worden een aanzienlijke overlap vertonen. Wanneer men bijvoorbeeld constateert dat het lezen van correspondentie weinig voorkomt hoeft dit niet te betekenen dat de vaardigheden, beschreven in de gedetailleerde descriptoren onder correspondentie lezen, niet in het examen vertegenwoordigd zijn. Zowel bij correspondentie lezen als bij lezen om informatie op te doen moeten kandidaten in staat zijn hoofdzaken en detailinformatie uit teksten te begrijpen. Het is aannemelijk dat wanneer een kandidaat over deze de vaardigheid beschikt bij het lezen van een tijdschriftartikel – zoals genoemd bij ‘lezen om informatie op te doen’ – hij of zij deze vaardigheid ook heeft bij het lezen van een correspondentietekst. Het verschil ligt dus niet in de leesvaardigheid, maar in de tekstsoort.
- Bij de vaststelling van de examenmodellen zou niettemin overwogen kunnen worden of het nodig is ontbrekende descriptoren meer expliciet in aanmerking te laten komen voor opname in de examens. Het gaat hierbij met name om zakelijk getinte correspondentie, oriënterend lezen en het lezen van instructies.

### **3.3 De inhoudelijke specificatie in relatie tot de schalen voor communicatieve competentie**

#### **3.3.1 Inleiding**

De niveaus in het ERK beschrijven een steeds grotere mate van taalvaardigheid. Taalleerders kunnen, naarmate hun niveau stijgt, met steeds meer tekstsoorten in steeds meer domeinen en situaties omgaan. Tevens groeit hun vaardigheid om met toenemende precisie met linguïstisch en cognitief steeds moeilijkere teksten uit de voeten te kunnen. Dit laatste aspect van de taalvaardigheid wordt met name beschreven in de schalen voor communicatieve competenties op p. 108-130 van het ERK en daarbinnen met name in de schalen voor communicatieve linguïstische competentie op p. 108-118.

In de volgende paragrafen wordt informatie gegeven over de linguïstische en cognitieve complexiteit van teksten en opgaven van de centrale examens leesvaardigheid Arabisch, Russisch, Spaans en Turks in relatie tot het ERK.

In paragraaf 3.3.2 wordt de werkwijze in deze fase van de specificatie toegelicht. In paragraaf 3.3.3 en 3.3.4 wordt verslag gedaan van de resultaten van de analyse van achtereenvolgens de leesteksten en de opgaven. Paragraaf 3.3.5 bevat een samenvatting en conclusies.

#### **3.3.2 Werkwijze**

In het ERK worden onder meer de inhoudelijke aspecten en de niveaus van taaltaken omschreven. De taalontwikkeling van een lager naar een hoger niveau kan beschreven worden als een toenemende vaardigheid om steeds meer taaltaken bij tekstsoorten van een steeds hogere moeilijkheidsgraad uit te voeren.

De vraag is in hoeverre de examens leesvaardigheid vanaf vmbo tot vwo conform de verwachtingen op grond van de standaardbepaling (zie hoofdstuk 4) teksten en opgaven bevatten die oplopen in moeilijkheidsgraad. Om deze vraag te kunnen beantwoorden is een ERK-gerelateerd beschrijvingsmodel nodig waarmee de teksten en opgaven kunnen worden beschreven.

Een onderzoek naar dergelijke dimensies in het ERK is uitgevoerd in het ‘*Dutch Grid-project*’ (Alderson e.a., 2004 en 2005). In dit Europese project is het ERK geanalyseerd op de mate waarin het duidelijke aanwijzingen en richtlijnen bevat voor de beschrijving en ontwikkeling van toetstaken op de verschillende ERK-niveaus. Alderson e.a. (2004, pag. 3) concluderen dat het ERK een bruikbaar instrument is, maar dat rechtstreekse richtlijnen voor toetsconstructie en beschrijving van toetsen op de verschillende niveaus niet zonder meer in het ERK te vinden zijn. Zij constateren de volgende problemen die met name voor de toetsconstructie op grond van het ERK gelden (Alderson e.a, p. 11):

1. Terminologische problemen: synoniemen of niet?
  2. Omissies, wanneer een concept of kenmerk dat nodig is voor testspecificatie eenvoudigweg ontbreekt.
  3. Inconsistenties, wanneer een kenmerk op het ene niveau wel en op een ander niveau niet wordt vermeld, waar een zelfde kenmerk voorkomt op twee verschillende niveaus, of wanneer op hetzelfde niveau een kenmerk in verschillende schalen verschillend wordt beschreven.
  4. Het ontbreken van definities, wanneer termen worden gepresenteerd, maar niet gedefinieerd.
- De problemen zijn er dus niet zozeer in gelegen dat beschrijvingscriteria in het ERK helemaal ontbreken, maar meer in het feit dat er sprake is van een gebrek aan explicitering, systematiek, consistentie en precisie, aspecten die voor toetsconstructie onontbeerlijk zijn.

Op basis van een grondige analyse van het ERK is er in het *Dutch Grid*-project een aan het ERK gerelateerd beschrijvingsmodel ontwikkeld voor lees- en luisteropgaven en -teksten dat op een meer systematische manier de relevante dimensies van het ERK probeert te beschrijven. Dit beschrijvingsmodel is *web-based* beschikbaar en toegankelijk via: [www.ling.lancs.ac.uk/CEFgrid](http://www.ling.lancs.ac.uk/CEFgrid).

Het *Dutch Grid* bevat beschrijvingsdimensies voor teksten en opgaven in lees- en luistertoetsen. De tekstdimensies kunnen onderverdeeld worden in een inhoudelijke categorie en in een categorie met betrekking tot cognitieve en linguïstische complexiteit. Tabel 3-1 geeft een overzicht van deze dimensies. Voor elke dimensie wordt in kolom 2 aangegeven of de beschrijvingssystematiek ervan rechtstreeks uit het ERK komt, een bewerking van het ERK is of uit een andere systematiek afkomstig is.

Tabel 3-1 Beschrijvingsdimensies van het Dutch Grid voor Lezen

Tekst dimensies	
	bron
Inhoudelijk	
Tekstbron	rechtstreeks uit ERK
Tekstsoort	DIALANG
Domein	rechtstreeks uit ERK
Onderwerp	rechtstreeks uit ERK
Cognitieve en linguïstische complexiteit	
Abstractie niveau	bewerkt obv ERK
Woordenschat	bewerkt obv ERK
Grammaticale complexiteit	bewerkt obv ERK
Tekstlengte	bewerkt obv ERK
ERK-niveau-inschatting	rechtstreeks uit ERK
Leesopgave dimensies	
Vraagtype	bewerkt obv ERK
Operaties	bewerkt obv ERK
ERK-niveau-inschatting	rechtstreeks uit ERK

In het huidige onderzoek is het *Dutch Grid* gebruikt om een eerste beschrijving te geven van de opgaven uit de examens leesvaardigheid. Deze beschrijving is ook gebruikt bij de selectie van examenopgaven voor de standaardisatie (zie hoofdstuk 4).

De analyses zijn uitgevoerd op de centrale examens leesvaardigheid Arabisch, Russisch, Spaans en Turks. Tabel 3.2 geeft een overzicht van de examens die in het onderzoek zijn opgenomen. Een lege cel geeft aan dat dit examen niet voor deze schoolsoort is gemaakt.

Tabel 3-2 Geanalyseerde examens

	VMBO-BB	VMBO-KB	VMBO-GT	havo	vwo
Arabisch	2004*		2004, 2005	2004, 2005	2004, 2005
Russisch				2004, 2005	2004, 2005
Spaans	2006	2006	2004, 2005	2004, 2005	2004, 2005
Turks		2003	2003, 2004	2003, 2004	2003, 2004

\* Proefexamen

Voor elk van de talen zijn de analyses uitgevoerd door 1 of 2 personen. Indien er 2 personen bij de analyses waren betrokken, heeft 1 persoon de analyses voor de vmbo-examens uitgevoerd, de andere voor de havo- en vwo-examens.

In de handleiding van het *Dutch Grid* wordt aanbevolen elke tekst en opgave door meerdere personen en in onderlinge discussie uit te laten voeren. Gezien het grote aantal teksten en opgaven (zie Tabel 3-3) is hier vanwege financiële en organisatorische problemen van afgezien. Niettemin achten we de resultaten voldoende waardevol om ze hier te rapporteren. De resultaten kunnen als indicatief beschouwd te worden. Er zijn duidelijke trends te onderkennen en de informatie is bruikbaar bij de verdere constructie van de examens.

Tabel 3-3 Aantal geanalyseerde leesteksten en –opgaven

examen	aantal teksten				aantal opgaven			
	Arabisch	Russisch	Spaans	Turks	Arabisch	Russisch	Spaans	Turks
vmbo-bb	13		17		30		44	
vmbo-kb			15	12			46	39
vmbo-gl/tl	27		31	25	88		87	83
havo	22	24	24	22	84	86	86	84
vwo	22	22	26	21	81	88	90	90
totaal	84	46	113	80	283	174	353	296
totaal generaal	323				1106			

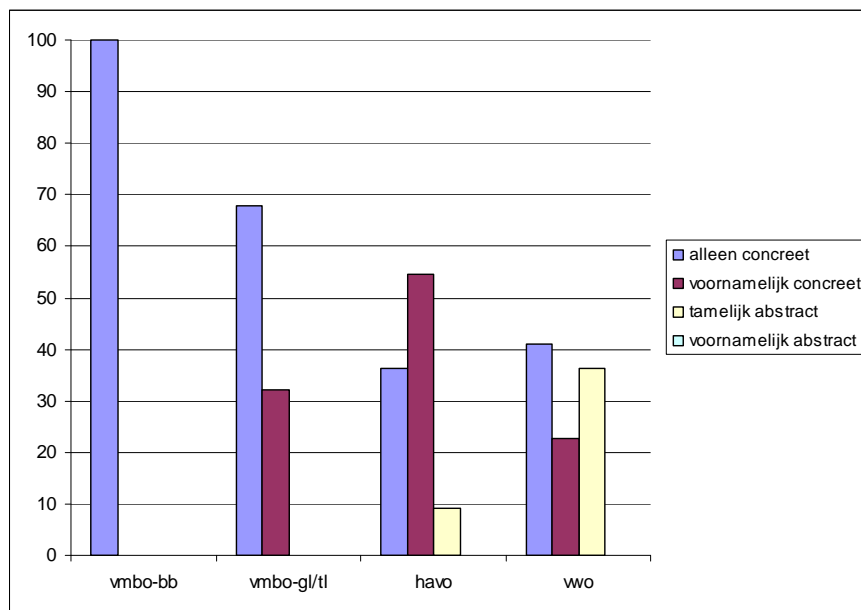
### 3.3.3 De linguïstische en cognitieve complexiteit van de teksten

In deze paragraaf wordt een overzicht gegeven van de linguïstische en cognitieve complexiteit van de examenteksten aan de hand van de mate van abstractie, de woordenschat, de grammatica, en de lengte van teksten.

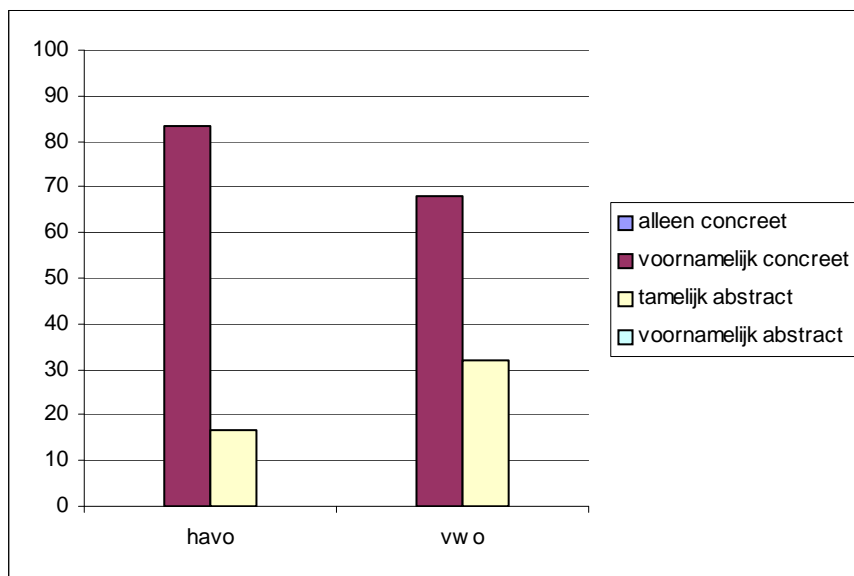
#### **Mate van abstractie van de teksten**

Wanneer men de omschrijvingen van de ERK-niveaus in ogenschouw neemt kan men hieruit opmaken dat de mate van abstractie van teksten toeneemt van niveau A1 tot niveau C2. Volgens de resultaten van de standaardbepaling neemt het niveaubereik van de centrale examens leesvaardigheid Arabisch, Russisch, Spaans en Turks toe van A2 tot B1/B2 (zie paragraaf 4.3). Willen de examens deze opbouw in het ERK reflecteren, dan moet deze toename van abstractie in de examens terugkomen. Figuur 3-13 t/m 3-16 geven een indicatie van de mate van abstractie van de teksten in de centrale examens leesvaardigheid van de vier onderzochte talen.

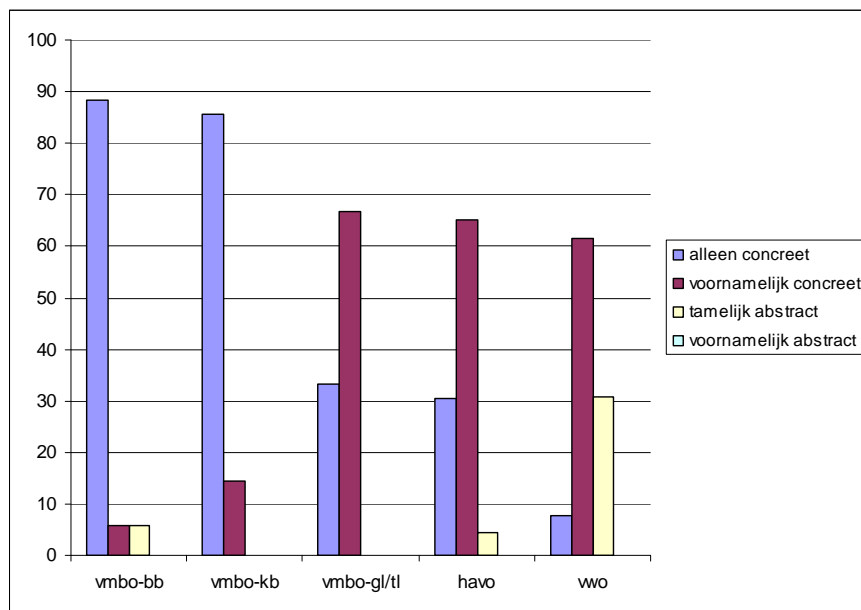
Figuur 3-13 Mate van abstractie van de examenteksten Arabisch in percentages



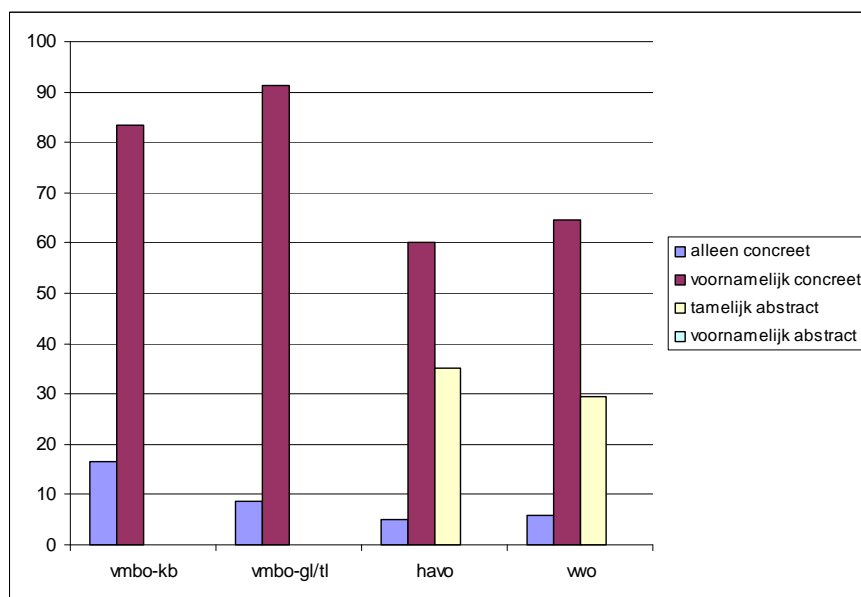
Figuur 3-14 Mate van abstractie van de examenteksten Russisch in percentages



