

Inhoudelijk Examineren

ONDERZOEK NAAR DE INHOUDSVALIDITEIT
VAN HET VWO EINDEXAMEN ENGELS

Student: Rob v/h Hof, 2154240
OPLEIDING: MASTER ENGELS @ FONTYS TILBURG (FLOT)
BEGELEIDER: DR. B. ROOSKEN
DATUM: 10-8-2022

Aanpassing

Dit is een aangepaste versie van mijn eindschrijft. School specifieke context is verwijderd, mede doordat dit een los spoor binnen het onderzoek was, dat niet werd samengevoegd met het examen spoor, noch tot een conclusie kwam.

Samenvatting

Tijdens het oefenen voor en toetsen van leesvaardigheid worden Cito examens benut. Ondanks dat dit al jaren gedaan wordt, zijn er twijfels onder collega's of deze examens voldoende de Engelse leesvaardigheid toetsen. Om de inhoudsvaliditeit van een examen of toets te bepalen, wordt er in de wetenschappelijke literatuur aangeraden om een specificatietabel op te stellen, waarin het onderwerp en het niveau van iedere vraag aangegeven wordt. Voor het examen ontbreekt het aan zo'n tabel.

Binnen dit onderzoek is geprobeerd om, achteraf, voor zes vwo-examens een specificatietabel op te stellen. Hiervoor is eerst gekeken naar wat de wetenschappelijke literatuur zegt over inhoudsvaliditeit, welke processen onderdeel zijn van leesvaardigheid, hoe deze getoetst kunnen worden, wat mogelijke onderwerpen zijn en hoe er verschil in niveau aangegeven kan worden. Tijdens het verkennen van de literatuur en het verrichten van de analyse bleek dat het bepalen van het niveau van een tekst uiterst complex, waarbij het ERK onvoldoende houvast biedt, ondanks dat het examen zich hier wel aan meet. Om een andere inschatting te krijgen van het niveau is ook gebruik gemaakt van de Coh-Metrix.

Uiteindelijk is deze informatie gebruikt om per examen een specificatietabel op te stellen en in te vullen. Hierna is ook een algemene specificatietabel voor het vwo-examen opgesteld om patronen en voorkeuren aan te kunnen geven. Hieruit blijkt dat de onderzochte examens een voorkeur hebben voor vragen met betrekking tot woordenschat en gevolgtrekking. Daarnaast wordt woordenschat meer beloond, wat impliceert dat de examens minder kijken naar tekstbegrip en meer naar woordkennis.

Inhoudsopgave

AANPASSING	2
SAMENVATTING	2
1. AANLEIDING	5
2. ONDERZOEKSVRAGEN	6
2.1 HOOFDVRAGEN	6
2.2 LITERATUURVRAGEN	6
3. LITERATUUR VERKENNING	7
3.1 INHOUDSVALIDITEIT	7
3.1.1 Doelen	9
3.1.2 Onderwerpen.....	9
3.1.3 Niveau	10
3.1.4 Vragen	10
3.1.5 Betrouwbaarheid	11
3.1.6 Controle	11
3.2 LEESVAARDIGHEID.....	12
3.2.1 Oefenen	14
3.2.2 Toetsing	15
3.2.3 ERK	15
3.3 EXAMEN	16
3.3.1 Korte historie.....	16
3.3.2 Doelen	17
3.3.3 Onderwerpen.....	17
3.3.4 Niveau	18
3.3.5 Vragen	19
3.4 RESULTAAT LITERATUURVRAGEN	19
4. METHODE VAN ONDERZOEK	21
4.1 BEPALEN VAN MOEILIJKEIDSGRAAD VAN TEKSTEN	21
4.2 BEPALEN VAN ONDERWERPEN	26
4.3 SPECIFICATIETABEL.....	29
5. PRAKTISCHE UITVOERING	30
6. PRAKTISCHE RELEVANTIE	33
7. RESULTATEN EN CONCLUSIES	34
7.1 RESULTATEN.....	34
7.1.1 Spreiding van niveau	41

7.2	<i>Verdeling van onderwerpen</i>	43
7.2	CONCLUSIES	46
7.2.1	<i>Niveau van de teksten</i>	46
7.2.2	<i>Verdeling van onderwerpen</i>	47
7.2.3	<i>Het examen</i>	48
7.2.4	<i>Combinatie tabel</i>	49
7.3	RESULTATEN HOOFDVRAAG	52
8.	DISCUSSIE EN AANBEVELINGEN	53
8.1	KANTTEKENINGEN	53
8.1.1	<i>Bij het examen</i>	53
8.1.2	<i>Bij dit onderzoek</i>	54
8.2	AANBEVELINGEN	55
8.2.1	<i>Voor het CvTE en Cito</i>	55
8.2.2	<i>Voor verder onderzoek</i>	56
	BIBLIOGRAFIE	58
	BIJLAGEN	65
1	VOORBEELDEN VAN VRAAGSTELLINGEN PER SOORT	66

1. Aanleiding

Om de leesvaardigheid in het Engels van de leerlingen te toetsen, wordt op school gebruik gemaakt van oude Cito examenteksten. In de onderbouw door vmbo-bb, -kb en -gt teksten in te zetten en in de bovenbouw havo of vwo-teksten. Zo is er een doorlopende leerlijn richting het eindexamen, waarbij ook tijdens de oefening en voorbereiding gebruikt wordt gemaakt van oude examens.

Nagenoeg na iedere summatieve leesvaardigheidstoets die afgenomen is, inclusief het eindexamen, rijst weer het vermoeden dat deze Cito examens wellicht niet voldoende de beheersing van de Engelse taal toetsen. Collega's zetten vraagtekens bij de validiteit van de Cito leesvaardigheid examens en specifiek bij de inhoudsvaliditeit ervan.

Een belangrijke reden hiervoor is dat ze cijfermatig voldoende te maken lijken te zijn met een beheersing van een beperkt aantal zeer toetsspecifieke strategieën. Hieronder wordt een bepaalde aanpak verstaan die specifiek gericht is op de toetsvorm en vraagstelling van Cito en hierbij niet ingaat op taalbegrip. Deze strategieën, inclusief toegepaste woordenlijsten, worden aangeboden in examentrainingen gegeven door externe organisaties, en in sommige gevallen zelfs geadviseerd aan leerlingen die dermate moeite hebben met de voorbereidende (oefen)toetsen, dat succes bij het eindexamen verre van gegarandeerd is.

Het is soms ook verbazingwekkend hoe een leerling, die tijdens de lessen en bij de schoolexamens al een goede grip op en beheersing van de taal heeft aangetoond, lager scoort op het examen dat te verwachten zou zijn op basis van die eerder aangetoonde vaardigheid, en vice versa. Kan het zijn dat taalvaardige leerlingen geen goede lezers zijn? Is er sprake van geluk of willekeur?

Tevens kan men zich afvragen of een voldoende ook echt een juiste niveau indicatie geeft, onder andere omdat Cito de normering (n-term) laat afhangen van het daadwerkelijke resultaat (een relatieve normering), waarbij gemiddeld goed gemaakte examens strenger beoordeeld worden dan gemiddeld slecht gemaakte. Anders gezegd, er wordt een correctie doorgevoerd op de moeilijkheid van het examen op basis van het resultaat, in plaats van dat het niveau vooraf vastgesteld wordt en hiermee de norm bepaald.

2. Onderzoeksvragen

2.1 Hoofdvragen

1. Toetst het examen vwo Engels met voldoende inhoudsvaliditeit de leesvaardigheid van leerlingen?

2.2 Literatuurvragen

1. Wat wordt er verstaan onder leesvaardigheid?
2. Wat wordt er verstaan onder inhoudsvaliditeit?
3. Hoe kan inhoudsvaliditeit toegepast worden op leesvaardigheid?
4. Hoe wordt de inhoudsvaliditeit van het examen vwo Engels bepaald?
 - 4.1 Binnen welke kaders moet het examen vwo Engels samengesteld worden?

3. Literatuur verkenning

3.1 Inhoudsvaliditeit

Iedere toets moet voldoen aan een aantal criteria. Maassen, den Otter, Wools, Hemker, Straetmans & Eggen (2014) hebben op basis van een reviewonderzoek en expertpanel vijf hoofdcategorieën geïdentificeerd:

1. Betrouwbaarheid
2. Generaliseerbaarheid
3. Validiteit
4. Gebruik toetsresultaat
5. Randvoorwaarden

Ze merken op dat *validiteit* vooral in theoretische artikelen genoemd werd, maar amper in empirisch onderzoek voorkwam. Dit terwijl wanneer het gaat om de kwaliteit van toetsen, validiteit en betrouwbaarheid vaak in één adem genoemd worden (van Berkel & Draaijer, 2011; Maassen & den Otter, 2015; Sanders & Brouwer, 2015; van Berkel, Bax, & Joosten-ten Brinke, 2017; Moelands, 2017; SLO, 2020). De twee concepten zijn zelfs dermate verweven met elkaar dat van Berkel et al. (2017) stellen dat wanneer een toets onbetrouwbaar is, deze onmogelijk inhoudsvalide kan zijn. Moelands (2017) merkt op dat bij toetsen in onderwijssituaties er een belangrijke rol wordt toegedeeld aan inhoudsvaliditeit.

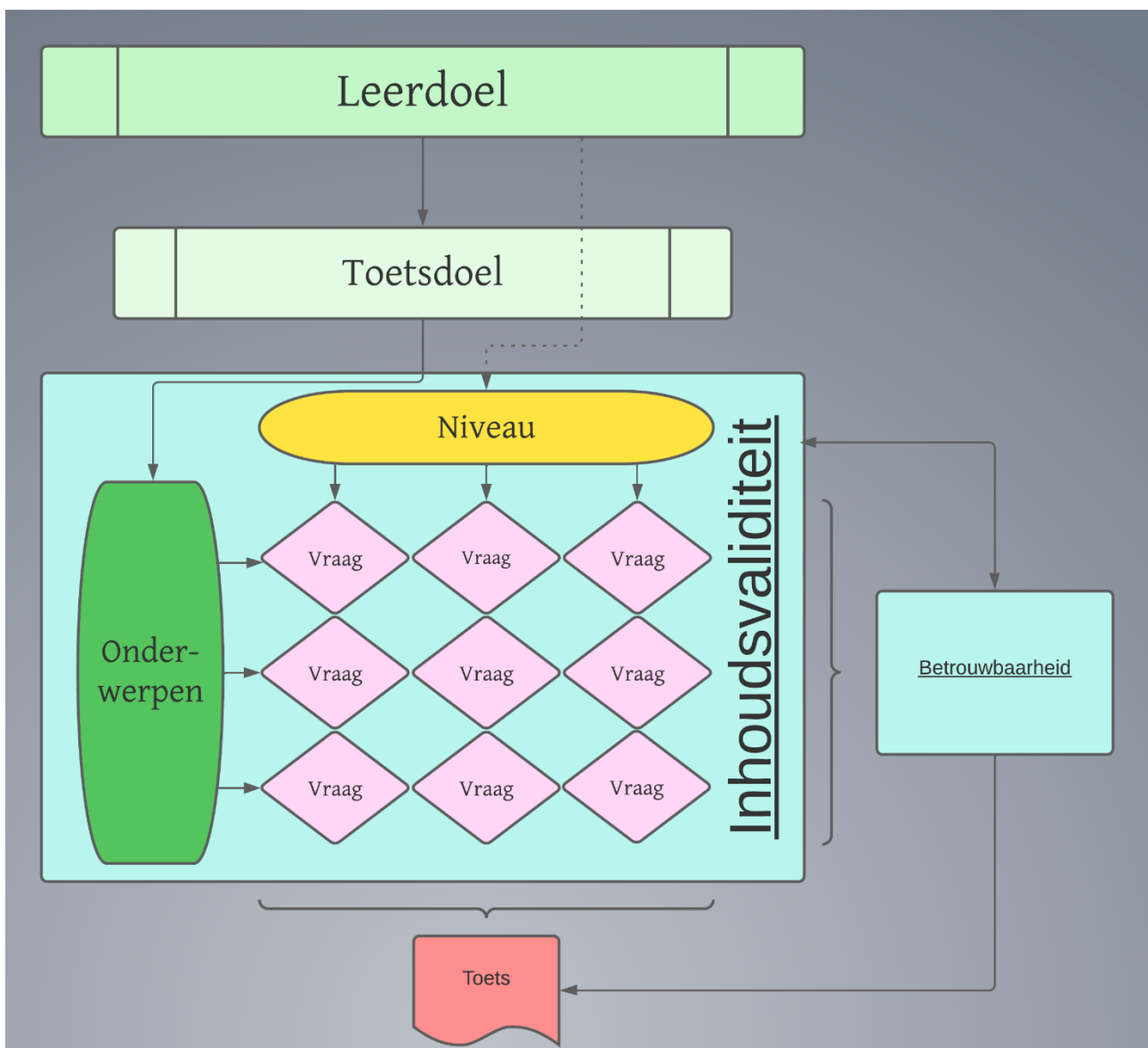
Inhoudsvaliditeit, of representativiteit, is de eigenschap dat de opgaven van een toets vertegenwoordigend zijn voor de kennis of vaardigheid die getoetst wordt op basis van de doelen uit het leergebied of studieonderdeel. Het is een bepaling over de relatie tussen de toetsvragen en de onderwijsdoelstellingen. Bij voldoende inhoudsvaliditeit zegt de uitslag van de toets (het resultaat van de kandidaat) ook iets over zijn of haar beheersing van de leerstof (Maassen, et al., 2014; van Berkel, Bax, & Joosten-ten Brinke, 2017).

In de meeste gevallen is het vooraf vaststellen van de inhoudsvaliditeit door een *specificatietabel* of *toetsmatrijs* op te stellen één van de eerste stappen in het gehele proces van toetsconstructie (van Berkel & Draaijer, 2011; van Berkel, Bax, & Joosten-ten Brinke, 2017). Het doel hiervan is om ervoor te zorgen dat er een goede spreiding is over de te toetsen onderwerpen en dit op verschillende niveaus gebeurt.

Voor het opstellen van de specificatietabel zijn er twee zaken nodig: onderwerpen en verschillende niveaus. De onderwerpen van de toets vloeien voort uit het toetsdoel dat op zijn beurt uit het overkoepelende leerdoel komt. (Maassen & den Otter, 2015; van Berkel, Bax, & Joosten-ten Brinke, 2017).

Niveauverschil wordt meestal aangebracht door het gebruik van een bepaalde taxonomie, zoals die van Bloom (1956), Romiszowski (1981) of het RTTI-model van Drost en Verra (2012) Hiermee wordt ook enigszins de moeilijkheid van de toets vooraf bepaald (Maassen & den Otter, 2015).

In figuur 1, hieronder, is schematisch weergegeven hoe doelen, onderwerpen, niveaus en vragen elkaar, inhoudsvaliditeit en betrouwbaarheid beïnvloeden. Daarna zal op ieder van deze facetten (exclusief inhoudsvaliditeit) dieper ingegaan worden.



Figuur 1: Schematische weergave hoe doelen, onderwerpen, niveaus en vragen van invloed zijn op inhoudsvaliditeit en betrouwbaarheid.

3.1.1 Doelen

Grote, globale leerdoelen worden veelal bepaald door het beleid van de school (Sanders, 2013) of door de overheid in de vorm van kerndoelen (Kerndoelen onderbouw voortgezet onderwijs, 2010) en eindtermen (Staatscourant van het Koninkrijk der Nederlanden, 2007).

De eerste stap in het samenstellen van een toets, is het exactere doel ervan te bepalen binnen de doorlopende leerlijn (Sanders, 2013; Maassen & den Otter, 2015; Moelands, 2017; van Berkel, Bax, & Joosten-ten Brinke, 2017). Sanders (2013) en Moelands (2017) schetsen beide een grove driedeling in toetsdoelen, of functies, namelijk ter beoordeling van het onderwijssysteem, ter beoordeling van de leerling en ter beoordeling van het onderwijsleerproces.

Maassen & den Otter (2015) hanteren alleen een verdeling tussen summatief (ter beoordeling van de leerling) en formatief (ter beoordeling van het onderwijsleerproces) als toetsdoel, met eronder enkele andere functies, zoals bij summatieve toetsen, classificeren of certificeren en, bij formatieve toetsen, het beoordelen van het leerproces of om van te leren (Sanders, 2013; Maassen & den Otter, 2015; Kuhlemeier & van Til, 2016).

Vooraf bij summatieve toetsing wordt er veel gesproken over de validiteit van de toets (Maassen, et al., 2014).

3.1.2 Onderwerpen

Vanuit de doelen kunnen geconcretiseerde onderwerpen opgesteld worden (Maassen & den Otter, 2015; van Berkel, Bax, & Joosten-ten Brinke, 2017). Deze onderwerpen vormen de eerste stap in het samenstellen van een specificatietabel, waarna per onderwerp besloten kan worden op welk niveau er een vraag over gesteld gaat worden (van Berkel & Draaijer, 2011; Maassen & den Otter, 2015; van Berkel, Bax, & Joosten-ten Brinke, 2017).

Om de leerdoelen zo concreet mogelijk te kunnen formuleren, moet er aandacht zijn voor welk gedrag de leerling moet tonen, ofwel welke handeling deze moet verrichten. Door werkwoorden te kiezen die observeerbaar zijn (Maassen & den Otter, 2015), of gemeten kunnen worden (Moelands, 2017) is het ook voor de leerling duidelijk wat deze moet doen. Te specifiek moeten de werkwoorden echter ook niet worden (Sanders, 2013). Dit

geeft de leerling nog enige keuzevrijheid in de manier waarop het doel aangetoond wordt.

De gekozen werkwoorden moeten in goede relatie staan tot de inhoud van de toets (Sanders, 2013; Moelands, 2017), welke op zijn beurt weer zo concreet mogelijk omschreven moet worden (Maassen & den Otter, 2015). Daarna kunnen nog specifieke voorwaarde gesteld worden, zoals het gebruik van hulpmiddelen of het aantal te geven voorbeelden.

