

Verslag lijn 2

Taal en niveau: Russisch havo/vwo en Spaans vmbo gl/tl

Vaardigheden: leesvaardigheid met focus op literaire teksten en fictie

Uitvoering: Harriet van Munster; Ad Winnubst

Redactie: Rosa Bartman; Danai Psathas; Noëlle Sommer

Inhoudsopgave

1. Aanleiding onderzoek	5
2. Eerder onderzoek.....	5
3. Onderzoeksvragen	8
4. Toetsoperationalisatie Russisch.....	9
5. Toetsoperationalisatie Spaans	13
6. Conclusie	18
7. Aanbevelingen	21
Referenties	22

1. Aanleiding onderzoek

Deze trial is onderdeel van lijn 2 van het onderzoeksproject ‘Talenexamens van de toekomst’ en richt zich op literaire teksten en fictie in high-stakes toetsen. Het doel van dit onderzoeksproject was o.a. het adviseren van vakvernieuwingscommissie MVT en een overtuigend idee geven van taken die wel of niet haalbaar zijn in een vernieuwd eindexamen.

De taken die in deze onderzoekslijn zijn ontwikkeld, sluiten aan bij concepteindtermen 16 en 18 van domein C (cultuurbewustzijn) uit het conceptexamenprogramma mvt (moderne vreemde talen) (SLO, 2024)¹:

- “16. De leerling doet ervaring op met verschillende vormen van fictie in de doeltaal.” (pp. 143-144)
- “18. Havo en vwo: De leerling toont inzicht in literaire teksten uit diverse cultuurhistorische perioden.” (pp. 148-149)

2. Eerder onderzoek

Literaire teksten worden met verschillende doelen ingezet in afsluitende toetsen en examens.

In navolging van Bloemert en Cats (2020, p. 402) hanteren we in dit verslag een brede definitie van de begrippen “literatuur” en “literaire teksten”. Zij omschrijven literatuur als “[...] alle vormen van creatief taalgebruik [...]. Hieronder vallen bijvoorbeeld: romans, gedichten, toneelstukken, speeches, songteksten, kinderboeken, films, fan fiction, oude literatuur, jeugdliteratuur, beeldromans, twihaiku, en instapoetry.” In dit verslag gebruiken we de termen “literaire tekst” en “fictie(tekst)” daarom als synoniemen. Op basis van een verkennende analyse van nationale en internationale afsluitende examens in zowel moedertaalonderwijs (L1) als vreemdetalenonderwijs (L2) kunnen grofweg twee typen examens waarin literaire teksten voorkomen, worden onderscheiden: (1) examens waarin literatuur(geschiedenis) wordt getoetst als afsluiting van onderwezen leerstof en (2) examens waarin literaire teksten worden aangeboden zonder specifieke, vereiste voorbereiding in de les met als doel om leesvaardigheid te toetsen.

Het Franse Baccalauréat is een voorbeeld van het eerste type. In deze examens worden teksten aangeboden in het Frans, waarbij open vragen worden gesteld over eerder onderwezen leerstof, zoals kennis van literatuurstromingen. Daarnaast komen samenvattings- en essayopdrachten aan de orde, waarbij de schrijfvaardigheid en taalverzorging van de leerling worden meegenomen in

¹ Op het moment dat dit onderzoek in uitvoering was, was het nog niet duidelijk of deze eindtermen in het schoolexamen of het centraal examen getoetst zouden worden. Inmiddels is gebleken dat eindtermen 16 in het schoolexamen getoetst zal worden, terwijl eindterm 18 in het centraal examen getoetst zal worden.

de beoordeling. Het Nederlandse centraal schriftelijk eindexamen is daarentegen een voorbeeld van het tweede type examen. Literaire teksten worden hier bij sommige mvt's ingezet om een uitspraak te doen over de leesvaardigheid van de leerling.

In 2020 is in een eerder stadium van het onderzoeksproject Talenexamens van de toekomst reeds een aantal taken ontworpen waarin literaire teksten centraal staan. Tabel 1 geeft een overzicht van deze taaltoetstaken. Bij deze taken lag de focus op het in 2020 gepubliceerde Companion Volume van het Europees Referentiekader (ERK) (Nederlandse Taalunie, 2022), waarin het oorspronkelijke ERK wordt uitgebreid met onder andere drie nieuwe schalen op het vlak van literaire teksten en fictie: *lezen als vrijetijdsbesteding, een persoonlijke reactie geven op creatieve teksten (waaronder literaire)* en *analyse van en kritiek op creatieve teksten (waaronder literatuur)*. Omdat bij de laatste twee schalen een productieve uitwerking in de doeltaal vaak een logische component is, bevatten de meeste taken in Tabel 1 dan ook een schrijfvaardigheidsonderdeel.

Tabel 1. Overzicht van ontwikkelde taken in Talenexamens van de toekomst 2020 rondom literaire teksten in high-stakes toetsen.

Taal, niveau, aantal leerlingen	Samenvatting taaltoetstaak	Evaluatie
Engels havo/vwo (B1/B2) 115 leerlingen	De taaltoetstaak bestaat uit het lezen van een literair stuk, waarop creatief en reflectief geschreven gereageerd moet worden. Het leesgedeelte wordt beoordeeld met een gesloten antwoordmodel waarbij de antwoorden juist of onjuist zijn. De schrijfpdracht wordt met een rubric beoordeeld, waar 0, 1 of 2 punten gehaald kunnen worden voor verschillende aspecten van schrijfvaardigheid.	De leerlingen vonden de taak pittig. Aandachtspunten volgens docenten: formulering van de opdrachten en de beoordeling.
Nederlands havo (3F) niet afgenomen	Leerlingen schrijven een korte tekst naar aanleiding van een literair verhaal met analysevragen. Het beoordelingsmodel dat hiervoor ontwikkeld is, is een globale comparatieve beoordeling op een aantal aspecten. De leesopdracht wordt beoordeeld met voorbeeldantwoorden die vergeleken kunnen worden met de antwoorden van de leerlingen.	-
Spaans vwo (B1/B2)	Een taak waarin alle vaardigheden werden gecombineerd. Als input voor de taak	De taak was wat te moeilijk voor de leerlingen. Dit hangt samen