Figuur 3-15 Mate van abstractie van de examenteksten Spaans in percentages



Figuur 3-16 Mate van abstractie van de examenteksten Turks in percentages



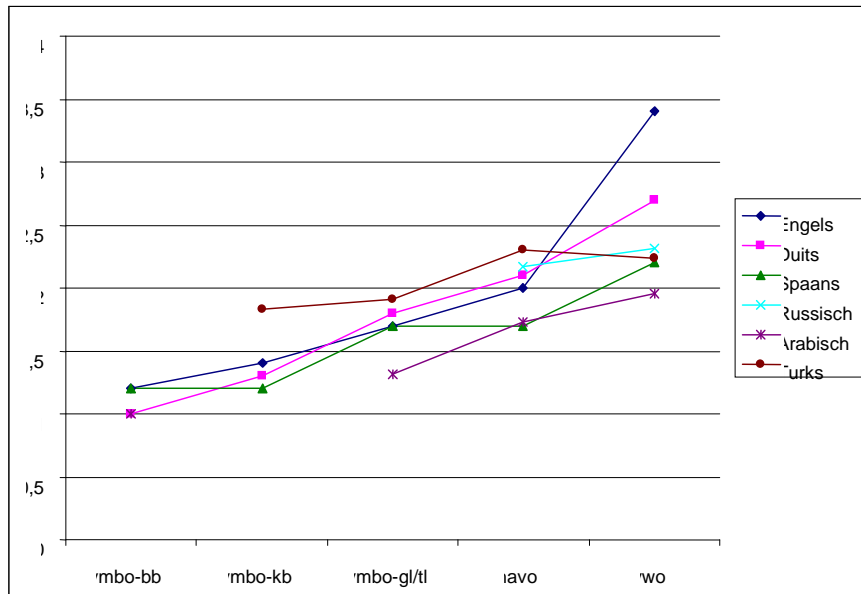
Het percentage teksten met een concrete inhoud neemt van vmbo tot vwo inderdaad af, terwijl er een toename is van teksten met een meer abstractere inhoud. Dit geldt voor alle schooltypes, hoewel er verschillen tussen de talen zijn.

Wanneer we aan de mate van abstractie de volgende waarden toekennen, kunnen we een abstractiescore berekenen voor elke schoolsoort.

Alleen concreet	1
Voornamelijk concreet	2
Tamelijk abstract	3
Voornamelijk abstract	4

Figuur 3-17 geeft deze abstractiescores weer voor de schoolsoorten. In deze grafieken zijn ook de resultaten weergegeven voor Engels en Duits. Deze zijn overgenomen uit Noijons e.a., 2006, pag. 25)

*Figuur 3-17 Abstractiescores van examenteksten lezen Arabisch, Russisch, Spaans, Turks, Engels en Duits*

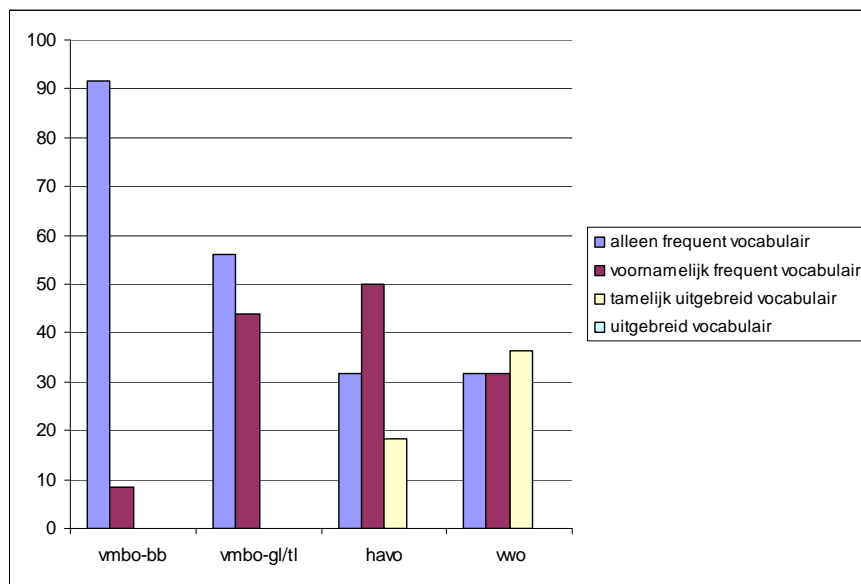


Alle examens blijken de in het ERK beschreven ontwikkeling te weerspiegelen. De abstractie van de examenteksten neemt toe van vmbo naar vwo, hoewel de mate waarin verschilt per taal. Vergeleken met Engels en Duits blijken de teksten in de hier geanalyseerde talen ook op vwo-niveau een meer concrete inhoud te hebben.

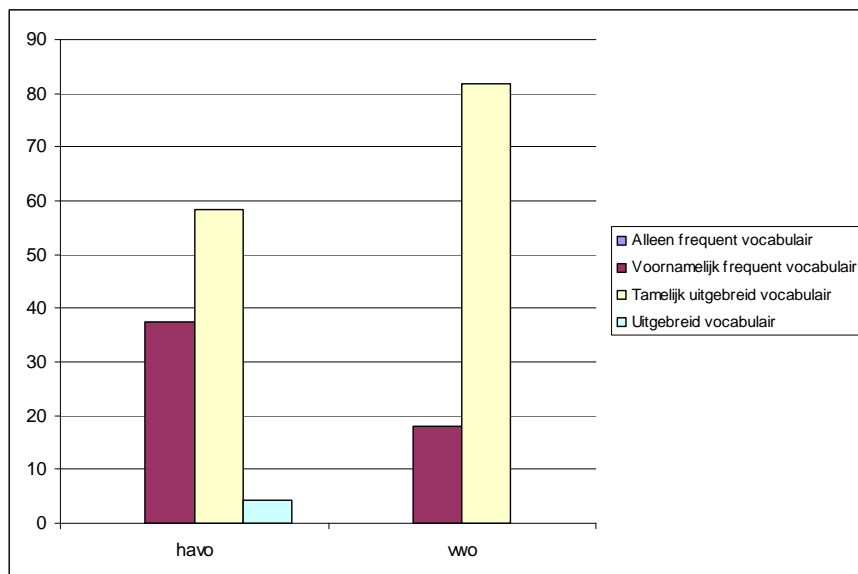
### **Woordenschat**

Volgens het ERK worden leesteksten, naarmate het ERK-niveau hoger worden, gekenmerkt door een steeds uitgebreidere woordenschat. Willen de examens deze opbouw in het ERK reflecteren, dan moet deze tendens zich dus ook in de examens van vmbo tot vwo voordoen. Figuur 3-18 t/m 3-21 geven de aard van de woordenschat in de examenteksten weer.

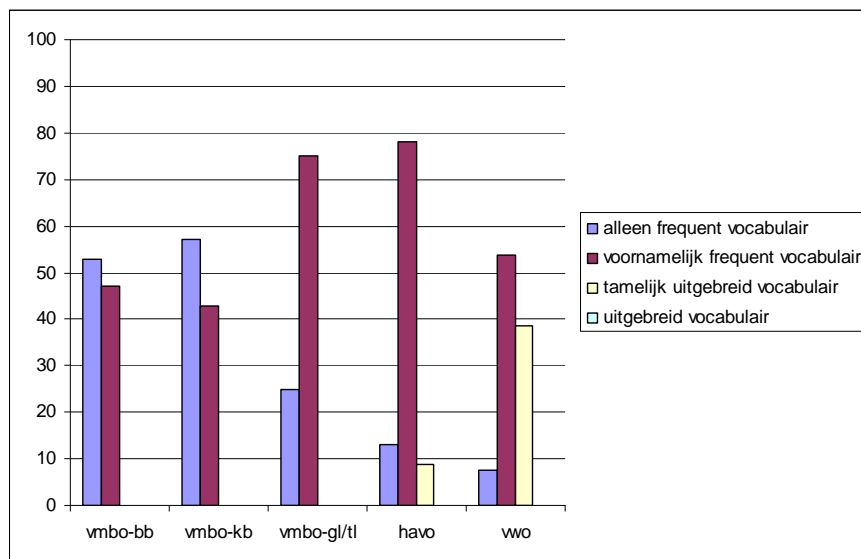
Figuur 3-18 Woordenschat van de examenteksten Arabisch in percentages



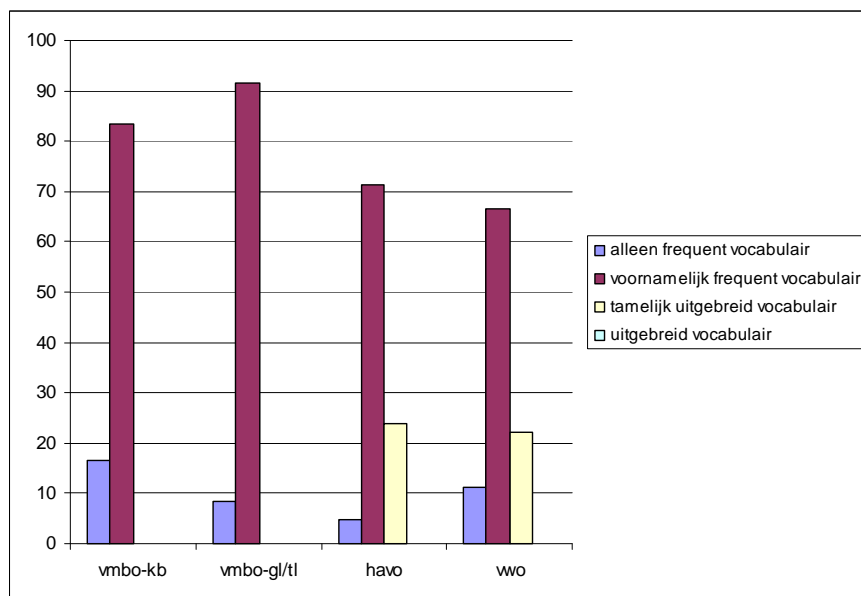
Figuur 3-19 Woordenschat van de examenteksten Russisch in percentages



Figuur 3-20 Woordenschat van de examenteksten Spaans in percentages



Figuur 3-21 Woordenschat van de examenteksten Turks in percentages



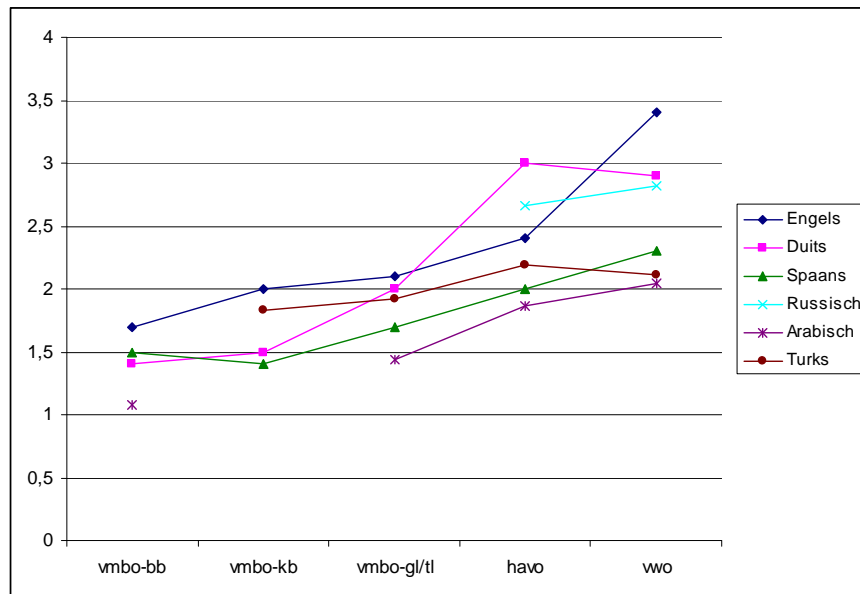
Het blijkt dat de in het ERK verwachte tendens in de examens wordt teruggevonden. De examenteksten bevatten vanaf vmbo tot vwo een steeds uitgebreidere woordenschat.

Wanneer we aan de gebruikte woordenschat de volgende waarden toekennen aan, kunnen we een woordenschatscore berekenen voor elke schoolsoort.

Alleen frequent vocabulaire	1
Voornamelijk frequent vocabulaire	2
Tamelijk uitgebreid vocabulaire	3
Uitgebreid vocabulaire	4

Figuur 3-22 geeft deze woordenschatcores weer voor de schoolsoorten. In deze grafieken zijn ook de resultaten weergegeven voor Engels en Duits. Deze zijn overgenomen uit Noijons e.a., 2006, pag. 26).

*Figuur 3-22 Woordenschatcores van examenteksten lezen Arabisch, Russisch, Spaans, Turks, Engels en Duits*



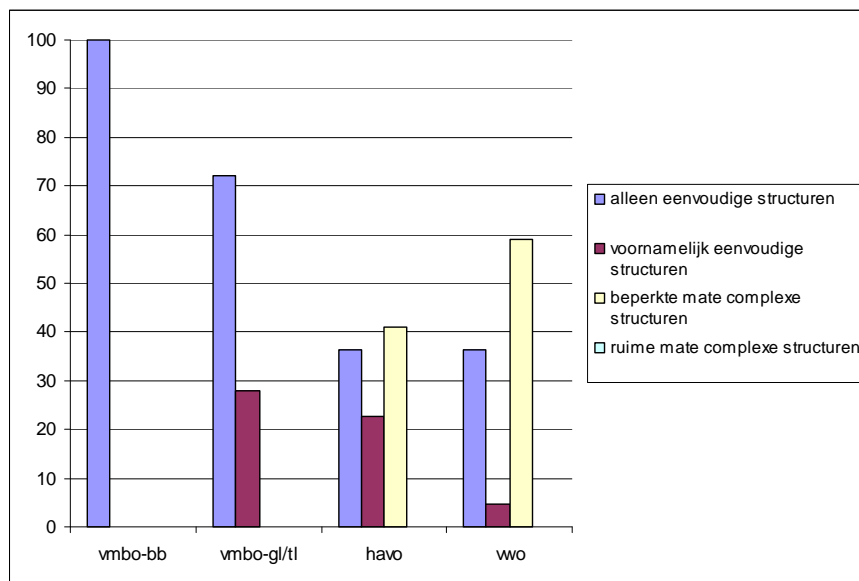
De examens blijken het ERK te weerspiegelen. Vmbo-examens bevatten vooral frequente woorden, terwijl de woordenschat in havo en vwo examens uitgebreider is. De mate waarin verschilt tussen de examens voor de diverse talen. De examenteksten Engels, Duits en Russisch bevatten de meest uitgebreide woordenschat.

