

Lezen en antwoorden bij de tekst met vragen geobserveerd

Een eye-trackstudie onder vwo 4-leerlingen

P. Rooijackers, G. van Silfhout, U. Schuurs, I. Mulders, en H. van den Bergh

Samenvatting

In leeslessen Nederlands in het secundair onderwijs wordt vaak de tekst met vragen gebruikt om leerlingen te toetsen en trainen in tekstbegrip. Daarbij is de gedachte doorgaans: wie goed vooraf leest, maakt de vragen ook goed. In deze studie wordt de relatie tussen het vooraf lezen van teksten en het beantwoorden van bijbehorende vragen onderzocht door analyse van de oogbewegingen. Vertrekpunt vormt de aanname in de Constructie-Integratietheorie van Kintsch (1998) dat lezers voor een dieper tekstbegrip meer aandacht besteden aan kernzinnen dan aan niet-kernzinnen. Er is onderzocht in hoeverre vwo 4-leerlingen (N = 16) tijdens vooraf lezen meer leestijd schenken aan kernzinnen dan aan niet-kernzinnen en in hoeverre dit verband houdt met het antwoorden. Uit de resultaten blijkt dat deelnemers tijdens vooraf lezen meer tijd besteden aan kernzinnen, maar op dieper tekstbegripsniveau kon geen relatie tussen vooraf lezen en antwoorden worden vastgesteld. Deelnemers lezen vooraf over het algemeen lineair en construeren pas tijdens het antwoordproces een dieper tekstbegrip. De vraag is daarmee in hoeverre de huidige inrichting van deze taak tekstbegrip bevordert en de dominantie ervan in het schoolvak Nederlands gerechtvaardigd is.

Kernwoorden: tekstbegrip, schoolvak Nederlands, eye-tracking, vwo, tekst met vragen

6 (in Vlaanderen het laatste jaar van de derde graad in het aso). De beschrijving van dit zogenaamde ‘niveau 4F’ kon echter nauwelijks worden ondersteund door empirische studies. De voorzitters van de Expertgroep spraken daarom in 2008 al van een “educated guess” (Robben, 2008).

De Expertgroep Taal stelde ook voor havo 5 normen op voor zakelijk lezen: ‘niveau 3F’, dat grofweg ook voor vwo 4-leerlingen geldt (het laatste jaar van de tweede graad in het aso). Leerlingen zouden vanaf dit niveau een grote variatie aan teksten “over opleidings- en beroepsgebonden en maatschappelijke onderwerpen” kunnen lezen, met “begrip voor geheel en details”. Een nadere beschrijving van niveau 3F en 4F is te vinden in Expertgroep Taal (2008).

De vraag is of we op dit moment enig inzicht hebben in het werkelijke tekstbegrip van de Nederlandse vwo-bovenbouwleerlingen. Belangrijkste dataverzameling over de vwo-lezer vormt het Nederlandse Centraal Schriftelijk Examen vwo waarvan we kunnen aannemen dat dit tekstbegrip toetst op niveau 4F (Van Onna, Straat, & Alberts, 2014): vwo-abituriënten zouden dus teksten kunnen begrijpen op niveau 4F. Onder deze laatste conclusie ligt de vaak in het schoolvak Nederlands gemaakte aanname dat wie goed vooraf heeft gelezen, ook de erop volgende vragen goed beantwoordt. Echter, de relatie tussen vooraf lezen en beantwoorden ligt in werkelijkheid aanmerkelijk complexer: we weten dat een vooraf gestelde taak zelf een sterke invloed heeft op het lezen (bijv. André, 1979; Hamilton, 1985; Yeari, Van den Broek, & Oudega, 2015; Cerdan, Vidal-Abarca, Martinez, Gilabert, & Gil, 2009, etc.). In internationaal onderzoek krijgen kandidaten dan ook meestal vooraf een taakcontext aangeboden voordat ze een tekst gaan lezen (PIRLS, PISA, IELTS etc.). Strikt genomen

1 Inleiding

1.1 Het tekstbegrip van de vwo-leerling

Het lezen van zakelijke teksten in de hoogste jaarlagen van het Nederlandse vwo is weinig bestudeerd. In 2010 stelde de Expertgroep Taal normen op voor zakelijk lezen voor vwo

zeggen de resultaten van het Centraal Examen Nederlands vwo alleen iets over of leerlingen op niveau 4F antwoorden bij een ‘tekst met vragen’. Over het tekstbegrip tijdens het lezen vóórdát ze de vragen beantwoordden, weten we vrijwel niets.