3.1.3 Niveau

Aandacht voor het niveau van de vragen is vooral om aandacht te kunnen schenken aan de moeilijkheid van de toets (Sanders, 2013; Maassen & den Otter, 2015). Het zorgt ook voor variatie binnen de toets. Daarnaast kan, door te kijken naar het resultaat van de leerling, niveauverschil helpen bij verder classificeren of preciezere feedback geven.

Om onderscheid in niveau aan te brengen, wordt veel gebruik gemaakt van de verdeling (kennis)reproductie – toepassing – inzicht (van Berkel & Draaijer, 2011; Maassen & den Otter, 2015; van Berkel, Bax, & Joosten-ten Brinke, 2017), gebaseerd op Drost en Verra (2012), maar ook de taxonomie van Bloom, in geüpdatete versie (Krathwohl, 2002; Sanders, 2013). Welke keuze gemaakt wordt, zal ook afhangen van de, vakspecifieke, leerdoelen (Knevel, 2013).

3.1.4 Vragen

Voorafgaand aan het opstellen van vragen moet er gekozen worden welke vorm de toets gaat aannemen. Open of gesloten vragen (Moelands, 2017; van Berkel, Bax, & Joosten-ten Brinke, 2017) zijn de meest voorkomende en “vertrouwde”, maar andere vormen, zoals een praktisch opdracht of (stage)verslag, zijn ook mogelijk (Moelands, 2017).

Bij iedere vraag(soort) moet er rekening gehouden worden met de tijd die het gemiddeld kost om deze te beantwoorden (van Berkel & Draaijer, 2011). Het stellen van meer vragen draagt veelal bij aan de betrouwbaarheid van de toets (van Berkel, Bax, & Joosten-ten Brinke, 2017), maar kost natuurlijk ook meer tijd om te maken en beoordelen.

Naast een heldere formulering van de vraag die aansluit bij de onderwerpen van de toets, zijn er per vraagsoort nog andere richtlijnen om de kwaliteit te waarborgen, zoals per vraag maar één leerdoel, geloofwaardige afleiders bij meerkeuze vragen, geen groot

verschil in woordgebruik of lengte, geen overbodige informatie, geen gebruik van extremen (zoals nooit, altijd) of vage omschrijvingen (soms, vaak), geen (onnodige) ontkenning, etc., etc. (Moelands, 2017; van Berkel, Bax, & Joosten-ten Brinke, 2017).

3.1.5 Betrouwbaarheid

Zoals eerder aangegeven, gaan inhoudsvaliditeit en betrouwbaarheid hand in hand. Onder betrouwbaarheid wordt doorgaans verstaan “de mate waarin de scores consistent, nauwkeurig en reproduceerbaar zijn” (Maassen, et al., 2014, p. 36).

Door meer vragen te stellen, wordt de betrouwbaarheid verhoogt (Sijtsma, 2011; van Berkel, Bax, & Joosten-ten Brinke, 2017), ervan uitgaande dat de vragen een goede spreiding over de onderwerpen van de toets en verdeling over niveaus hebben. Het voordeel van een langere toets is ook dat de standaard meetfout zal afnemen, wat de betrouwbaarheid verhoogt (Sijtsma, 2011).

3.1.6 Controle

Vrijwel alle bovengenoemde onderwerpen komen terug in het *COTAN beoordelingssysteem voor de kwaliteit van tests* (Evers, Lucassen, Meijer, & Sijtsma, 2010), wat kan helpen om te controleren of aan iedere stap in het proces is voldaan. Als alternatief is de gids door van Berkel en Draaijer (2011) of de brochure van Maassen en den Otter (2015) goed bruikbaar, zij het minder uitgebreid, of de checklist van Moelands (2017) of SLO (2020).

In hun boek *Toetsen in het hoger onderwijs* (2017) geven van Berkel, Bax, & Joosten-ten Brinke korte checklists per onderdeel.

Alle documenten zijn er ter ondersteuning van de toetsmaker opdat deze kwalitatief goede toetsen kan produceren.

3.2 Leesvaardigheid

Het OECD definieert *reading literacy* als “[het] begrijpen van, gebruiken van, evalueren van, reflecteren op en zich bezighouden met teksten teneinde eigen doelen te bereiken, het ontwikkelen van de eigen kennis en potentieel en om deel te kunnen nemen aan de maatschappij¹” (OECD, 2019, p. 28). Zij willen hiermee duidelijk maken dat lezen, in hun ogen, meer is dan alleen het begrijpen van dat wat je leest, maar ook wat je ermee doet. Ze benoemen dat de handeling van lezen bestaat uit het gebruik maken van een veelvoud aan cognitieve en taalkundige processen, waaronder decoderen, woordkennis, grammatica, kennis over tekststructuren en ook metacognitieve vaardigheden op het gebied van leesstrategieën.

Veel onderzoek is erop gericht om deze verschillende vaardigheden zo goed mogelijk te identificeren en hun bijdragen aan leesvaardigheid te begrijpen en kwantificeren, zie bijv. (Perfetti, 1999; Scarborough, 2001; Cromley & Azevedo, 2007; Ahmed, et al., 2016). Dit gaat allemaal verder in detail dan het eenvoudige beeld van leesvaardigheid door Gough en Tunmer (1986) wat slechts decoderen en begrip onderscheidt.

Veel terugkerende componenten zijn:

- Woordenschat
- Syntaxis en semantiek
- Leesstrategieën
- Kennis van tekststructuur
- Voorkennis

Met deze selectie wordt voorbijgegaan aan vaardigheden zoals fonologische kennis, decoderen, vloeiend lezen en motivatie. Deze, en andere, vaardigheden spelen wel een rol in leesvaardigheid, maar, met uitzondering van motivatie, vooral bij jonge kinderen (Perfetti, 1999; Scarborough, 2001) en minder bij adolescenten.

Hieronder volgt een uiterst beknopte uitleg van de bovengenoemde componenten, zoals ze ook verder in het onderzoek gebruikt zullen worden.

¹ Vertaling door auteur.

Oorspronkelijke tekst: “*Reading literacy is understanding, using, evaluating, reflecting on and engaging with texts in order to achieve one’s goals, to develop one’s knowledge and potential and to participate in society.*”

Woordenschat

“Woordenschat wordt gedefinieerd als de kennis van woordbetekenissen van leerlingen” (van Gelderen, 2018, p. 7). Wanneer een leerling niet weet wat een woord betekent, is het komen tot een dieper begrip van de tekst lastiger. Onderzoek van o.a. Schmitt, Jiang & Grabe (2011) en Trapman, van Gelderen, van Schooten & Hulstijn (2017) impliceert dat zo’n 95% tot 98% van de woorden bekend moet zijn om de tekst goed te kunnen doorgronden.

De relatie tussen woordenschat en tekstbegrip is wederzijds, wat wil zeggen dat een ruime woordenschat bijdraagt aan beter tekstbegrip, goed tekstbegrip kan leiden tot het leren van nieuwe woorden, een beperkte woordenschat tekstbegrip hindert en gelimiteerd tekstbegrip ervoor zorgt dat er minder nieuwe woorden geleerd worden uit de betreffende tekst (Oakhill, Cain, & Carsten, 2019).

Syntaxis en semantiek

Woorden kunnen een andere betekenis krijgen wanneer deze is een andere zin of context gebruikt worden. Bijvoorbeeld een woord als *bank*² (zitmeubel of financiële instelling) of *kantelen* (meervoud van kanteel of werkwoord). Lezers kunnen nieuwe betekenissen van woorden leren door ze in context te lezen (Ricketts, Bishop, Pimperton, & Nation, 2011). Er is ook een duidelijk verband tussen semantische kennis en tekstbegrip, waarbij de vaardigheid om woordbetekenis op te maken vanuit de context voorspellend is voor tekstbegrip in het algemeen (Mimeau, Ricketts, & Deacon, 2018).

Leesstrategieën

Verschillende leesstrategieën dragen bij aan het begrijpen van teksten. Een van de meeste complete en doorontwikkelde modellen hiervan is het *Construction-Integration* (CI) model van Kintsch (1998). Binnen het constructie stuk van de term wordt er een zogeheten *situatie model* gemaakt van de gelezen tekst ter bevordering van het begrijpen ervan. Met het integratie gedeelte wordt al bestaande kennis toegepast op wat er gelezen is. Vervolgens is er een continue wisselwerking tussen alle facetten waardoor de lezer nieuwe kennis opdoet en de tekst (steeds) beter begrijpt.

Andere veelvoorkomende strategieën zijn o.a. (Duke, Pearson, Strachan, & Billman, 2011):

² Voorbeeld geïnspireerd door Perfetti (1999)

- Leesdoel stellen
- Voorspellen
- Monitoren van begrip
- Gevolgtrekken
- Visualiseren

Kennis van tekststructuur

Teksten zijn doorgaans hiërarchisch opgebouwd (woord – zin – paragraaf – tekst) met wederkerende patronen, zoals oorzaak-gevolg of vergelijking-verschil, die herkenbaar kunnen zijn door middel van het gebruik van bijv. tussenkoppen of andere lexicale elementen. Het juist identificeren van deze structuren helpt bij het betere construeren van het mentale situatie model en daarmee tekstbegrip (Bogaerds-Hazenberg, Evers-Vermeul, & van den Bergh, 2020).

Voorkennis

Het CI-model (Kintsch, 1998) omschrijft o.a. hoe voorkennis over het onderwerp van de tekst benut wordt om de tekst als geheel beter te begrijpen. Onderzoek toont aan dat beperkte voorkennis een hinder vormt voor het begrijpen van de tekst en hoe juist het hebben van voorkennis een betere voorspeller is voor tekstbegrip dat “goed” of “slecht” kunnen lezen (Willingham, 2009). Vanaf een drempelwaarde van 33,5% neemt begrip sterker toe (O'Reilly, Wang, & Sabatini, 2019), wat wil zeggen dat er minimaal 33,5% aan onderwerp specifieke woorden goed begrepen moeten worden voorafgaand aan het lezen van de tekst, om de tekst goed te begrijpen. Meer woordkennis, in brede zin, vooraf zorgt voor meer begrip.

3.2.1 Oefenen

Leesvaardigheid is, zoals het woord al aangeeft, een vaardigheid. Stokking (2001) omschrijft een vaardigheid als “[het] vermogen tot het leveren van een bepaalde prestatie of het uitvoeren van een bepaalde taak” (p. 16). Hierbij wordt aangenomen dat de prestatie stabiel is, ondanks wisselende omstandigheden. Om voldoende beheersing te toetsen, moet ook de toevallige juist of onjuist in acht genomen worden. Zeker wanneer de directe vaardigheid niet zichtbaar is, maar alleen af te leiden valt uit het resultaat of een product, zoals bij leesvaardigheid het geval is (Stokking, 2001; Ahmed, et al., 2016).

Stanley en Williamson (2017) beargumenteren dat een vaardigheid niet uitsluitend een opsomming is van (feiten) kennis, maar ook iets dat gradaties kent die door oefening verbeterd kunnen worden. Door te begrijpen op welk onderdeel van leesvaardigheid er een misverstand in het begrijpen van de tekst ontstaat, kan er een gerichte oplossing gezocht worden (Perfetti, 1999; Cromley & Azevedo, 2007; Mimeau, Ricketts, & Deacon, 2018; Bogaerds-Hazenberg, Evers-Vermeul, & van den Bergh, 2020).

Kortom, er kan gesproken worden over verschillende beheersingsniveaus, welke waarschijnlijk aan te tonen zijn door onderliggende vaardigheden of kennis aan te tonen via het maken van een product. Dit impliceert dat het mogelijk is om over inhoudsvaliditeit te spreken bij leesvaardigheidstoetsen.

3.2.2 Toetsing

Het goed toetsen van leesvaardigheid blijkt lastig (Kamalski, Sanders, Lentz, & van den Bergh, 2005), complex (Fletcher, 2006) en heeft vaak een nauwe focus op een deelvaardigheid (Afflerbach, Cho, Crassas, & Kim, 2018). Daarnaast is het interpreteren van hoe goed iemand een tekst begrijpt op basis van toetsing afhankelijk van de exacte vraagstelling van de toets (Kamalski, Sanders, Lentz, & van den Bergh, 2005; Cutting & Scarborough, 2006).

Om leesvaardigheid in brede zin, enigszins in overeenstemming met de definitie van het OECD, te toetsen, zou er meer formatief, procesgericht en tijdens de les gemonitord moeten worden (Afflerbach, Cho, Crassas, & Kim, 2018; Houtveen, van Steensel, & de la Rie, 2019).

3.2.3 ERK

Het ERK heeft voor algemene leesvaardigheid een verdeling over verschillende taalbeheersingsniveaus, oplopend van pre-A1 tot en met C2. De omschrijving van de vaardigheid die de gebruiker zou moeten beheersen is vooral gericht op de lengte en complexiteit van de tekst en het taalgebruik erbinnen (North, Goodier, & Piccardo, 2018), al worden er geen gedetailleerde voorbeelden gegeven (Natova, 2019). Hierna worden een vijftal specifiekere “doelen” benoemd:

- Correspondentie lezen
- Oriënterend lezen
- Lezen ter informatie en argumentatie

- Instructies lezen
- Als vrijetijdsbesteding

Voor zowel de algemene omschrijven als onder ieder van de vijf “doelen” zijn er specifiekere ‘*can do*’ stellingen opgesteld. Deze kleinere doelen zijn omschreven op zo’n manier dat ze duidelijker aantoonbaar zouden moeten zijn. Echter ontbreekt het aan specifieke manieren waarop het toetsbaar zou zijn. Daarnaast ontbreekt het bij de hogere niveaus (C1/C2) geregeld aan descriptoren.

De *Syllabus Centraal Examen* (College voor Toetsen en Examens, 2019) neemt deze globale descriptoren, vertaald, vrijwel direct over in de toetsspecificatie van het eindexamen.

3.3 Examen

Het eindexamen als geheel krijgt de nodige kritiek. Zo is het niet duidelijk of het examen wel voldoende betrouwbaar is en dit is afhankelijk van de methode waarop dit bepaald wordt (Kuhlemeier & van Til, 2016), wordt de kwaliteit niet onafhankelijk beoordeeld (Scheerens, Brouwer, Sanders, Veldkamp, & van der Vegt, 2019), geeft het correctiemodel geen ruimte voor vakinhoudelijke interpretatie (van Haperen, 2022), zijn de eisen gebaseerd op een systeem dat niet onomstreden is (Kwakernaak & de Vries, 2006; Kwakernaak, 2010; Kwakernaak, 2012), en ten aanzien van het vwo-examen Engels heeft Tillema-Kortman (2005) aangetoond dat deze niet louter de beheersing van de Engelse taal toetst, maar dat ook niet talige factoren een rol spelen.

3.3.1 Korte historie

Sinds 1845 kent Nederland een vorm van het Centraal Examen (CE) dat momenteel verplicht³ is om een diploma te behalen in het voortgezet onderwijs. Het College voor Toetsen en Examen (CvTE) is verantwoordelijk voor de inhoud van de verschillende examens, stelt deze jaarlijks (opnieuw) vast en beschrijft deze in een syllabus. Sinds ruwweg 1970 is het Centraal Instituut voor Toetsontwikkeling (Cito) betrokken bij de productie en afname van de centrale eindexamens (Kuhlemeier & van Til, 2016; Scheerens, Brouwer, Sanders, Veldkamp, & van der Vegt, 2019). Cito stelt tegenwoordig alle eindexamens samen, nog altijd onder de verantwoordelijkheid van het CvTE. In 2007 zijn er eindtermen bepaald (Staatscourant van het Koninkrijk der Nederlanden,

³ Enkele zeer specifieke omstandigheden nagelaten, zoals in schooljaar 2019-2020 ten tijde van de coronapandemie.

2007), waar tot 2021 mee is gewerkt. Op dit moment is SLO bezig met een actualisering van de kerndoelen en examenprogramma's (Teunis, B., 2022).

Ten aanzien van de inhoud van een examen anno 2021 kan het volgende gesteld nog worden:

- De eindtermen liggen vast.
- CvTE stelt een inhoudelijke syllabus op.
- Cito ontwerpt een examen dat toetst of de eindtermen inhoudelijk behaald zijn.

3.3.2 Doelen

Het centraal examen heeft een afsluitingsfunctie (Kuhlemeier & van Til, 2016), ook wel certificering genoemd (Moelands, 2017). Hiermee zou moeten worden bepaald “of een leerling een onderwijstraject met succes doorlopen heeft” (Moelands, 2017, p. 6). Het centraal examen is geen onderdeel meer van het leerproces en verschaft het geen informatie die door docenten gebruikt kan worden om hun lessen aan de desbetreffende groep invulling te geven.

De *syllabus centraal examen 2021 MVT* stelt dat “het centraal examen betrekking heeft op domein A, leesvaardigheid” (College voor Toetsen en Examens, 2019, p. 9). De globale eindterm is als volgt omschreven:

Domein A: Leesvaardigheid

1. De kandidaat kan:

- aangeven welke informatie relevant is, gegeven een vaststaande behoefte;
- de hoofdgedachte van een tekst(gedeelte) aangeven;
- de betekenis van belangrijke elementen van een tekst aangeven;
- relaties tussen delen van een tekst aangeven;
- conclusies trekken met betrekking tot intenties, opvattingen en gevoelens van de auteur.

Afbeelding 1: Eindterm moderne vreemde talen, havo en vwo.

Het betreft hier vijf algemene doelen, waarbij concrete manieren van het aantonen ervan bewust nog niet benoemd zijn. Het doel van het eindexamen mag aangenomen worden te zijn het vaststellen van de mate waarin de kandidaat de vijf omschreven subdoelen kan aantonen.

3.3.3 Onderwerpen

Om tot specifiekere onderwerpen te komen vanuit het globale doel, volgt het CvTE in hun syllabus (College voor Toetsen en Examens, 2019) de bewoording van het ERK,

zoals vertaald door Fasoglio, de Jong, Trimbos, & Beeker (2015). Ze hebben zowel de globale beheersingsniveaus overgenomen, als specifiekere ‘can do’-descriptoren.