18 leerlingen	dienden de trailer van een film en een interview met de auteur van een boek dat verfilmd is. Deze input moest vervolgens worden verwerkt in een spreek- of schrijfpdracht (naar keuze van de leerling).	met het onderwijs, waarin niet alle vaardigheden al op het door de taak gevraagde niveau waren.
Russisch havo/vwo 2 leerlingen	Taaltoetstaak bestaande uit twee delen: 1) Drie gedichten met begrips- en vergelijkingsvragen, zowel op het gebied van leesvaardigheid als schrijfstrategieën, en zowel meerkeuzevragen als open vragen. 2) Video met spoken word en commentaren op de video. Getoetst worden: - luistervaardigheid, aan de hand van de gesproken tekst in de videoclip; - leesvaardigheid, aan de hand van commentaren op de videoclip; - schrijfvaardigheid, aan de hand van schrijfpdrachten.	Als er verder naar deze taak gekeken gaat worden, is het belangrijk om aandacht te geven aan een sluitend beoordelingsmodel met eventueel meerdere beoordelaars en de technische mogelijkheden voor deze taak. Daarnaast is het belangrijk om te kijken naar de relatie tussen beeld en geluid en of die eventueel afleidend zou kunnen werken. Er zijn dus relatief veel zaken waar rekening mee gehouden zal moeten worden. Deze taak is daarom wellicht minder geschikt voor verdere ontwikkeling dan sommige andere taken die zijn ontwikkeld.

In het huidige onderzoek verkennen we hoe de eindtermen die in sectie 1 zijn genoemd, binnen de kaders van de huidige centrale examens mvt getoetst zouden kunnen worden. In sommige mvt-examens worden op de niveaus vmbo-gl/tl, havo en vwo naast zakelijke leesteksten uit kranten of tijdschriften al literaire teksten aangeboden, met als doel om de leesvaardigheid van leerlingen te toetsen. Op dit moment worden de literaire teksten meestal aan de hand van gesloten en semi-open vragen (doorgaans beweringenvragen, waarbij een leerling van iedere bewering in een tabel moet aangeven of deze juist of onjuist is). In de meeste examens gaat het om één literaire tekst van de in totaal 10 tot 16 teksten die in een mvt-examen zitten.

Het bevragen van literaire teksten in het centraal examen heeft mogelijk een psychometrisch voordeel ten opzichte van enkel het bevragen van zakelijke teksten. Een beperkte analyse van tia's (toets- en itemanalyses) van de centrale examens voor de talen Engels, Frans en Duits uit 2019 tot 2023 suggereert dat literaire teksten leesvaardigheid iets beter zouden kunnen meten dan andere tekstsoorten². Het gemiddelde onderscheidende vermogen (gemeten aan de hand van de item-testcorrelatie *rit*) van items bij literaire teksten lag namelijk hoger (0,37) dan het gemiddelde onderscheidende vermogen van alle items in de onderzochte examens (0,31). Daarnaast was de gemiddelde moeilijkheid (p-waarde) van de literaire items nagenoeg

² Dit zou verder onderzocht moeten worden, om de suggestie steviger te onderbouwen.

hetzelfde (0,62) als de gemiddelde moeilijkheid van alle items in de verschillende examens die in de analyse zijn meegenomen (0,63).

Daarnaast zou het opnemen en bevragen van literaire teksten de examens inclusiever kunnen maken. Cito (2024) heeft een kleinschalig onderzoek uitgevoerd naar het opnemen van literaire teksten of fictie in het centraal examen Nederlands op niveau vmbo-gl/tl. Uit dit beperkte onderzoek onder 43 leerlingen blijkt dat het zinvol is om naast reguliere, zakelijke teksten ook literaire teksten op te nemen in leesvaardigheidsexamens. Leerlingen presteerden gemiddeld genomen vergelijkbaar op vragen bij een reguliere, zakelijke tekst als vragen bij een literaire tekst. Uit nadere analyse bleek echter dat meisjes gemiddeld beter presteerden op opgaven bij de literaire tekst dan de reguliere tekst, terwijl jongens gemiddeld juist beter presteerden op opgaven bij de reguliere, zakelijke tekst. Het opnemen van verschillende tekstsoorten kan dus voorkomen dat bepaalde groepen nadelige effecten ondervinden van een bepaalde tekstsoort in een examen.

3. Onderzoeksvragen

De hoofdvragen voor lijn 2 luiden:

- Op welke manier kan het construct leesvaardigheid worden gemeten in een high-stakes examen aan de hand van literaire teksten?
- Wat vinden docenten en leerlingen op vmbo-gl/vmbo-tl/havo/vwo van de inzet van literaire teksten in een leesvaardigheidstoets?

Om de eerste hoofdvraag te beantwoorden, zijn drie deelvragen geformuleerd:

- Welke tekstsoorten passen bij de doelgroepen (leerlingen op vmbo-gl/tl enerzijds en leerlingen op havo/vwo anderzijds)?
- In hoeverre kunnen er bij literaire teksten opgaven geconstrueerd worden die passen binnen de huidige richtlijnen van het centraal schriftelijk examen?
- In hoeverre kunnen opgaven bij literaire teksten beoordeeld worden volgens het huidige correctievoorschrift voor centraal schriftelijke examens?

In dit verslag bespreken we nu eerst een toetstaak die voor Russisch (havo/vwo) is ontwikkeld door een toetsdeskundige Russisch. Daarna bespreken we een toetstaak die voor Spaans (vmbo-gl/tl) is ontwikkeld door een toetsdeskundige Spaans. Op basis van de resultaten uit de afname van deze toetstaken beantwoorden we de hoofd- en deelvragen. De toetstaken die in dit verslag worden gepresenteerd, zijn in diverse aspecten vernieuwend ten opzichte van het huidige centraal schriftelijk examen. Allereerst bestaan deze toetstaken volledig uit literaire teksten, terwijl de huidige mvt-examens vaak maar één literaire tekst bevatten (van de in totaal 10-16 teksten in een examen). Daarnaast bestaan de ontwikkelde toetstaken uit diverse opgavenvormen, terwijl literaire teksten op dit moment doorgaans bevragd worden aan de

hand van klassieke meerkeuzevragen of beweringenvragen. Tot slot is de toetstaak voor Spaans vmbo-gl/tl vernieuwend, omdat er momenteel geen literaire teksten worden bevroegd op dit niveau.