Om in deze twee kwesties – het tekstbegrip van de vwo bovenbouw-leerling en de relatie tussen vooraf lezen en beantwoorden bij de tekst met vragen – meer duidelijkheid te brengen, onderzoeken we in deze studie de relatie tussen het vooraf lezen en het beantwoorden. We volgden daarom de oogbewegingen van Nederlandse vwo bovenbouw-leerlingen terwijl ze korte zakelijke teksten lazen én terwijl zij bijbehorende vragen over de tekst beantwoordden.

1.2 Relatie tussen vooraf lezen en het antwoord

Het leesgedrag van een expert-lezer is redelijk scherp in kaart gebracht. Bepalend is met name de studie van Pressley en Afflerbach (1995) waarin 38 hardopdenkstudies zijn geanalyseerd om de algemene kenmerken van een volwassen lezer op te kunnen stellen. Het blijkt dat een goede lezer zich vooral kenmerkt door de inzet van tekstbegrip bevorderende leesstrategieën, zoals leesdoelen stellen en voorkennis activeren tijdens het lezen. Deze bevindingen hebben inmiddels hun weg gevonden in tal van overzichtsstudies over tekstbegrip (Duke & Pearson, 2002; Perfetti, Landi, & Oakhill, 2005; McNamara, 2007, etc.).

Het natuurlijke leesproces is lastig in kaart te brengen. In de jaren tachtig en negentig maakten onderzoekers vooral gebruik van hardopdenkprotocollen om zicht te krijgen op het verloop van het leesproces en de rationale erachter (vgl. Pressley & Afflerbach, 1995). Maar belangrijk nadeel van hardopdenkprotocollen is dat de verbalisatie tijdens het lezen het natuurlijke handelingsproces danig verstoort (Gerjets, Kammerer, & Werner, 2011).

Vanaf de jaren negentig wordt onder andere daarom oogbewegingenregistratie steeds vaker in onderzoek ingezet om het lezen te registreren, al richten deze studies zich wel vaak op deelaspecten van het lezen (bijv.

Cozijn, 2000; Van Silfhout, 2014; Yeari et al., 2015). Eye-tracking registreert waarop en hoelang ogen fixeren. De methode gaat ervan uit dat verschillen in de duur van fixaties verschillen in cognitieve verwerkingstijd vertegenwoordigen; bij ‘saccades’, oogverplaatsingen, wordt de aandacht verschoven (Duchowski, 2007; Holmqvist & Andersson, 2017). Eye-tracking verstoort het natuurlijke leesproces minder dan hardop denken en kent bovendien inmiddels weinig technische beperkingen. Eye-trackingdata verstrekken evenwel geen rationale bij het waargenomen leesgedrag (Cozijn, 2000; Bax, 2013).

Het is tegenwoordig duidelijk dat vooraf lezen niet eenduidig de kwaliteit van een leesproduct bepaalt. Een taak die vooraf wordt gegeven, heeft zelf juist een belangrijk effect op het lezen (bijv. André, 1979; Cerdan et al., 2009; Hamilton, 1985; Lorch, Lorch, & Mogan, 1987; Narvaez, Van den Broek, & Barron-Ruiz, 1999). Een taak met open vragen zet lezers bijvoorbeeld meer aan tot diepgaand begrip dan een taak met gesloten vragen (Magliano, Millis, Ozuru, & McNamara, 2007; Ozuru, Briner, Kurby, & McNamara, 2013). En Yeari et al., (2015) vonden dat lezers bij een samenvattingsopdracht meer tijd besteedden aan hoofdzaken, dan bij een taak met gesloten vragen.