De syllabus benadrukt dat de teksten overwegend betrekking hebben op de schalen ‘oriënterend lezen’ en ‘lezen ter informatie en argumentatie’ (College voor Toetsen en Examens, 2019, p. 16).

Het door Kwakernaak (Kwakernaak & de Vries, 2006; Kwakernaak, 2012) beschreven probleem van vermenging van leesdoelen met specifieke teksten en hiermee schalen pogen in te delen, is ook overgenomen. In het koppelingsverslag tussen de toenmalige examens en het ERK, erkende de auteurs dit probleem al (Noijons & Kuijper, 2006).

3.3.4 Niveau

“Ondanks het feit dat de eindterm leesvaardigheid voor de havo- en vwo-examens gelijkluidend is, betekent dit niet dat de examens dezelfde moeilijkheidsgraad hebben. De specificaties per examen uitgaande van ERK-niveaus [laten] zien dat dit ook niet het geval is” (College voor Toetsen en Examens, 2019, p. 9).

Volgens het eigen onderzoek (Fasoglio, de Jong, Trimbos, & Beeker, 2015) ligt het niveau van het vwo-examen rond C1. Een voldoende voor het examen toont zeker B2 aan, maar er is een ruime voldoende nodig om C1 aan te tonen (College voor Toetsen en Examens, 2019). Dit is een herijking van het eerdere koppelingsonderzoek van Noijons en Kuijper (2006), waar de grens lag op B1/B2 niveau, wat ook te maken had met de beperkingen van dat onderzoek (o.a. slechts één examen geanalyseerd).

In de constructieopdracht van 2014 (College voor Toetsen en Examens, 2014) wordt nog verwezen naar het koppelingsonderzoek van Noijons en Kuijper, om een gemiddelde niveauscore⁴ voor de opdrachten aan te duiden, wat vervolgens gebruikt wordt om een indicatie van het niveaubereik aan te geven (80% B2 opgaven, 20% C1).

Het exact bepalen van het ERK-niveau lijkt giswerk te blijven en sterk afhankelijk van aannames van de persoon die de inschatting maakt (Kwakernaak, 2010; Verheyen, van Til, Keuning, & Feskens, 2015), mede omdat gedetailleerde omschrijvingen van waar een tekst op het niveau aan zou moeten voldoen ontbreken (Natova, 2019).

⁴ 4,3, waarbij 4 = B2 en 5 = C1. De opgaven behoren tussen B2 en C1 in te liggen.

3.3.5 Vragen

De constructieopdracht (College voor Toetsen en Examens, 2014) geeft zeer specifiek aan welke soort vragen in het examen mogen staan en in welke verhoudingen. Zo dient 60% van de scorepunten te behalen te zijn door het correct beantwoorden van ‘traditionele’ meerkeuze vragen (met 3 tot 6 alternatieven) en meerkeuze invul (*cloze*) vragen.

De overige 40% van de scorepunten mogen verdeeld worden over andere vraagsoorten, zoals combinatievragen, ordeningsvragen, open vragen, of citeeropdrachten, waarbij er een evenwichtige verdeling wordt nagestreefd.

3.4 Resultaat literatuurvragen

1. Wat wordt er verstaan onder leesvaardigheid?

Deze vraag kent geen kort, eenvoudig of eenduidig antwoord. Het is mogelijk om er een definitie aan te geven zoals het OECD (2019) heeft gedaan, of om het schematisch weer te geven zoals het SLO (van Gelderen, 2018) heeft gedaan met figuur 2.



Figuur 2: Wiel van leesvaardigheid en deeltaarigheden, van Gelderen 2018

Het ERK (North, Goodier, & Piccardo, 2018) heeft een selectie gemaakt van taakgerichte doelen en vaardigheden. De domein A omschrijving van het CvTE (2019) is in tegen stelling tot de ERK selectie juist breed omschreven.

Vaak wordt er gesproken over het begrijpen van teksten, maar hoe dit meetbaar gemaakt moet worden, of waar het van afhankelijk is, verschilt per auteur.

2. Wat wordt er verstaan onder inhoudsvaliditeit?

Uit [3.1 inhoudsvaliditeit](#), p 8:

Inhoudsvaliditeit, of representativiteit, is de eigenschap dat de opgaven van een toets vertegenwoordigend zijn voor de kennis of vaardigheid die getoetst wordt op basis van de doelen uit het leergebied of studieonderdeel. Het is een bepaling over de relatie tussen de toetsvragen en de onderwijsdoelstellingen. Bij voldoende inhoudsvaliditeit zegt de uitslag van de toets (het resultaat van de kandidaat) ook iets over zijn of haar beheersing van de leerstof (Maassen, et al., 2014; van Berkel, Bax, & Joosten-ten Brinke, 2017).

3. Hoe kan inhoudsvaliditeit toegepast worden op leesvaardigheid?

Zoals in [3.2.2 toetsing](#) besproken is dit lastig. Doordat er keuzes gemaakt worden in de toetsconstructie, vaak om zich te richten op bepaalde deelvaardigheden, is het niet altijd even makkelijk om algemeen begrip van een tekst te toetsen, vooral niet in een examen setting.

4. Hoe wordt de inhoudsvaliditeit van het examen vwo Engels bepaald?

Deze wordt niet bepaald. Cito geeft wel een toets en item analyse, maar dit is geen specificatietabel en komt pas achteraf.

4.1 Binnen welke kaders moet het examen vwo Engels samengesteld worden?

De constructie opdracht van het CvTE (2014) stelt strenge kaders met betrekking tot vraagsoorten en de verdeling hiervan. Daarnaast zijn er richtlijnen ten aanzien van de onderwerpen die aan bod mogen komen, gericht op maatschappelijke taboes en niveau specifieke thema's.

4. Methode van onderzoek

Om de inhoudsvaliditeit van het vwo-examen Engels te bepalen zal van zes recente vwo-examens een specificatietabel opgesteld worden. De volgende examens zullen worden geanalyseerd: 2018 tijdvak 1 & 2, 2019 tijdvak 1 & 2, 2021 tijdvak 1 & 2. De reden om zes examens te benutten, is om te zien of er patronen te identificeren zijn. Zo kan er een betere vergelijking gemaakt worden tussen wat er in de voorbereiding in brede zin nodig is voor het examen, in tegen stelling tot wat er voor één specifiek examen nodig zou zijn geweest. De verwachting is om op deze manier ook niet tegen dezelfde beperking aan te lopen als Noijons en Kuijper (2006).

Er is voor de tijdvak 1 examens gekozen omdat Cito hiervan een toets- en itemanalyse beschikbaar stelt. De tijdvak 2 examens zorgen voor meer data met de verwachting dat hiermee uitschieters beter te verklaren zullen zijn (regressie naar het gemiddelde, (Galton, 1886)).

Voor het opstellen van de specificatietabellen, moeten het niveau en de onderwerpen van de examens bepaald worden (zoals uitgelegd in [3.1 Inhoudsvaliditeit](#)), zodat daarna de vragen ingedeeld kunnen worden.

4.1 Bepalen van moeilijkheidsgraad van teksten

Het objectief bepalen van de moeilijkheidsgraad van een willekeurige tekst is niet eenvoudig (Benjamin, 2012; Hacquebord & Lenting-Haan, 2012; Nelson, Perfetti, Liben, & Liben, 2012; Velleman & van der Geest, 2014; Kauchak, Leroy, & Hogue, 2017; Crossley, Skalicky, & Dascalu, 2019; Natova, 2019). Verschillende formules pogen (lexicale) factoren zoals (gemiddelde) woordlengte, (gemiddelde) zinslengte, tekstlengte, (gemiddeld) aantal lettergrepen, dekkingen/graad/frequentie woorden, grammaticale constructies, woordsoorten, verbindingswoorden en verwijzingen te benutten om de moeilijkheid van de tekst te voorspellen. Sommige voorspellingen komen goed overeen met de inschattingen van experts (Hacquebord & Lenting-Haan, 2012; Nelson, Perfetti, Liben, & Liben, 2012) of van lezers van de tekst (Nelson, Perfetti, Liben, & Liben, 2012; Kauchak, Leroy, & Hogue, 2017; Crossley, Skalicky, & Dascalu, 2019), of met de resultaten van begripsvragen door lezers bij geijkte toetsen (Nelson, Perfetti, Liben, & Liben, 2012).

Er zijn nog geen wetenschappelijk onderzochte formules die een goede inschatting maken van het ERK-niveau van een tekst. Dit is ook lastig daar het ERK geen concrete eisen heeft opgesteld per niveau waar een tekst aan moet voldoen (North, Goodier, & Piccardo, 2018; Natova, 2019). Natova (2019) heeft geprobeerd om meer concrete eisen te formuleren, en ook een vergelijking tussen de Flesch Reading Ease score (Flesch, 1949) en het ERK getoetst, zoals eerder gedaan door Waller, Horak en Tyrell (2016) en de website Linguapress⁵.

Er zijn wel tools in ontwikkeling en een aantal websites online die aanknopingspunten geven voor het ERK-niveau van een tekst. Veelal wordt er gebruik gemaakt van vergelijkingen met bestaande woordlijsten per ERK-niveau, eventueel aangevuld met algoritmes die woorden die niet op deze lijsten voorkomen, pogen te koppelen aan woorden die er wel op staan (Cathoven, 2022). Daarbovenop kan er gebruik worden gemaakt van herkenning van bijzinnen of grammaticale structuren die gekoppeld zijn aan ERK-niveaus (Cambridge University Press, 2015; Cathoven, 2022). De beperking blijft wel dat het vooral om lexicale factoren gaat en niks zegt over welke mate van achtergrondkennis of gevolgtrekking nodig is om de tekst werkelijk te doorgronden.

Vorderingen in het gebruik van *natural language processing* (NLP) zijn gericht op het beter analyseren van de samenhang binnen teksten om zo een inschatting van complexiteit te kunnen maken die beter overeenkomt met zowel de inschatting van lezers als het getoetst begrip (McNamara & Graesser, 2012; Crossley, Skalicky, & Dascalu, 2019). Voor L2-lezers komen deze voorspellingen ook beter overeen met de moeite die ze ogenschijnlijk lijken te hebben, al is geen enkele formule perfect voorspellend (Nahatame, 2021).

Dit onderzoek maakt gebruik van een combinatie van drie methoden om een inschatting te krijgen van het ERK-niveau van de teksten uit de examens.

1. De Coh-Metrix
2. De Flesch Reading Ease Score
3. Websites die ERK-niveau schatten

⁵ <https://linguapress.com/teachers/flesch-kincaid.htm>

Coh-Metrix

Voor de Coh-Metrix (McNamara & Graesser, 2012) is gekozen omdat deze een makkelijk interpreteerbaar resultaat geeft op een schaal van 0 tot 100, waarbij een hogere waarde een eenvoudigere tekst impliceert. Daarnaast komt de inschatting van de Coh-Metrix vrij goed overeen met hoe studenten daadwerkelijk presteerden (Nelson, Perfetti, Liben, & Liben, 2012; McNamara & Graesser, 2012; Nahatame, 2021) en geeft de Coh-Metrix ook voor L2-lezers een goede inschatting van de moeilijkheid van een tekst (McNamara & Graesser, 2012; Nahatame, 2021). De Arizona State University heeft de door hun ontwikkelde tool beschikbaar gesteld voor dit onderzoek. De (aangepaste) webversie is beschikbaar via <http://141.225.61.35/cohmetrix2017>.

Flesch Reading Ease score

Om de Flesch Reading Ease (FRE) score te berekenen is het nodig om het aantal woorden van de tekst te bepalen, het aantal zinnen en het aantal lettergrepen. Dit is een taak die prima geautomatiseerd kan. Hierom worden er vijf websites benut om de FRE te berekenen, waarna er een gemiddelde score berekend zal worden. Het betreft de volgende websites:

1. charactercalculator.com/flesch-reading-ease/
2. www.online-utility.org/english/readability_test_and_improve.jsp
3. www.webfx.com/tools/read-able/check.php
4. www.textinspector.com/
5. readabilityformulas.com/free-readability-formula-tests.php

De reden om specifiek deze vijf websites te benutten, heeft te maken met gebruiksgemak (teksten zijn makkelijk te plakken in een tekst vak), snel resultaat (direct zichtbaar of na één klik), positie binnen zoekresultaten via Google (op *flesch reading ease score calculator*) en eerdere bekendheid door de auteur.

De reden om vijf websites te gebruiken, is omdat er kleine verschillen kunnen zijn in hoe de algoritmes die erachter schuilen de verschillende leestekens interpreteren. Zo kunnen er meer of minder zinnen, woorden, of lettergrepen onderscheiden worden, waardoor er, consistent binnen de website, verschillen kunnen ontstaan. Door de waardes te middelen is de verwachting dat uitschieters, als deze niet dermate extreem zijn dat ze beter genegeerd kunnen worden, verzacht worden. Met deze gemiddelde FRE-score wordt het ERK-niveau benaderd vergelijkbaar met de methode in het

onderzoek van Natova (2019), maar met als aanpassing dat de tientallen-waardes geen dubbel niveau aangeven.

Tabel 1: *ERK-niveau in vergelijking met FRE-score volgens Natova en Linguapress*

ERK-niveau	Flesch Reading Ease score
A1	>90
A2	80 – 89
B1	70 – 79
B2	60 – 69
C1	50 – 59
C2	0 – 49

ERK-websites

Als derde manier van het inschatten van het niveau van de teksten, zullen een aantal websites benut worden die zeggen een inschatting te kunnen maken van het ERK-niveau van een tekst. Het betreft de volgende websites:

1. www.roadtogrammar.com/textanalysis/
2. www.textinspector.com/
3. <https://hub.cathoven.com/>

Zoals eerder aangegeven, maken deze websites gebruik van woordlijsten en aanvullende algoritmen om tot een niveau indicatie te komen. De uitzondering hierbij is textinspector.com, welke gebruikt maakt van tot wel twintig verschillende indicaties om tot een voorspelling te komen, elk op basis van wetenschappelijk onderzoek. De reden om, weer, meerdere websites te gebruiken is om uitschieters te kunnen compenseren en niet afhankelijk te zijn van één methode, inclusief de meetfouten die daarin kunnen zitten.

Alle informatie die zo verzameld wordt, FRE-scores die leiden tot een ERK-indicatie gepaard met nog drie inschattingen van het ERK-niveau van de tekst, wordt gecombineerd om tot een gemiddelde ERK-inschatting te komen van de tekst. Om het gemiddelde te berekenen wordt eerst aan iedere ERK-inschatting een numerieke waarde toegekend. Deze waarde is te zien in tabel 2. Het gemiddelde van de vier inschattingen wordt ook met deze tabel vergeleken om tot een gemiddelde ERK-inschatting te komen. Omdat waarschijnlijk de inschattingen van elkaar zullen

verschillen, wordt het gemiddelde eerst afgerond op het dichtstbijzijnde tiental voorafgaand aan de vergelijking.

Tabel 2: Numerieke toekenning per ERK-inschatting

ERK-inschatting	Waarde
B1	10
B2	20
C1	30
C2	40

Voorbeeld

Stap 1

Om het hierboven beschreven proces te verduidelijken, kan een voorbeeld helpen. Stel dat de [vijf genoemde websites](#) voor het berekenen van de FRE de volgende scores geven bij een voorbeeld tekst:

53,38	54,4	58,5	52,29	58,5
-------	------	------	-------	------

Deze waarden worden vervolgens gemiddeld tot 55,41 wat volgens tabel 1 overeenkomt met een C1 inschatting qua moeilijkheid.

Stap 2

De drie [genoemde ERK-inschattingswebsites](#) geven de volgende inschattingen:

C1	C2	B2,8
----	----	------

Ieder van deze inschattingen wordt eerst opgezet naar een numerieke waarde volgens tabel 2: 30 – 40 – 28. Hier komt de inschatting vanuit stap 1 bij, ook omgezet (C1 → 30).

Van deze vier waarde wordt het gemiddelde berekend⁶: 32 wat vervolgens wordt afgerond naar 30. Na vergelijking met tabel 2 krijgt de tekst een inschatting van C1.

Kanttekening

Door gemiddelde te nemen, te converteren van een FRE score naar ERK-inschatting, dit weer om te zetten naar een numerieke waarde én een afronding te maken van een gemiddelde, is het goed mogelijk dat er teksten die op de grens tussen B2 en C1 of C1 en C2 liggen verkeerd ingedeeld worden. Dit is echter eenvoudig te controleren door de afronding pas later toe te passen of achterwege te laten en zo te zien of er verschillen optreden.

⁶ $(30+40+28+30) / 4$

4.2 Bepalen van onderwerpen

Het vwo-examen Engels heeft geen onderwerpen zoals fotosynthese of de Gouden Eeuw die op verschillende niveaus bevraagd worden. De ERK ‘*can do*’-descriptor (College voor Toetsen en Examens, 2019) zouden binnen een specificatietabel de functie van onderwerpen kunnen vervullen, maar hier zitten enkele problemen aan. Naast de al eerder beschreven, zie [3.3.3 Onderwerpen](#), problemen bij het overnemen van omschrijvingen vanuit het ERK en problemen met de indeling van, en beperking binnen, de schalen, zijn er nog twee bezwaren tegen het gebruik van de ‘*can do*’-descriptor als onderwerpen om een specificatietabel op te stellen voor het vwo-examen Engels.

Ten eerste, het taalgebruik is dermate vaag dat specifieke vragen eraan toekennen lastig zal zijn. Bijvoorbeeld onder oriënterend lezen staat er “kan meer complexe advertenties begrijpen” (College voor Toetsen en Examens, 2019, p. 14), maar geen uitleg over wat “meer complexe” advertenties zijn. En zelfs begrijpen is voor meerdere interpretaties vatbaar. Dit bezwaar komt overeen met de algemene omschrijving van tekstniveaus zoals opgemerkt door Natova (2019).

Ten tweede, is aan de ‘*can do*’-descriptor al direct een niveau gekoppeld, dus de bovengenoemde “complexere advertenties” zou alleen op B2 niveau mogelijk zijn, en niet lager of hoger.