4. Toetsoperationalisatie Russisch

4.1 Overzicht taak

De voor Russisch ontwikkelde taaltoetstaak toetst leesvaardigheid aan de hand van uiteenlopende literaire teksten in de breedste zin van het woord: een strip, gedichten, liedteksten en een Facebookpost. Deze teksten worden bevroegd zoals nu in de centrale examens gebruikelijk is, d.w.z. met meerkeuzevragen, beweringenvragen en kort-open-antwoordvragen, waarbij moet worden opgemerkt dat beweringenvragen vooral gebruikelijk zijn bij vmbo en minder bij havo en vwo. De taak is bedoeld voor havo- en vwo-leerlingen Russisch. In de taaltoetstaak wordt onder meer gevraagd naar de gedachtegang van de auteur (tekstballonnetjes invullen in een strip), stilistische keuzes van de auteur ('Welk gedicht is in rijm geschreven?'), tekstdoel ('Welk gedicht is voor kinderen geschreven?'), impliciete tekstinhoud en gevoelens die de auteur uitdrukt. In totaal waren 19 scorepunten te behalen met 12 items.

4.2 Ontwikkeling

De taaltoetstaak is geconstrueerd naar model van de huidige centrale examens op basis van de Constructieopdracht Talen vmbo, havo, vwo. De constructieopdracht is een document dat door het College voor Toetsen en Examens (CvTE) wordt opgesteld en dat omschrijft aan welke eisen de door Cito ontwikkelde examens moeten voldoen. Bij deze taaltoetstaak was de insteek om alleen leesvaardigheid te toetsen, maar dit construct kan niet helemaal los worden gezien van cultuurbewustzijn. Een leerling met kennis van literatuur, cultuur en geschiedenis van het land van de doeltaal zal immers waarschijnlijk beter scoren dan een minder cultuurbewuste leerling. Literatuurteksten lenen zich goed voor vragen die verder gaan dan vragen naar 'wat er staat' en voor vragen die een beroep doen op hogere denkvaardigheden.

Als inspiratiebron zijn voorafgaand aan de constructie van de taak het Russische equivalent van het Nederlandse centraal examen en het Language and Literature-examen van International Baccalaureate (Language B) bekeken. Over het algemeen geldt dat de tekstfragmenten uit dergelijke examens van te hoog niveau zijn voor examenkandidaten Russisch in Nederland en dat de overgrote meerderheid van de opgaven uit deze examens ook andere vaardigheden toetst dan enkel leesvaardigheid. Om die reden zijn vergelijkbare opgaven over het algemeen niet inzetbaar in de huidige kaders van het Nederlandse centraal examen. Toch zijn enkele

elementen uit bovenstaande examens meegenomen naar deze taaltoetstaak, bijvoorbeeld het opnemen van opgaven die betrekking hebben op meerdere vergelijkbare teksten (vergelijken van teksten), het herkennen van stijlmiddelen, zoals vergelijking, retorische vraag en overdrijving en het koppelen van karakterbeschrijvingen aan personages uit een verhaal.

Het bleek een uitdaging te zijn om literaire teksten te vinden op het juiste niveau. We passen literaire teksten namelijk, op details na, niet aan, aangezien het doel is om de teksten op een zo authentiek mogelijke wijze te presenteren aan de kandidaten. Het vraagt daarom een grote tijdsinvestering om teksten te vinden die qua niveau geschikt zijn voor examenkandidaten Russisch. De taaltoetstaak bevat teksten op B1/B2-niveau, maar eigenlijk is dit niveau net te hoog voor de kandidaten. Hiervoor is echter gecompenseerd in het constructieproces van de opgaven.

Een bijkomend probleem is dat het gros van de literaire teksten niet in een examen kan worden gebruikt, omdat er zaken in staan die mogelijk gevoelig liggen (voor bepaalde groepen). Het is eigen aan literatuur dat deze de randjes van 'wat kan' en 'wat niet kan' opzoekt. De preambule, onderdeel van de constructieopdracht die het College voor Toetsen en Examens (CvTE) jaarlijks opstelt, schrijft echter voor dat gevoelige inhoud vermeden dienen te worden in de centrale examens. Als literatuur een prominentere plaats krijgt in het centraal examen, moet worden heroverwogen of de preambule onverminderd strak moet worden nageleefd.

Wanneer je eenmaal geschikt materiaal hebt gevonden, bleek het goed mogelijk om verschillende soorten opgaven te construeren. Goede literaire teksten hebben een hoge informatiedichtheid, bevatten meningen, emoties, een expliciete dan wel impliciete boodschap en zijn om deze redenen uitstekend bevraagbaar. Het combineren van verschillende teksten levert nog meer vraagmogelijkheden op. Daarnaast zorgt het gebruik van strips en beeldmateriaal zowel voor een prettige 'look and feel' van het examen als voor een visuele ondersteuning bij lastigere teksten.

In de ontwikkelde toetstaak zijn de ERK-descriptoren in Tabel 1 aangehouden. De taaltoetstaak bevatte diverse teksten. De eerste opgave ging over een strip van vier panelen, waarbij bij één paneel de tekst is weggelaten. In een meerkeuzevraag moest de leerling de juiste tekstballon kiezen bij dit paneel. De tweede opgave bevatte een verhaal met illustraties, waarbij de titel van het verhaal was weggelaten en de leerling in een meerkeuzevraag de juiste titel moest kiezen. In de volgende opgave las de leerling korte gedichten en songteksten met eenzelfde thema. In deze opgave moest de leerling een vergelijkende analyse maken van deze teksten aan de hand van meerkeuzevragen en open vragen. Hiermee wordt een beroep gedaan op de hogere denkvaardigheden van de leerling, passend bij de ERK-descriptor 'Kan beschrijven op welke manier meerdere werken verschillen in hun behandeling van hetzelfde thema.' Er worden vragen gesteld over de door de auteur gekozen stilistische vorm ('Welk van de teksten is niet in rijm geschreven?'), over schrijfstrategieën ('Citeer een voorbeeld van beeldspraak.') en over

genre ('Welke tekst is voor kinderen bedoeld?'). De laatste opgave ging over een Facebookpost, waarbij leerlingen onder andere in een meerkeuzevraag de juiste toon van de tekst moesten kiezen.