De tekst met vragen lijkt het zakelijk lezen in het klaslokaal Nederlands in de vwo-bovenbouw te domineren (Canton, Aler, Heemskerk, Van der Westen, & Willemsen, 2013; Meestringa & Ravesloot, 2012). Deze leestaak en -toets kent in het schoolvak Nederlands een lange traditie (Hulshof, Kwakernaak, & Wilhelm, 2015) en huidige Nederlandse bovenbouwmethodes als *Nieuw Nederlands* en *Op niveau* leggen er een sterk accent op. Ook het landelijk Centraal Schriftelijk Examen Nederlands vwo vertrekt vanuit deze taak en heeft daarmee naar alle waarschijnlijkheid een sterk terugslag-effect op de inrichting van het leesonderwijs in het schoolvak Nederlands (Meestringa & Ravesloot, 2012). De grondslag van deze taak is meestal dezelfde. Leerlingen moeten, zonder uitgebreide taakcontext vooraf, een tekst grondig bestuderen, met oog voor hoofdzaken, hoofdgedachte, structuur en alineafuncties;

daarna volgt een aantal vragen, en de tekst blijft dan steeds raadpleegbaar.

Er is in Nederland momenteel een duidelijke onvrede over deze situatie: de tekst met vragen zou leiden tot een oppervlakkige bestudering van de tekst vooraf en ze zou zelf centraal zijn komen te staan in didactiek en vakinhoud (SLO, 2016; SLO, 2017; Nederlands Nu! & Sectiebestuur Nederlands LT, 2018; zie eerder al Hoogeveen & Bonset, 1998). Deze onvrede bereikte in Nederland de afgelopen jaren herhaaldelijk de media (bijv. De Bruin, 2018). De empirische onderbouwing bij dit standpunt is beperkt (SLO, 2017), al zijn er een aantal studies waaruit blijkt dat een score bij de tekst met vragen slechts in beperkte mate het diepere begrip van een tekst weerspiegelt (Kamalski, 2007; Land, 2009).

In deze studie willen we de relatie tussen het vooraf lezen van een tekst en het beantwoorden van vragen verkennen door middel van eye-tracking, dat het natuurlijke lezen immers dicht benadert.

Om de kwaliteit van het vooraf lezen en beantwoorden te beschrijven, gaan we uit van de Constructie-Integratietheorie van Kintsch (1998), vertrekpunt van vele Nederlandse en internationale leesonderzoeken (vgl. McNamara & Magliano, 2009). Volgens Kintsch (1998) is een tekst een netwerk van hiërarchisch samenhangende kleine, betekenisvolle beweringen (proposities). Een lezer is steeds op zoek naar deze samenhang en construeert tussen woorden en zinnen verbanden (een microstructuur) en tussen alinea's (een macrostructuur). Dit constructieproces mondt dan volgens Kintsch uit in een 'tekst-basis', een vrij letterlijke mentale representatie van de tekst, en in een 'situatiemodel', een model waarin de lezer de wereld van de tekst opneemt in zijn eigen voorkennis, opvattingen etc. Op dit laatste niveau kent een lezer aan een tekst betekenis toe en legt hij onuitgesproken verbanden.

Een cruciale – en voor *eye-tracking* bijzonder relevante – aanname in de constructie-integratietheorie van Kintsch is dat lezers de inhoudelijke structuur van een tekst herkennen en reconstrueren en van meet af aan hoofd- en bijzaken onderscheiden. Volgens

Kintsch kennen aandachtige lezers in hun zoektocht naar samenhang vrijwel automatisch aan tekstgedeelten met een hoog topisch gehalte extra belang toe (Kintsch, 1998, m.n. p. 174). Meerdere onderzoeken hebben aangetoond dat geoefende lezers aan kernzinnen meer leestijd besteden dan aan detailzinnen, zelfs al bij de eerste lezing ervan (Pressley & Afflerbach, 1995; Kintsch, 1998; Mross, 1989 (geciteerd in Kintsch, 1998); Lorch, Lorch, & Matthews, 1985; Hyönä & Niemi, 1990; Yeari et al., 2015). Selectief oog hebben voor de inhoudsstructuur van de tekst is dan ook een prominente strategie van de expert-lezer (bijv. Duke & Pearson, 2002; McNamara, 2007; Perfetti et al., 2005).

