

Jozefien Loman & Jolien Wouters
Centrum voor Taal en Onderwijs
Contact: fien.loman@kuleuven.be
jolien.wouters@kuleuven.be

Krachtige evaluatie van begrijpend lezen: complexiteit en variatie in toetsvragen

1. Inleiding

De leesvaardigheid en het leesplezier van Vlaamse leerlingen gaan achteruit. Zowel het PISA-onderzoek uit 2018 als de PIRLS-resultaten uit 2016 en 2021 tonen dat aan. De resultaten maken duidelijk dat er nood is aan een effectieve begrijpend-leesinstructie en aan versterking van de vakdidactische competenties van leraren. Wat de meest effectieve didactische aanpak voor het begrijpend-leesonderwijs inhoudt, werd de voorbije jaren uitvoerig bestudeerd. Zo werden, in opdracht van de Vlaamse Onderwijsraad (Vlor), vijf didactische principes geformuleerd voor een krachtige, effectieve en eigentijdse begrijpend-leesdidactiek in het basisonderwijs. Samen zorgen deze didactische principes of sleutels ervoor dat leerlingen op een actieve manier betekenis toekennen aan een geschreven tekst (Merchie e.a. 2019).

Het blijft natuurlijk niet alleen bij deze sleutels. Wie een krachtige begrijpend-leesdidactiek in de klaspraktijk wil implementeren, moet weten welk publiek hij voor zich heeft. Verschillende studies benadrukken daarom het belang van het monitoren en evalueren van de verschillende aspecten van begrijpend lezen (Houtveen e.a. 2019; Merchie e.a. 2019). De evaluatie van begrijpend lezen moet immers beantwoorden aan de complexiteit ervan: niet alleen moeten vaardigheden als ‘het tekstbegrip’ worden gemonitord, ook aspecten als ‘leesstrategieën’ en ‘leesmotivatie’ moeten in rekening worden gebracht (Houtveen e.a. 2019).

Door het ontwikkelen van een leerlingvolgsysteem voor begrijpend lezen voor leerlingen van zes tot en met twaalf jaar tracht ‘Sleutel 6’ hierop in te spelen: een ‘zesde sleutel’ die de begrijpend-leesontwikkeling van leerlingen breed in beeld brengt, en leraren houvast geeft om sterker begrijpend-leesonderwijs te bieden, voor alle leerlingen.

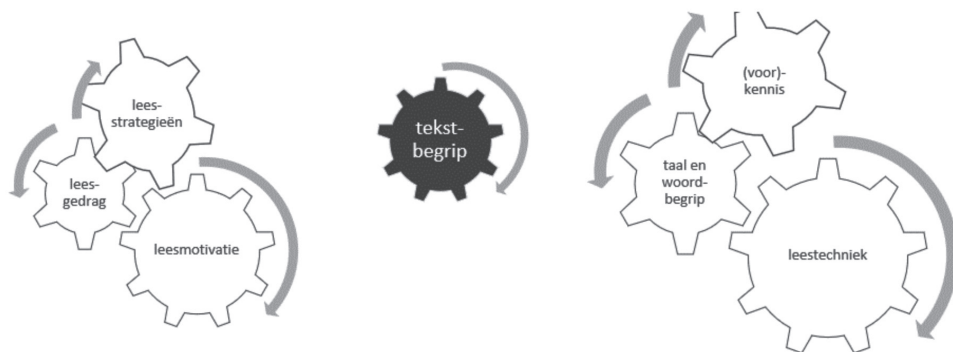
2. Sleutel 6: een digitaal leerlingvolgsysteem voor begrijpend lezen

In samenwerking met uitgeverij Plantyn ontwikkelt het Centrum voor Taal en Onderwijs (KU Leuven) een leerlingvolgsysteem voor begrijpend lezen in het basisonderwijs. We werken aan een wetenschappelijk onderbouwd instrument dat de begrijpend-leesontwikkeling in kaart brengt en handvatten geeft om het didactisch handelen van leraren te versterken. De onderzoeksgroep krijgt voor het realiseren van dit project financiële ondersteuning van het Vlaams Agentschap voor Innoveren en Ondernemen (VLAIO).

Sleutel 6 bevat drie componenten:

1. *screening-tools*: gestandaardiseerde toetsen voor een globale beeldvorming;
2. *breedbeeld-tools*: een brede waaier aan soorten instrumenten voor verbreding en verdieping van de beeldvorming;
3. *feedback voor leraren*: resultaten van screening- en breedbeeld-tools, en concrete tips om met leerlingen aan de slag te gaan.