Om toch een specificatietabel te kunnen opstellen, wordt binnen dit onderzoek gebruik gemaakt van handvatten die het PISA 2018 *Assessment and Analytical Framework* onderzoek geeft (OECD, 2019), gekoppeld aan de eindtermen van domein A: Leesvaardigheid (College voor Toetsen en Examens, 2019) verder ondersteund door wat de wetenschappelijke literatuur als toetsbare factoren ziet die een goede voorspeller kunnen zijn voor de mate van tekstbegrip.

De OECD richt zich bij het meten van leesvaardigheid op een aantal specifieke cognitieve processen gekoppeld aan tekstverwerking, te weten: *read fluently*, *locate information*, *understand*, en *evaluate and reflect* (OECD, 2019, p. 33). Wat betreft het vwo-examen kan *read fluently* (vloeiend lezen) buiten beschouwing gelaten worden, daar hier geen vragen op gericht zijn en er (ruim) voldoende tijd is om teksten te (her)lezen. De andere drie processen zijn hoofdgroepen, met elk nog deelprocessen. Deze vormen de basis van de onderwerpen voor de specificatietabel en worden hieronder verder uitgewerkt.

Tabel 3: tekstverwerkingsprocessen volgens PISA 2018 Reading framework (OECD, 2019, p 33)

Hoofdproces	Deelproces
<i>Locate information</i>	<i>Access and retrieve information within a text</i>
	<i>Search and select relevant text</i>
<i>Understand</i>	<i>Represent literal meaning</i>
	<i>Integrate and generate inferences</i>
<i>Evaluate and reflect</i>	<i>Assess quality and credibility</i>
	<i>Reflect on content and form</i>
	<i>Detect and handle conflict</i>

Gezien de samenstelling van het vwo-examen Engels, waarbij er uitsluitend vragen over een enkele tekst gesteld worden en geen vergelijking tussen teksten nodig is, kunnen twee deelprocessen, *search and select relevant text* en *detect and handle conflict*, verder buiten beschouwing gelaten worden.

Access and retrieve information

Om de juiste, specifieke informatie binnen een tekst te vinden, wordt er een beroep gedaan op hoe goed de lezer de taak begrijpt, diens kennis van tekstopbouw (tussenkopjes, paragrafen), en kundigheid in het beoordelen van de relevantie van een stuk tekst. Vooral dit laatste komt overeen met de domein A omschrijving “aangeven welke informatie relevant is, gegeven een vaststaande behoefte”. Vanuit de wetenschappelijke literatuur is dit het beste te koppelen aan kennis van tekststructuur (Bogaerds-Hazenbergh, Evers-Vermeul, & van den Bergh, 2020) en/of leesstrategieën (Kintsch, 1998; Kintsch & Rawson, 2005; Duke, Pearson, Strachan, & Billman, 2011).

Om te bepalen of een vraag binnen deze categorie valt, moet het antwoord te vinden zijn in een aantal woorden, frase of numerieke waarden binnen de tekst. Hierbij is het niet nodig om een groter tekstbegrip te hebben dan de passage waar het antwoord in staat.

Represent literal meaning

Bij deze categorie gaat het om het begrijpen van enkele zinnen of een korte passage. Dit sluit aan bij domein A “de betekenis van belangrijke elementen van een tekst aangeven”, met de veronderstelling dat het belang niet zelf ingeschat hoeft te worden. Zoals omschreven, lijkt dit vooral een beroep te doen op de woordenschat van de lezer (Schmitt, Jiang, & Grabe, 2011; Trapman, van Gelderen, van Schooten, & Hulstijn, 2017; Oakhill, Cain, & Carsten, 2019).

Om te bepalen of een vraag specifiek de lezer vraagt naar de letterlijke betekenis van een passage, ofwel woordkennis, zal worden gekeken of het juist beantwoorden van de vraag afhankelijk is van het kennen van één of meerdere specifieke woorden, zonder dat hierbij om gevolgtrekking, tekststructuur of iets anders gevraagd wordt. Veelal zal het antwoord een parafrase zijn van dat wat er in de tekst staat.

Integrate and generate inferences

Hierbij gaat het om taken die om begrip vragen van gehele passages of complexere samenhang (zoals ruimtelijke, chronologische of oorzakelijke verbanden). Doorgaans is het product het aangeven van de (hoofd)gedachte van een stuk tekst, een samenvatting, of een (nieuwe) titel. Domein A beschrijft dit als “de hoofdgedachte van een tekst(gedeelte) aangeven”, maar ook “relaties tussen delen van een tekst aangeven” sluit hierbij aan. Dit proces gaat om de kundigheid in gevolgtrekking van de lezer (Perfetti & Stafura, 2015; Ahmed, et al., 2016; Kendeou, McMaster, & Christ, 2016), gecombineerd met kennis van tekststructuur (Bogaerds-Hazenberg, Evers-Vermeul, & van den Bergh, 2020).

Vragen die specifiek naar de (hoofd)gedachte van een stuk tekst vragen, horen hierbij. Belangrijk is dat het antwoord niet letterlijk, of geparafraseerd, in de tekst mag voorkomen.

Assess quality and credibility

Vaardige lezers kunnen ook de kwaliteit en geloofwaardigheid van een tekst beoordelen, bijvoorbeeld of de informatie valide, actueel, accuraat en/of onbevooroordeeld is. Domein A heeft geen equivalent hierin, waardoor verondersteld mag worden dat hier geen vragen in gesteld zullen worden.

Reflect on content and form

Vaardige lezers moeten ook kunnen reflecteren op de kwaliteit van schrijven en de schrijfstijl van de auteur. Hieronder valt het beoordelen van hoe de inhoud en vorm van de tekst samenhangen met het doel van de schrijver en diens opvattingen. Domein A omschrijft dit als “conclusies trekken met betrekking tot intenties, opvattingen en gevoelens van de auteur.” Het draait hierbij weer om het kunnen interpreteren van informatie die niet letterlijk in de tekst staat, wat het beste aansluit bij gevolgtrekking (Ahmed, et al., 2016; Kendeou, McMaster, & Christ, 2016).

Vragen die vragen naar de intentie van de schrijver, emotie of emotioneel belang van woorden, en schrijfstijl vallen hieronder.

Het is mogelijk dat tijdens het uitvoeren van deze analyse vragen niet geclassificeerd kunnen worden, onder meerdere componenten passen, of het toevoegen van een extra categorie nodig blijkt. Indien dit het geval blijkt te zijn, zal dit later worden verwerkt.

4.3 Specificatietabel

Nu bepaald is hoe de niveaus ingedeeld zullen worden en naar welke onderwerpen gekeken zal worden, is het mogelijk een algemene specificatietabel voor het vwo-examen Engels op te stellen. Dit is tabel 4 hieronder.

Tabel 4: *Specificatietabel voor leesvaardigheid examens vwo op basis van eindtermen, OECD en literatuur*

<u>Onderwerp</u>			<u>Moeilijkheidsgraad van de tekst</u>		
			B2	C1	C2
OECD/PISA	Domein A	Component	Coh-Metrix waarde		
Access and retrieve information	Aangeven welke informatie relevant is, gegeven een vaststaande behoefte	Kennis van tekststructuur & leesstrategieën			
Represent literal meaning	De betekenis van belangrijke elementen van een tekst aangeven	Woordenschat			
Integrate and generate inferences	De hoofdgedachte van een tekst(gedeelte) aangeven	Gevolgtrekking			
	Relaties tussen delen van een tekst aangeven	Kennis van tekststructuur			
Reflect on content and form	Conclusies trekken met betrekking tot intenties, opvattingen en gevoelens van de auteur	Gevolgtrekking			

Het is van belang om ook te vermelden dat vrijwel alle literatuur aangeeft dat iedere individuele meting onderhevig is aan meetfouten en slechts een deel van het geheel van tekstbegrip meet (Francis, et al., 2006) en producten imperfect het proces weergeven (Cutting & Scarborough, 2006; Fletcher, 2006; van den Broek & Espin, 2012; Kendeou, McMaster, & Christ, 2016; Afflerbach, Cho, Crassas, & Kim, 2018).

Factoren zoals voorkennis, motivatie en een variatie aan leesstrategieën, die een belangrijke rol spelen in tekstbegrip (Willingham, 2009; Kendeou, McMaster, & Christ, 2016), zijn buiten beschouwing gelaten, omdat deze binnen de beperkingen van de vorm van het huidige examen niet tot nauwelijks toetsbaar zijn en ook niet opgenomen zijn in de OECD omschrijvingen of eindtermen.

5. Praktische uitvoering

De zes examens zijn door de onderzoeker geanalyseerd tussen 22 mei en 5 augustus 2022 zoals omschreven in [4.1](#) (bepalen van moeilijkheidsgraad van teksten) en [4.2](#) (bepalen van onderwerpen).

Bij het beoordelen van de moeilijkheidsgraad van de teksten is nog één extra methode gebruikt om het ERK-niveau te verfijnen: duidelijke zichtbare verdelingen van de gebruikte woorden over de ERK-niveaus volgens Cambridge en Oxford woordenlijsten. Hiervoor is via textinspector.com de vergelijking met het *English Vocabulary Profile Online*⁷ benut en via oxfordlearnersdictionaries.com/text-checker/ een vergelijking met de Oxford 5000. Op momenten dat het mogelijk was om 90% of meer van de woorden in te delen, werd het hoogste niveau dat hiervoor nodig was als extra indicatie benut. Wanneer 89% ingedeeld werd, werd het hoogste niveau grotendeels meegerekend, onder de 89% werd niks meegenomen. Zo kreeg ongeveer 40% van de teksten twee extra indicaties voor het ERK om mee te middelen.

Van de websites die gebruikt zijn voor het inschatten van het ERK-niveau, viel op dat textinspector.com vaak hoger dan het gemiddelde inschatte. Dit was echter niet dermate veel hoger dat het een vertekend beeld gaf⁸.

Ook bij de websites die gebruikt zijn voor het bereken van de Flesch Reading Ease score was er variatie, meer nergens meer de twee keer de standaarddeviatie aan afwijking en geen website gaf consistent een afwijkende waarde, waardoor er geen reden was om de scores niet mee te nemen in de berekening.

Tijdens het analyseren viel op dat er redelijk een connectie te maken was tussen de Coh-Metrix waarde en het ERK-niveau, maar het was geen perfecte correlatie. Globaal viel wel te stellen dat wanneer de Coh-Metrix waarde hoger was, het ERK-niveau lager was, wat klopt met de opzet van de meetmethoden.

Tijdens het plaatsen van de vragen binnen de onderwerpen bleek dat sommige vragen niet exact bij een onderwerp van tabel 4 te plaatsen waren. Het ging om vragen waarbij een woord uit de tekst was weggelaten en er gevraagd werd welk woord dit zou kunnen

⁷ <http://englishprofile.org/wordlists/evp>

⁸ De gemiddelde score van alle teksten mét [textinspector](http://textinspector.com) is ongeveer 33 (C1) en zonder is het 31 (C1). De standaarddeviatie verandert van 5,48 naar 5,53.

zijn geweest. Dit soort *cloze* vragen verlangen van de lezer dat hij de “logica” van de zin kan herstellen (Kleijn, Pander Maat, & Sanders, 2019), waarbij het een beroep doet op woordenschat, kennis van syntaxis en semantiek. Voor enkele *cloze* vragen, waarbij de antwoorden bestonden uit woorden die tekstuele verbanden aanduiden zoals *what is more, among other things*, of *rhetorically*, is ook kennis van tekststructuur nodig, waardoor dit een eigen categorie verdient.

Een tweetal open vragen zijn aan twee categorieën toegeschreven.

Uiteindelijk ziet de specificatietabel binnen dit onderzoek eruit zoals tabel 5 op pagina 32. Voor de overzichtelijkheid zijn de Coh-Metrix waarde zo gegroepeerd dat er geen lege kolommen ontstaan, maar er nog wel onderscheid in moeilijkheid te maken is. Bij de specificatietabellen per examen is de exacte waarde gebruikt.

Tabel 5: Blanco specificatietabel binnen dit onderzoek

				Moeilijkheidsgraad van de tekst (ERK & Coh-Metrix 3.0)									
Onderwerp				C2			C1			B2	B1		
OECD	Domein A	Component		10 - 40	41 - 50	51 - 65	10 - 49	50 - 60	61 - 80	70 - 90	65 - 70	Punten	%
1	Access and retrieve information	Aangeven welke informatie relevant is	Leesstrategieën & tekststructuur										
2	Represent literal meaning	De betekenis van belangrijke elementen van een tekst aangeven	Woordenschat										
3	Integrate and generate inferences	De hoofdgedachte van een tekst(gedeelte) aangeven	Gevolgtrekking										
4		Relaties tussen delen van een tekst aangeven	Tekststructuur										
5a			Cloze: woordenschat, syntaxis en semantiek										
5b			Cloze: woordenschat, syntaxis en semantiek, tekststructuur										
6	Reflect on content and form	Conclusies trekken met betrekking tot intenties, opvattingen en gevoelens van de auteur	Gevolgtrekking										
Aantal punten												0	
percentage v.d. punten													

6. Praktische relevantie

Op basis van de specificatietabellen van de zes vwo-examens is het waarschijnlijk dat aangegeven kan worden welke (deel)vaardigheden en kennis voornamelijk nodig zijn om deze succesvol te maken. Secties Engels kunnen deze informatie gebruiken, naast de vergelijking met het huidige curriculum, om aanpassingen te maken, indien gewenst.

Andere talensecties zouden eenzelfde afweging kunnen maken, onder de aanname dat een Cito examen Duits, Frans of Spaans niet veel verschilt van een examen Engels.

Uiteindelijk zouden ook Cito de analyse kunnen benutten om hun examens te finetunen of andere aanpassingen te maken, zoals het toevoegen van stappen in het samenstellen van de examens om vooraf al beter op inhoudsvaliditeit te toetsen. Al is dit laatste onwaarschijnlijk gezien de weerstand vanuit Cito en het CvTE om de dialoog aan te gaan ten aanzien van de inhoud en de constructie van de examens (van Haperen, 2022).

7. Resultaten en conclusies

In totaal zijn er 71 teksten met de daarbij horende 246 vragen onderzocht. Deze 246 vragen waren bij elkaar 278 punten waard. Van acht van de teksten was de oorspronkelijke bron niet meer te achterhalen. Zes van de teksten waren een gedeelte van een eerste hoofdstuk van een boek. De overige 57 teksten konden teruggevonden worden op het internet. 48 van de 71 teksten (67%) zijn op enige wijze aangepast. Het kon zijn dat enkele woorden waren aangepast, toegevoegd, verwijderd of verduidelijkt, maar in enkele gevallen betrof het ook het (significant) inkorten van de originele tekst, of het aanpassen van een woord of zin waar een vraag over werd gesteld. Enkele keren was er ook een spellingsfout gemaakt in een (achter)naam.

7.1 Resultaten

Voor elk van de zes onderzochte examens is een specificatietabel opgesteld. Deze specificatietabellen (tabel 6 tot en met tabel 11) zijn te vinden op pagina 35 tot en met 40.

De waarden binnen de opgestelde specificatietabellen zijn het aantal te behalen punten per onderwerp bij teksten van de betreffende moeilijkheidsgraad binnen het specifieke examen. Het gaat niet om het aantal vragen. De keuze voor aantal punten boven aantal vragen is omdat het een beter beeld geeft van waar de nadruk ligt voor de toets makers en waar leerlingen zich het beste op voor kunnen bereiden. Sommige vragen leveren ook meerdere punten op, waardoor ze een groter aandeel hebben in de uiteindelijke score.

Ter illustratie: in het 1^e tijdvak examen van 2018 was er één (1) punt te verdienen voor een vraag binnen onderwerp 1 (*Access and retrieve information // Aangeven welke informatie relevant is // Leesstrategieën & tekststructuur*) bij een tekst van niveau C2 met Coh-Metrix waarde 44.

Geheel rechts staat hoeveel punten er binnen elk onderwerp in totaal te verdienen waren binnen het betreffende examen en welk percentage van het totaal aantal punten dit is. Zo waren er voor onderwerp 1 in totaal 6 punten te verdienen voor 12,5% van het totaal.

Geheel onderaan staat hoeveel punten er binnen het betreffende niveau te verdienen waren, inclusief welk percentage van het totaal dit is. Zo waren er op ERK-niveau C2 met Coh-Metrix score 44 in totaal 9 punten te verdienen voor 18,8% van het totaal.