1.3 Onderzoekshypothesen

Doel van deze studie is enerzijds meer zicht te krijgen op het tekstbegrip van vwo bovenbouw-leerlingen, en anderzijds op het verband tussen het vooraf lezen en het vragen beantwoorden bij de dominerende taak in het Nederlandse tekstbegripsonderwijs: de tekst met vragen. Uitgangspunt zal zijn dat lezers, zoals vaak bij deze taak verondersteld, de tekst eerst aandachtig bestuderen en daarbij proberen een adequaat situatiemodel op te bouwen: ze zullen duidelijk aan kernzinnen meer leestijd schenken dan aan niet-kernzinnen. Daarnaast veronderstellen we een sterke samenhang tussen het lezen vooraf en het vragen beantwoorden.

We onderzoeken dit bij vwo 4-leerlingen, leerlingen die bij tamelijk complexe teksten onderscheid zouden moeten kunnen aanbrengen tussen hoofd- en bijzaken – dat stelt althans de Expertgroep Taal (2008) over lezers op niveau 3F (vwo 4-leerlingen).

Voor het *vooraf lezen* verwachten we allereerst dat vwo bovenbouw-leerlingen over het algemeen adequate tekstbases en situatiemodellen zullen opbouwen: ze zullen dus meer verwerkingstijd (leestijd) wijden aan kernzinnen dan aan niet-kernzinnen.

Ten aanzien van het *beantwoorden* vertrekken we vanuit de hypothese dat de participanten tijdens vooraf lezen minimaal een adequate tekstbasis hebben opgebouwd en dus weten waar het antwoord op een vraag te vinden is: ze zullen aanzienlijk meer leestijd

besteden aan de tekstdelen die voor het antwoord relevant zijn dan aan de tekstdelen die hiervoor irrelevant zijn. Ten tweede: de participanten hebben tijdens het vooraf lezen een adequaat situatiemodel opgebouwd en zullen dus snel en adequaat antwoorden.

Voor de *relatie tussen het vooraf lezen en het beantwoorden* formuleren we twee verwachtingen. Ten eerste: lezers die tijdens het vooraf lezen langer naar kernzinnen keken en korter naar niet-kernzinnen, zullen effectiever antwoorden: ze zullen langer kijken naar de tekstdelen die voor het antwoord relevant zijn en korter naar tekstdelen die irrelevant zijn, vergeleken met lezers die kernzinnen niet positief onderscheiden van niet-kernzinnen (cf. Rouet, 2006). Ten tweede: lezers die langer naar kernzinnen kijken en korter naar niet-kernzinnen, zullen vaker een goed antwoord op een vraag geven.

2 Methode

2.1 Participanten en experimentopzet

Twintig vwo 4-leerlingen van één grote collegiale middelbare school in het zuiden van Nederland werden aselect gekozen. Drie van de gevraagde leerlingen besloten om uiteenlopende redenen niet te participeren. De zeventien deelnemende leerlingen hadden geen problemen met hun zicht. Zeven jongens namen deel, tien meisjes. De gemiddelde leeftijd was 16;3 jaar (sd = 6 maanden). Alle leerlingen waren moedertaalsprekers Nederlands. Achteraf bleken de data van één participant geheel onbruikbaar, hoogstwaarschijnlijk vanwege het dragen van softlenzen, die de oogdetectie vaak belemmeren (vgl. Holmqvist & Andersson, 2017). Bij drie participanten werden data gedeeltelijk (doorgaans één tekstlezing) uitgesloten vanwege technische problemen.