### **Grammatica**

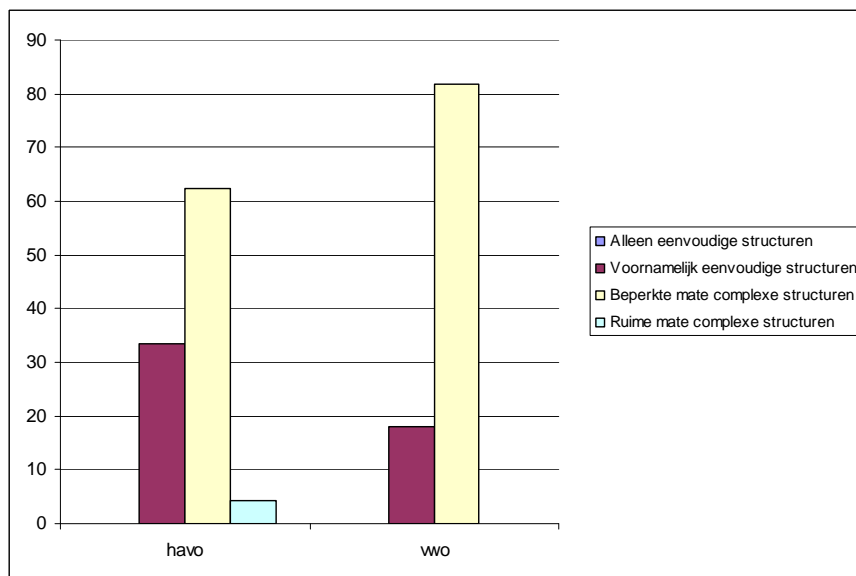
Het ERK veronderstelt een toename van de grammaticale complexiteit van de leesteksten naarmate het ERK-niveau hoger wordt.

Figuren 3-23 tot en met 3-26 geven de grammaticale complexiteit van de leesteksten in de centrale examens leesvaardigheid weer.

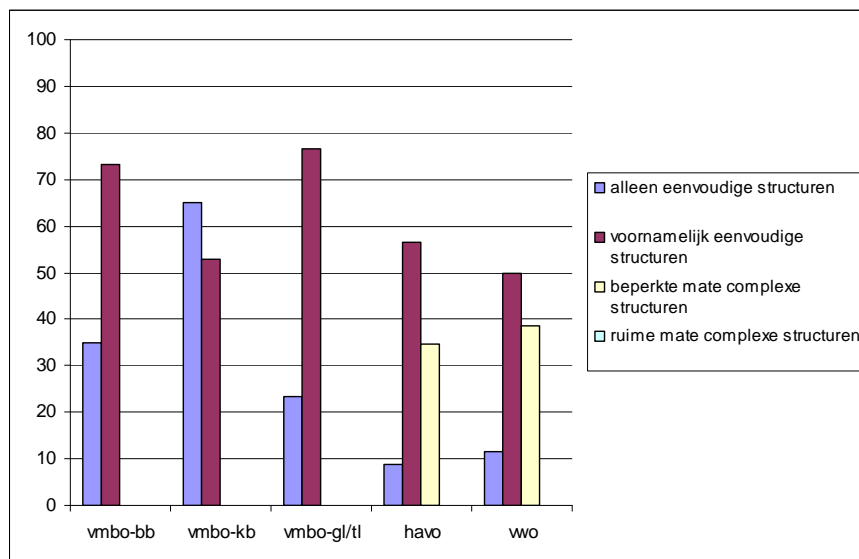
Figuur 3-23 Grammaticale complexiteit van de examenteksten Arabisch in percentages



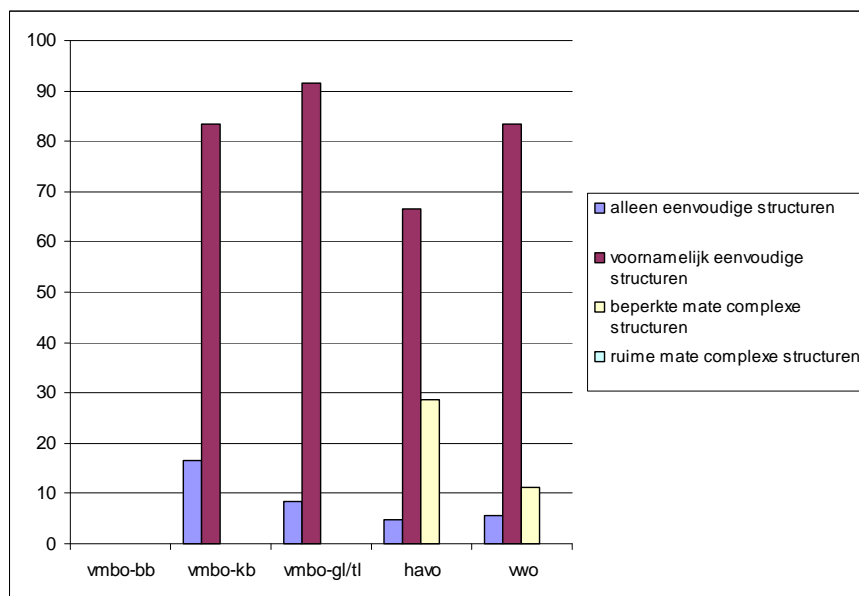
Figuur 3-24 Grammaticale complexiteit van de examenteksten Russisch in percentages



Figuur 3-25 Grammaticale complexiteit van de examenteksten Spaans in percentages



Figuur 3-26 Grammaticale complexiteit van de examenteksten Turks in percentages



Het blijkt dat de in het ERK verwachte tendens in de examens wordt teruggevonden.

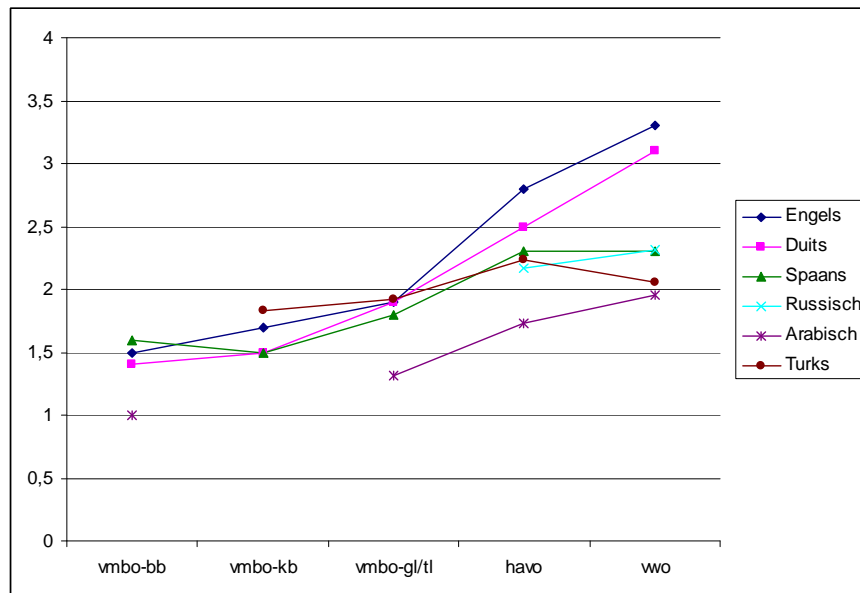
De examenteksten bevatten – met onderlinge verschillen tussen de talen – vanaf vmbo tot wvo meer complexe grammaticale structuren.

Wanneer we de aan de mate van grammaticale complexiteit de volgende waarden toekennen, kunnen we een grammaticascore berekenen voor de examens van elke schoolsoort.

Alleen eenvoudige structuren	1
Voornamelijk eenvoudige structuren	2
Beperkte mate complexe structuren	3
Ruime mate complexe structuren	4

Figuur 3-27 geeft deze grammaticascores weer voor de examens leesvaardigheid van de verschillende schoolsoorten. In deze grafieken zijn ook de resultaten weergegeven voor Engels en Duits. Deze zijn overgenomen uit Noijons & Kuijper (2006, pag. 27).

*Figuur 3-27 Grammaticascores van examenteksten lezen Arabisch, Russisch, Spaans, Turks, Engels en Duits*



De examens blijken het ERK te weerspiegelen. Vmbo-examens bevatten meer teksten met eenvoudige grammaticale structuren, de vwo-teksten meer complexe structuren, terwijl de havo-teksten een tussenpositie innemen. De examens Duits en Engels maken echter met name op vwo-niveau meer gebruik van complexere structuren dan de examens Arabisch, Russisch, Spaans en Turks.

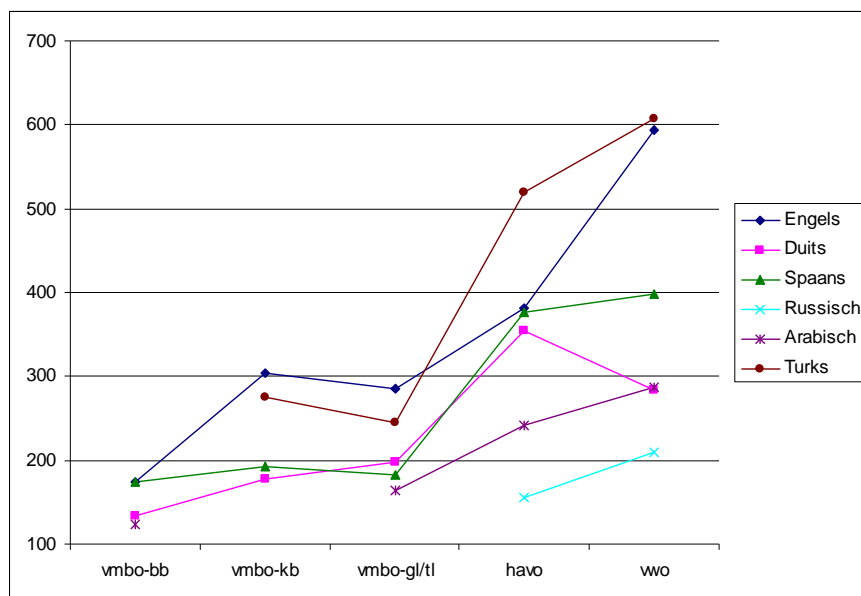
### **Tekstlengte**

Een andere dimensie in het ERK die samenhangt met een toenemende leesvaardigheid naarmate het niveau stijgt, is de lengte van teksten die men moet kunnen begrijpen. Tabel 3-5 t/m en figuur 3-29 geven de tekstlengte van de examenteksten weer voor de verschillende talen. Ter vergelijking zijn in figuur 3-28 ook hier de gegevens voor Engels en Duits opgenomen. Uit Noijons e.a., 2006, pag. 29.

Tabel 3-4 Gemiddelde tekstlengte per schoolsoort in aantal woorden per schoolsoort

	VMBO-BB	VMBO-KB	VMBO-GT-LT	havo	vwo
<b>Arabisch</b>					
Gem.	124		163	241	286
Sd	92,9		133,2	210,1	247,3
Aantal teksten	13		27	22	22
<b>Russisch</b>					
Gem.				156	209
Sd				107	126
Aantal teksten				24	22
<b>Spaans</b>					
Gem.	174	193	182	376	398
Sd	100	119	126	226	236
Aantal teksten	17	15	31	24	26
<b>Turks</b>					
Gem.		275	246	519	607
Sd		174	158	140	154
Aantal teksten		11	24	22	20

Figuur 3-28 Gemiddelde tekstlengte in aantal woorden per schoolsoort



De examenteksten in de examens leesvaardigheid worden voor elke taal van vmbo tot vwo langer. Het vwo en havo hebben gemiddeld langere teksten dan de vmbo-examens. De verschillen tussen talen zijn echter aanzienlijk. Zo zijn de teksten op het vwo bij Engels en Turks eenderde langer dan bij Spaans en twee keer langer dan bij Duits en Arabisch.

### 3.3.4 De complexiteit van de taken

Uit de vorige paragraaf blijkt dat de toename van linguïstische en cognitieve complexiteit in de schalen voor communicatieve competentie van het ERK in de centrale examens leesvaardigheid worden gereflecteerd door de cognitieve en linguïstische kenmerken van de examenteksten. Maar ook de taken die lezers bij een tekst moeten uitvoeren lopen volgens het ERK en de Taalprofielen op in complexiteit. Het *Dutch Grid* geeft ook hiervoor een beschrijvingskader dat

gebaseerd is op het ERK. De leesopgaven kunnen daarmee beschreven worden volgens de dimensies vraagtype en operaties. De aantallen beschreven opgaven zijn te vinden in tabel 3.3.

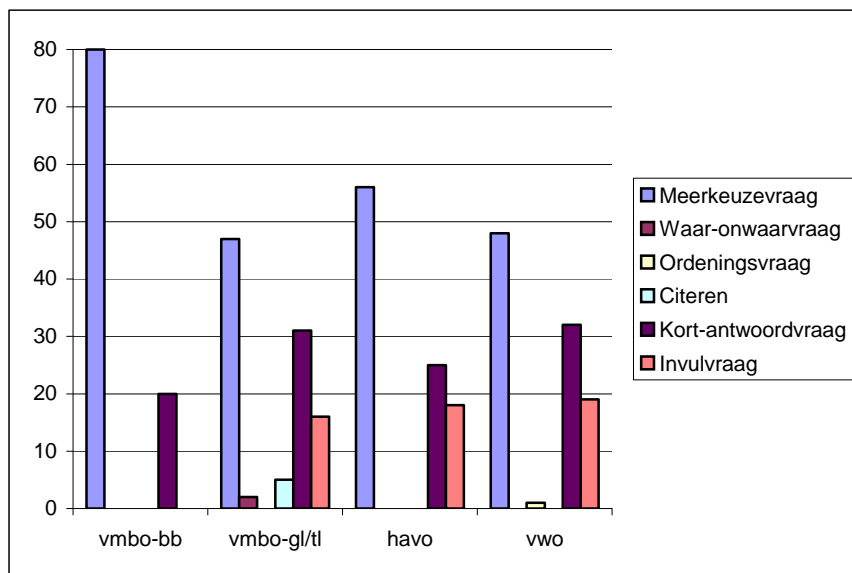
### Vraagtypen

Het *Dutch Grid* onderscheidt de volgende vraagtypen. In het ERK wordt geen relatie aangegeven tussen niveaus en vraagtypen.

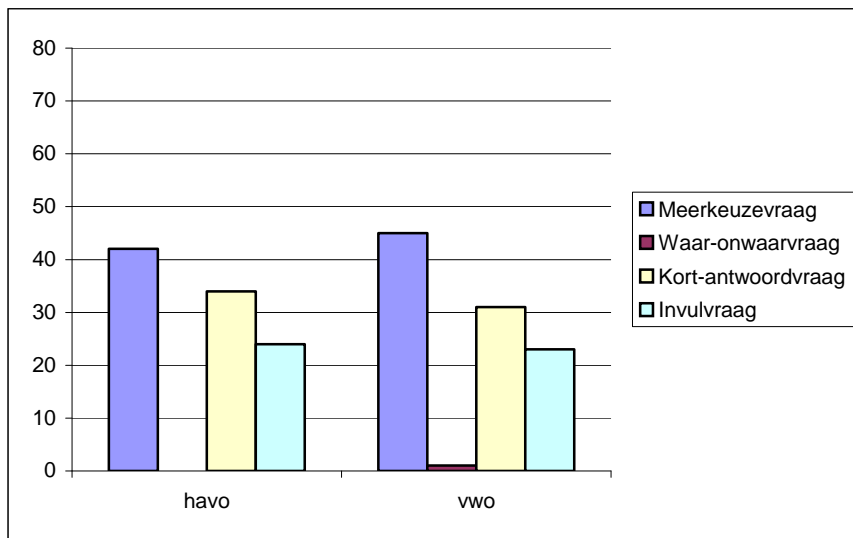
Antwoordtype	Vraagtype
Gesloten antwoord	Meerkeuzevraag
	Waar-onwaarvraag
	Combinatievraag
	Orderingsvraag
	Citeren
Kort open antwoord	Kort-antwoordvraag
	Cloze-vraag
	Invulvraag
	Aanvulvraag
	Samenvatting aanvullen
Lang open antwoord	Essay
	Samenvatting
	Onderbouwen
	Anders

Figuur 3-29 t/m 3.32 geven de percentuele verdeling van de vraagtypen in de centrale examens leesvaardigheid Arabisch, Russisch, Spaans en Turks weer.