Tabel 1. Overzicht van ERK-descriptoren die zijn gebruikt in de ontwikkeling van de toetstaak voor Russisch havo/vwo.

Categorie	Niveau	Descriptor
Algemene leesvaardigheid	B2	Kan in hoge mate zelfstandig lezen, past de eigen leesstijl en leessnelheid aan verschillende teksten en doeleinden aan, en maakt selectief gebruik van toepasselijke naslagwerken. Beschikt over een grote actieve leeswoordenschat, maar kan enige moeite hebben met weinig voorkomende idiomatische uitdrukkingen.
	B1	Kan eenvoudige gedichten en liedjesteksten begrijpen waarin een directe taal en stijl zijn gehanteerd.
Lezen als vrijetijdbesteding	B1	Kan beschrijvingen van plaatsen, gebeurtenissen, expliciet geuite gevoelens en perspectieven begrijpen in verhalende teksten, gidsen en tijdschriftartikelen die in veelgebruikte, alledaagse taal zijn geformuleerd.
	B1	Kan, al of niet met behulp van een woordenboek, het plot volgen van verhalen, eenvoudige romans en stripverhalen als ze een duidelijke verhaallijn hebben en in veelvoorkomende, alledaagse taal zijn opgesteld.
Analyse en kritische waardering van creatieve teksten	B1	Kan de essentiële thema's en personages in korte verhalende teksten over vertrouwde situaties die alleen veelvoorkomende alledaagse taal bevatten.
	B2	Kan twee werken met elkaar vergelijken wat betreft thema's, personages en scènes, overeenkomsten en contrasten verkennen en het belang van de verbanden ertussen uitleggen.
	B2	Kan beschrijven op welke manier meerdere werken verschillen in hun behandeling van hetzelfde thema.

4.3 Afname en beoordeling

Op 1 februari 2024 is de taaltoetstaak afgenomen onder twee leerlingen in één klas van twee docenten. Beide leerlingen waren geen moedertaalsprekers van het Russisch. De taaltoetstaak kon worden beoordeeld aan de hand van een correctievoorschrift, met daarin de juiste antwoorden op de meerkeuzevragen en kort-open-antwoordvragen. De toetstaak kon zo door de deelnemende docenten zelf worden beoordeeld. De beoordeelde toetstaken zijn echter niet geretourneerd aan de onderzoekers.

Zowel de deelnemende leerlingen als de deelnemende docenten vulden na de afname van de taak een vragenlijst in. In de vragenlijst voor leerlingen werd gevraagd om op een 5-puntsschaal te beoordelen hoe leuk en moeilijk ze de taak vonden. In gesloten vragen konden de leerlingen aangeven welk onderdeel ze het leukst en het moeilijkst vonden, en of de taak elementen bevatte waarmee ze in de lessen nog niet hadden geoefend. Tot slot was er een open vraag waarin de leerling zijn of haar mening over de taak kon geven.

In de vragenlijst voor docenten werd onder andere gevraagd om op 5-puntsschalen te beoordelen in hoeverre de taak als leuk en moeilijk werd ervaren door de leerlingen, of de inhoud van de taak passend was bij de leefwereld van de leerlingen, of de moeilijkheidsgraad passend was bij de leerlingen, of de opzet van de taak nieuw was voor de leerlingen en of een dergelijke taak passend zou zijn voor een centraal examen.

4.4 Resultaten

Omdat de vragenlijsten door slechts één docent en één leerling zijn ingevuld, is een kwantitatieve analyse van deze data weinig zinvol. We omschrijven hier daarom op kwalitatieve wijze het beeld dat we op basis van de vragenlijsten kunnen vormen.

De docent die de vragenlijst heeft ingevuld, gaf aan dat zijn leerling de teksten in de toetstaak leuker vond dan de reguliere, zakelijke examenteksten. Daarnaast gaf de docent aan de inhoud van de taak passend te vinden bij de leefwereld van de leerlingen en was de moeilijkheidsgraad volgens de docent ook passend, al gaf deze docent ook aan dat leerlingen wel veel onbekende woorden moesten opzoeken in het woordenboek. Dit schreef de docent echter voornamelijk toe aan de beperkte woordenschat van de leerlingen, en niet aan de moeilijkheidsgraad van de toetstaak. De docent gaf verder aan dat de taak compleet nieuw was voor de leerlingen, omdat de taak een stripverhaal bevatte. Tot slot gaf de docent aan de toetstaak erg geschikt te vinden voor een centraal examen, omdat deze teksten het examen interessanter en veelzijdiger zouden maken. Toch gaf de docent ook aan van mening te zijn dat men moet oppassen met literaire teksten, omdat ze doorgaans op verschillende manieren te interpreteren zijn. Deze mening werd nog eens extra onderstreept doordat deze docent liet weten bij één opgave, horend bij de vier

gedichten/songteksten, een ander antwoord dan in het correctievoorschrift werd voorgeschreven ook acceptabel te vinden.

Uit correspondentie met de docent bleek daarnaast dat beeldmateriaal bij de teksten niet alleen als leuk ervaren wordt, maar ook als nuttig, omdat plaatjes en afbeeldingen helpen de context beter te begrijpen. De teksten zonder illustraties werden door de leerlingen als moeilijk ervaren. Plaatjes lijken moeilijkere teksten dus toegankelijker te maken.

De vragenlijst die door de leerling is ingevuld, bevestigt het beeld van de docent grotendeels: de leerling gaf aan de taak inderdaad leuk te vinden, maar tegelijkertijd gaf de leerling ook aan de taak aan de moeilijke kant te vinden. Deze leerling gaf aan de opgave bij het beeldverhaal de leukste opgave te vinden. Dit was tevens een element waar de leerling nog nooit mee had geoefend in de lessen. De analyse van de vier gedichten/liedteksten in opgave 3 werd door deze leerling het moeilijkst gevonden.