De combinatie van de screening- en breedbeeld-tools zorgt ervoor dat de leraar aandacht heeft voor en informatie verzamelt over het volledige ‘construct’ van begrijpend lezen. Door middel van de screening-tools komt de leraar te weten hoe het staat met het tekstbegrip van de leerling. Aanvullend kan hij met de breedbeeld-tools informatie verzamelen over ‘de leesstrategieën’, ‘het leesgedrag’, ‘de leesmotivatie’, ‘de (voor)kennis’, ‘de leestechniek’ en ‘het taal- en woordbegrip’ van de leerling(en). Op basis van deze informatie kan de leraar doelgericht actie ondernemen op niveau van de leerling, de klas of de school.



Afbeelding 1 – Construct van begrijpend lezen in 'Sleutel 6'.

3. Variatie en complexiteit van toetsvragen in de screening-tools: een deelonderzoek van Sleutel 6

3.1 Onderzoeksopzet

De screening-tools in de Sleutel-6-tool bevatten een variatie aan tekstsoorten- en types en aan vraagtypes. Die variatie bouwen we in om zo goed mogelijk te variëren in de moeilijkheidsgraad van de toetsvragen en om een doorlopende lijn te realiseren. Daarnaast biedt die variatie extra informatie aan de leraren bij de diagnose en analyse van de toetsresultaten. Op welke vragen scoort de leerling zwak? Waar kan ik harder op inzetten?

Maar die moeilijkheid of complexiteit van toetsvragen is niet makkelijk te beschrijven. Want wat maakt een toetsvraag nu precies complex? Welke criteria spelen een rol in de prestaties van leerlingen?

Op basis van de bestaande literatuur creëerden we in een eerste fase een zo volledig mogelijk kader van de complexiteit van toetsvragen bij begrijpend lezen. We baseerden ons hiervoor onder andere op Robinson, Weir, de kritische factoren voor lezen (*Language and Reading Research Consortium*), de invloed van tekstsoort en vraagtype (Eason e.a.), het construct 'begrijpend lezen' (Merchie e.a 2019.; Kintsch), de taxonomieën van Bloom, enz. Om dat theoretische kader verder aan te vullen en te concretiseren, voerden we een deelonderzoek uit, waarbij we input vroegen aan een groep beoordeelaars (23 leesexperten en 14 leraren).

3.2 Onderzoeksvragen

Voor dit deelonderzoek stelden we de volgende onderzoeksvragen op:

1. Welke criteria bepalen volgens experts en leraren de variatie in moeilijkheidsgraad van toetsvragen in ‘Sleutel 6’?
2. Welke van de criteria kunnen worden opgenomen in een complexiteitsmeter, met als doel leraren handvatten geven voor gerichte ondersteuning van leerlingen?

Aan de hand van een paarsgewijze vergelijking beoordelen de twee groepen beoordelaars de toetsvragen van het vierde leerjaar (of groep 6). De beoordelaars kiezen in een online tool de toetsvraag die volgens hen het meest complex is en formuleren daarbij op basis van welke criteria ze hun keuze hebben gemaakt.

3.3 Methode: comparatief beoordelen

Bij het beoordelen van complexe competenties is ‘comparatief beoordelen’ een aangewezen methode. Bij ‘holistisch beoordelen’ zal immers een bepaald onderdeel bij de ene beoordelaar harder doorwegen dan bij een andere beoordelaar, waardoor de betrouwbaarheid van de beoordeling daalt. Ook ‘analytisch beoordelen’ (bijvoorbeeld: aan de hand van criterialijsten of *rubrics*) laat nog steeds ruimte open voor interpretatie van de criteria tussen beoordelaars. Bovendien zal niet elke beoordelaar even streng of mild beoordelen, en kunnen volgorde-effecten een rol spelen in het beoordelingsproces. Bij comparatief beoordelen vergelijk je twee producten en stel je je de vraag welke van de twee beter/moeilijker/... is. Met voldoende vergelijkingen per beoordelaar en voldoende beoordelaars, kom je tot betrouwbare en valide resultaten (Goossens 2018).

De paarsgewijze vergelijking door de beoordelaars gebeurde via *no more marking*, een online tool die de beoordelaars net zoveel vergelijkingen voorschotelt totdat een bepaald saturatiepunt bereikt wordt waarop de resultaten valide en betrouwbaar zijn.

Op basis van de resultaten van de beoordelaars konden we de toetsvragen ordenen op moeilijkheidsgraad en vergelijken met de resultaten van de leerlingen tijdens het toen afgeronde vooronderzoek. Tijdens een digitale focusgroep bespraken we de resultaten.

4. Vaststellingen op basis van verkennende analyses

Uit de paarsgewijze vergelijkingen kwamen enkele algemene bevindingen naar boven. Zo stelden we vast dat elke beoordelaar een eigen terminologie hanteert, die vaak gelinkt is aan een bepaald theoretisch kader. Er wordt onder andere verwezen naar de manier van lezen (“*De vraag rechts gaat over zoekend lezen*”), naar de eindtermen (“*Dit lijkt mij eerder verwerkingsniveau structureren*”), naar informatie die al dan niet letterlijk te vinden is (“*Voor de linkse vraag moeten de leerlingen expliciete en impliciete info uit de tekst combineren*”) en naar de spreiding van informatie (“*Die info staat dicht bij elkaar in de tekst*”).