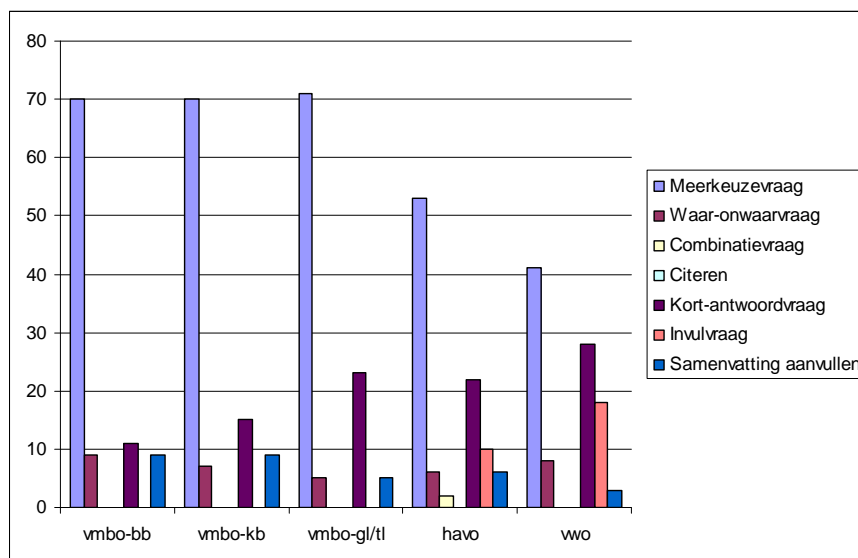
*Figuur 3-29 Vraagtypen in de Examens leesvaardigheid Arabisch in percentages*



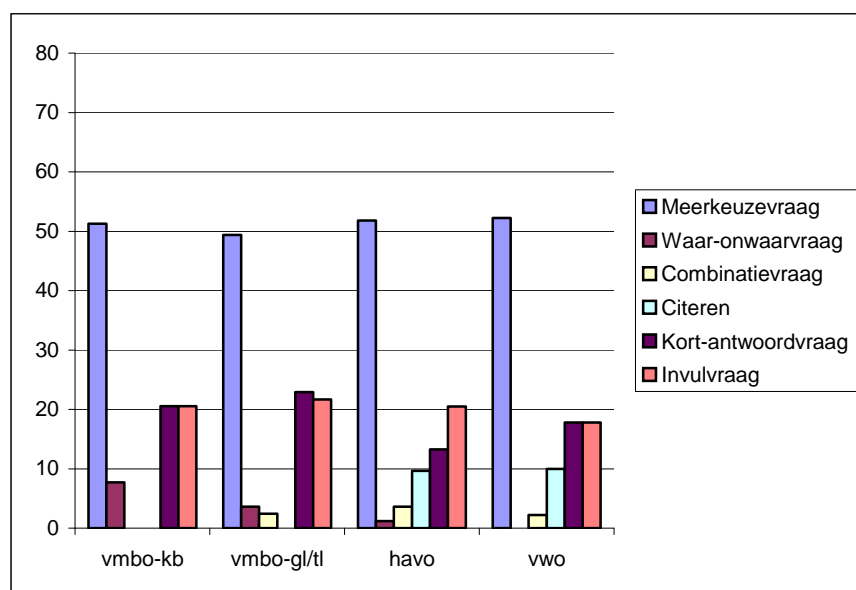
Figuur 3-30 Vraagtypen in de Examens leesvaardigheid Russisch in percentages



Figuur 3-31 Vraagtypen in de Examens leesvaardigheid Spaans in percentages



Figuur 3-32 Vraagtypen in de Examens leesvaardigheid Turks in percentages



In de examens komt een variatie aan gesloten en korte open antwoordvragen voor. Er is geen duidelijke relatie tussen schoolsoort en vraagtype. Een dergelijke relatie wordt in het ERK ook niet verondersteld.

### Operaties

Het ERK veronderstelt dat een lezer naarmate hij een hoger ERK-niveau beheerst een uitgebreider repertoire aan leesoperaties moet kunnen uitvoeren.

Het beschrijvingskader van het *Dutch Grid* voor de leesoperaties bevat drie dimensies:

- De taakdimensie onderscheidt drie soorten operaties: herkennen, het maken van inferenties en evalueren
- De expliciteitsdimensie beschrijft of de informatie waarnaar gevraagd wordt expliciet of impliciet in de tekst is terug te vinden
- De inhoudsdimensie beschrijft waarnaar in de opgave wordt gevraagd.

Hieronder wordt de uitwerking van deze dimensies in het *Dutch Grid* weergegeven. Deze drie dimensies zijn onafhankelijk van elkaar en kunnen in alle combinaties voorkomen.

Taakdimensie	Expliciteitsdimensie	Inhoudsdimensie
Herkennen	Expliciet	Hoofdzaken / grote lijnen Details Opinie
Inferenties maken	Impliciet	Houding van de schrijver Conclusies Communicatieve doel
Evalueren		Tekststructuur / relaties tussen tekstdelen

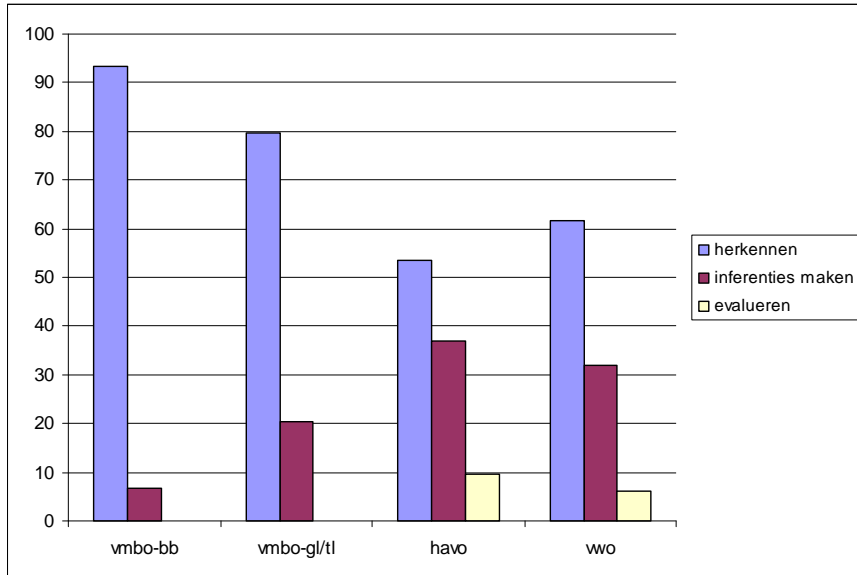
In deze paragraaf worden de opgaven in de centrale examens leesvaardigheid aan de hand van deze taakdimensies beschreven.

### Taakdimensies

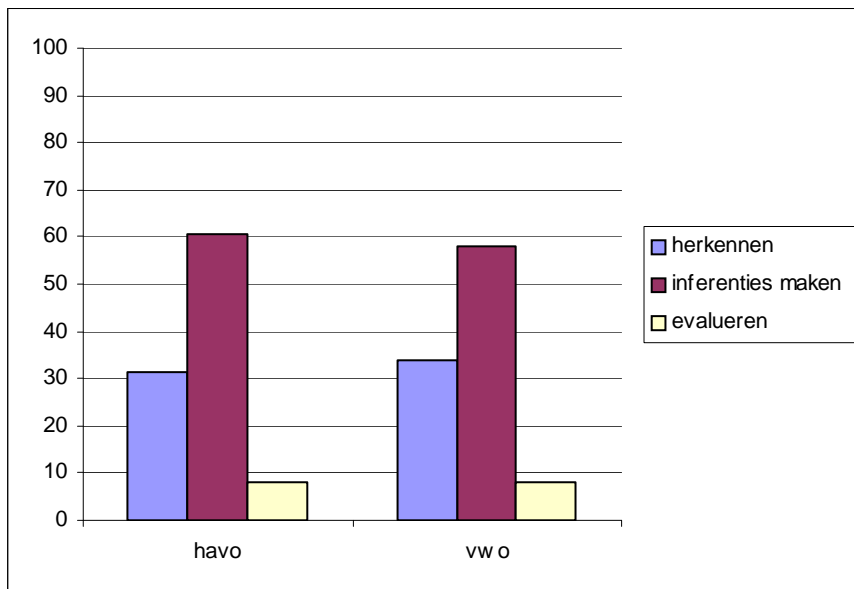
Uit de schalen in het ERK valt op te maken dat, naarmate het ERK-niveau hoger wordt, er een toename is van leesopgaven die gericht zijn op het maken van inferenties en evaluaties op grond van de tekst en een afname van opgaven die directe herkenning van informatie beogen te meten.

Figuur 3-33 t/m 3-36 geven een overzicht van deze taakdimensie van de examenopgaven.

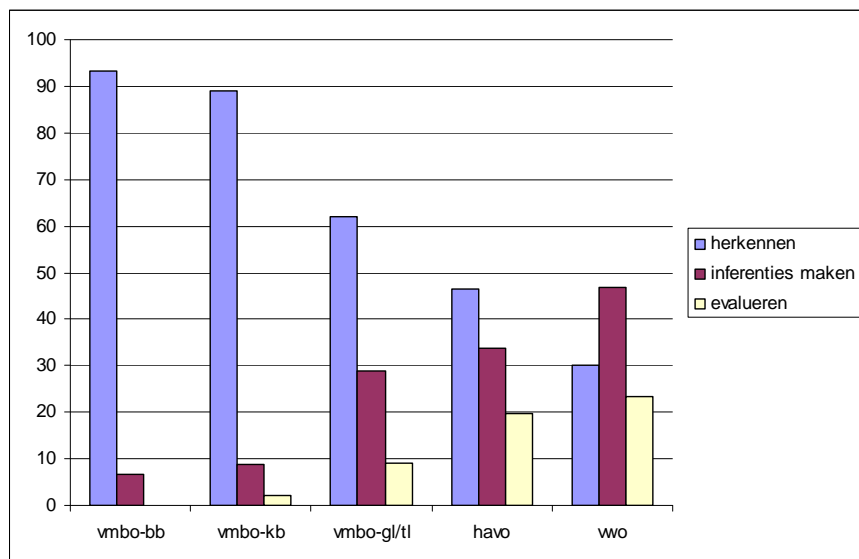
Figuur 3-33 Taakdimensies van de leesopgaven Arabisch per schoolsoort in percentages



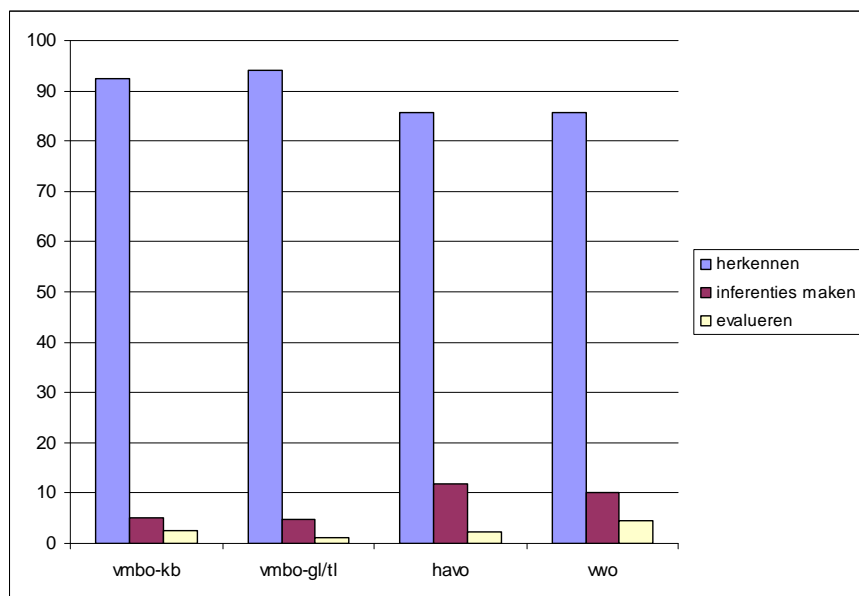
Figuur 3-34 Taakdimensies van de leesopgaven Russisch per schoolsoort in percentages



Figuur 3-35 Taakdimensies van de leesopgaven Spaans per schoolsoort in percentages



Figuur 3-36 Taakdimensies van de leesopgaven Turks per schoolsoort in percentages

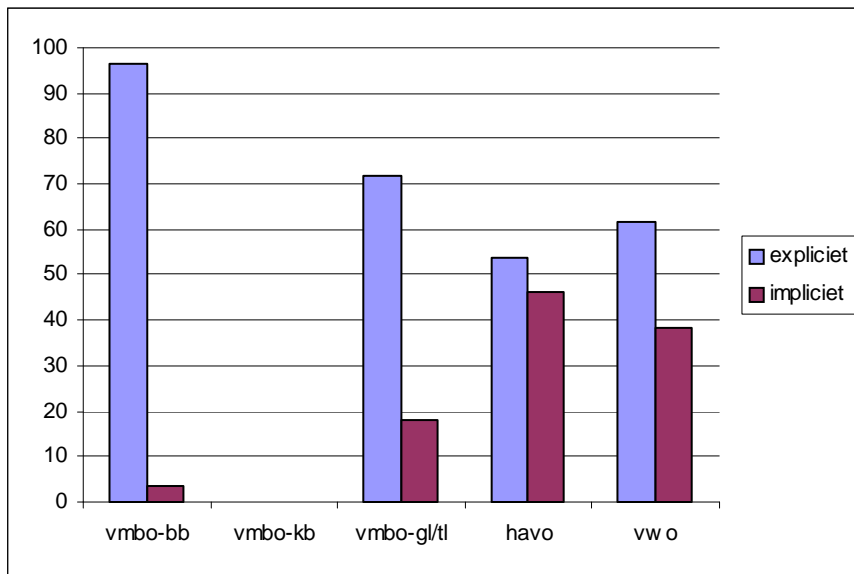


De tendens in het ERK is bij Arabisch en Spaans duidelijk in de examens terug te vinden. Bij de vmbo-examens is meer sprake van herkenning, terwijl de opgaven in de havo- en vwo-examens een beroep doen op het maken van inferenties. Alleen bij Turks is deze tendens nauwelijks aanwezig. Herkennen van informatie blijft hier op alle niveaus dominant. havo en vwo Russisch weerspiegelen het patroon bij Arabisch en Spaans.

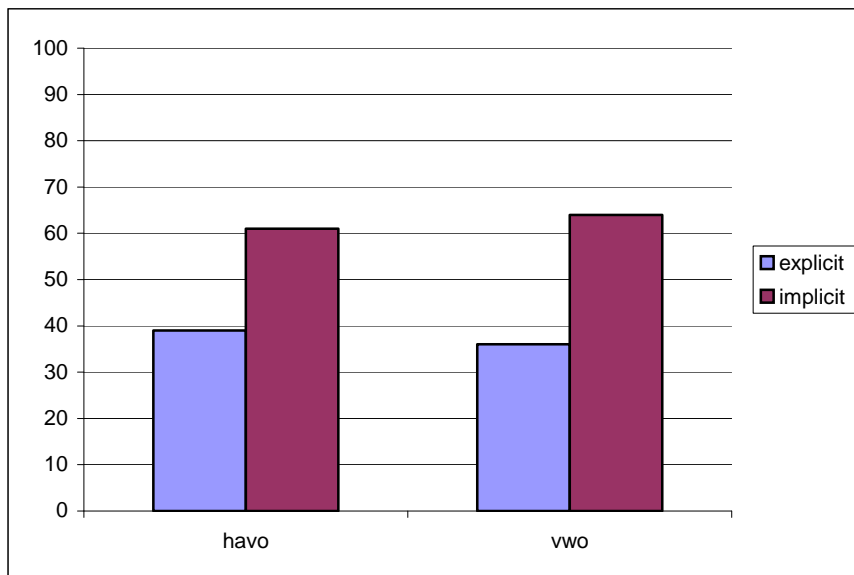
### **De expliciteitsdimensie**

Figuur 3-37 t/m 3-40 geven weer in hoeverre de leesopgaven betrekking hebben op expliciete of impliciete informatie.