5. Toetsoperationalisatie Spaans

5.1 Overzicht taak

Deze taaltoetstaak Spaans vmbo toetst leesvaardigheid bij examenklasleerlingen vmbo-gl/tl aan de hand van literaire teksten. De taaltoetstaak bevatte vier teksttypen: raadsels, een gedicht, een verhaal en een korte strip. Hierbij werden meerkeuzevragen, matchingvragen en kort-open-antwoordvragen gesteld. In totaal moesten leerlingen tien items voor 15 scorepunten maken in 45 minuten. Deze taaltoetstaak is zo opgezet dat deze in principe aan te passen is naar andere niveaus door andere typen literaire teksten te gebruiken, afhankelijk van het niveau van de doelgroep.

5.2 Ontwikkeling

Bij de ontwikkeling van de taaltoetstaak was het ERK onderliggend. De ERK-descriptoren die in Tabel 2 staan, werden in deze taak getoetst. Tevens is er bij de ontwikkeling van de taak rekening gehouden met de kaders van de huidige centrale examens mvt. Ter inspiratie is er gekeken naar voorbeelden van de toetsing van leesvaardigheid aan de hand van literaire teksten in de andere centrale examens mvt en in buitenlandse examens leesvaardigheid, specifiek die van PISA en International Baccalaureate.

Slechts weinig opgaven uit de buitenlandse examens PISA en International Baccalaureate waren geschikt voor de huidige taaltoetstaak. Bij de PISA-examens werden meerdere voorbeelden gevonden van opgaven bij literaire teksten waarin gevraagd werd naar een mening of een interpretatie. De antwoordmodellen bij deze vragen zijn uitgebreid en geven veel

voorbeelden, maar deze zijn niet uitputtend. Binnen de kaders van de huidige centrale examens mvt zouden dergelijke lang-open-antwoordvragen en de bijbehorende uitgebreide antwoordmodellen te ingewikkeld worden. Het is nu namelijk al lastig om een sluitend correctievoorschrift te maken voor kort-open-antwoordvragen, omdat leerlingen soms antwoorden geven die niet tot nauwelijks te voorzien zijn.

Het zoeken naar geschikte literaire teksten op vmbo-gl/tl bleek een uitdaging. Er is met name online gezocht, maar ook in de thuiscollectie van de toetsdeskundige Spaans die deze taak heeft ontwikkeld. De informatieve teksten die doorgaans op gl/tl-niveau worden ingezet in het examen worden meestal ingekort en vereenvoudigd. Bij literaire teksten is dit echter veel minder of niet mogelijk. Uiteindelijk zijn vier literaire teksten geselecteerd voor de toetstaak. In de verhalende tekst is hier en daar wel wat vereenvoudigd en zijn een aantal voetnoten toegevoegd om onbekende woorden uit te leggen.

Hoewel over het algemeen gekozen is voor bekende vraagtypes (matchingvragen, meerkeuzevragen, een beweringenvraag en een aantal kort-open-antwoordvragen), zijn in de opgaven in de huidige taaltoetstaak toch vernieuwende elementen opgenomen, zoals het vragen naar het thema van een verhaal. Ook zijn er opgaven geconstrueerd die al wel in de centrale examens mvt voorkomen, maar nog niet bij het huidige niveau. Zo komt het terugplaatsen van zinnen in de tekst (in dit geval het terugplaatsen van quotes in een stripverhaal) bij vwo al voor, maar is dit vernieuwend voor het niveau vmbo gl/tl. Ook is gepoogd goede kort-open-antwoordvragen te construeren, die momenteel slechts beperkt worden ingezet in het centraal examen Spaans vmbo gl/tl.

Tabel 2. Overzicht van ERK-descriptoren die zijn gebruikt in de ontwikkeling van de toetstaak voor Spaans vmbo-gl/tl.

Categorie	Niveau	Descriptor
Algemene leesvaardigheid	A2	Kan korte, eenvoudige teksten begrijpen die uit vaak voorkomende woorden bestaan, waaronder ook een aantal woorden uit de gemeenschappelijke internationale woordenschat.
	A2+	Kan korte, eenvoudige teksten begrijpen over vertrouwde zaken van concrete aard, die zijn geschreven in veelgebruikte alledaagse of werkgebonden taal.
Lezen als vrijetijdbesteding	A2	Kan begrijpen wat er in een fotoverhaal (bijvoorbeeld in een lifestyle magazine) gebeurt en zich een beeld vormen van hoe de personages zijn.

	A2+	Kan genoeg begrijpen om korte, eenvoudige verhalen en strips te lezen die over vertrouwde, concrete situaties gaan en waarin veelvoorkomende, alledaagse taal wordt gebruikt.
	B1	Kan, al of niet met behulp van een woordenboek, het plot volgen van verhalen, eenvoudige romans en stripverhalen als ze een duidelijke verhaallijn hebben en in veelvoorkomende, alledaagse taal zijn opgesteld.
Identificeren van aanwijzingen en afleiden van interpretaties (mondeling en schriftelijk)	A2	Kan de betekenis en functie van onbekende standaarduitdrukkingen afleiden op basis van hun plaats in de tekst (bijvoorbeeld aan het begin of het eind van een brief).
	A2	Kan getallen, data, namen, eigennamen enzovoort benutten om het onderwerp van een tekst te bepalen.
	A2+	Kan herkenning van bekende woorden gebruiken om de betekenis af te leiden van onbekende woorden in korte uitdrukkingen die in alledaagse situaties worden gebruikt.
	B1	Kan de waarschijnlijke betekenis van onbekende woorden in een tekst afleiden door de samenstellende delen waaruit ze zijn opgebouwd te identificeren (bijvoorbeeld stam, lexicale elementen, voor- en achtervoegsels).
	B1	Kan een redenering of een reeks van gebeurtenissen in een verhaal volgen door zich op veelvoorkomende logische verbindingswoorden (bijvoorbeeld, echter, omdat) en temporele verbindingswoorden (bijvoorbeeld daarna, eerder) te concentreren.
	B1	Kan een korte verhalende tekst bekijken of beluisteren en voorspellen wat daarna gaat gebeuren.
	B1	Kan, op basis van koppen en titels, conclusies trekken of voorspellingen doen over de inhoud van een tekst.