Participanten (N = 16) bestudeerden op het computerscherm vier teksten, variërend in lengte van 240 tot 290 woorden. Na bestudering van de tekst volgden op het scherm telkens vier of vijf vragen. Bij het maken van de vragen bleef de tekst steeds naast de vraag raadpleegbaar zoals de praktijk is in het schoolvak Nederlands. Na elk tekstblok met

vragen volgde een korte mogelijkheid tot pauze. De volgorde van de stimuli was voor alle deelnemers gelijk. Gedurende de hele *online* sessie volgde een eye-tracker de oogbewegingen van de deelnemers. Na het *eye-tracking*-experiment vulden de leerlingen een vragenlijst in, waarin naar enkele algemene gegevens werd gevraagd maar ook naar een inschatting van de motivatie en concentratie gedurende het experiment. Ten slotte volgde een kort retrospectief *stimulated recall*-interview: leerlingen werd hun antwoordgedrag tijdens de laatste set met vragen getoond en werd gevraagd hierop te reflecteren.

2.2 De stimuli

De inhoud en opzet van bovenbouwmethodes Nederlands en van het landelijk Centraal Schriftelijk Examen Nederlands vwo waren leidend in het samenstellen van de stimuli. We selecteerden uit landelijke dagbladen en tijdschriften vier korte opiniërende en uiteenzettende teksten, van maximaal 300 woorden lengte. De teksten liepen in moeilijkheidsgraad op van niveau 2F tot 4F. Bij elke tekst construeerden we vervolgens een acht- tot tiental vragen van wisselend niveau. Een screeningspaneel, bestaande uit twee gepromoveerde docenten taalbeheersing en twee geroutineerde bovenbouwdocenten Nederlands, screeende daarop al het materiaal uitvoerig en stelde het correctiemodel vast. Een papieren pretest-afname onder vier vwo 4-klassen volgde. Op basis van de leerlingantwoorden werden ten slotte per tekst vier of vijf items geselecteerd. Selectiecriteria waren in volgorde van gewicht: a) de kwaliteit van de vraagstelling (eenduidigheid, helderheid), b) de waarschijnlijkheid dat een vraag afwijkend oogbewegingengedrag genereerde tussen lezers met een goed en fout antwoord (baseerden leerlingen met een fout antwoord niet steeds hun antwoord op dezelfde tekstpassage als leerlingen met een goed antwoord?), en ten slotte c) statistische indicatoren (p-waarden en rit's).

De vragen werden zo geconstrueerd dat ze op zowel de tekstbasis als het situatiemodel betrekking hadden. Bij het merendeel van de vragen werd daarom in de stamvraag een voor de vraag relevant citaat uit de brontekst

opgenomen, zonder vermelding van alineaf of regelnummer. Zo werd een beroep gedaan op de door de lezer opgestelde tekstbasis. Bij het construeren van de vragen werden *elaboration questions* en *bridging inference questions* geformuleerd, beide in open vraagvorm. *Elaboration questions* vragen om een tekstpassage te verbinden met de achtergrondkennis van een lezer; *bridging inference questions* vereisen dat een lezer twee of meer tekstpassages inhoudelijk verbindt. Deze twee vraagtypen hebben op het situatiemodel betrekking (Cerdan et al., 2009; Kintsch, 1998; Land, 2009; O'Reilly & McNamara, 2007).

Door deze tweeledige inrichting van het item kon bij het beantwoorden worden verwacht dat er uiteenlopende oogbewegingspatronen optraden: leerlingen met een redelijke tot goede tekstbasis zullen het citaat uit de stamvraag direct in de tekst weten te lokaliseren, terwijl leerlingen met een matige tekstbasis de passage moeilijker of niet vinden. Leerlingen met een redelijke tekstbasis én een adequaat situatiemodel zullen hun aandacht geheel of grotendeels schenken aan de voor beantwoording noodzakelijke passages of zullen direct een antwoord geven, zonder de tekst te raadplegen (vgl. Rouet, 2006).

Omdat we willen weten of deelnemers tijdens het vooraf lezen meer aandacht schenken aan kernzinnen, werd aan het screeningspaneel verzocht de kernzinnen in de vier geselecteerde teksten aan te wijzen. Daartoe ontvingen ze een uitgebreide schriftelijke instructie vooraf. De overeenstemming tussen de vier leden van dit paneel hierin bleek hoog (Cronbachs $\alpha = .82$). Daarnaast gaf het paneel bij elke vraag aan welke tekstpassages noodzakelijk zijn om de vraag correct te beantwoorden. Hierin bleken de vier leden van het paneel zeer sterk te overeenstemmen (Cronbachs $\alpha = .96$).