2018 – 1

Tabel 6: Specificatietabel voor vwo-examen Engels 2018, tijdvak 1

				Moeilijkheidsgraad van de tekst (ERK & Coh-Metrix 3.0)												
Onderwerp				C2		C1	C2		C1				B2			
OECD	Domein A		Component	44	47	48	49	50	56	57	58	60	73	77	Punten	%
1	Access and retrieve information	Aangeven welke informatie relevant is	Leesstrategieën & tekststructuur	1	1						1	2		1	6	12,5%
2	Represent literal meaning	De betekenis van belangrijke elementen van een tekst aangeven	Woordenschat	2				2,5		4				4	12,5	26,0%
3	Integrate and generate inferences	De hoofdgedachte van een tekst(gedeelte) aangeven	Gevolgtrekking	2	2	1	1			1	2				9	18,8%
4		Relaties tussen delen van een tekst aangeven	Tekststructuur	1	1		1	0,5		1	1			1	6,5	13,5%
5a			Cloze: woordenschat, syntaxis en semantiek	1	1			3	1	1	1	1			9	18,8%
5b			Cloze: woordenschat, syntaxis en semantiek, tekststructuur	1							1				2	4,2%
6	Reflect on content and form	Conclusies trekken met betrekking tot intenties, opvattingen en gevoelens van de auteur	Gevolgtrekking	1									2		3	6,3%
Aantal punten				9	5	1	2	6	1	7	6	3	2	6	48	
percentage v.d. punten				18,8%	10,4%	2,1%	4,2%	12,5%	2,1%	14,6%	12,5%	6,3%	4,2%	12,5%		

2018 – 2

Tabel 7: Specificatietabel voor vwo-examen Engels 2018, tijdvak 2

				Moeilijkheidsgraad van de tekst (ERK & Coh-Metrix 3.0)													
Onderwerp				C1	C2		C1						B1				
OECD	Domein A	Component		12	24	43	49	52	53	55	57	59	63	65	69	Punten	%
1	Access and retrieve information	Aangeven welke informatie relevant is	Leesstrategieën & tekststructuur				2		2				2		1	7	15,6%
2	Represent literal meaning	De betekenis van belangrijke elementen van een tekst aangeven	Woordenschat		2		4					0,5		3	3	12,5	27,8%
3	Integrate and generate inferences	De hoofdgedachte van een tekst(gedeelte) aangeven	Gevolgtrekking				1		1		3		1	1		7	15,6%
4		Relaties tussen delen van een tekst aangeven	Tekststructuur						2		1	0,5	1			4,5	10,0%
5a			Cloze: woordenschat, syntaxis en semantiek				2	2	1			1	1		7	15,6%	
5b			Cloze: woordenschat, syntaxis en semantiek, tekststructuur		1	1		1				1			4	8,9%	
6	Reflect on content and form	Conclusies trekken met betrekking tot intenties, opvattingen en gevoelens van de auteur	Gevolgtrekking														
																	</

Tabel 8: Specificatietabel voor vwo-examen Engels 2019, tijdvak 1

				Moeilijkheidsgraad van de tekst (ERK & Coh-Metrix 3.0)												
Onderwerp				C2	C1	C2				C1	C2	C1		B2		
OECD	Domein A	Component		23	38	43	46	49	50	50	52	57	59	83	Punten	%
1	Access and retrieve information	Aangeven welke informatie relevant is	Leesstrategieën & tekststructuur					1	1		1			2	5	11,1%
2	Represent literal meaning	De betekenis van belangrijke elementen van een tekst aangeven	Woordenschat					2	3		3		1	4	13	28,9%
3	Integrate and generate inferences	De hoofdgedachte van een tekst(gedeelte) aangeven	Gevolgtrekking		1	1			2	1	1	1			7	15,6%
4		Relaties tussen delen van een tekst aangeven	Tekststructuur		1			1	1		1			1	5	11,1%
5a			Cloze: woordenschat, syntaxis en semantiek	1			2	1			2	1		7	15,6%	
5b			Cloze: woordenschat, syntaxis en semantiek, tekststructuur				1		1					2	4,4%	
6	Reflect on content and form	Conclusies trekken met betrekking tot intenties, opvattingen en gevoelens van de auteur	Gevolgtrekking				1	2	1	1	1				6	13,3%
Aantal punten				1	2	1	4	7	9	2	7	3	2	7	45	
percentage v.d. punten				2,2%	4,4%	2,2%	8,9%	15,6%	20,0%	4,4%	15,6%	6,7%	4,4%	15,6%		

Tabel 9: Specificatietabel voor vwo-examen Engels 2019, tijdvak 2

				Moeilijkheidsgraad van de tekst (ERK & Coh-Metrix 3.0)										
Onderwerp				C2		C1		C2	C1	C2	C1			
OECD	Domein A	Component		33	50	52	54	58	61	62	62	64	Punten	%
1	Access and retrieve information	Aangeven welke informatie relevant is	Leesstrategieën & tekststructuur	1	2		1	1				1	6	13,0%
2	Represent literal meaning	De betekenis van belangrijke elementen van een tekst aangeven	Woordenschat	1	1		3	3	2	1		2	13	28,3%
3	Integrate and generate inferences	De hoofdgedachte van een tekst(gedeelte) aangeven	Gevolgtrekking			1	1	1	2				6	13,0%
4		Relaties tussen delen van een tekst aangeven	Tekststructuur		1				1				2	4,3%
5a			Cloze: woordenschat, syntaxis en semantiek		1		1			1			3	6,5%
5b			Cloze: woordenschat, syntaxis en semantiek, tekststructuur			1	1	1		1		2		6
6	Reflect on content and form	Conclusies trekken met betrekking tot intenties, opvattingen en gevoelens van de auteur	Gevolgtrekking		2		2		3	1		1	10	21,7%
Aantal punten				2	8	2	9	5	9	3		6	2	46
percentage v.d. punten				4,3%	17,4%	4,3%	19,6%	10,9%	19,6%	6,5%		13,0%	4,3%	

Tabel 10: Specificatietabel voor vwo-examen Engels 2021, tijdvak 1

				Moeilijkheidsgraad van de tekst (ERK & Coh-Metrix 3.0)												
Onderwerp				C2		C1	C2		C1	C2	C1					
OECD	Domein A	Component		17	44	48	50	52	54	57	61	61	68	77	Punten	%
1	Access and retrieve information	Aangeven welke informatie relevant is	Leesstrategieën & tekststructuur				1			3	1				5	10,4%
2	Represent literal meaning	De betekenis van belangrijke elementen van een tekst aangeven	Woordenschat	3	3	1		1	1				1	3	13	27,1%
3	Integrate and generate inferences	De hoofdgedachte van een tekst(gedeelte) aangeven	Gevolgtrekking		2		1	2		2	1			1	9	18,8%
4		Relaties tussen delen van een tekst aangeven	Tekststructuur		1		1		2						4	8,3%
5a			Cloze: woordenschat, syntaxis en semantiek		1		1	1		2		1	1		7	14,6%
5b			Cloze: woordenschat, syntaxis en semantiek, tekststructuur		1			1							2	4,2%
6	Reflect on content and form	Conclusies trekken met betrekking tot intenties, opvattingen en gevoelens van de auteur	Gevolgtrekking		1		3	1			1		1	1	8	16,7%
Aantal punten				3	9	1	7	6	3	7	3	1	3	5	48	
percentage v.d. punten				6,3%	18,8%	2,1%	14,6%	12,5%	6,3%	14,6%	6,3%	2,1%	6,3%	10,4%		

2021 – 2

Tabel 11: Specificatietabel voor vwo-examen Engels 2021, tijdvak 2

				Moeilijkheidsgraad van de tekst (ERK & Coh-Metrix 3.0)												
Onderwerp				C2	C1	C2	C1				C2	C1	B2			
OECD	Domein A	Component		32	47	47	49	53	55	56	57	59	62	74	Punten	%
1	Access and retrieve information	Aangeven welke informatie relevant is	Leesstrategieën & tekststructuur		1		1	1	1	1					5	10,9%
2	Represent literal meaning	De betekenis van belangrijke elementen van een tekst aangeven	Woordenschat	2		2	1	1			3		1	2	12	26,1%
3	Integrate and generate inferences	De hoofdgedachte van een tekst(gedeelte) aangeven	Gevolgtrekking	2	2		2	2		1	2		1		12	26,1%
4		Relaties tussen delen van een tekst aangeven	Tekststructuur	1	1		2								4	8,7%
5a			Cloze: woordenschat, syntaxis en semantiek		2	1			1	1					5	10,9%
5b			Cloze: woordenschat, syntaxis en semantiek, tekststructuur				1								1	2,2%
6	Reflect on content and form	Conclusies trekken met betrekking tot intenties, opvattingen en gevoelens van de auteur	Gevolgtrekking		2		1		1		2	1			7	15,2%
Aantal punten				5	8	3	8	4	2	3	8	1	2	2	46	
percentage v.d. punten				10,9%	17,4%	6,5%	17,4%	8,7%	4,3%	6,5%	17,4%	2,2%	4,3%	4,3%		

7.1.1 Spreiding van niveau

ERK

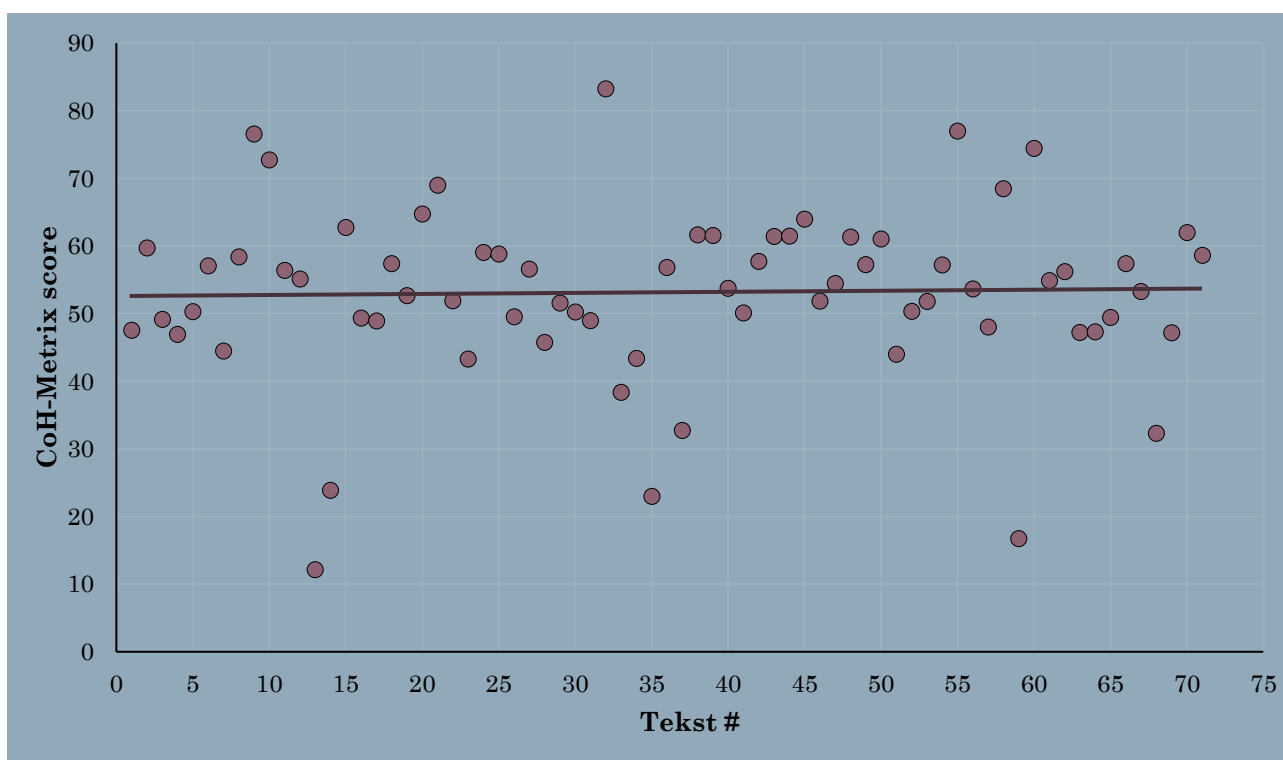
Er zijn vier verschillende ERK-niveaus van teksten onderscheiden: B1, B2, C1 en C2. Hoeveel teksten per niveau, per examen en hoeveel in totaal en welk percentage dit van alle onderzochte teksten is, staat in tabel 12.

Tabel 12: Aantal teksten per ERK-niveau per examen

Examen	ERK-Niveau			
	B1	B2	C1	C2
2018 – 1		2	5	4
2018 – 2	1		8	4
2019 – 1		1	5	6
2019 – 2			7	4
2021 – 1			6	6
2021 – 2		1	6	5
Som	1	4	37	29
% van totaal	1,41%	5,63%	52,11%	40,85%

Coh-Metrix

De verdeling van teksten volgens de Coh-Metrix is logischerwijs breder dan volgens de ERK-niveaus, omdat de schaal van de Coh-Metrix meer waarden bevat. Grafiek 1 is een spreidingsdiagram van alle 71 teksten op chronologische volgorde en hun Coh-Metrix waarde. Hierin is een trendlijn toegevoegd om de gemiddelde waarde aan te geven (53).



Grafiek 1: Spreiding van teksten volgens Coh-Metrix waarde

In navolging van tabel 12, is in tabel 13 de verdeling van Coh-Metrix niveaus over de examens weergegeven.

Tabel 13: Aantal teksten per Coh-Metrix bin per examen

Examen	Coh-Metrix bin							
	11 - 20	21 - 30	31 - 40	41 - 50	51 - 60	61 - 70	71 - 80	80+
2018 – 1				5	4		2	
2018 – 2	1	1		3	5	3		
2019 – 1		1	1	5	4			1
2019 – 2			1	1	4	5		
2021 – 1	1			3	4	3	1	
2021 – 2			1	4	5	1	1	
Som	2	2	3	21	26	12	4	1
% van totaal	2,82%	2,82%	4,23%	29,58%	36,62%	16,90%	5,63%	1,41%

Doordat er op twee manieren naar het niveau van de teksten gekeken is, is het ook mogelijk om hiertussen een overzicht te maken. Dit is gedaan in tabel 14, welke aangeeft hoeveel teksten er voorkwamen binnen een 5-punts bereik van de Coh-Metrix waarde in vergelijking met het ERK-niveau.

Tabel 14: Aantal teksten per combinatie Coh-Metrix bin en ERK-niveau

Coh-Metrix score	B1	B2	C1	C2	Totaal
11 – 15			1		1
16 – 20				1	1
21 – 25				2	2
26 – 30					-
31 – 35				2	2
36 – 40			1		1
41 – 45				4	4
46 – 50			4	13	17
51 – 55			8	3	11
56 – 60			13	2	15
61 – 65			8	2	10
66 – 70	1		1		2
71 – 75		2			2
76 – 80		1	1		2
81 – 85		1			1
Totaal	1	4	37	29	71

7.2 *Verdeling van onderwerpen*

[Bijlage 1](#) geeft een verzameling van de exacte vraagstelling per onderwerp. Hieronder staan per onderwerp voorbeelden van de meest voorkomende formuleringen.

1. Access and retrieve information // Aangeven welke informatie relevant is // Leesstrategieën & tekststructuur
 - Citeer ...
 - In welke alinea wordt ... ? Noteer het nummer van deze alinea op het antwoordblad.
2. Represent literal meaning // De betekenis van belangrijke elementen van een tekst aangeven // Woordenschat
 - Geef van (elke van) de volgende beweringen aan of deze overeenkomt met ...
 - Which of the following {is true}?
3. Integrate and generate inferences // De hoofdgedachte van een tekst(gedeelte) aangeven // Gevolgtrekking
 - What is the main point ... ?
 - Which of the following is in line with ... ?
 - Which of the following is true according to ... ?
4. Integrate and generate inferences // Relaties tussen delen van een tekst aangeven // Tekststructuur
 - In welke volgorde stonden bovenstaande {alinea's | zinnen} ?
 - What is the function of paragraph ... ?
5. Integrate and generate inferences // Relaties tussen delen van een tekst aangeven
 - Which of the following fits the gap in ... ?
6. Reflect on content and form // Conclusies trekken met betrekking tot intenties, opvattingen en gevoelens van de auteur // Gevolgtrekking
 - How can the tone ... ?
 - How does the writer introduce ... ?
 - Which of the following reflects {the writer's opinion} ?

Bij onderwerp 5 is het nog van belang dat het onderscheid tussen soort A en B gemaakt wordt door de antwoordmogelijkheden. Afbeelding 2 geeft een voorbeeld van een vraag binnen onderwerp 5a en afbeelding 3 voor onderwerp 5b. Beide vragen komen uit het 2019 examen, 1^e tijdvak.

- 1p **1** Which of the following fits the gap in paragraph 1?
- A** ethical principles
 - B** national electricity consumption
 - C** population size
 - D** socio-economic circumstances

Afbeelding 2: Voorbeeld van vraag binnen onderwerp 5a

- 1p **24** Which of the following fits the gap in paragraph 12?
- A** Admittedly,
 - B** In retrospect,
 - C** Instead,
 - D** Nevertheless,
 - E** That's why

Afbeelding 3: Voorbeeld van vraag binnen onderwerp 5b

Voor elk van de zeven uiteindelijk onderscheidde onderwerpen is vast te stellen hoeveel vragen erbinnen gesteld zijn en hoeveel punten deze opleverden. In tabel 15 is te zien hoeveel vragen er per onderwerp per examen gesteld zijn, en in tabel 16 hoeveel punten per onderwerp per examen er te verdienen waren.

Om de tabellen wat compacter te houden, is gewerkt met nummering van de onderwerpen in plaats van een volledige omschrijving. Hieronder staat het overzicht van die nummering en onderwerpen, zoals ze ook gebruikt zijn in de specificatietabellen.

1. Access and retrieve information // Aangeven welke informatie relevant is // Leesstrategieën & tekststructuur
2. Represent literal meaning // De betekenis van belangrijke elementen van een tekst aangeven // Woordenschat
3. Integrate and generate inferences // De hoofdgedachte van een tekst(gedeelte) aangeven // Gevolgtrekking
4. Integrate and generate inferences // Relaties tussen delen van een tekst aangeven // Tekststructuur
5. Integrate and generate inferences // Relaties tussen delen van een tekst aangeven //
 - Cloze: woordenschat, syntaxis en semantiek
 - Cloze: woordenschat, syntaxis en semantiek, tekststructuur
6. Reflect on content and form // Conclusies trekken met betrekking tot intenties, opvattingen en gevoelens van de auteur // Gevolgtrekking

Tabel 15: Aantal vragen per onderwerp per examen

Aantal vragen per onderwerp							
Examen	1	2	3	4	5a	5b	6
2018 – 1	5	6,5	9	6,5	9	2	3
2018 – 2	7	7,5	7	4,5	7	4	3
2019 – 1	5	8	7	5	7	2	6
2019 – 2	6	9	6	2	3	5	10
2021 – 1	4	10	9	3	7	2	7
2021 – 2	5	8	12	4	5	1	7
Som	32	49	50	25	38	16	36
% van totaal	13,01%	19,92%	20,33%	10,16%	15,45%	6,50%	14,63%

Tabel 16: Aantal punten per onderwerp per examen

Aantal punten per onderwerp							
Examen	1	2	3	4	5a	5b	6
2018 – 1	6	12,5	9	6,5	9	2	3
2018 – 2	7	12,5	7	4,5	7	4	3
2019 – 1	5	13	7	5	7	2	6
2019 – 2	6	13	6	2	3	6	10
2021 – 1	5	13	9	4	7	2	8
2021 – 2	5	12	12	4	5	1	7
Som	34	76	50	26	38	17	37
% van totaal	12,23%	27,34%	17,99%	9,35%	13,67%	6,12%	13,31%

7.2 Conclusies

7.2.1 Niveau van de teksten

Met betrekking tot het niveau van de teksten kan voor de onderzochte examens binnen de gebruikte methodiek het volgende gesteld worden.