5.3 Afname en beoordeling

Idealiter was de taaltoetstaak online afgenomen. Er is gepoogd de taak in de omgeving van Microsoft Forms te zetten, maar deze omgeving bleek te beperkt, onder andere omdat deze niet geschikt is voor matchingvragen of beweringenvragen. De taak is daarom op papier afgenomen onder tien leerlingen uit vmbo-gl/tl 4, negentien leerlingen uit havo 4 elementair en tien leerlingen uit vwo elementair.³ De leerlingen vmbo-gl/tl hebben gemiddeld 3,5 jaar Spaans gevolgd, gemiddeld 2,5 uur per week. De havo- en vwo-leerlingen Spaans elementair hebben een half jaar Spaans gevolgd, gemiddeld 3 uur per week. Bij de afname waren drie docenten Spaans uit het voortgezet onderwijs betrokken. Na de afname vulden zowel de leerlingen als de docenten vragenlijsten in om de toetstaak te evalueren. Hiervoor is dezelfde vragenlijst gebruikt als voor de evaluatie van de Russische toetstaak (zie Sectie 4.3).

De taak is beoordeeld aan de hand van een correctievoorschrift, de gangbare beoordelingsmethodiek van de huidige centrale examens mvt. De toetsdeskundige die deze taak heeft ontwikkeld, heeft de ingevulde antwoordbladen van de leerlingen vmbo-gl/tl 4 nagekeken en geanalyseerd. Eén docent heeft de antwoordbladen van haar havo/vwo-leerlingen zelf nagekeken en een overzicht van de leerlingscores naar de toetsdeskundige gemaild. De antwoorden van deze leerlingen zijn echter niet meegestuurd en worden daarom in de bespreking van de resultaten niet meegenomen. Daarnaast heeft de betrokken toetsdeskundige de door de leerlingen en docenten ingevulde vragenlijsten geanalyseerd.

5.4 Resultaten

Leerlingscores

In totaal waren er bij de toetstaak 15 scorepunten te behalen. In Tabel 3 staat een overzicht van de gemiddelde score die leerlingen behaalden op de toetstaak, uitgesplitst per schoolniveau. Ook staat in de tabel weergegeven hoe lang iedere groep leerlingen gemiddeld over de taak deed. Uit Tabel 3 is op te maken dat de vmbo-leerlingen redelijk scoorden op de toetstaak: zij behaalden gemiddeld de helft van het totaal aantal te behalen scorepunten. De havo-leerlingen scoorden gemiddeld lager dan de vmbo-leerlingen, wat te maken zou kunnen hebben met het feit dat deze leerlingen pas een half jaar Spaans volgen. Opvallend genoeg scoorden de vwo-leerlingen gemiddeld het hoogst. Deze leerlingen hadden op het moment van afname ook slechts een half jaar Spaans gevolgd. Het schoolniveau, de inzet van compenserende strategieën en kennis van de wereld zouden een verklaring zijn voor deze hoge gemiddelde score.

³ De mvt's Arabisch, Italiaans, Russisch, Spaans en Turks kunnen als verkort traject (elementaire variant) worden gevolgd vanaf klas 4 van de havo of het vwo (SLO, 2007).

De leerlingen kregen 45 minuten de tijd voor deze taak. Uit tabel 3 blijkt dat leerlingen gemiddeld genomen minder dan 45 minuten nodig hadden om de taak te voltooien. De ingeschatte toetsduur was dus realistisch.

Tabel 3. Gemiddelde score en toetsduur per deelnemende groep leerlingen.

Groep	Gemiddelde score (max. 15)	Gemiddelde duur in minuten (max. 45)
vmbo-gl/tl 4 ($n = 10$)	7,60	40
havo 4 ($n = 19$)	6,10	35
vwo 4 ($n = 10$)	10,10	42
Gewogen gemiddelde	7,51	38

Evaluatie leerlingen

De leerlingevaluatie is door 38 van de in totaal 39 deelnemende leerlingen ingevuld. Tabel 4 toont hoe leuk en hoe moeilijk de leerlingen de taak in zijn geheel vonden. De leerlingen gaven de taak een gemiddelde score van 2,74 op een Likert-schaal met de schaalpunten 'helemaal niet leuk' (1), 'niet zo leuk' (2), 'redelijk leuk' (3), 'leuk' (4) en 'heel leuk' (5). Uit Tabel 4 blijkt dat vwo-leerlingen de taak gemiddeld genomen iets leuker vonden dan de vmbo-gl/tl- en havo-leerlingen.

Verder gaven de leerlingen aan de taak over het algemeen moeilijk te vinden: de taak kreeg een gemiddelde score van 2,22 op een Likert-schaal met de schaalpunten 'heel moeilijk' (1), 'moeilijk' (2), 'niet moeilijk/niet makkelijk' (3), 'makkelijk' (4) 'heel makkelijk' (5), zoals te zien is in Tabel 4. Opvallend was dat de havo-leerlingen de taak gemiddeld genomen het moeilijkst vonden, terwijl vmbo-leerlingen de taak gemiddeld het makkelijkst vonden.

Tabel 4. Gemiddeld oordeel voor de leukheid en moeilijkheid van de taak per deelnemende groep leerlingen.

Groep	Gemiddeld oordeel leukheid	Gemiddeld oordeel moeilijkheid
vmbo-GL/TL 4 ($n = 10$)	2,70	2,70
havo 4 ($n = 19$)	2,57	1,89
vwo 4 ($n = 10$)	3,10	2,40
Gewogen gemiddelde	2,74	2,23

Ook konden leerlingen in de vragenlijst aangeven welke tekst ze het leukst vonden en welke tekst ze het moeilijkst vonden. Op de vraag welke tekst de leerlingen het leukst vonden, waren de antwoorden zeer verdeeld. Wel bleek dat leerlingen het verhaal het moeilijkst vonden. Verder gaven leerlingen aan de opgave bij het gedicht het makkelijkst te vinden, maar uit de beoordeling van de antwoorden van de vmbo-leerlingen, bleek dat slechts één leerling deze opgave juist had. De open vragen bij het verhaal werden het moeilijkst gevonden. Tot slot gaven leerlingen aan dat vooral de raadsels en strip nieuw waren en dat ze met het lezen van dit soort teksten nog niet hadden geoefend in de lessen.