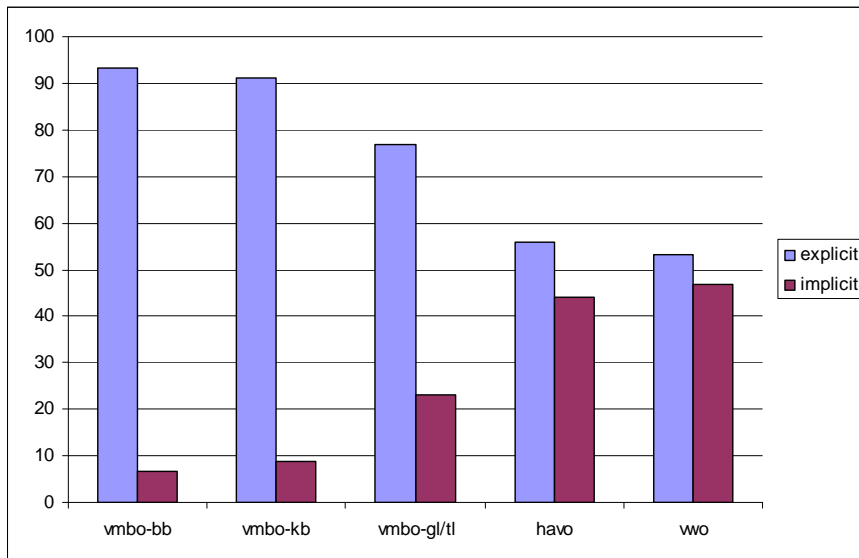
Figuur 3-37 Aard van de informatie waar opgaven Arabisch betrekking op hebben in percentages



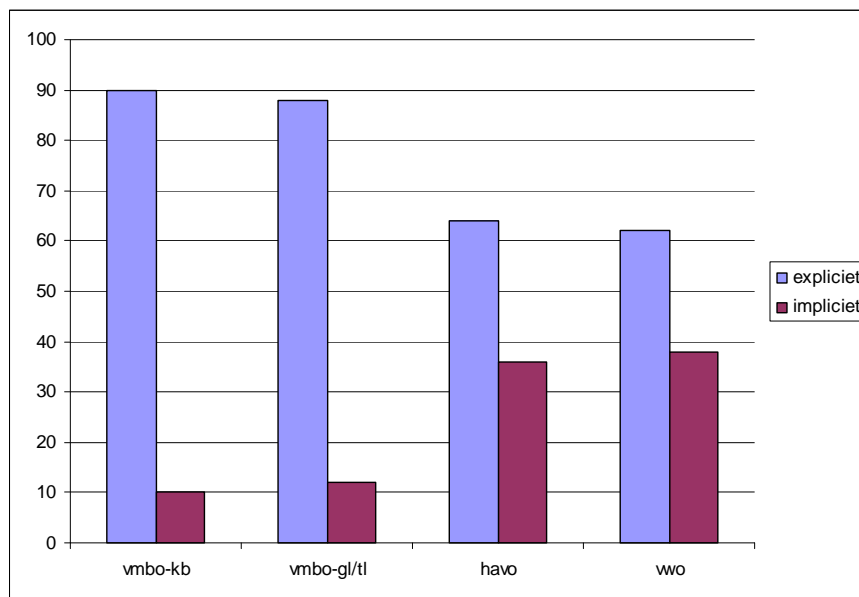
Figuur 3-38 Aard van de informatie waar opgaven Russisch betrekking op hebben in percentages



Figuur 3-39 Aard van de informatie waar opgaven Spaans betrekking op hebben in percentages



Figuur 3-40 Aard van de informatie waar opgaven Turks betrekking op hebben in percentages

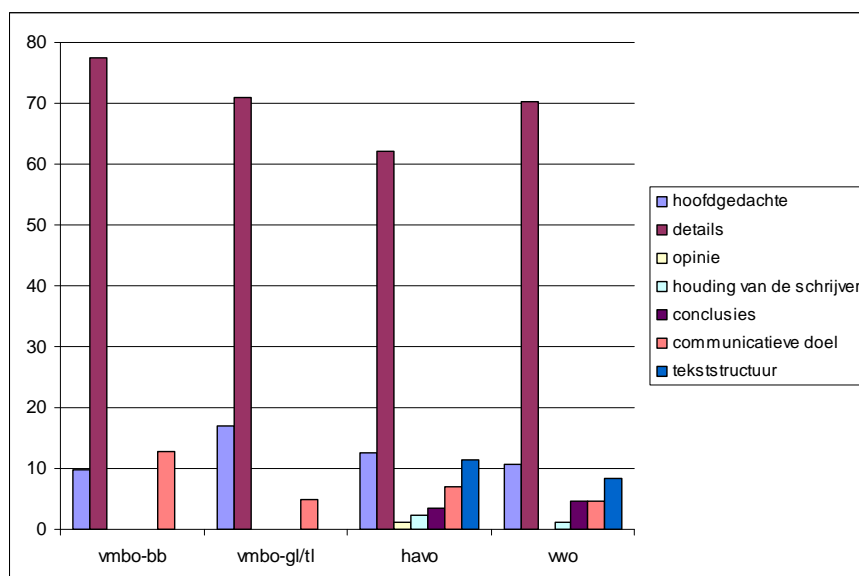


Voor alle examens geldt dat de informatie op VMBO-niveau vooral of uitsluitend expliciet in te tekst te vinden is, terwijl in havo- en wvo-examens meer gevraagd wordt naar impliciete informatie.

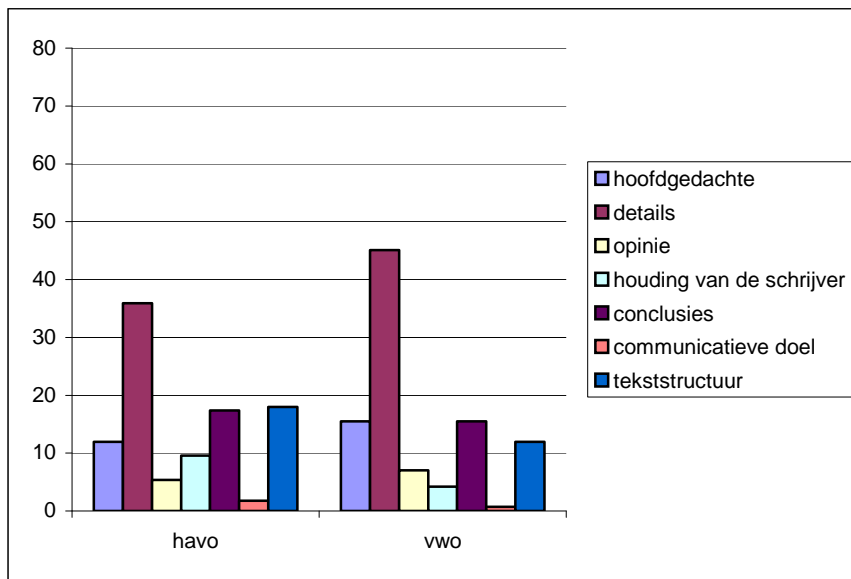
### Inhoudsdimensie

De inhoud waarop leesopgaven betrekking hebben wordt in het ERK voor de diverse niveaus beschreven. Hoewel het ERK dergelijke inhoudelijke beschrijvingen bevat, zijn ze niet consistent en is een duidelijke relatie met de niveaus niet op voorhand te leggen. Er is wel een impliciete relatie, doordat in de 'can-do statements' deze inhoudelijk beschrijving vaak (maar niet altijd) gekoppeld wordt aan termen die duiden op de eenvoud of complexiteit van de te lezen teksten. De enige duidelijke doch impliciete veronderstelling van het ERK is dat naarmate het niveau hoger wordt, de variëteit van wat iemand uit een tekst moet begrijpen groter wordt en dat de teksten linguïstisch en cognitief complexer worden. Figuur 3-41 t/m 3-44 geven een beschrijving van de inhoud van de operaties van de leesopgaven voor de diverse schooltypen.

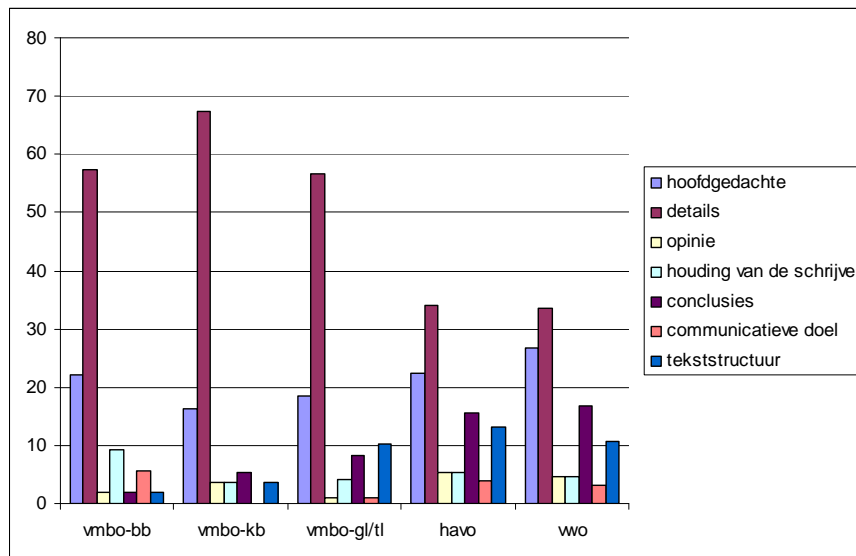
Figuur 3-41 Inhoud van de operaties waar opgaven Arabisch betrekking op hebben in percentages



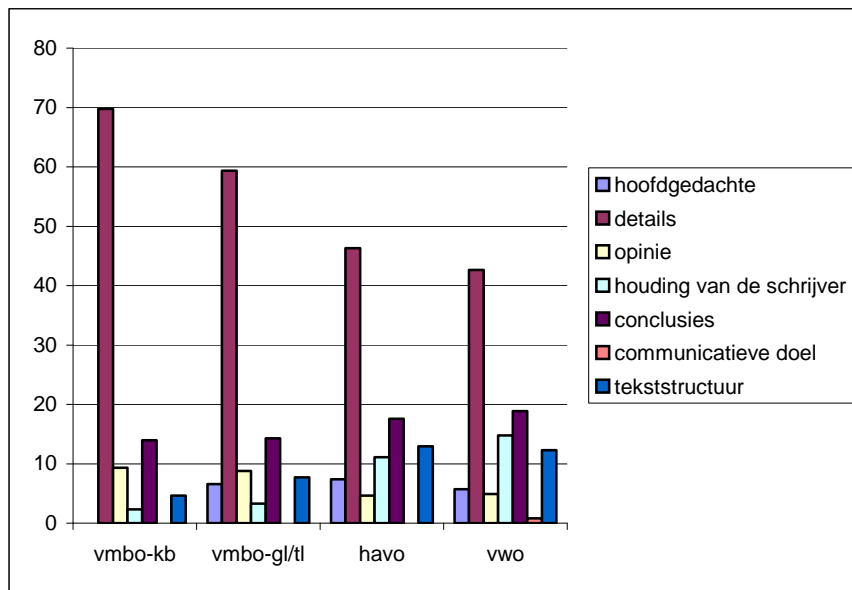
Figuur 3-42 Inhoud van de operaties waar opgaben Russisch betrekking op hebben in percentages



Figuur 3-43 Inhoud van de operaties waar opgaben Spaans betrekking op hebben in percentages



Figuur 3-44 Inhoud van de operaties waar opgaven Turks betrekking op hebben in percentages



Voor alle talen vertonen de examens een toename van de diversiteit van de inhoud van de operaties van vmbo naar vwo, hetgeen overeenkomt met de impliciete strekking van de niveaubeschrijvingen in het ERK. Er is – behalve bij Arabisch – een relatieve afname van vragen naar details te bespeuren.

### 3.3.5 Conclusies bij de specificatie van teksten en taken

De standaardbepaling voor de examens leesvaardigheid moderne vreemde talen (zie hoofdstuk 4) toont aan dat van vmbo-bb tot vwo de centrale examens leesvaardigheid op hogere ERK-niveaus betrekking hebben.

Het ERK veronderstelt bij stijging van het niveau een toename van de linguïstische en cognitieve complexiteit van leesteksten die taalleerders moeten kunnen begrijpen en eveneens van leesvaardigheidstaken die zij moeten kunnen uitvoeren. Deze ontwikkeling wordt men name beschreven in de schalen voor communicatieve competenties (ERK, p. 108–130).

De uitgevoerde inhoudelijke beschrijving van de examens leesvaardigheid Arabisch, Russisch, Spaans en Turks laat zien dat deze toename van linguïstische en cognitieve complexiteit in de examens voor alle talen met name op tekstniveau is terug te vinden. Daarbij worden er binnen deze algemene tendens verschillen tussen de talen gevonden. De teksten worden grammaticaal complexer, de abstractie van de teksten neemt toe en het woordgebruik wordt steeds uitgebreider en gevarieerder. De vwo-examens Arabisch, Russisch, Spaans en Turks zijn in linguïstisch en cognitief opzicht wat minder complex dan de examens Engels en Duits.

De beschrijvingen van de opgaven wijzen erop dat, naarmate het niveau van de examens hoger wordt, examenkandidaten een grotere verscheidenheid aan operaties moeten kunnen uitvoeren.

## 4 Standaardisatie

### 4.1 Inleiding

In dit hoofdstuk wordt verslag gedaan van de werkwijze en de resultaten van de standaardisatieprocedure. Paragraaf 4.2 bevat een beschrijving van het beoordelingsproces en in paragraaf 4.3 wordt verslag gedaan van de analyses die zijn uitgevoerd ter bepaling van de kwaliteit van de gevolgde beoordelingsprocedure. Tenslotte worden in paragraaf 4.4 de resultaten van de standaardbepaling voor achtereenvolgens Arabisch, Russisch, Spaans en Turks gepresenteerd.

De Handleiding benadrukt dat, alvorens aan het standaardiseren te beginnen, het noodzakelijk is dat er een *communis opinio* ontstaat ten aanzien van de betekenis van de ERK-niveaus bij degenen die meewerken aan de standaardisering, met name onder taalspecialisten in andere sectoren, regio's en landen (in dit project: anderen dan de specialisten bij Cito). De Handleiding onderscheidt vier stadia in deze fase:

1. Familiarisatie: een proces dat vergelijkbaar is met dat in fase 1 dat ook met familiarisatie wordt aangeduid. Op dit terrein zijn in het kader van het koppelingproject geen verdere activiteiten ondernomen, behalve bij het proces van standaardbepaling (zie hieronder bij punt 4).
2. Training met gestandaardiseerde voorbeelden bij de productieve vaardigheden. In het kader van het huidige koppelingsonderzoek zijn hier geen verdere activiteiten ondernomen.
3. Standaardiseren van taaluitingen. In het koppelingsonderzoek bij de centrale examens leesvaardigheid was dit proces niet relevant.
4. Standaardbepaling. De Handleiding onderscheidt hierbij twee fases:
  - Het beoordelingsproces
  - Data-analyse ter validering van de standaarden.

### 4.2 Het Beoordelingsproces

Hieronder volgt een kort overzicht van de verschillende stappen in dit proces.

- *Bepaling van de doelen van de beslisprocedure*

Het doel van het beoordelingsproces is dat beoordelaars het minimale ERK-niveau bepalen dat een kandidaat moet hebben om een bepaalde taaltoets met succes uit te voeren. Met andere woorden, om voor ieder examen de cesuur te bepalen waarbij van een kandidaat gezegd kan worden dat hij een ERK-niveau heeft behaald dat relevant is voor het doel van het betreffende examen.