Zoals te zien in [tabel 12](#), is er een voorkeur voor C1 teksten boven alle andere, met 37 teksten – of 52% – van het totaal. Daarna waren er meer C2 teksten, 29 of 41%. Hiermee lijken de examens moeilijker te zijn dan het beoogde B2/C1 niveau, dat naar boven kwam in eerder vergelijkingsonderzoek (Noijons & Kuijper, 2006; Fasoglio, de Jong, Trimbos, & Beeker, 2015). Het kan zijn dat dit verschil aan de methode van niveaubeoordeling binnen dit onderzoek heeft gelegen, of eerdere inschatting waren toch te laag, zoals Feskens, Keuning, van Til & Verheyen (2014; Verheyen, van Til, Keuning, & Feskens, 2015) onderzochten.

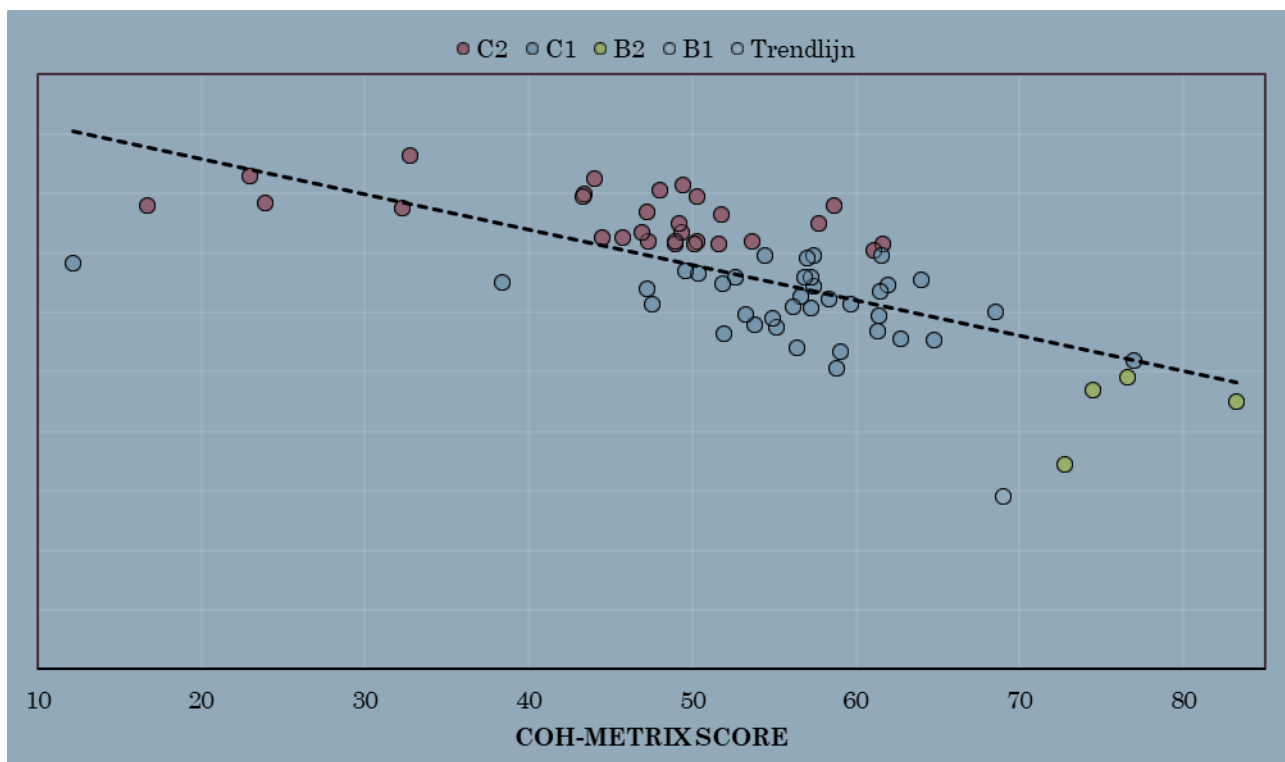
Het grootste gedeelte van de teksten kreeg een Coh-Metrix score tussen de 46 en 65 (52 teksten, 73%, [tabel 14](#)).

Uit tabel 14 valt ook een Coh-Metrix bereik op te maken waarbinnen het waarschijnlijk is dat een tekst een C1 of C2 niveau inschatting kan krijgen. Van de 32 teksten met een Coh-Metrix score tussen de 12 en 52 kregen er 24 een C2 inschatting tegenover 8 een C1. Van de 34 teksten met een Coh-Metrix score tussen de 52 en 69 kregen er 5 een C2 inschatting en 28 een C1. In tabel 17 is dit naast elkaar zichtbaar gemaakt.

Tabel 17: Waarschijnlijk ERK-niveau op basis van Coh-Metrix score

Coh-Metrix score	ERK-niveau
10 – 52	C2
52 – 69	C1
69 – 84	B2

Globaal valt te stellen dat hoe hoger de Coh-Metrix score, hoe waarschijnlijker het is dat het ERK-niveau lager zal zijn, wat klopt met de verwachtingen. Dit is weergegeven in grafiek 2 door alle teksten op zowel Coh-Metrix scores als ERK-niveau te indiceren.



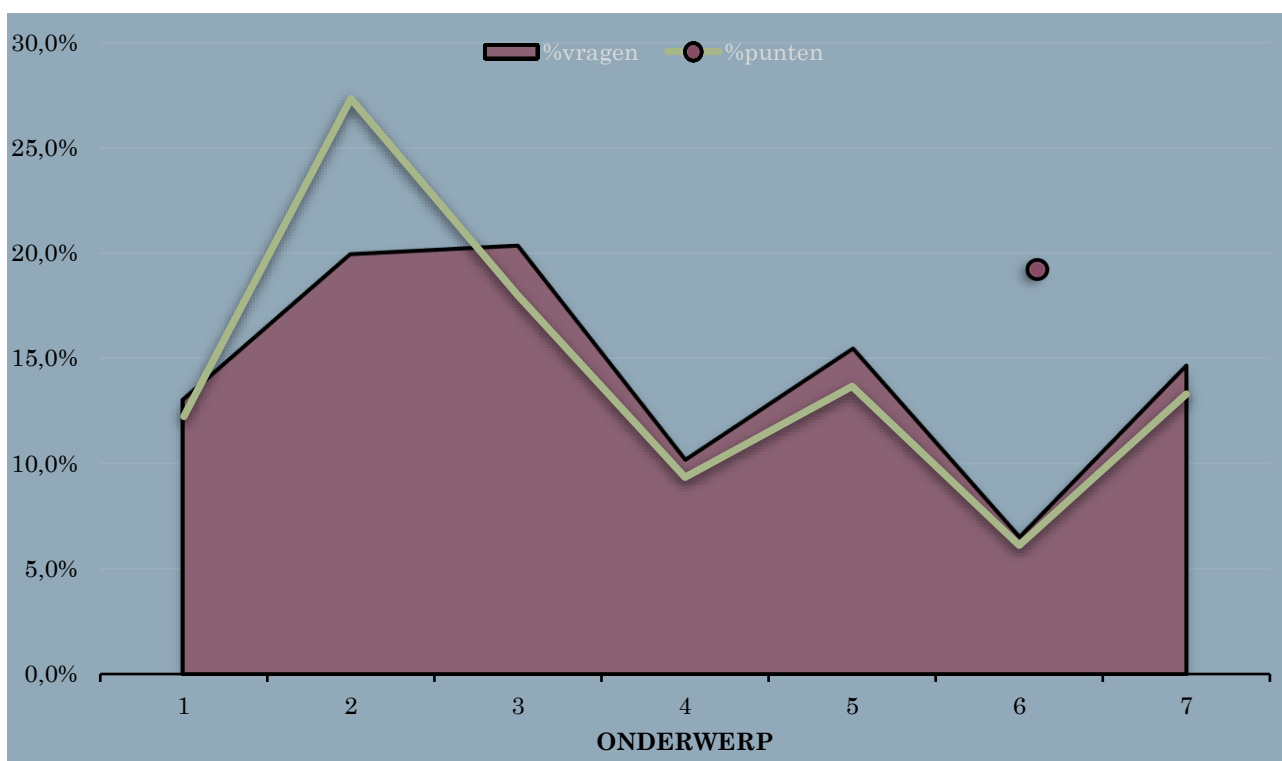
Grafiek 2: Verdeling van teksten volgens ERK-Niveau en Coh-Metrix score

7.2.2 Verdeling van onderwerpen

In tabel 15 en 16 is al zichtbaar dat er geen evenredige verdeling over de onderwerpen is. Dit is geheel in lijn met zowel de constructie opdracht die Cito krijgt (College voor Toetsen en Examens, 2014) als wat het OECD nastreefde (OECD, 2019). In grafiek 3 is te zien wat de verhouding is tussen welk percentage van de vragen binnen ieder onderwerp viel en welk percentage van de punten dit opleverde. Hier springt onderwerp 2 (*Represent literal meaning // De betekenis van belangrijke elementen van een tekst aangeven // Woordenschat*) uit met een significant groter aandeel in de punten (27,3%) dan in het aantal vragen (19,9%).

Met betrekking tot het percentage vragen, krijgt naast onderwerp 2, onderwerp 3 (*Integrate and generate inferences // De hoofdgedachte van een tekst(gedeelte) aangeven // Gevolgtrekking*) de meeste aandacht, 20,3% van de 246 vragen. Met 18% van de punten is het geen extra uitschieter.

Dit impliceert dat het vwo-examen Engels met betrekking tot de benoemde onderwerpen vooral nadruk legt op woordenschat en gevolgtrekking, waarbij woordenschat extra beloond wordt.



Grafiek 3: Verhouding tussen %vragen per onderwerp en %punten per onderwerp

7.2.3 Het examen

Tabel 18 geeft de verdeling van vragen en punten over de verschillende ERK-niveaus.

Tabel 18: Aantal vragen en punten per ERK-niveau

ERK-Niveau	Aantal vragen	% van totaal	Aantal punten	% van totaal
B1	3	1,22%	4	1,48%
B2	12	4,88%	17	6,10%
C1	115	46,75%	128	46,06%
C2	116	47,15%	129	46,37%
Totaal	246	100%	278	100%

Volgens de constructie opdracht (College voor Toetsen en Examens, 2014) zou 80% van de vragen op B2 niveau moeten zijn en 20% op C1. Dit lijkt niet het geval te zijn. Nu blijft het zo dat de ERK-niveaus nogal breed zijn. Met de Coh-Metrix is het mogelijk om meer en fijner onderscheid te maken. Tabel 19 geeft de verdeling van aantal vragen en punten weer per 5 puntsinterval van de Coh-Metrix score.

Op basis van de hoeveelheid teksten (tabel 14) is het logisch dat het bereik van 46 – 50 (meest waarschijnlijk C2) eruit springt met meer vragen (72) en punten (79). Hier waren immers ook de meeste teksten (17). Het valt te verwachten dat het bereik 56 – 60 (meest waarschijnlijk C1), met 15 teksten, vergelijkbaar zou zijn, maar met 47 vragen en 53 punten, is dat niet zo.

Tabel 19: Aantal vragen en punten per Coh-Metrix bin

Coh-Metrix score	Aantal vragen	% van totaal	Aantal punten	% van totaal
11 – 15	1	0,4%	1	0,4%
16 – 20	2	0,8%	3	1,1%
21 – 25	4	1,6%	5	1,8%
26 – 30	-	-	-	-
31 – 35	6	2,4%	7	2,5%
36 – 40	2	0,8%	2	0,7%
41 – 45	18	7,3%	20	7,2%
46 – 50	72	29,3%	79	28,4%
51 – 55	40	16,3%	43	15,5%
56 – 60	47	19,1%	53	19,1%
61 – 65	32	13,0%	36	12,9%
66 – 70	6	2,4%	7	2,5%
71 – 75	3	1,2%	4	1,4%
76 – 80	8	3,3%	11	4,0%
81 – 85	5	2,0%	7	2,5%
Totaal	246	100%	278	100%

7.2.4 Combinatie tabel

Het is ook mogelijk om een groot gedeelte van de informatie hierboven te verwerken in één examen-specificatietabel. Dit is gedaan in tabel 20 op de volgende pagina. Deze tabel geeft een combinatie van het totaal aantal te verdienen punten per onderwerp en niveau met een gemiddelde aantal punten dat het onderwerp in totaal oplevert.

Door kleuren te benutten is het in een oogopslag zichtbaar waar de nadruk met betrekking tot zowel onderwerpen als niveau in de onderzochte examens lag. Blauwe vlakken betekenen dat die combinatie van onderwerp en niveau meer dan gemiddeld + één standaarddeviatie ($=5 + 5,32$) voor kwam. Voor paarse vlakken geldt dat die combinatie meer dan gemiddeld + tweemaal de standaarddeviatie voor kwam.

Groene vlakken hebben geen extra betekenis, alleen dat die combinatie van onderwerp en niveau voorkwam. Zo waren er bij vijf van de zes examens twee punten te verdienen binnen onderwerp 1 (*Access and retrieve information // Aangeven welke informatie relevant is // Leesstrategieën & tekststructuur*) op ERK-niveau C2 met een Coh-Metrix waarde tussen de 41 en 50, voor een totaal van 10. In het verlengde hiervan, waren er geheel geen vragen binnen onderwerp 1 gesteld bij teksten op ERK-niveau C1 met een Coh-Metrix waarde tussen de 10 en 49.

Tabel 20: Combinatie tabel van totaal te verdienen aantal punten en gemiddelde

				Moeilijkheidsgraad van de tekst (ERK & Coh-Metrix 3.0)									
Onderwerp				C2			C1			B2	B1		
OECD	Domein A		Component	10 - 40	41 - 50	51 - 65	10 - 49	50 - 60	61 - 80	70 - 90	65 - 70	Punten	%
1	Access and retrieve information	Aangeven welke informatie relevant is	Leesstrategieën & tekststructuur	1	10	3		13	3	3	1	5,67	12,23%
2	Represent literal meaning	De betekenis van belangrijke elementen van een tekst aangeven	Woordenschat	8	19,5	9	2	12,5	12	10	3		
3	Integrate and generate inferences	De hoofdgedachte van een tekst(gedeelte) aangeven	Gevolgtrekking	2	15	5	2	19	7			8,33	17,99%
4		Relaties tussen delen van een tekst aangeven	Tekststructuur	1	10,5	3	1	6,5	2	2		4,33	9,35%
5a			Cloze: woordenschat, syntaxis en semantiek	1	12	2	1	18	4			6,33	13,67%
5b			Cloze: woordenschat, syntaxis en semantiek, tekststructuur	1	7	1	1	4	3			2,83	6,12%
6	Reflect on content and form	Conclusies trekken met betrekking tot intenties, opvattingen en gevoelens van de auteur	Gevolgtrekking	1	12	5		10	7	2		6,17	13,31%
Aantal punten				2,50	14,33	4,67	1,17	13,83	6,33	2,83	0,67	46,33	
percentage v.d. punten				5,43%	30,92%	10,02%	2,55%	29,80%	13,71%	6,10%	1,48%		

Geheel rechts staat hoeveel punten er binnen elk onderwerp gemiddeld te verdienen waren. Onderwerp 1 levert gemiddeld 5,67 punten op binnen een examen, onderwerp 2 12,67, etc. Daarnaast staat nog het percentage van het totaal aantal gemiddelde punten dat dit is.

Geheel onderaan staat hoeveel punten er gemiddeld binnen het betreffende niveau te verdienen waren. Zo was er op niveau B2 gemiddeld 2,83 punten te verdienen, wat 6,10% van het totaal is.

Uit deze tabel is ook op te maken dat op C2 niveau er meer nadruk ligt op woordenschat, en op C1 niveau meer op gevolgtrekking en de *cloze* opgaven. Woordenschat levert als geheel het hoogste percentage punten op, wat impliceert dat het examen meer op dit onderwerp toetst, dan op tekstbegrip. Door de spreiding die er is, is het wel noodzakelijk voor een leerling om meerdere onderwerpen goed te maken, wil deze een voldoende behalen.

7.3 Resultaten hoofdvraag

Aan het begin van dit onderzoek werd de volgende vraag gesteld:

1. Toetst het examen vwo Engels met voldoende inhoudsvaliditeit de leesvaardigheid van leerlingen?

In de conclusies hierboven staan, verspreid, antwoorden, maar ten behoeve van de helderheid worden ze hier samengevat.

1. De onderzochte examens hebben een ongelijke verdelingen over de onderwerpen van domein A (afbeelding 1). Vragen richting het aangeven van de betekenis van belangrijke elementen van een tekst (onderwerp 2) krijgen meer punten toegekend dan andere onderwerpen. Daarna wordt er vooral om gevolgtrekking gevraagd. Wanneer deze verhoudingen naast de OECD suggesties (OECD, 2019, p. 42) gelegd worden, ontbreken niet alleen vragen waarbij meerdere teksten gelezen moeten worden, maar wordt er ook niet voldoende aandacht besteedt aan gevolgtrekking en het beoordelen van de kwaliteit van teksten. Op basis hiervan kan worden gesteld dat het eindexamen Engels onvoldoende inhoudsvaliditeit heeft.