2.3 Procedure

Bij de afname op locatie is een mobiele SMI-eye-tracker gebruikt (versie 3.6) met een samplefrequentie van 250 Hz; bij de oogbewegingenregistratie van de eye-tracker is het rechteroog genomen. Als parameters voor de eventdetectie zijn conform Nyström en

Holmqvist (2010) een minimale fixatieduur van 40 milliseconden gekozen en een minimale saccadeduur van 15 milliseconden. De SMI eye-tracker kent op basis van saccadedetectie events toe; deze saccades werden toegekend vanaf een snelheidsdrempelwaarde van 40°/sec, een courante waarde bij leesonderzoek (vgl. Holmqvist & Andersson, 2017).

De deelnemers ontvingen vooraf een uitgebreide toelichting op het experiment door de onderzoeker. Daarin werd de werking van de eye-tracker kort uiteengezet en werd benadrukt dat het een algemeen onderzoek naar de leesontwikkeling van vwo-leerlingen betrof. De onderzoeker beklemtoonde daarnaast zeer uitdrukkelijk dat de getoonde teksten terdege bestudeerd dienden te worden, alsof er een toets over zou volgen. Daarna volgde een kalibratie- en validatieprocedure die, in verband met de enigszins ongebruikelijk lange tijdsduur van het *eye-track*-experiment, bij elke tekst herhaald werd. Deelnemers bepaalden zelf het tempo waarin de stimuli gepresenteerd werden en konden naar eigen believen naar een volgend scherm schakelen door de spatiebalk in te drukken; terugkeren naar een vorig scherm was niet mogelijk. Er werd vooraf geen maximumduur aan het experiment gesteld; deelnemers konden zelf bepalen hoeveel tijd zij aan lezing en beantwoording besteedden.

De zeventien deelnemers besteedden gemiddeld 27.5 minuten (sd = 4.7 min.) aan het *online* deel van het experiment. Het hele experiment met introductie, pauze en afronding duurde doorgaans een zestigtal minuten. De participanten gaven in de controlevragenlijst aan dat ze zich doorgaans tijdens het experiment ruim voldoende geconcentreerd voelden (op een tienpuntsschaal: gem. = 7.0; sd = 1.2) en ruim voldoende gemotiveerd (gem. = 7.5; sd = 1.1).

2.4 Analyse

Lezen vooraf

Voorafgaand aan de analyse werd de kwaliteit van de oogbewegingsdata bekeken en werden alle fixaties handmatig gecontroleerd via het programma Fixation (Cozijn, 2006). De kwaliteit van de data bleek over het algemeen hoog.

Om te kunnen onderzoeken in hoeverre vwo 4-leerlingen onderscheid maken tussen kernzinnen en detailzinnen, werden de vier teksten opgedeeld in *areas of interest* (aoi's), waarbij zinnen afzonderlijk gecodeerd werden als kernzinnen en niet-kernzinnen. Daarop brachten we twee factoren in kaart die de leestijden van deze aoi's kunnen beïnvloeden. Ten eerste: het ligt voor de hand dat deelnemers langere aoi's langer zullen bekijken dan kortere aoi's. Daarom werd per aoi de hoeveelheid karakters (letters incl. interpunctie) berekend. Ten tweede: lezers zullen doorgaans de eerste zinnen van een alinea langer bekijken dan de vervolgzinnen (Rayner, 1998). Daarom codeerden we de kernzinnen die vooraan in een alinea stonden tot een afzonderlijke categorie. Deze beide factoren zullen in de data-analyse worden verdisconteerd.

Voor de vier teksten werden zo in totaal 51 aoi's vastgesteld, waarvan er 18 betrekking hadden op kernzinnen en 33 op niet-kernzinnen; een kernzin-aoi bestond daarbij doorgaans uit één enkele zin, terwijl een niet-kernzin-aoi's vaak meerdere zinnen in de tekst overspande. Van de 18 kernzin-aoi's stonden er zes voorop in de alinea. De gemiddelde lengte van een tekst-aoi was 137 karakters.