Het algoritme bij de standaardbepaling dat in het koppelingsonderzoek is gebruikt, wordt hieronder kort beschreven. De data worden verzameld via de zogenaamde "mandjes"-methode. Iedere beoordelaar wordt gevraagd om iedere leesopgave in een gemerkt "mandje" te plaatsen waarvan de naam overeenkomt met het laagste niveau waarop die leesopgave dient beheerst te worden. Er zijn vijf mandjes, genaamd A1, A2, B1, B2 en C1+, die overeenkomen met de niveaus die zijn verbonden met de examenprogramma's (C1+ betreft de niveaus C1 en C2). Als een leesopgave geplaatst wordt in mandje B1, dan betekent dit dat volgens de beoordelaar iemand op het niveau van B1 de leesopgave dient te beheersen en per implicatie wordt beheersing van deze leesopgave verondersteld bij alle personen met een hoger niveau (B2 en hoger). Echter, van iemand met het niveau A2 (of lager) kan niet worden verwacht dat hij de leesopgave beheerst. Deze methode van standaardbepaling is ontwikkeld voor het project DIALANG (2002). De methode is gekozen omdat zij goed werkbaar is en betrouwbare en bruikbare resultaten oplevert. Daarnaast waren verschillende leden van het koppelingproject bekend met deze methode.
- *Selectie van leesopgaven*

Idealiter zouden *alle* leesopgaven uit de centrale examens leesvaardigheid Arabisch, Russisch, Spaans en Turks beoordeeld dienen te worden tijdens de standaardbepaling. Dit zou betekend

hebben dat de beoordelaars per taal (met uitzondering van Russisch, waarvoor alleen havo en vwo-examens geproduceerd worden) ca. 200 leesopgaven zouden hebben moeten beoordelen. Het is duidelijk dat dit een te zware taak zou zijn geweest om tijdens één sessie te doen. Er zijn daarom representatieve selecties gemaakt uit ieder examen op een bepaald niveau. Bij de steekproef werden de toetsmatrijzen van de examens gebruikt. Selectiecriteria waren:

- Teksttype (beschrijvend, betogend etc.)
- Tekstlengte
- Type Leesopgave (meerkeuze, open)
- Operaties (reproductie, conclusie, voorspelling etc.)
- Vragen naar tekstinhoud (hoofdgedachte, detail)

De leesopgaven die op deze manier waren geselecteerd werden in een randomvolgorde geplaatst zodat de beoordelaars voor elk leesopgave afzonderlijk zouden moeten bepalen welk minimumniveau noodzakelijk was om de leesopgave juist te beantwoorden. In feite werden de *teksten* (met de bijbehorende opgaven) in een randomvolgorde geplaatst omdat anders bij iedere vraag de bijbehorende tekst opnieuw gereproduceerd had moeten worden. Daarnaast dienden de opgaven ook in een logische volgorde bij een tekst te worden geplaatst, zoals in het examen. Op deze manier werd voor Arabisch een totaal van 82 leesopgaven geselecteerd, voor Russisch 80 opgaven, voor Spaans 80 opgaven en voor Turks 77 leesopgaven.

- *Selectie van beoordelaars*

Ter verhoging van de validiteit van de standaardbepaling zijn beoordelaars gezocht onder docenten aan scholen voor voortgezet onderwijs (waarvoor de examens bestemd waren) en van WO- en HBO-instellingen waar docenten worden getraind. Andere beoordelaars zijn gerekruteerd uit het zakenleven en bij particuliere taleninstituten.

Er was een aantal redenen waarom medewerkers aan het project waren uitgesloten van deelname aan de standaardbepaling. Sommige medewerkers waren bijvoorbeeld betrokken geweest bij de constructie van de betrokken opgaven. Hoewel het nuttig was geweest te zien hoe zij (hun) opgaven zouden koppelen aan de ERK-niveaus, zou hun bekendheid met de leesopgaven toch waarschijnlijk hebben geresulteerd in vertekende resultaten.

- *Training van de beoordelaars*

De beoordelaars werden op dezelfde manier vertrouwd gemaakt met het ERK als in de fase van de familiarisatie. De beoordelaars hebben uitvoerig gediscussieerd over de relevantie van het ERK. Vervolgens kregen de beoordelaars een aantal oefenopgaven waarmee ze de mandjesprocedure konden oefenen. Voor elke opgave werd de volgende vraag gesteld:

Geef s.v.p. bij elke opgave aan welk niveau (A1, A2, B1, B2 of C1+) minimaal vereist is om de opgave goed te kunnen maken. (*Omcirkel achter elke opgave het getal in de kolom met het antwoord van uw keuze*).

Tekst	Opgave	Niveau				
		A1	A2	B1	B2	C1+
1	1	1	2	3	4	5

De training van de beoordelaars werd in de vier taalgroepen uitgevoerd en werd geleid door leden van het project. Gedurende de discussies in de taalgroepen bereikten de deelnemers overeenstemming over de minimumniveaus bij een aantal oefenopgaven. Deze leesopgaven werden niet meegenomen in de feitelijke standaardbepaling.

- *De beoordelingsessies*

Na de training kregen de beoordelaars de reeksen teksten en opgaven aangeboden in de randomvolgorde die hierboven is toegelicht. De beoordelaars hadden twee tot drie uur nodig om alle opgaven te beoordelen met behulp van een beoordelingsformulier waarvan een deel hierboven is afgebeeld. Beoordelaars hebben niet geklaagd over te grote werkdruk. Integendeel: een groot aantal beoordelaars bood aan om ook deel te nemen aan standaardbepaling bij andere vaardigheden.

- *Procedures dataverzameling*  
De beoordelingsformulieren zijn opgehaald en de data zijn op optisch leesbare formulieren overgebracht. De volgende data zijn verzameld: ID beoordelaar, taal en beoordeling (1 tot 5, behorend bij A1 tot C1+) per leesopgave.

### 4.3 Data-analyse ter validering van de standaarden

De volgende fase in de procedure van de standaardbepaling vormde de data-analyse om de nauwkeurigheid van de standaarden te kunnen bepalen. De data-analyse bestaat uit twee procedures:

1. *Bepaling van beoordelaarsovereenstemming*
2. *Bepaling van minimumscores bij relevante ERK-niveaus voor ieder examen*

#### *Bepaling van beoordelaarsovereenstemming*

Allereerst zijn de beoordelaarsovereenstemming en de beoordelaarsbetrouwbaarheid berekend. Eerst worden hieronder per taal de resultaten gegeven bij het totale aantal opgaven dat is beoordeeld. Daarna zijn de leesopgaven hergegroepeerd en wordt de beoordelaarsbetrouwbaarheid (Cronbach's alfa), de beoordelaarsovereenstemming en de gemiddelde score van het minimumniveau (1=A1, 2=A2, 3=B1, 4=B2, 5=C1) gegeven over beoordelaars en leesopgaven voor elk examen.

#### **Arabische leesopgaven**

Aantal beoordelaars: 16

Totaal aantal items: 82

Alle leesopgaven beoordeeld in randomvolgorde

Geschatte variantie componenten:

Leesopgaven (p):41%

Beoordelaars (b):05%

Residu (pb,e):54%

Beoordelaarsbetrouwbaarheid: .92

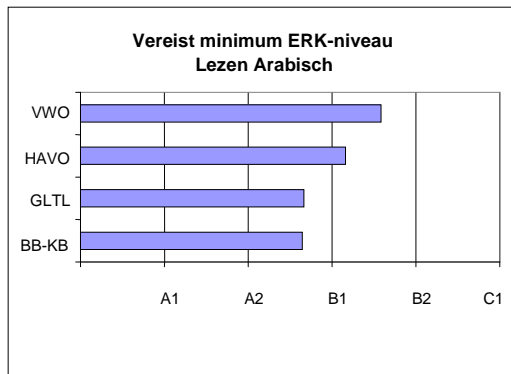
Beoordelaarsovereenstemming (Rho2): .92

*Tabel 4-1 Beoordelaarsbetrouwbaarheid, beoordelaarsovereenstemming en gemiddeld vereiste minimumniveau per examen (Arabisch)*

Examen	N leesopgaven	Beoordelaarsbetrouwbaarheid ( $\alpha$ )	Beoordelaarsovereenstemming (Rho2)	Gemiddeld vereiste minimum niveau
BB	10	.84	.80	2,6
GL/TL	30	.94	.93	2,7
havo	20	.81	.77	3,2
vwo	20	.63	.57	3,6

De in de laatste kolom gegeven waarden bij het gemiddeld vereiste minimumniveau komen overeen met de niveaus 1=A1, 2=A2, 3=B1, 4=B2, 5=C1. Het gemiddeld vereiste minimumniveau wordt ook in de volgende figuur weergegeven.

Figuur 4-1 Gemiddeld vereiste minimumniveau (Arabisch)



### Russische leesopgaven

Aantal beoordelaars: 15

Totaal aantal items: 80

Alle leesopgaven beoordeeld in randomvolgorde

Geschatte variantie componenten:

Leesopgaven (p):25%

Beoordelaars (b):14%

Residu (pb,e):61%

Beoordelaarsbetrouwbaarheid.86

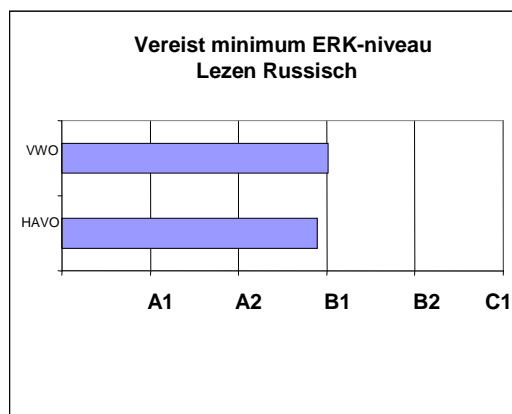
Beoordelaarsovereenstemming (Rho2):.84

Tabel 4-2 Beoordelaarsbetrouwbaarheid, beoordelaarsovereenstemming en gemiddeld vereiste minimumniveau per examen (Russisch)

Examen	N leesopgaven	Beoordelaarsbetrouwbaarheid ( $\alpha$ )	Beoordelaarsovereenstemming (Rho2)	Gemiddeld vereiste minimum niveau
havo	40	.88	.84	2,9
vwo	40	.85	.82	3,0

De in de laatste kolom gegeven waarden bij het gemiddeld vereiste minimumniveau komen overeen met de niveaus 1=A1, 2=A2, 3=B1, 4=B2, 5=C1. Het gemiddeld vereiste minimumniveau wordt ook in de volgende figuur weergegeven.

Figuur 4-2 Gemiddeld vereiste minimumniveau (Russisch)



### Spaanse leesopgaven

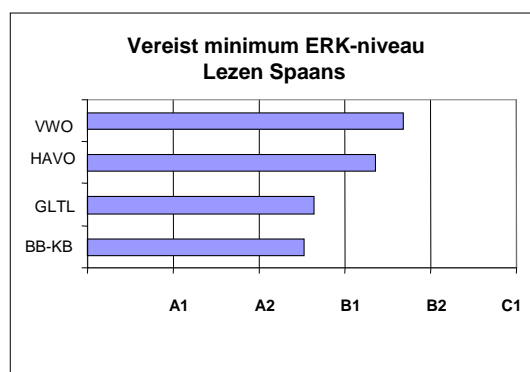
Aantal beoordelaars: 16  
Totaal aantal items: 80  
Alle leesopgaven beoordeeld in randomvolgorde  
Geschatte variantie componenten:  
Leesopgaven (p):50%  
Beoordelaars (b):03%  
Residu (pb,e):47%  
Beoordelaarsbetrouwbaarheid.94  
Beoordelaarsovereenstemming (Rho2):.94

Tabel 4-3 Beoordelaarsbetrouwbaarheid, beoordelaarsovereenstemming en gemiddeld vereiste minimumniveau per examen (Spaans)

Examen	N leesopgaven	Beoordelaarsbetrouwbaarheid ( $\alpha$ )	Beoordelaarsovereenstemming (Rho2)	Gemiddeld vereiste minimum niveau
BB	12	.75	.67	2,5
GL/TL	24	.84	.82	2,6
havo	22	.88	.88	3,6
vwo	22	.93	.92	3,7

De in de laatste kolom gegeven waarden bij het gemiddeld vereiste minimumniveau komen overeen met de niveaus 1=A1, 2=A2, 3=B1, 4=B2, 5=C1. Het gemiddeld vereiste minimumniveau wordt ook in de volgende figuur weergegeven.

Figuur 4-3 Gemiddeld vereiste minimumniveau (Spaans)



### Turkse leesopgaven

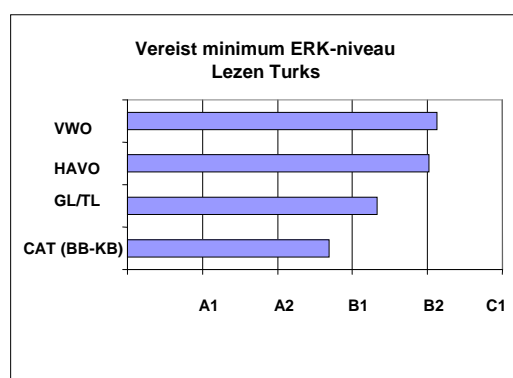
Aantal beoordelaars: 13  
Totaal aantal items: 77  
Alle leesopgaven beoordeeld in randomvolgorde  
Geschatte variantie componenten:  
Leesopgaven (p):51%  
Beoordelaars (b):10%  
Residu (pb,e):39%  
Beoordelaarsbetrouwbaarheid.94  
Beoordelaarsovereenstemming (Rho2):.93

Tabel 4-4 Beoordelaarsbetrouwbaarheid, beoordelaarsovereenstemming en gemiddeld vereiste minimumniveau per examen (Turks)

Examen	N leesopgaven	Beoordelaarsbetrouwbaarheid ( $\alpha$ )	Beoordelaarsovereenstemming (Rho2)	Gemiddeld vereiste minimum niveau
CAT	17	.84	.75	2,7
GL/TL	30	.82	.77	3,3
havo	16	.92	.90	4,0
vwo	14	.88	.82	4,1

De in de laatste kolom gegeven waarden bij het gemiddeld vereiste minimumniveau komen overeen met de niveaus 1=A1, 2=A2, 3=B1, 4=B2, 5=C1. Het gemiddeld vereiste minimumniveau wordt ook in de volgende figuur weergegeven.

Figuur 4-4 Gemiddeld vereiste minimumniveau (Turks)



#### 4.3.1 Discussie analyse beoordelaarsovereenstemming

- De beoordelaarsovereenstemming bij alle opgaven die aan de beoordelaars werden voorgelegd was bij alle drie de talen meer dan .90 (over alle niveaus gezamenlijk), behalve bij Russisch waar deze .86 was. Dit lijkt er op te wijzen dat de beoordelaars in voldoende mate overeenstemmen over het vereiste minimum ERK-niveau dat per opgave nodig is om deze met succes te beantwoorden.
- Bij de analyse valt op te merken dat de toewijzing van items aan verschillende niveaus niet slechts wordt verklaard door verschillen in items en beoordelaars, maar in enkele gevallen door niet verklaarde effecten (meetfout). Echter, de beoordelaarsovereenstemming en de beoordelaarsbetrouwbaarheid is dermate hoog (meer dan .90) dat wij deze variantie kunnen verwaarlozen.
- De beoordelaars hebben de opgaven uit het examen op het laagste niveau (bb) geplaatst aan het lage einde van de ERK-schaal en ze hebben de opgaven die uit de examens op de hogere niveaus zijn genomen (havo en vwo) aan het hoogste einde van de ERK-schaal geplaatst. Dit niveau-onderscheid komt overeen met de spreiding in de Nederlandse examenniveaus waar bb het laagste niveau is en vwo het hoogste.
- Beoordelaars van buiten lijken met de toetsdeskundigen overeen te stemmen over de moeilijkheidsgraad van opgaven.
- Data-analyse toont aan dat de beoordelaars van mening zijn dat een oplopend ERK-niveau nodig is om de opgaven uit de examens van oplopend niveau met succes te beantwoorden.
- Bij een aantal talen ontlopen de verschillen in veronderstelde vereiste minimum ERK-niveaus bij examens van verschillend schoolniveau elkaar weinig: Arabisch BB en GL/TL; Russisch: havo en vwo; Turks: havo en vwo.

#### 4.4 Het bepalen van de minimumscores voor relevante ERK-niveaus bij ieder examen

Bij het koppelingsonderzoek van de examens Frans, Duits en Engels was de volgende stap in de data-analyse het bepalen van de minimumscores op een examen die een kandidaat diende te behalen om te kunnen claimen dat hij zich op een bepaald ERK-niveau bevond. Daarbij was het interessant deze score te vergelijken met de door CEVO bepaalde cesuur voor voldoende/ onvoldoende bij ieder examen van een bepaald jaar en deze cesuurscore te interpreteren in termen van ERK-niveaus.