8. Discussie en aanbevelingen

8.1 Kanttekeningen

8.1.1 Bij het examen

Door beperkingen binnen de vorm van het examen (collectieve afname) is het meten van voorkennis, leesmotivatie, leesstrategieën en andere factoren niet mogelijk. Daarnaast is het niet duidelijk wat met de opgedane kennis naderhand gedaan wordt. Derhalve is het examen haast bij voorbaat al niet geschikt voor het meten van leesvaardigheid volgens OECD-definitie. Vooral het niet rekening houden met voorkennis is een ernstig bezwaar volgens Willingham (2009), die aangeeft dat juist voorkennis bepalend is voor de mate waarin je een tekst kunt begrijpen, wat ook ondersteund wordt ander onderzoek (Kamalski, Sanders, & Lentz, 2008), en nieuwe informatie kunt verwerken en behouden. Hierop voortbordurend stelt Ringenaldus (2018) dat het beter zou zijn wanneer voor het examen MVT bepaalde onderwerpen vooraf aangekondigd zouden worden. Niet alleen om voorkennis op te doen, maar ook om bijbehorende woordenschat op te doen.

Een andere grote beperking is het ERK zelf. Dit maakt al aannames in de definitie van leesvaardigheid, waar Cito vervolgens door “gebonden” is. Het is opmerkelijk dat hier tekstsoorten (correspondentie en instructies) gecombineerd worden met leesdoelen (oriënteren, informeren, vrijetijdsbesteding), waarbij allerlei andere, en zelfs eerdergenoemde, tekstsoorten buiten beschouwing worden gelaten, evenals andere mogelijke doelen om te lezen. Kwakernaak (2012) zette ook al zijn vraagtekens bij dit verschijnsel in een poging te achterhalen welke opvatting van leesvaardigheid er achter de centraal schriftelijke examens zit (Kwakernaak & de Vries, 2006). Wanneer het fundament waarop het Cito moet bouwen al scheuren heeft, is het wellicht niet zo vreemd dat het huis wankelt.

Ook het voorschrift in vraagsoorten waar Cito door gebonden is (College voor Toetsen en Examens, 2014), is opmerkelijk. Waarom specifiek 60% meerkeuze vragen? Al sinds 1984 is duidelijk dat de manier van toetsen van invloed is op het resultaat van de leerling (Shohamy, 1984). Een andere manier vraagstelling, bijvoorbeeld een sorteertaak (Kamalski, Sanders, Lentz, & van den Bergh, 2005), kan beter zijn om leesvaardigheid in het Engels te toetsen, of een aanpassing in de manier van *cloze* toetsen die aantoonbaar beter begrip toetst (Kleijn, Pander Maat, & Sanders, 2019).

Klingner (2004) pleit zelfs voor een grote variatie aan vraagstellingen en taken om begrip te toetsen, al zal ook dit in een examen setting niet makkelijk of praktisch haalbaar zijn.

8.1.2 Bij dit onderzoek

Ondanks dat er gezocht is naar een zo objectief mogelijke manier van beoordelen van tekstniveau en het classificeren van vragen binnen onderwerpen blijft de exacte invulling van de specificatietabellen en indelen van vraagsoorten een inschatting van één persoon. Om de validiteit van dit onderzoek zo veel mogelijk te garanderen is er voor het beoordelen van het tekstniveau gebruik gemaakt van een veelvuldig, wetenschappelijk getoetst instrument (de Coh-Metrix) en meerdere websites ter onderlinge verificatie, waarbij er meerdere wetenschappelijke onderzoeken ter ondersteuning gebruikt zijn. Hoewel de keuzes gemaakt zijn door de onderzoeker, zijn dit geen arbitrair gekozen middelen en de verantwoording is terug te lezen in [hoofdstuk 4](#).

Het is de verwachting van de auteur dat wanneer een ander dit onderzoek zou herhalen, er in de indeling van de vragen geen significante verschillen zullen optreden. Vooral omdat de omschrijving van vraagstelling binnen onderwerpen door de OECD (2019) helder genoeg is dat een één-op-één vergelijking tussen het CITO examen en de OECD kaders weinig aan interpretatie vanuit de lezer vraagt. Bijvoorbeeld bij een vraag als *“What is ‘dirty’ about the word ‘gentrifier’, according to the text? According to critics it stands for”*⁹ is het duidelijk dat er naar de betekenis van een woord binnen context gevraagd wordt, zonder dat verder tekstbegrip noodzakelijk is.

Wat betreft het inschatten van het ERK-niveau van de teksten, zou er wel een verschil op kunnen treden. De ERK-niveaus zijn dermate onspecifiek, dat een tekst die op de grens van B2 en C1 ligt door verschillende beoordelaars anders geïndiceerd kan worden. Zelfs wanneer meerdere lezers dit zouden doen, is overeenstemming verre van gegarandeerd en het bepalen van een “gemiddelde moeilijkheid” is vervolgens aan de onderzoekers en potentieel arbitrair. Zo ontstaat met het weglaten van de afronding die binnen dit onderzoek gedaan is bij het berekenen van het gemiddelde ERK-niveau een andere verdeling tussen B2, C1 en C2 teksten. Er zouden dan negen B2 teksten en elf

⁹ examen 2021, tijdvak 1, tekst 4, vraag 6

C1 teksten meer zijn, tegenover 21 C2 teksten minder. Met deze verdeling komt het examen nog altijd niet aan de beoogde 80/20 verhouding tussen B2 en C1.

Er is ook niet gekeken of de vraagstelling voldoet aan voorwaarden die gesteld worden aan een goede meerkeuze vraag (van Berkel, Bax, & Joosten-ten Brinke, 2017), zoals geen onnodige negatieve formulering, geen betrekking op details, of geen overlapping in alternatieven. Terwijl dit juist een verklaring zou kunnen zijn waarom specifieke vragen significant slechter gemaakt zijn binnen een examen, of waarom docenten van mening zijn dat begrip onvoldoende getoetst wordt. Eveneens is er niet gekeken naar de volgorde waarin vraagsoorten of tekstniveaus elkaar opvolgen en of dit mogelijk van invloed is op het resultaat.

Er is ook geen vergelijking gemaakt met andere leesvaardigheidstoetsen, zoals de Cambridge, TOEFL, of ACT toetsen, die ook middels meerkeuze vragen tekst- en taalbegrip toetsen en, met uitzondering van ACT, aangeven zich aan ERK-niveaus te houden.

8.2 Aanbevelingen

8.2.1 Voor het CvTE en Cito

Op basis van de aangehaalde literatuur, de analyse van de examens en het opstellen van de specificatietabellen zouden het CvTE en Cito baat kunnen vinden in de volgende aanbevelingen.

1. Door vooraf een specificatietabel op te stellen, wordt het noodzakelijk om vast te leggen wat het niveau van de teksten is en, bij voorkeur, hoe dit bepaald wordt. Daarnaast zullen de onderwerpen nauwkeuriger omschreven moeten worden, wat zal helpen om vraagstelling gericht te maken.
 - a. Om te helpen tekstniveau te bepalen kan een eigen systeem opgesteld worden om dit zo objectief mogelijk te beoordelen. Als (tijdelijk) alternatief kan ook een wetenschappelijk onderbouwde methode, zoals de Coh-Matrix, benut worden.
 - b. Wat betreft onderwerpen/vraagstelling kan afgestapt worden van ERK-omschrijvingen die onvoldoende specifiek en toetsbaar zijn. Door over te stappen naar OECD omschrijvingen, waar heldere, onderbouwde eisen en voorbeeld opdrachten beschikbaar zijn, is er ook een extern document dat kan helpen bij het meten van de kwaliteit van het examen.

2. Gezien het belang van achtergrondkennis bij leesvaardigheid zou er een betere voorbereiding getroffen kunnen worden wanneer (een selectie van) onderwerpen vooraf aangekondigd zouden worden.
 - a. Wanneer tijdens de lessen al aandacht wordt besteed aan het onderwerp en bijbehorende woordenschat, zullen teksten minder vaak aangepast of verduidelijkt hoeven te worden. Het is dan mogelijk om 100% authentieke teksten te gebruiken.
3. Een grotere variatie in taken en teksten kan ook helpen om de toetsing van leesvaardigheid betrouwbaarder te maken.
 - a. Hierbij kan gedacht worden aan taken waarbij meerdere teksten vergeleken moeten worden of specifieke informatie gevonden moet worden in één uit meerdere teksten.
 - b. Naast taken gericht op meerdere teksten, kan er ook gebruik worden gemaakt van taken gericht op het beoordelen van teksten.
 - c. Ook zouden meer tekstsoorten benut kunnen worden, wat de mogelijkheden, binnen het vooraf vastgestelde onderwerp, uitbreidt.
4. Wanneer het vast percentage vraagsoorten wordt losgelaten, is er meer ruimte om vragen te construeren die begrip toetsen.
 - a. Om toch houvast te bieden kan er een verdeling gemaakt worden over de onderwerpen (zie adviespunt 1).

8.2.2 Voor verder onderzoek

Door de beperkingen van dit onderzoek is het aannemelijk dat vervolgonderzoek gedaan kan worden dat hier specifiek op gericht is.

1. Om tekstniveaus goed en objectief te kunnen beoordelen op ERK-niveau is te adviseren om precieze eisen en voorwaarden te omschrijven. Vervolgens zou dit getoetst kunnen worden door experts, gebruikers en middels opdrachten.
2. De (*multiple choice*) vragen van de vwo-examens zouden nader onderzocht kunnen worden, onder andere in welke mate ze voldoen aan de bijvoorbeeld de voorwaarden zoals gesteld door van Berkel, Bax, & Joosten-ten Brinke (2017).
 - a. Er kan een verklaring gezocht worden voor waarom sommige vragen, op basis van Cito data, meer fout gemaakt zijn dan goed.

- b. De verhouding tussen de vraagsoorten kan getoetst worden. Is dit werkelijk 60% aan “algemeen” meerkeuze en *cloze* en 40% andersoortig (combinatie, ordening, open, citeer) en heeft dit laatste een evenwichtige verdeling?
- 3. Wat betreft tekstsoorten, -onderwerpen en manieren van begrip toetsen kan er een vergelijking gemaakt worden met andere, gestandaardiseerde, leesvaardigheid examens, zoals die van het Britse Cambridge, de Amerikaanse ACT, of het Franse staatsexamen Engels (BAC).

Bibliografie

- Afflerbach, P., Cho, B.-Y., Crassas, M. E., & Kim, J.-Y. (2018). Best Practices in reading assessment. In L. M. Morrow, & L. B. Gambrell, *Best Practices in Literacy Instruction* (pp. 337-359). New York: Guilford.
- Ahmed, Y., Frances, D. J., York, M., Fletcher, J. M., Barnes, M., & Kulesz, P. (2016). Validation of the direct and inferential mediation (DIME) model of reading comprehension in grades 7 through 12. *Contemporary Educational Psychology*, 68-82.
- Benjamin, R. G. (2012). Reconstructing Readability: Recent Developments and Recommendations in the Analysis of Text Difficulty. *Educational Psychology Review*, 63-88.
- Biggs, J. (1996). Enhancing Teaching through Constructive Alignment. *Higher Education*, 347-364.
- Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *The Taxonomy of educational objectives, handbook I: The Cognitive domain*. New York: David McKay Co., Inc.
- Bogaerds-Hazenberg, S. T., Evers-Vermeul, J., & van den Bergh, H. (2020). A Meta-Analysis on the Effects of Text Structure Instruction on Reading Comprehension in the Upper Elementary Grades. *Reading Research Quarterly*, 435-462.
- Cambridge University Press. (2015). *English Grammar Profile Online*. Retrieved from EnglishProfile: <https://www.englishprofile.org/english-grammar-profile/egp-online>
- Cathoven. (2022, 8 3). *CEFR Checker*. Retrieved from Cathoven: <https://www.cathoven.com/en/cefr-checker/>
- College voor Toetsen en Examens. (2014). Vakspecifieke deel MVT examens havo/vwo. *Constructieopdracht Talen 2014*. Utrecht: College voor Toetsen en Examens.
- College voor Toetsen en Examens. (2019). *Moderne Vreemde Talen VWO - Syllabus Centraal Examen 2021 Arabisch, Duits, Engels, Frans, Russisch, Spaans, Turks*. Utrecht: College voor Toetsen en Examens. Retrieved from <https://www.examenblad.nl/examen/engels-vwo-2/2021>

- Cromley, J. G., & Azevedo, R. (2007). Testing and Refining the Direct and Inferential Mediation Model of Reading Comprehension. *Journal of Educational Psychology*, 311-325.
- Crossley, S. A., Skalicky, S., & Dascalu, M. (2019). Moving beyond classic readability formulas: new methods and new models. *Journal of Research in Reading*, 541-561.
- Cutting, L. E., & Scarborough, H. S. (2006). Prediction of Reading Comprehension: Relative Contributions of Word Recognition, Language Proficiency, and Other Cognitive Skills can depend on how Comprehension is Measured. *Scientific Studies of Reading*, 277-299.
- Drost, M., & Verra, P. (2012). *Handboek RTTI*. Bodegraven: Docentplus.
- Duke, N. K., Pearson, P., Strachan, S. L., & Billman, A. K. (2011). Essential Elements of Fostering and Teaching Reading Comprehension. In S. J. Samuels, & A. Farstrup, *What research has to say about reading instruction* (4e ed. ed., pp. 51-93). Newark, NJ: International Reading Association.
- Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *COTAN Beoordelingssysteem voor de kwaliteit van tests*. Amsterdam: NIP/COTAN.
- Fasoglio, D., de Jong, K., Trimbos, B., & Beeker, A. (2015). *Taalprofielen 2015: herziene versie van Taalprofielen 2004*. Enschede: SLO.
- Feskens, R., Keuning, J., van Til, A., & Verheyen, R. (2014). *Prestatiestandaarden voor het ERK in het eindexamenjaar. Een internationaal ijkingsonderzoek*. Arnhem: Cito.
- Flesch, R. (1949). *The Art of Readable Writing*. New York: Harper & Row.
- Fletcher, J. M. (2006). Measuring Reading Comprehension. *Scientific Studies of Reading*, 323-330.
- Francis, D. J., Snow, C. E., August, D., Carlson, C. D., Miller, J., & Iglesias, A. (2006). Measures of Reading Comprehension: A Latent Variable Analysis of the Diagnostic Assessment of Reading Comprehension. *Scientific Studies of Reading*, 301-322.

- Galton, F. (1886). Regression Towards Mediocrity in Hereditary Stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 246-263.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, Reading, and Reading Disability. *Remedial and Special Education*, 6-10.
- Hacquebord, H., & Lenting-Haan, K. (2012). Kunnen we de moeilijkheid van teksten meten? Naar concrete maten voor de referentieniveaus. *Levende Talen Tijdschrift*, 14-24.
- Houtveen, A., van Steensel, R., & de la Rie, S. (2019). *De vele kanten van leesbegrip: Literatuurstudie naar onderwijs in begrijpend lezen in opdracht van het Nationaal Regieorgaan Onderwijsonderzoek en de Inspectie van het Onderwijs*. Rotterdam: Erasmus Universiteit.
- Kamalski, J., Sanders, T., & Lentz, L. (2008). Coherence Marking, Prior Knowledge, and Comprehension of Informative and Persuasive Texts: Sorting Things Out. *Discourse Processes*, 323-345.
- Kamalski, J., Sanders, T., Lentz, L., & van den Bergh, H. (2005). Hoe kunt je het beste meten of een leerling een tekst begrijpt? Een vergelijkend onderzoek naar vier methoden. *Levende Talen Tijdschrift*, 3-9.
- Kauchak, D., Leroy, G., & Hogue, A. (2017). Measuring Text Difficulty Using Parse-Tree Frequency. *Journal of the Association for Information Science*, 2088-2100.
- Kendeou, P., McMaster, K. L., & Christ, T. J. (2016). Reading Comprehension: Core Components and Processes. *Reading, Writing, and Language*, 62-69.
- Kerndoelen onderbouw voortgezet onderwijs*. (2010, 09 17). Retrieved from Rijksoverheid.nl:
<https://www.rijksoverheid.nl/documenten/besluiten/2010/09/17/kerndoelen-onderbouw-voortgezet-onderwijs>
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge university press.
- Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling, & C. Hulme, *The Science of Reading: A Handbook* (pp. 211-226). Oxford: Blackwell Publishing Ltd.

- Kleijn, S., Pander Maat, H., & Sanders, T. (2019). Cloze testing for comprehension assessment: The HyTeC-cloze. *Language Testing*, 553-572.
- Klingner, J. K. (2004). Assessing Reading Comprehension. *Assessment for Effective Intervention*, 59-70.
- Knevel, R. (2013). Taxonomiën zijn hot...en handig! *Toets!*, 1-10.
- Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*, 212-218.
- Kuhlemeier, H., & van Til, A. (2016). *Toetsen op School Voortgezet onderwijs*. Arnhem: Stichting Cito Instituut voor Toetsontwikkeling.
- Kwakernaak, E. (2010). Het ERK Hinkt. *Levende Talen Magazine*, 4-8.
- Kwakernaak, E. (2012). Opvattingen over leesvaardigheid. Kwakernaak reageert op Hoeflaak. *Levende Talen Magazine*, 23-26.
- Kwakernaak, E., & de Vries, E. (2006). School- en eindexamens Duits: De CEVO-syllabus. *Levende Talen Magazine*, 93(7), 39-41.
- Maassen, N., & den Otter, D. (2015). *Eerste hulp bij Toetsen: Grip op toetskwaliteit*. Enschede: RCEC, Universiteit Twente.
- Maassen, N., den Otter, D., Wools, S., Hemker, B., Straetmans, G., & Eggen, T. (2014). *Kwaliteit van toetsen binnen handbereik. Een reviewstudie van onderzoek en onderzoeksresultaten naar de kwaliteit van toetsen*. RCEC, Cito.
- McNamara, D. S., & Graesser, A. C. (2012). Coh-Metrix: An automated tool for theoretical and applied natural language processing. In P. M. McCarthy, & C. Boonthum-Denecke, *Applied Natural Language Processing: Identification, Investigation and Resolution* (pp. 188-205). IGI Global.
- Mimeau, C., Ricketts, J., & Deacon, S. (2018). The Role of Orthographic and Semantic Learning in Word Reading and Reading Comprehension. *Scientific Studies of Reading*, 384-400.
- Moelands, H. (2017). *Toetsconstructieproces in 8 stappen. Een handleiding voor toetsconstructeurs met informatie over de acht stappen van het toetsconstructieproces*. Cito.