Evaluatie docenten

Twee van de drie deelnemende docenten hebben de vragenlijst ingevuld. Beide docenten zijn werkzaam in het vmbo op gl/tl-niveau. Zij gaven aan de indruk te hebben dat hun leerlingen de taak redelijk leuk vonden: ze gaven een 3,5 op een Likert-schaal met de schaalpunten 'helemaal niet leuk' (1), 'niet zo leuk' (2), 'redelijk leuk' (3), 'leuk' (4) en 'heel leuk' (5). Dit is hoger dan het gemiddelde oordeel voor de leukheid van de taak dat leerlingen gaven (2,74). Daarnaast hadden de docenten de indruk dat hun leerlingen de taak aan de moeilijke kant vonden: ze gaven een 2,5 op een een Likert-schaal met de schaalpunten 'heel moeilijk' (1), 'moeilijk' (2), 'niet moeilijk/niet makkelijk' (3), 'makkelijk' (4) en 'heel makkelijk' (5). Het docentenoordeel voor de moeilijkheid van de taak is vergelijkbaar met het oordeel dat de leerlingen gaven (2,23). In de vragenlijst gaven de vmbo-docenten echter ook aan de moeilijkheidsgraad van de taak passend bij het niveau van hun leerlingen te vinden. Bij de interpretatie van deze resultaten moet rekening worden gehouden met het feit dat de docentenevaluatie alleen door vmbo-docenten is ingevuld, terwijl de leerlingevaluatie door vmbo-, havo- en vwo-leerlingen is ingevuld.

De docenten waren verder van mening dat de materialen goed aansloten bij de leefwereld van de leerlingen. Ook gaven de docenten aan dat de opzet en inhoud van de taak redelijk nieuw was voor de leerlingen, maar vonden ze de taak en materialen een goede manier om leesvaardigheid te toetsen. Tot slot stonden de docenten redelijk positief tegenover het inzetten van dergelijke literaire teksten in een centraal examen.

6. Conclusie

Naar aanleiding van de toetsoperationalisatie en de afnamegegevens beantwoorden we hier de onderzoeksvragen. De onderzoeksvragen waren als volgt:

- Op welke manier kan het construct leesvaardigheid worden gemeten in een high-stakes examen aan de hand van literaire teksten?
- Wat vinden docenten en leerlingen op vmbo-gl/vmbo-tl/havo/vwo van de inzet van literaire teksten in een leesvaardigheidstoets?

Om de eerste onderzoeksvraag te beantwoorden, zijn drie deelvragen geformuleerd:

- Welke tekstsoorten passen bij de doelgroepen (leerlingen op vmbo-gl/tl enerzijds en leerlingen op havo/vwo anderzijds)?
- In hoeverre kunnen er bij literaire teksten opgaven geconstrueerd worden die passen binnen de huidige richtlijnen van het centraal schriftelijk examen?
- In hoeverre kunnen opgaven bij literaire teksten beoordeeld worden volgens het huidige correctievoorschrift voor centraal schriftelijke examens?

Welke tekstsoorten passen bij de doelgroepen (leerlingen op vmbo-gl/tl enerzijds en leerlingen op havo/vwo anderzijds)?

In de toetstaak die voor Russisch (havo/vwo) is ontworpen, kwamen een strip, een beeldverhaal, gedichten en songteksten en een Facebookpost aan bod. In de Spaanse toetstaak (vmbo-gl/tl) kwamen raadsels, een gedicht, een verhaal en een strip aan bod. Uit de docentenevaluaties bleek dat de deelnemende docenten deze teksten passend vonden bij de belevingswereld van de doelgroepen.

De toetsdeskundigen die deze taken hebben ontworpen, gaven aan dat het lastig was om literaire teksten te vinden van een geschikt niveau voor leerlingen op vmbo-, havo- en vwo-niveau. Omdat de centrale examens authentieke teksten moeten bevatten, kunnen literaire teksten niet tot nauwelijks worden aangepast. Toch vonden de deelnemende docenten het niveau van de toetstaken voor Spaans en Russisch goed bij hun leerlingen aansluiten.

In hoeverre kunnen er bij literaire teksten opgaven geconstrueerd worden die passen binnen de huidige richtlijnen van het centraal schriftelijk examen?

De toetsdeskundige die de toetstaak voor Russisch heeft ontwikkeld, constateerde dat het een uitdaging is om teksten te vinden die qua inhoud binnen de preambule passen. De preambule schrijft namelijk voor dat gevoelige inhoud vermeden dienen te worden in de centrale examens. Literaire werken zijn echter vaak bedoeld zijn om te schuren en te emotioneren en zoeken vaak de randjes op van 'wat kan' en 'wat niet kan'. Voor deze taak is het uiteindelijk toch gelukt om materiaal te vinden dat volgens de preambule is toegestaan.

Momenteel worden literaire teksten in de centrale examens mvt overwegend bevraagd aan de hand van traditionele meerkeuzeopgaven en beweringenvragen. In de taken die hier voor Russisch en Spaans zijn ontwikkeld, is er geëxperimenteerd met andersoortige opgavenvormen, zoals kort-open-antwoordvragen, sorteervragen ('Zet onderstaande tekstballonnetjes in de juiste volgorde') en matrixvragen, waarin de leerling voor vier tekstelementen moest aangeven in welke teksten (van in totaal vier teksten) dit element voorkwam. Op basis van de toetstaken die in dit onderzoek zijn ontwikkeld, kan worden aangenomen dat deze vernieuwende opgavenvormen ook in centrale examens inzetbaar zouden moeten zijn, voor zover deze binnen de constructieopdracht passen.

In hoeverre kunnen opgaven bij literaire teksten beoordeeld worden volgens het huidige correctievoorschrift voor centraal schriftelijke examens?

In combinatie met het huidige correctievoorschrift is het goed mogelijk om gesloten vragen te stellen bij literaire teksten. Ook kort-open-antwoordvragen zijn mogelijk bij literaire teksten, al zal het correctievoorschrift hier mogelijk wat meer aandacht behoeven dan bij een zakelijke tekst. De toetsdeskundige Spaans zag namelijk dat het correctievoorschrift niet voldoende sluitend was bij de ontwikkelde toetstaak voor Spaans: de leerlingen gaven antwoorden die niet in het correctievoorschrift stonden, maar die bij nader inzien ook als goed antwoord konden worden aangemerkt. De toetsdeskundige Russisch had een vergelijkbare ervaring, maar dan bij een gesloten vraag in plaats van bij een open vraag. Dit gegeven is waarschijnlijk in verband te brengen met het feit dat literaire teksten vaak voor meerdere interpretaties open staan. Bij zowel open als gesloten vragen is het dus zaak om te voorkomen dat meerdere antwoorden juist zijn, of een correctievoorschrift op te stellen waarin voldoende ruimte is om meerdere antwoorden (gedeeltelijk) juist te rekenen.