Zoals boven al was aangegeven, hebben de beoordelaars geen volledige examens voorgelegd gekregen, maar een representatieve selectie van opgaven uit de beschikbare examens van een bepaald jaar. Omdat de leerlingaantallen per examen te gering waren, konden (anders dan bij de eerder onderzochte talen) geen analyses volgens de itemresponstheorie (IRT) worden gedaan. In het vorige onderzoek stelde het IRT-model ons in staat om standaarden (cesuren) over te zetten naar examens van verschillende jaren, zolang er via een overlappende relatie tussen die examens waren aangebracht. Omdat we in de huidige studie geen gebruik konden maken van het IRT-model, zijn wij niet zonder meer in staat om standaarden die we bij de subsets van items kunnen bepalen over te zetten naar andere (sub)sets. We gaan er echter vanuit dat onze steekproeven van opgaven uit de verschillende examens een getrouwe afspiegeling vormt van het gehele examen waaruit de steekproef getrokken was, en gebruiken deze aanname om voor dat specifieke examen van dat specifieke jaar een relevante ERK-cesuur bepalen. Zie voor een meer uitvoerige behandeling van deze problematiek bijlage 2.

Indien wij een simpele random trekking zouden hebben uitgevoerd, dan was het mogelijk geweest dat items op sommige ERK-niveaus niet zouden voorkomen in de steekproef. Om het risico hierop te beperken hebben we op basis van inhoudelijke kenmerken een aantal soorten items onderscheiden. In het examenmodel (zie bijlage 1) voor ieder examen staat aangegeven hoe een examen dient te zijn opgebouwd. Het model schrijft voor welk type items in welke verhouding dient voor te komen (bijv. 30% open vragen en 70% gesloten vragen). Ook schrijft het model voor op welke vaardigheden in welke verhouding de items betrekking moeten hebben (bijv.: 20% conclusies trekken). Het aantal items dat werd gekozen, was proportioneel aan het aantal items van elke soort. Hierdoor is verdere weging van de data onnodig.

Stel nu dat de deskundigen bij een bepaalde steekproef hebben bepaald dat van de 20 items uit de steekproef er voor 2 items een minimaal niveau van A1 vereist was om de items goed te maken, bij 14 items een niveau A2 en bij 4 items een niveau B1, dan kunnen wij voor de relevante cesuur A1/A2 berekenen wat de score moet zijn op deze subset van het examen. Als wij kunnen stellen dat de gevonden percentages items op de niveaus A1, A2 en B1 bij de steekproef de percentages van het gehele examen goed vertegenwoordigen, dan kunnen wij ook de minimale score berekenen op de bij de steekproef behorende examen die nodig is om te kunnen claimen dat iemand het niveau A2 heeft gehaald.

Omwille van de vergelijkbaarheid van de gegevens per examen geven wij de genoemde minimumscores in onderstaande tabellen aan in termen van het vereiste minimumpercentage juiste antwoorden op een geheel examen. Om te claimen dat een kandidaat het voor dat examen relevante ERK-niveau heeft gehaald, moet de kandidaat het gegeven percentage juiste antwoorden hebben gegeven. In de onderstaande tabellen geven we ook nog aan wat het (door de CEVO bepaalde) percentage goede antwoorden moet zijn om een *voldoende* voor het examen te behalen.

Tabel 4-5 Minimumscores bij relevante ERK-niveaus in percentages van het totaal aantal opgaven per examen Arabisch

Examen	Relevante Niveau(s)	Minimum percentage goede antwoorden	CEVO cesuur	
			2003	2004
BB/KB	A2	43%	Niet berekend <sup>1</sup>	Niet berekend <sup>2</sup>
GL/TL	A2	43%	60%	50%
havo	B1	78%	53%	53%
vwo	B1	45%	50%	53%

We zien dat voor het GL/TL examen een leerling die volgens de CEVO een voldoende heeft behaald ook het relevante niveau A2 heeft behaald. Dat geldt niet voor het havo-examen, waar een leerling een voldoende heeft gehaald bij 53% goede antwoorden, maar om B1 te claimen 78% juist moet hebben. Bij het vwo examen heeft een leerling die volgens de CEVO een voldoende heeft gehaald, ook het relevante niveau B1 bereikt.

Tabel 4-6 Minimumscores bij relevante ERK-niveaus in percentages van het totaal aantal opgaven per examen Russisch

Examen	Relevante niveau(s)	Minimum percentage goede antwoorden	CEVO cesuur	
			2005	2006
havo	A2 en B1	27% en 81%	53%	53%
vwo	A2 en B1	20% en 76%	53%	53%

We zien dat de twee examens Russisch volgens de deskundigen elkaar niet veel in moeilijkheidsgraad ontlopen. Om op het havo examen op het ERK-niveau A2 te kunnen claimen, dient de kandidaat 27% van de opgaven goed te hebben en 81% goed om B1 te claimen. Bij een score op de CEVO cesuur kan men globaal zeggen dat men een niveau heeft tussen A2 en B1 in. Een vergelijkbaar beeld is te zien bij het vwo-examen.

Tabel 4-7 Minimumscores bij relevante ERK-niveaus in percentages van het totaal aantal opgaven per examen Spaans

Examen	Relevante niveaus(s)	Minimum percentage goede antwoorden	CEVO cesuur	
			2004	2005
BB/KB	A2	45%	Niet berekend <sup>3</sup>	Niet berekend <sup>4</sup>
GL/TL	A2	40%	50%	50%
havo	B1	55%	52%	50%
vwo	B1 en B2	33% en 92%	58%	56%

<sup>1</sup> Proefexamen

<sup>2</sup> Proefexamen

<sup>3</sup> Proefexamen

<sup>4</sup> Proefexamen

Bij het examen GL/TL heeft een leerling die een voldoende heeft behaald volgens de CEVO-cesuur (minimaal 50% juiste antwoorden) ook het relevante ERK-niveau behaald. Bij het havo-examen komt de grens voldoende (52%) nagenoeg overeen met het relevante ERK-niveau (55%). Bij het vwo heeft de leerling de kans om te bewijzen dat hij B2 heeft behaald. Hij moet dan wel aanzienlijk beter (92%) scoren dan de score bij een voldoende volgens de CEVO-cesuur (56-58%).

Tabel 4-8 *Minimumscores bij relevante ERK-niveaus in percentages van het totaal aantal opgaven per examen Turks*

Examen	Relevante niveaus	Minimum percentage goede antwoorden	CEVO cesuur	
			2003	2004
CAT	A2	35% <sup>5</sup>	Niet berekend <sup>6</sup>	Niet berekend
GL/TL	B1	61%	53%	58%
havo	B1 en B2	16% en 80%	53%	53%
vwo	B1 en B2	17% en 66%	53%	53%

Een leerling haalt op het GL/TL examen een voldoende als hij minimaal 53% van de opgaven juist heeft. Als hij 61% van de opgaven juist heeft beantwoord, heeft hij daarbij ook het relevante ERK-niveau B1 behaald. De examens havo en vwo ontlopen elkaar niet veel in moeilijkheidsgraad volgens de deskundigen. Slechts 16-17% goede antwoorden zijn nodig om het niveau B1 te claimen. Echter, dat is nauwelijks een relevant niveau, wat ook te zien is aan de eisen die de CEVO stelt voor een voldoende. Maar om het niveau B2 te kunnen claimen moet de kandidaat wel weer hoger scoren (80% en 66%).

#### 4.4.1 Discussie bepaling minimumscores voor relevante ERK-niveaus

- Wanneer de items waarvan experts het niveau hebben vastgesteld, opnieuw zouden worden voorgelegd aan kandidaten, dan kan het gebeuren dat een kandidaat anders handelt dan op grond van het oordeel van deskundigen zou kunnen worden verwacht: een kandidaat geeft foute antwoorden op relatief makkelijke vragen en goede antwoorden op relatief moeilijke vragen. Zoals eerder gezegd, hebben wij door de kleine aantallen kandidaten onvoldoende data bij de huidige examens tot onze beschikking om van het IRT-model gebruik te kunnen maken. In het vorige onderzoek bij de examens Frans, Duits en Engels, waar het model wél toepasbaar was, bleek dat dit fenomeen zich overigens niet voordeed. Zie voor een meer uitvoerige verantwoording de bijlage 2.
- Bij de berekening van de minimumscores bleek, zoals ook al eerder vermeld bij de analyse van de beoordelaarsovereenstemming, dat examens voor verschillende schoolsoorten elkaar niet veel in moeilijkheidsgraad ontlopen (Arabisch: BB en GL/TL; Russisch: havo en vwo; Turks: havo en vwo).
- Met name bij de examens Turks op havo en vwo-niveau lijkt de leerling een zeer hoge score te moeten behalen om het relevante ERK-niveau te kunnen claimen.
- Het vergelijken over de talen heen is problematisch. De betreffende talen hebben in het Nederlandse onderwijsveld een verschillende, eigen status, die ook niet goed vergelijkbaar is met de onderling meer overeenkomende status van de talen Frans, Duits en Engels. Ook de leerlingpopulaties die aan de examens Arabisch, Russisch, Spaans en Turks deelnemen, verschillen onderling nogal. Niettemin is het opvallend dat voor Turks havo en vwo het meest relevante ERK-niveau ligt op of voorbij B2. Dit niveau wordt bij de andere vreemde talen alleen door de experts vereist voor Engels, en dan alleen nog op het vwo.

<sup>5</sup> Bij de steekproef van opgaven in de standaardbepaling

<sup>6</sup> In een computeradaptieve toets (CAT) worden geen scores berekend (percentage juist beantwoorde opgaven); derhalve is er ook geen sprake van cesuren. In plaats daarvan wordt met behulp van IRT een vaardigheidsniveau vastgesteld waarbij een kandidaat is geslaagd.

## 5 Samenvatting, conclusies en aanbevelingen

In opdracht van het Ministerie van OCW heeft Cito een project uitgevoerd dat de volgende doelstellingen had:

- Het inhoudelijk analyseren van de examenopgaven leesvaardigheid voor Arabisch, Spaans en Turks vmbo en havo/vwo aan de hand van de criteria van de 'Dutch Grid';
- Het organiseren en uitvoeren van een standaardsetting procedure voor de examens leesvaardigheid Arabisch, Spaans en Turks vmbo en havo/vwo.

Op verzoek van het onderwijsveld zijn de examens Russisch zonder meerkosten voor de opdrachtgever in dit onderzoek meegenomen.

### 5.1 Samenvatting

In dit rapport is verslag gedaan van de door Cito uitgevoerde inhoudelijke specificatie en de standaardisatie van de centrale examens leesvaardigheid in de moderne vreemde talen Arabisch, Russisch, Spaans en Turks. Daarbij zijn de procedures gevolgd die zijn beschreven in de concepthandleiding die is gepubliceerd door de Raad van Europa. Om tot een verantwoorde specificatie en standaardisatie te komen, dienden betrokken medewerkers eerst vertrouwd gemaakt te worden met het ERK in de fase van de familiarisatie. Ook nadien, toen deskundigen van buiten Cito gevraagd is het minimum ERK-niveau te bepalen van opgaven die in de leesvaardigheidsexamens voorkomen, zijn deze deskundigen eerst vertrouwd gemaakt met het ERK.

Door medewerkers aan het project is tijdens het specificatieproces een inhoudsanalyse gemaakt van de leesexamens. De vraag was in hoeverre de leesexamens vanaf vmbo tot vwo conform de verwachtingen daadwerkelijk teksten en opgaven bevatten die oplopen in ERK-niveau. Om deze vraag te kunnen beantwoorden is een ERK-gerelateerd beschrijvingsmodel gebruikt waarmee de teksten en opgaven zijn beschreven. Hierbij dient op gemerkt te worden dat voor Russisch slechts examens op twee niveaus worden geproduceerd: havo en vwo.

Het is mogelijk gebleken voor ieder van de leesvaardigheidsexamens voor de verschillende schooltypen aan te geven wat het gemiddelde minimum-leesvaardigheidsniveau van een kandidaat moet zijn in termen van het ERK om met succes de vragen in een examen te beantwoorden. Daarnaast hebben wij kunnen aangeven wat voor ieder van de gebruikte examens de minimum-percentages juiste antwoorden dienen te zijn om relevante ERK-niveaus te behalen. Wij hebben deze minimumpercentages per examen afgezet tegen de CEVO-cesuur voldoende-onvoldoende. Omdat de aantallen leerlingen die in de talen Arabisch, Russisch, Spaans en Turks examen doen, relatief gering zijn, zijn onze conclusies niet altijd vergelijkbaar met die bij een vergelijkbaar onderzoek naar de examens in de talen Frans, Duits en Engels.

### 5.2 Conclusies

Hieronder geven wij een overzicht van de belangrijkste conclusies die wij in dit onderzoek hebben kunnen trekken. Wij tekenen daarbij aan dat het onze taak was om de bestaande centrale examens leesvaardigheid Arabisch, Russisch, Spaans en Turks te koppelen aan het ERK. Het is aan anderen om hier consequenties aan te verbinden voor de inhoud en het niveau van de examenprogramma's en de centrale examens leesvaardigheid.

1. Het volgen van alle gesuggereerde stappen in de Handleiding is een tijdrovend en kostbaar proces. Echter, Cito heeft inmiddels veel ervaring opgedaan in deze materie en heeft – mede in overleg met de Raad van Europa – de procedures hier en daar kunnen vereenvoudigen.
2. Het koppelingsproces is een goede manier om de inhoud en de statistische kwaliteiten van een examen kritisch te onderzoeken.

3. Een globale beschouwing van de opgaven in de examens Arabisch, Russisch, Spaans en Turks bevestigt het beeld van de analyses die eerder uitgevoerd waren voor de examens Frans, Duits en Engels dat de opgaven voornamelijk gericht zijn op *lezen om informatie op te doen*. Dit heeft te maken met de voorschriften in de examenmodellen, die juist aan dit soort lezen het grootste gewicht toekennen.
4. De examens bevatten een variatie aan tekstbronnen, tekstsoorten en onderwerpen die in het ERK vermeld worden. Deze grote variatie in teksten in de examens maakt een uitgebreide revisie van de huidige examens minder noodzakelijk.
5. De examenteksten weerspiegelen de in het ERK veronderstelde toename van linguïstische en cognitieve complexiteit bij het stijgen van het niveau van de examens. De variatie in leestaken die kandidaten bij leesteksten moeten kunnen uitvoeren neemt toe van vmbo naar vwo.
6. De vwo-examens Arabisch, Russisch, Spaans en Turks zijn in linguïstisch en cognitief opzicht wat minder complex dan de examens Engels en Duits (voor Frans is dit niet uitgezocht).
7. Deskundige beoordelaars van buiten hebben op basis van het ERK een niveau-onderscheid tussen opgaven aangebracht dat overeenkomt met de spreiding in de Nederlandse examenniveaus.
8. De standaardbepaling toont aan dat van vmbo-bb tot vwo de centrale examens leesvaardigheid op steeds hogere ERK-niveaus betrekking hebben.
9. Bij een aantal talen ontlopen de verschillen in veronderstelde vereiste minimum ERK-niveaus bij examens van verschillend schoolniveau elkaar weinig.
10. De CEVO-cesuren (voldoende/onvoldoende) vallen in veel gevallen *niet* samen met de door deskundigen bepaalde cesuren bij relevante ERK-niveaus.

### 5.3 Aanbevelingen

1. Bij de vaststelling van de examenmodellen zou overwogen kunnen worden of het nodig is ontbrekende descriptoren uit het ERK meer expliciet in aanmerking te laten komen voor opname in de examens. Het gaat hierbij met name om zakelijk getinte correspondentie, oriënterend lezen en het lezen van instructies.
2. Het verdient aanbeveling met behulp van het ERK en het *Dutch Grid* het onderscheid in moeilijkheidsgraad tussen examens in één taal op verschillende niveaus te herzien.