- Nahatame, S. (2021). Text Readability and Processing Effort in Second Language Reading: A Computational and Eye-Tracking Investigation. *Language Learning*, 1004-1043.
- Natova, I. (2019). Estimating CEFR reading comprehension text complexity. *The Language Learning Journal*, 699-710.
- Nelson, J., Perfetti, C., Liben, D., & Liben, M. (2012). *Measures of Text Difficulty: Testing their Predictive Value for Grade Levels and Student Performance*. New York: NY: Student Achievement Partners.
- Noijons, J., & Kuijper, H. (2006). *De koppeling van de centrale examens leesvaardigheid moderne vreemde talen aan het Europees Referentiekader*. Arnhem: Cito.
- North, B., Goodier, T., & Piccardo, E. (2018). *Common European Framework of Reference for Languages: learning, teaching, assessment*. Council of Europe.
- Oakhill, J. V., Cain, K., & Carsten, E. (2019). Reading Comprehension and Reading Comprehension Difficulties. In D. A. Kilpatrick, R. M. Joshi, & R. K. Wagner, *Reading Development and Difficulties: Bridging the Gap Between Research and Practice* (pp. 83-115). Cham, Zwitserland: Springer International Publishing.
- OECD. (2019). *PISA 2018 Assessment and Analytical Framework*. Paris: OECD Publishing.
- O'Reilly, T., Wang, Z., & Sabatini, J. (2019). How Much Knowledge is too Little? When a Lack of Knowledge Becomes a Barrier to Comprehension. *Psychological Science*, 1344-1351.
- Perfetti, C. (1999). Comprehending written language: a blueprint of the reader. In C. M. Brown, & P. Hagoort, *The Neurocognition of Language* (pp. 167-208). Oxford: Oxford University Press.
- Perfetti, C., & Stafura, J. (2015). Comprehending implicit meanings in text without making inferences. In E. J. O'Brien, A. E. Cook, & R. F. Lorch, Jr, *Inferences during Reading* (pp. 1-18). Cambridge: Cambridge University Press.
- Ricketts, J., Bishop, D. V., Pimperton, H., & Nation, K. (2011). The Role of Self-Teaching in Learning Orthographic and Semantic Aspects of New Words. *Scientific Studies of Reading*, 47-70.

- Ringenaldus, M. (2018, 6 1). *Hoe valide is het eindexamen Duits?* Retrieved from Didactief: <https://didactiefonline.nl/artikel/hoe-valide-is-het-eindexamen-duits>
- Romiszowski, A. J. (1981). *Designing Instructional Systems. Decision Making in Course Planning and Curriculum Design*. London: Kogan Page Ltd.
- Sanders, P. (2013). *Toetsen op School*. Arnhem: Cito.
- Sanders, P., & Brouwer, A. (2015). *RCEC Beoordelingssysteem voor de Kwaliteit van Studietoetsen en Examens*. RCEC.
- Scarborough, H. S. (2001). Connecting Early Language and Literacy to Later Reading (Dis)abilities: Evidence, Theroy, and Practice. In S. B. Neuman, & D. K. Dickinson, *Handbook of Early Literacy Research, Volume 1*. New York: The Guilford Press.
- Scheerens, J., Brouwer, A., Sanders, P., Veldkamp, B., & van der Vegt, A. (2019). *Fundamentele vragen over examens en toetsing*. Utrecht: Oberon.
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The Percentage of Words Known in a Text and Reading Comprehension. *The Modern Language Journal*, 26-43.
- Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing*, 147-170.
- Sijtsma, K. (2011). Studietoetsen: zinvolle inhouds of betrouwbare score? *Examens*, 5-8.
- SLO. (2020, Augustus 6). *Criteria voor de kwaliteit van een toets*. Retrieved from SLO: <https://www.slo.nl/handreikingen/havo-vwo/shared/kwaliteit-schoolexamen/criteria-kwaliteit/>
- Staatscourant van het Koninkrijk der Nederlanden*. (2007). Retrieved from Overheid.nl: <https://zoek.officielebekendmakingen.nl/stcrt-2007-111-p16-SC80845.html>
- Stanley, J., & Williamson, T. (2017). Skill. *Noûs*, 713-726.
- Stokking, K. (2001). Leren van Vaardigheden. In A. Wald, & J. van der Linden, *Leren in Perspectief* (pp. 15 - 30). Leuven - Apeldoorn: Garant.
- Teunis, B. (2022). *Actualisatie kerndoelen en examenprogramma's*. Retrieved from SLO: <https://www.slo.nl/thema/meer/actualisatie-kerndoelen-examenprogramma/>

- Tillema-Kortman, M. (2005). De T2-specificiteit van het Centraal Schriftelijk Eindexamen Engels onderzocht. *Levende Talen Tijdschrift*, 18-27.
- Trapman, M., van Gelderen, A., van Schooten, E., & Hulstijn, J. (2017). Reading Comprehension Level and Development in Native and Language Minority Adolescent Low Achievers: Roles of Linguistic and Metacognitive Knowledge and Fluency. *Reading & Writing Quarterly*, 239-257.
- van Berkel, H., & Draaijer, S. (2011). Gids voor toetsontwikkeling. *Examens*, 1-8.
- van Berkel, H., Bax, A., & Joosten-ten Brinke, D. (2017). *Toetsen in het hoger onderwijs* (4e, herziene druk ed.). Houten: Bohn Stafleu van Loghum.
- van den Broek, P., & Espin, C. A. (2012). Connecting Cognitive Theory and Assessment: Measuring Individual Differences in Reading Comprehension. *School Psychology Review*, 315-325.
- van Gelderen, A. (2018). *Begrijpend lezen: wat is dat? De componenten die een rol spelen bij begrijpend lezen*. Enschede: SLO.
- van Haperen, T. (2022, juni). *Brief aan CvTE en CITO over slecht examen economie vwo 2022*. Retrieved from [tonvanhaperen.com: https://tonvanhaperen.com/2022/06/10/brief-aan-cvte-en-cito-over-slecht-examen-economie-vwo-2022/](https://tonvanhaperen.com/2022/06/10/brief-aan-cvte-en-cito-over-slecht-examen-economie-vwo-2022/)
- Velleman, E., & van der Geest, T. (2014). Online test tool to determine the CEFR reading comprehension level of text. *Procedia Computer Science*, 350-358.
- Verheyen, R., van Til, A., Keuning, J., & Feskens, R. (2015). ERK-NIVEAU VAN LEERLINGEN ONDERGEWAARDEERD. Uitkomsten van een internationaal ijkingsonderzoek. *Levende Talen Magazine*, 4-8.
- Waller, D., Horak, T., & Tyrell, S. (2016). Establishing difficulty level consistency across texts in exams at four levels using lexical profiling. *CALR Linguistics Journal*.
- Willingham, D. T. (2009). *Why Don't Students like School? A Cognitive Scientist Answers Questions about how the Mind works and what it means for your Classroom*. San Francisco: Jossey-Bass.

Bijlagen

1. Voorbeelden van vraagstellingen per soort.

1 Voorbeelden van vraagstellingen per soort

1.1 1: Access and retrieve information // Aangeven welke informatie relevant is // Leesstrategieën & tekststructuur

- Aan welke twee voorwaarden moest het LED-tapijt volgens de tekst voldoen om dit succesvol te kunnen? Noteer de twee voorwaarden op het antwoordblad.
- Wat is de onderliggende reden voor Camerons belastingmaatregel volgens Yvonne Roberts? Citeer de eerste twee woorden van de zin uit alinea 4-9 waar het antwoord te vinden is.
- Aan welke voorwaarde genoemd in alinea 8-11 moet worden voldaan om deze aanpak daadwerkelijk te doen slagen? Citeer deze voorwaarde of antwoord in het Nederlands.
- Welk zinsdeel beschrijft het bezwaar dat Gareth Jones tegen dit gebruik heeft? Citeer dit zinsdeel.
- Hoe wordt deze manier van denken eerder in de tekst genoemd? Citeer dit.
- Welke twee factoren verhinderen dit nu, volgens Amanda Casey? Geef antwoord in het Nederlands.
- In welke alinea wordt een voorbeeld gegeven van een film die erin is geslaagd om verandering op gang te brengen? Noteer het nummer van deze alinea op het antwoordblad.
- Komt het aspect van het ontbreken van gezagsregels al in een eerdere zin van de recensie aan de orde? Zo nee, antwoord 'Nee'. Zo ja, citeer de eerste twee woorden van de zin waarin dit naar voren komt.
- Wordt in de recensie een mogelijke reden gegeven? Zo nee, antwoord "Nee". Zo ja, noteer het nummer van de alinea waarin de reden gegeven wordt op het antwoordblad.
- How many explanations for this are given in the text?
- Which of the following sentences from paragraph 6 to 9 best identifies the problem that will be solved when people embrace this new technique?

1.2 2: Represent literal meaning // De betekenis van belangrijke elementen van een tekst aangeven // Woordenschat

- According to the I-person in paragraph 2, which of the following is true of "a profound experience of art"?
- Bepaal voor elke van de onderstaande beweringen of deze wel of niet overeenkomt met de inhoud van alinea 7-9.
- Citeer twee van de drie andere verwoordingen die de schrijver gebruikt voor "avian influenza".
- Geef van de volgende beweringen aan of ze overeenkomen met de inhoud van alinea's 4 en 5.
- Geef van de volgende beweringen aan of ze overeenkomen met de opvattingen of observaties van Gopnik (alinea's 5 tot en met 7).

- Geef van de volgende kwalificaties aan of de voorzitter van Fondation Ipsen eraan moet voldoen.
- Geef van elke van de volgende beweringen aan of deze overeenkomt met de tekst.
- Geef van elke van de volgende beweringen aan of deze wel of niet overeenkomt met de inhoud van alinea 3 en 4.
- Geef van elke van de volgende omschrijvingen aan of daarvan een voorbeeld wordt gegeven in de tekst.
- Geef van elke van de volgende personen en instanties aan of ze volgens Rosamund Urwin tot zondebok zijn gemaakt op Twitter.
- How can this question be rephrased?
- Welk bijkomend voordeel zou deze maatregel die Peter Wilby voorstelt kunnen hebben voor deze buurten?
- Welk woord aan het einde van alinea 3 benadrukt dat Mr Moody onder normale omstandigheden de situatie in het hotel anders had beoordeeld? Noteer dit woord op het antwoordblad.
- What do lines 60-79 make clear about the flight into France?
- What does the example of The Royal Bank of Scotland make clear? (paragraph 5)
- What does the writer try to make clear here?
- What is 'dirty' about the word 'gentrifier', according to the text? According to critics it stands for
- What is it a good example of?
- What is the purpose of the form Sofia has to fill in (paragraph 8)?
- Which of the following becomes clear about moles in this letter?
- Which of the following descriptions refers to the Fondation Ipsen, judging from the text?
- Which of the following does "it" refer to?
- Which of the following is true, judging from paragraph 1?
- Which of the following observations about anthropomorphism is mentioned in this article?
- Which three of the following can be identified as today's blacksmiths, farriers and buggy-whip makers according to paragraphs 5-7?
- Why does Frank Hung bring up his street?
- Leg in je eigen woorden uit wat bedoeld wordt met "this urge".
- Wat adviseert Edward Tikoft om te doen in plaats van innoveren? Geef antwoord in je eigen woorden.
- Which paragraph contains an actual example of such a conflict?
- Wie of wat is dit "monster"? Geef antwoord in het Nederlands of in het Engels.

1.3 3: Integrate and generate inferences // De hoofdgedachte van een tekst(gedeelte) aangeven // Gevolgtrekking

- According to Gopnik, which of the following best describes this model?
- What becomes clear from paragraph 2?
- What does paragraph 3 make clear?

- What does the writer claim in paragraphs 6 and 7?
- What fear is expressed in this text?
- What is George MacDonald's main point?
- What is paragraph 3 meant to make clear?
- What is the EU trying to do in the US, according to Roger Waite?
- What is the intended effect of these two things?
- What is the main point made about the British seaside in paragraph 2?
- What is true according to paragraphs 2-4?
- What message does Anne McElvoy want to convey in paragraph 4?
- Which conclusion does paragraph 4 lead up to?
- Which of the following about Anne McElvoy is true based on paragraphs 1-3?
- Which of the following best illustrates the point the writer wishes to make in paragraph 4?
- Which of the following claims about the I-person is in line with paragraph 3?
- Which of the following is in accordance with paragraph 2?
- Which of the following is in line with paragraph 2?
- Which of the following is the point made in the first paragraph?
- Which of the following is true according to paragraphs 4-5?
- Which of the following is true of Professor Asher Cashdan, according to paragraph 7?
- Which of the following is true, judging from paragraphs 5-8?
- Which of the following quotations expresses the main point made in this article?
- Which of the following quotations illustrates the point made in paragraph 6?
- Which of the following sentences sums up the article?
- Which of the following statements about "12 Years a Slave" is in line with the contents of paragraphs 1 and 2?
- Which of the following statements best substantiates the conclusion in the last paragraph?
- Which of the following statements is in accordance with paragraphs 4 and 5?
- Which of the following statements is in line with Tim Entwistle's letter?
- Which of the headings below is the original subheading of this article?

1.4 4: Integrate and generate inferences // Relaties tussen delen van een tekst aangeven // Tekststructuur

- Geef bij elke plaats aan welke zin daar hoort. Noteer de letter van de zin achter elk nummer op het antwoordblad.
- How can paragraphs 6-8 be characterised best?
- How does Colvin's remark in paragraph 2 relate to his opinion as presented in paragraph 1?
- How does paragraph 2 relate to paragraph 1?
- In welke volgorde stonden bovenstaande alinea's in de tekst?
- In welke volgorde stonden de zinnen in alinea 6? Zet de letters in de juiste volgorde.
- Leg uit waar "hier" en "daar" naar verwijzen.

- Paragraph 3 consists of two parts. With which sentence does the second part of the paragraph start?
- What is the function of paragraph 3 with regard to the tech companies?
- What is the main function of the reference to 'Romeo' (lines 37-39)?
- Which of the following aspects didn't live up to expectations either?
- Which of the following is the other half?
- Which of the following is true about paragraph 5?
- Which of the following would best describe the purpose of paragraph 4?
- Which would be a suitable alternative beginning for paragraph 4?
- Why does Landesman mention Bradley Myles? (paragraphs 3 and 4)
- Why does the writer provide these examples?
- Leg in je eigen woorden uit wat bedoeld wordt met "this urge".
- Wat adviseert Edward Tikoft om te doen in plaats van innoveren? Geef antwoord in je eigen woorden.
- Which paragraph contains an actual example of such a conflict?
- Wie of wat is dit "monster"? Geef antwoord in het Nederlands of in het Engels.

1.5 5a: Relaties tussen delen van een tekst aangeven // Integrate and generate inferences // Cloze: woordenschat, syntaxis en semantiek

- Noteer de juiste cijfers achter de letters op het antwoordblad om de volgorde aan te geven waarin de woorden in de tekst hebben gestaan.
- Which of the following fits both gaps?
- Which of the following fits the gap in paragraph 1?
- Which of the following fits the gap in the last sentence?
- Which of the following fits the gap in the text?
- Which of the following series of qualifications fits the string of gaps in paragraph 6?

1.6 5b: Relaties tussen delen van een tekst aangeven // Integrate and generate inferences // Cloze: woordenschat, syntaxis en semantiek, Tekststructuur

- Geef bij elke plaats aan welk van de citaten daar hoort. Noteer de letter van het citaat achter het nummer op het antwoordblad.
- Which of the following fits the gap at the beginning of paragraph 3?
- Which of the following was the original first sentence of paragraph 4?

1.7 6: Reflect on content and form // Conclusies trekken met betrekking tot intenties, opvattingen en gevoelens van de auteur // Gevolgtrekking

- Based on the author's views as expressed in the article, which of the following adjectives would best modify the noun "attempt" in the title?
- Bepaal voor de volgende stijlfiguren of deze wel of niet gebruikt worden in dit artikel.
- Geef van de volgende citaten uit de tekst aan van welke van de bovenstaande stijlfiguren ze een voorbeeld zijn.
- Geef van de volgende opmerkingen aan of ze spottend bedoeld zijn.
- How can the tone of paragraph 4 be characterised best?

- How could this quotation be characterised best?
- How does the writer introduce the subject in paragraphs 1 and 2?
- How should this remark be interpreted?
- Welke van de vijf besproken films is volgens Nigel Andrews het minst geslaagd? Schrijf de titel van deze film op.
- What attitude does Lee Nason express towards this group of people?
- What does the writer want to make clear with this remark?
- What does the writer wish to convey in paragraph 7?
- What is the writer's goal in paragraph 4?
- What opinion does the writer give in this final remark?
- Which of the following best describes the tone of the letter?
- Which of the following echoes Tom Vanderbilt's personal view on taste as expressed in paragraph 5?
- Which of the following expressions captures Simon Boyle's criticism of Amanda Vanstone?
- Which of the following opinions can be ascribed to Matthew Engel based on paragraphs 8 and 9?
- Which of the following paragraphs does not contain any explicit criticism of the book?
- Which of the following paragraphs lacks a critical note on Bryson's work?
- Which of the following reflects criticism of the "Financial Fair Play regime" (paragraph 1)?
- Which of the following reflects the author's opinion in paragraph 8?
- Which of the following statements is in line with Michael Eustace's opinion?
- Which of the following words reflects Chris Skillern's attitude to Gareth Jones' argument best?
- Why did Lewis put the word "developed" in italics?
- Why does the writer characterise Tinder in this way? He wants to make clear that using Tinder
- Wordt er door Felix Gillette kritiek geuit op het boek dat hij hier recenseert? Zo nee, antwoord "Nee". Zo ja, noteer het nummer van de alinea waarin dit gebeurt.