Ondanks de verschillende kaders, kunnen we zowat alle beoordelingscriteria linken aan aspecten van een model dat het proces van begrijpend lezen in kaart brengt. Dit model werd ontwikkeld door de *RAND Reading Study Group* (2002), en later verfijnd in het wetenschappelijk rapport naar de vijf didactische sleutels voor een effectieve begrijpend-leesdidactiek (Merchie e.a. 2019). We geven het model hieronder in vereenvoudigde versie weer:

Model	Voorbeeld beschrijving beoordelaars
De lezer voert	<i>“Het woord ‘attributen’ zal moeilijkheden opleveren bij bepaalde leerlingen”.</i>
binnen een <i>socio-culturele context</i>	<i>“Zou voorkennis niet bij veel leerlingen een rol spelen? Veel ouders spelen immers padel, denk ik...”.</i>
activiteiten uit	<i>“Deze vraag vereist intensief lezen”.</i>
met een <i>tekst</i>	<i>“De leerlingen moeten een moeilijk gedicht interpreteren”.</i>
om deze te begrijpen.	

Tabel 1 – Model om het proces van begrijpend lezen in kaart te brengen.

Uit de focusgroep bleek verder dat de beoordelaars:

- niet altijd goed in staat zijn om de voorkennis van leerlingen in te schatten;
- de capaciteiten van leerlingen om met visualisaties om te gaan onderschatten (bijvoorbeeld: ‘details waarnemen’ of ‘onthouden wat op een afbeelding staat’);
- niet altijd weten waarop leraren momenteel al sterk inzetten (en dus op welke soorten teksten/vragen leerlingen al vaak oefenen in de klascontext);
- worden afgeleid door de ordening die bepaalde kaders lijken te suggereren (bijvoorbeeld: beoordelen is niet altijd moeilijker dan structureren).

5. Next steps: uitwerking van *tags* voor complexiteit

Dankzij het deelonderzoek naar complexiteit hebben we een vrij volledig beeld van de kaders die worden gebruikt om complexiteit te beoordelen. Het maakte ook duidelijk dat er heel wat overlap tussen de verschillende kaders bestaat. Toch blijken kenmerken die voor experts opvallend of belangrijk zijn, niet allemaal even goede voorspellers van moeilijkheid te zijn.

Op basis van onze bevindingen, hebben we een hele reeks van *tags* kunnen formuleren bij toetsvragen. Hoe meer *tags* kunnen toegekend worden aan een bepaalde toetsvraag, hoe moeilijker deze vraag voor leerlingen zal zijn. Zo kunnen we onze toetsvragen vergelijken, een doorlopende lijn realiseren en bieden we de leraar extra ondersteuning voor analyse en diagnose. De *tags* hebben zowel te maken met ‘de tekst’ als met ‘de op-

dracht' en 'de (denk)activiteit' die door de lezer moeten worden uitgevoerd. Ze kunnen ook worden gelinkt aan de verschillende onderdelen van het complexe construct van begrijpend lezen (zie: Afbeelding 1).

6. Besluit

We voerden een deelonderzoek uit naar de criteria die de complexiteit van begrijpend-leesopdrachten in toetsen voor leerlingen kunnen verklaren. We hadden daarvoor twee onderzoeksvragen (zie 3.2).

We kunnen besluiten dat leraren en experts niet sterk zijn in het inschatten van de moeilijkheidsgraad van toetsvragen, maar dat hun gezamenlijke reflecties wel kunnen leiden tot het formuleren van *tags* of criteria die kunnen worden toegekend aan toetsvragen. Algemeen geldt dat hoe meer een toetsvraag aantikt op de criteria, hoe complexer deze is voor de leerlingen. Door de criteria samen te brengen in een eenvoudig model, gelinkt aan het construct van begrijpend lezen, kunnen leraren besluiten trekken over de ondersteuning die ze leerlingen moeten bieden om betere begrijpend-lezers te worden.

Referenties

- Goossens, M. (2018). 'Eerlijk en doeltreffend beoordelen met paarsgewijze vergelijking. Een toepassing van D-PAC'. Antwerpen: FSW, Universiteit Antwerpen.
- Houtveen, A., R. van Steensel & S. de la Rie (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'. Den Haag/Utrecht: NRO/Inspectie van het Onderwijs.
- Merchie, E., S. Gobyn, E. De Bruyne, F. De Smedt, M. Schiepers, M. Vanbuel, P. Ghesquière, K. Van den Branden & H. Van Keer (2019). 'Effectieve, eigentijdse begrijpend leesdidactiek in het basisonderwijs. Wetenschappelijk eindrapport van een praktijkgerichte literatuurstudie'. Gent: UGent.