## 6 Referenties

- Alderson, J. Charles (2004), Neus Figueras, Henk Kuijper, Günter Nold, Sauli Takala, Claire Tardieu, *The Development of Specifications for Item Development and Classification within the Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Final Report of the Dutch ERK Construct Project*.
- CEVO (2006), *Examenprogramma Moderne Vreemde Talen*, Utrecht.
- Council of Europe (2001), *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge University Press, Cambridge.
- Council of Europe (2003), Language Policy Division, *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching and Assessment. Manual Preliminary Pilot Version*. Council of Europe, Strasbourg.
- Dutch CEFR *Grid Reading / Listening*, [www.lancs.ac.uk/fss/projects/grid/](http://www.lancs.ac.uk/fss/projects/grid/).
- Liemberg, E., Meijer, D. (ed.) (2004), *Taalprofielen*. NaB-MVT, Enschede.
- Noijons, José & Henk Kuijper (2006), De koppeling van de centrale examens leesvaardigheid moderne vreemde talen aan het Europees Referentiekader. Cito, Arnhem.
- Noijons, José (2006b) & Henk Kuijper, *Leesvaardigheidsexamens moderne vreemde talen in Europees verband*, Cito, Arnhem.
- SLO (2006), *Handreiking schoolexamen moderne vreemde talen*, Enschede.



**Bijlagen**



## **Bijlage 1 Samenvatting van de examenmodellen Arabisch, Russisch, Spaans en Turks voor vmbo en havo/vwo**

### **Eindtermen voor BB/KB en GT**

Et 1: Aangeven of een tekst, gegeven een bepaalde informatiebehoefte, relevante informatie bevat.  
Et 2: hoofdgedachte aangeven  
Et 3: betekenis van belangrijk element aangeven  
Et 4: gegeven uit teksten vergelijk+ conclusies trekken  
Et 5: verbanden tussen delen v.e. tekst herkennen en aangeven

**BB** Totaal ca.. 35 vragen  
**Eindterm 1** 10%  
**Eindterm 2+ 3** 85%  
**Eindterm 4** 5%

**KB** Totaal ca. 40 vragen  
**Eindterm 1** 2 vragen (min)  
**Eindterm 2** 2 (min)  
**Eindterm 3** 8 (min)  
**Eindterm 4** 2 (min)  
**Eindterm 5** 1 (min)

**GI/TL** Totaal ca. 45 vragen  
**Eindterm 1** 2 vragen (min)  
**Eindterm 2** 2 vragen (min)  
**Eindterm 3** 3 vragen (min)  
**Eindterm 4** 2 vragen (min)  
**Eindterm 5** 1 vraag (min)

Gebruikelijk is om de aantallen voor et 1,4,5 het minimum te nemen en de rest van het examen te vullen met et 2 en et 3.

#### Voorschrift voor soort vragen:

Meerkeuze en invul (cloze) vragen (max 8) 65%  
Open vragen 35%

**havo+vwo** Totaal ca. 45 vragen

Et 1: Aangeven of een tekst, gegeven een bepaalde informatiebehoefte, relevante informatie bevat.  
Et 2: hoofdgedachte aangeven  
Et 3: betekenis van belangrijk element aangeven  
ET 4: onderdelen van een tekst benoemen en verband tussen delen aangeven  
ET 5: conclusies trekken mbt taalgebruik/schrijfdoel/opvattingen van schrijver

**Eindterm 1:** 2 vragen (min)  
**Eindterm 2** 3 vragen  
**Eindterm 3** verdeeld over het examen  
**Eindterm 4** 4 vragen (min)  
**Eindterm 5** 2 vragen (min)

Gebruikelijk is om de aantallen voor et 1, 4, 5 het minimum te nemen en de rest van het examen te vullen met et 2 en et 3.

#### Voorschrift voor soort vragen:

Meerkeuze en invul (cloze) vragen 60%  
Open vragen 40%

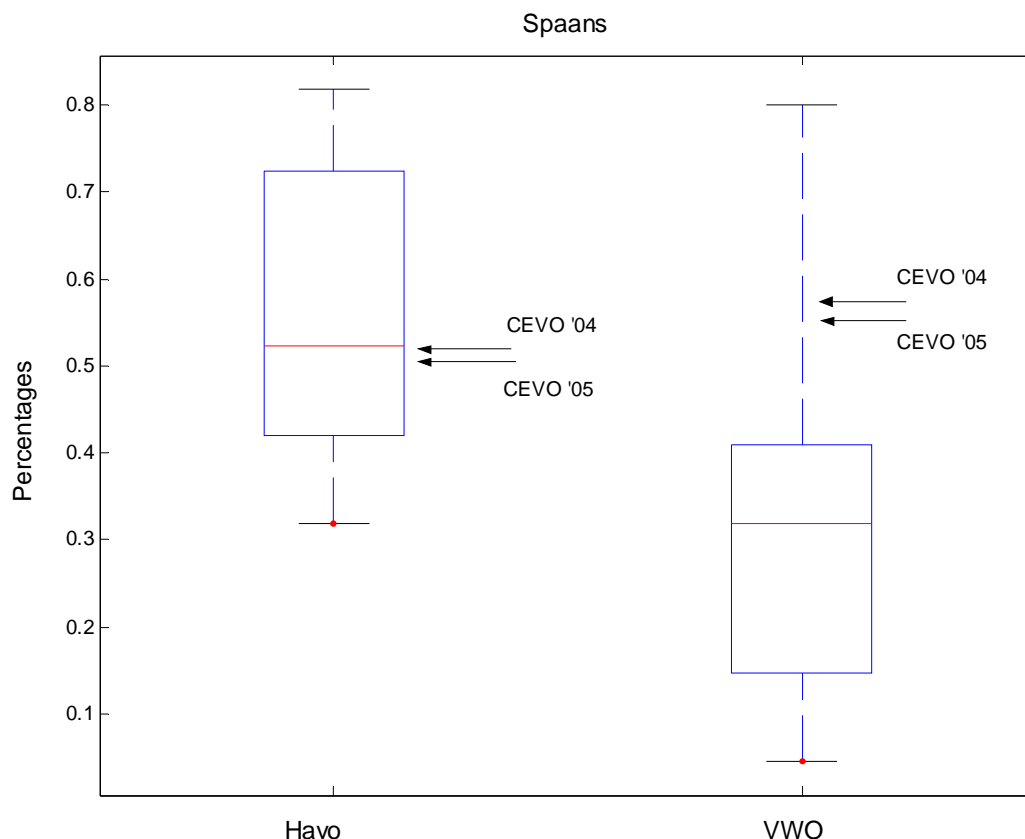
## Bijlage 2 De bepaling van minimum scores voor relevante ERK-niveaus

### Opzet van de analyse

In de onderhavige studie wordt gebruik gemaakt van expertoordelen. Zoals eerder beschreven geven beoordelaars bij elk item aan op welk ERK niveau het item beheerst moet worden. Zo kunnen we bij elke beoordelaar vaststellen welke items beheerst moeten worden op, bijvoorbeeld, niveau A2. We kijken naar het percentage items dat beheerst moet worden op het niveau waarop het betreffende examen geacht wordt te meten. Elke beoordelaar levert een percentage zodat we bij elk examen een aantal percentages kunnen berekenen. Hierboven zijn de gemiddelde percentages per examen gerapporteerd. De beoordeelde items zijn een representatieve steekproef van de beschikbare items en daardoor beschouwen we deze percentages als schattingen van het percentage items dat beheerst moet worden van alle beschikbare items dan wel van een examen dat is samengesteld met een representatieve steekproef van de beschikbare items.

### Overeenstemming tussen beoordelaars

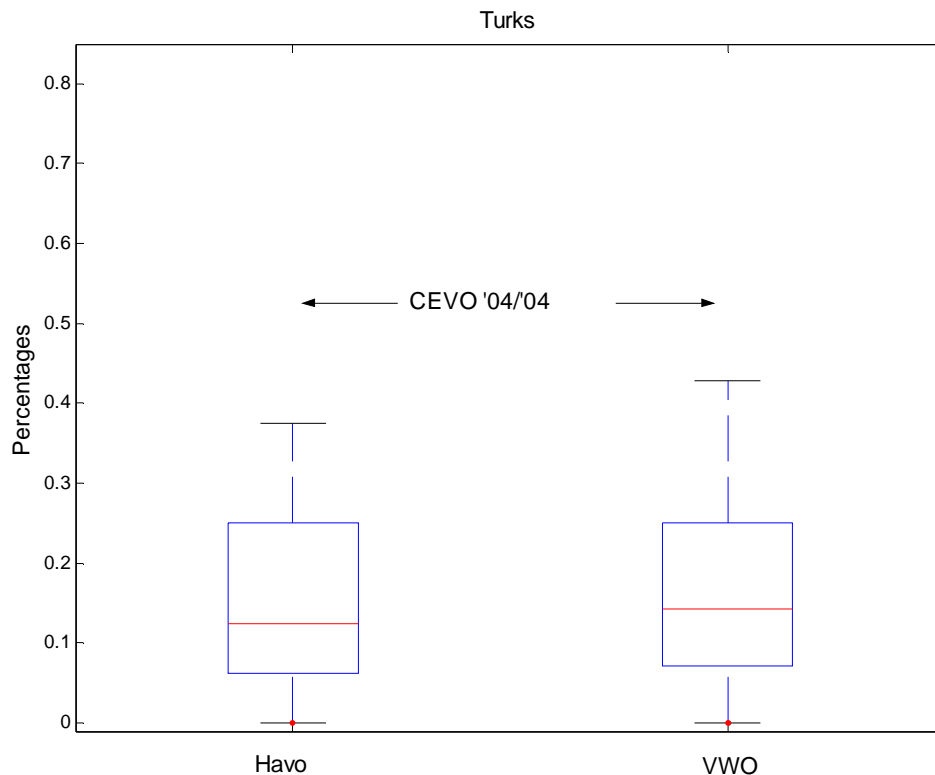
Het rapporteren van gemiddelde percentages doet geen recht aan de variatie tussen beoordelaars. Als de percentages van verschillende beoordelaars sterk variëren dan moeten we concluderen dat we er niet in geslaagd zijn om vast te stellen welk niveau beheerst moet worden om de items te kunnen maken. Daarnaast kunnen we, op basis van het gemiddelde percentage, niet goed vaststellen of en in welke mate de CEVO cesuur afwijkt van de percentages die zijn gevonden in de huidige studie.



Om zowel het gemiddelde percentage als de variatie tussen beoordelaars zichtbaar te maken kunnen we zogenaamde doos en bakkebaard (engels: *box and whisker*) plots gebruiken. We gebruiken ze om enkele resultaten toe te lichten.

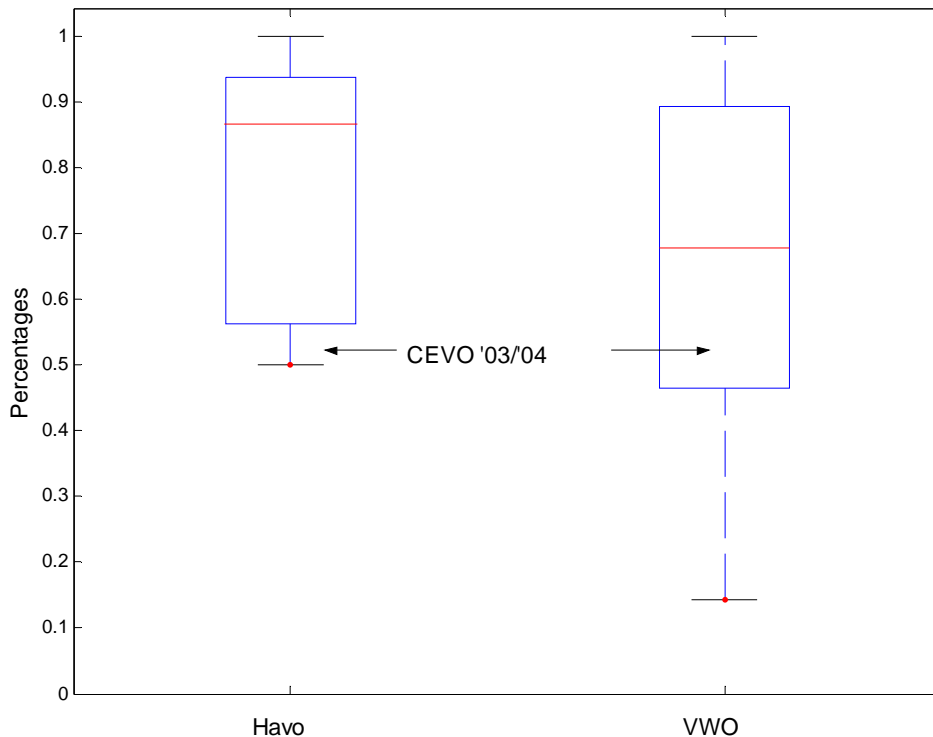
In bovenstaande figuur zien we een doos en bakkebaard plot van de percentages die zijn gevonden bij examens Spaans voor havo en vwo. De doos geeft het bereik aan van de meeste percentages.

De bakkebaarden geven aan waar de rest van de percentages lagen.<sup>7</sup> Beoordelaars met afwijkende percentages liggen buiten de bakkebaarden. Het gemiddelde percentage is aangegeven met een horizontale lijn in de doos. Tenslotte is de plaats van de CEVO cesuur aangegeven in de plot. Allereerst zien we dat de beoordelaars het niet volledig eens waren. Voor de havo examens Spaans variëren de percentages tussen 0,42 en 0,72. De CEVO cesuur ligt voor het examen vwo Spaans buiten de doos en dat wil zeggen dat ze meer afweken dan de variatie tussen beoordelaars.



Tenslotte zien we in bovenstaande figuur de resultaten voor het examen Turks voor havo en vwo. De percentages variëren tussen de 0,25 en 0,05. De CEVO cesuren waren duidelijk hoger dan wat volgens de beoordelaars noodzakelijk is voor de beheersing van niveau B1. Kijken we naar de percentages nodig om niveau B2 te beheersen krijgen we onderstaande figuur. In deze figuur is te zien dat de CEVO cesuren meer overeenkomen met hetgeen nodig is om niveau B2 te kunnen claimen. Merk op dat de overeenstemming tussen beoordelaars niet groot was voor het vwo examen.

<sup>7</sup> Om precies te zijn liggen de randen van de doos bij de onderste 25% en de bovenste 25%.



### Meetfouten

Wanneer de items waarvan experts het niveau hebben vastgesteld, opnieuw zouden worden voorgelegd aan kandidaten, dan kan het gebeuren dat een kandidaat inconsistent is in die zin dat hij of zij items van een lager niveau fout beantwoordt en van een hoger niveau goed. Hiervoor kunnen twee oorzaken zijn:

1. De items en de kandidaten zijn niet op dezelfde schaal te representeren. De toets ordent kandidaten op een andere wijze dan het ERK.
2. Meetfout: de variatie in de prestaties van de kandidaten zorgt voor willekeurige afwijkingen.

De tweede oorzaak is van weinig belang voor het gebruik van de resultaten. Het percentage goed beantwoorde items van een kandidaat blijft een schatting van het percentage goed op de gehele verzameling items. Als een A2 kandidaat bijvoorbeeld, 80% van de items goed zou moeten maken dan kunnen we stellen dat een kandidaat met meer dan 80% goede antwoorden volgens de toets tenminste het niveau A2 heeft.

De eerste oorzaak: de *moeilijkheid* van de items en de *vaardigheid* van de kandidaten zijn niet op één schaal te representeren, blijft een probleem. Of dit fenomeen zich voordoet, kan worden onderzocht met behulp van het IRT-model. Zoals eerder gezegd, hebben wij door de kleine aantallen kandidaten onvoldoende data bij de huidige examens tot onze beschikking om van het IRT-model gebruik te kunnen maken. Echter, in het vorige onderzoek bij de examens Frans, Duits en Engels, waar het model wél toepasbaar was, bleek dat dit fenomeen zich niet voordeed.

### Opmerking over de CEVO cesuur

Het is opvallend dat de CEVO cesuur tendeert net boven de 50% te liggen terwijl de in de huidige studie vastgestelde percentages die behoren bij een ERK-niveau variëren van examen tot examen. In onze studie lijkt de CEVO cesuur niet gerelateerd te zijn aan het ERK niveau. Dit was ook te verwachten: het ERK is immers later verschenen dan de N-norm waarop de CEVO-cesuren zijn bepaald. Er is derhalve geen formele relatie tussen het relevante ERK-niveau van een examen en de CEVO-cesuur.