Wat vinden docenten en leerlingen op vmbo-gl/vmbo-tl/havo/vwo van de inzet van literaire teksten in een leesvaardigheidstoets?

De taak voor Russisch is door slechts één docent en één leerling geëvalueerd. Beiden reageerden positief op de taak. In de vragenlijst gaf de leerling die de vragenlijst heeft ingevuld aan het beeldverhaal de leukste tekst uit de taak te vinden. Tegelijkertijd gaf deze leerling aan nog nooit met beeldverhalen en stripverhalen te hebben geoefend. Hoewel de docent aangaf de moeilijkheidsgraad goed te vinden passen bij de leerlingen, vond de leerling die de vragenlijst heeft ingevuld de taak wel aan de moeilijke kant. De docent was tot slot kritisch over de uitvoerbaarheid van literaire teksten in een centraal examen, omdat literaire teksten vaak voor meerdere interpretaties open zijn en hiermee in het correctievoorschrift rekening moet worden gehouden.

De taak voor Spaans is door twee docenten en 38 leerlingen geëvalueerd. Hoewel de docenten de het idee hadden dat de leerlingen de taak redelijk leuk vonden, vonden de leerlingen de gemiddeld genomen niet zo leuk tot redelijk leuk. Hierin was een groot verschil zichtbaar tussen leerlingen van de verschillende doelgroepen: vwo-leerlingen vonden de taak het leukst, gevolgd door vmbo-leerlingen en tot slot havo-leerlingen. Er was een grotere overeenstemming tussen de docenten en de leerlingen op het vlak van de moeilijkheid van de taak, waarbij zowel docenten als leerlingen aangaven de taak iets aan de moeilijke kant te vinden, al gaven de docenten tegelijkertijd ook aan dat ze de moeilijkheidsgraad van de taak goed vonden passen bij het niveau de leerlingen. Verder gaven de docenten aan dat zij literaire teksten een goede manier om leesvaardigheid te toetsen vonden en waren ze positief tegenover het inzetten van literaire teksten in een centraal examen, wat bij Spaans op de niveaus havo en vwo ook al gebeurt, maar nog niet op vmbo gl/tl.

Al met al reageren docenten en leerlingen overwegend positief op opgaven bij literaire teksten en op het inzetten van literaire teksten in een centraal examen, al plaatsen docenten ook kritische kanttekeningen bij de beoordeling van literaire teksten in een centraal examen. Omdat literaire teksten doorgaans meerdere interpretaties toelaten, moet hiermee in het beoordelingsmodel rekening worden gehouden.

7. Aanbevelingen

Vinden van geschikt bronmateriaal

Bij de ontwikkeling van de toetstaken bleek het een uitdaging om geschikte literaire teksten uit de Russischtalige of Spaanstalige wereld te vinden. De toetsdeskundigen hebben bij de ontwikkeling van deze taak vooral via internet gezocht en in hun eigen boekencollecties gezocht naar geschikt bronmateriaal. Een van de toetsdeskundigen opperde dat het bezoeken van boekhandels in landen waar de doeltaal wordt gesproken meer geschikt bronmateriaal zou kunnen opleveren. Ook zou het benaderen van Russisch- of Spaanstalige scholen in binnen- en buitenland zou aanknopingspunten voor geschikt bronmateriaal kunnen opleveren.

Specifiek voor de examens Spaans zou de bijdrage in Levende Talen Magazine van Marloes Mekenkamp (2024) een startpunt voor interessant en geschikt bronmateriaal kunnen bieden. In dit artikel geeft zij suggesties voor Spaanstalige poëziewerken die in geïntegreerde taal- en cultuurlessen Spaans gebruikt zouden kunnen worden. Er zou nader kunnen worden onderzocht of deze zich lenen voor gebruik in examens. Toch is er hierbij ook voorzichtigheid geboden, omdat het door dit artikel goed mogelijk is dat deze werken door sommige docenten Spaans uitgebreid zijn behandeld in de lessen, waardoor bepaalde groepen leerlingen een voordeel zouden kunnen hebben als deze poëziewerken in een centraal examen zouden worden opgenomen.

Constructieve afstemming

Bij constructieve afstemming draait het om een goede samenhang tussen leerdoelen, leerresultaten, onderwijsactiviteiten en toetsing (Biggs, 1996). Wanneer literaire teksten een sterkere component zouden krijgen in de mvt-examens, zullen zaken als literatuurgeschiedenis, literaire begrippen (genre, stijl, retoriek) en het (vergelijkend) analyseren van literaire teksten een grotere rol moeten krijgen in de lesprogramma's mvt.

Referenties

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32, 347-364. <https://doi.org/10.1007/BF00138871>

Bloemert, J., & Cats, E. (2020). Literatuur: Geen vak apart. In S. Dönszelmann, C. van Beuningen, A. Kaal & R. de Graaff (Eds.), *Handboek vreemdetalendidactiek: Vertrekpunten vaardigheden vakinhoud* (pp. 401-414). Uitgeverij Coutinho.

Cito. (2024). *Onderzoek naar het opnemen van literaire/fictie teksten in het cse Nederlands voor vmbo-gt*. [Intern document]

Mekenkamp, M. (2024). Poëzie als interculturele ontmoetingsplek: Inspiratie uit de Spaanstalige wereld. *Levende Talen Magazine*, 111(1), 6-10.

SLO. (2007). *Handreiking schoolexamen moderne vreemde talen elementair havo/vwo: Arabisch, Italiaans, Russisch, Spaans, Turks*. <https://www.slo.nl/@4450/handreiking-19/>

SLO. (2024). *Toelichtingsdocument conceptexamenprogramma's moderne vreemde talen*. <https://www.slo.nl/thema/meer/actualisatie-kerndoelen-examenprogramma/actualisatie-examenprogramma/examenprogramma-moderne-vreemde-talen/@24200/toelichtingsdocument-mvt/>

Neem voor meer informatie over de specifieke taken en het onderzoek contact op met:
rosa.bartman@cito.nl