Voor het vooraf lezen van de teksten zijn vijf statistische modellen gespecificeerd. In het eerste model wordt een algemeen gemiddelde leestijd, de variantie tussen leerlingen, de variantie tussen teksten en de residuele variantie geschat. In de volgende modellen worden effecten een voor een toegevoegd zodat we het verschil in passing van de modellen eenduidig kunnen interpreteren. De aoi's verschillen in lengte; daarom wordt in het tweede model het effect van de hoeveelheid karakters op de leestijd van een aoi geschat. (Analyses vooraf hebben geleerd dat de relatie tussen kijktijd en lengte van de aoi niet-lineair is, en om die reden is niet alleen het lineaire maar ook het kwadratische lengte-effect geschat.) In het derde model schatten we de gemiddelde leestijden voor kernzinnen en niet-kernzinnen afzonderlijk. Omdat lezers voorop geplaatste zinnen in een alinea meestal meer leestijd schenken, wordt in het

vierde model het effect van deze alinea-openende kernzinnen afzonderlijk meegenomen. In het vijfde en laatste model ten slotte wordt toegestaan dat het lengte-effect varieert tussen kernzinnen en niet-kernzinnen.

Deze vijf modellen worden toegepast op de (natuurlijke logaritme van de) *first-pass dwell time* en de (natuurlijke logaritme van de) totale kijktijd per aoi, bekende leesmaten in de *eye-tracking*-literatuur. De *first-pass dwell time* geeft de totaaltijd weer van alle fixaties binnen een gecodeerde regio vóóordat er achteruit of vooruit wordt gekeken naar een andere regio, en zou daarmee indicatief zijn voor de eerste verwerking van een zin. De totale kijktijd geeft de totaaltijd van alle fixaties binnen een gecodeerde regio weer tijdens het gehele leesproces, verrekend dus alle regressies naar die regio toe, en incorporeert daarmee eerder complexere cognitieve processen zoals het oplossen van begripsproblemen (Holmqvist & Andersson, 2017).

Antwoord

De door participanten gegeven antwoorden bij de vragen werden gecorrigeerd door een van de onderzoekers aan de hand van het door het screeningspaneel vastgestelde correctiemodel; bij twijfel werden antwoorden voorgelegd aan leden van het screeningspaneel.

Beantwoorden

Om het oogbewegingengedrag tijdens het beantwoorden in kaart te brengen, ontwierpen we een tweede serie van *areas of interest*. De vraag zelf werd allereerst telkens in één aoi ondergebracht. Het screeningspaneel was vooraf gevraagd aan te geven welke tekstelementen voor de beantwoording relevant waren. Op basis hiervan konden we per vraag nog twee soorten aoi's onderscheiden: a) tekst-aoi's die voor beantwoording van de vraag relevant zijn, b) tekst-aoi's die voor beantwoording van de vraag irrelevant zijn. Omdat de lengte van een aoi waarschijnlijk van invloed zal zijn op de leestijd, bepaalden we voor elke aoi de hoeveelheid karakters. Ten slotte werd voor alle aoi's afzonderlijk de totale kijktijd berekend. De *first-pass dwell time* is hier overigens niet bruikbaar: ze ver-

onderstelt een lineaire leeswijze en het beantwoorden kenmerkt zich juist door scannend zoekgedrag en een interactie tussen tekst en vraag (Brunfaut & McCray, 2015; Rouet, 2006).

Voor het beantwoorden ontstonden bij een tekst minimaal zeven, maximaal elf tekst-aoi's van minimaal één, maximaal vier tekst-regels. In totaal werden er 33 tekst-aoi's onderscheiden, waarvan er naargelang de vraag doorgaans één of twee relevant waren voor de beantwoording. De achttien vragen zelf werden ondergebracht in achttien aparte aoi's. De gemiddelde lengte van een aoi bij het beantwoorden was 198 karakters.

Voor het beantwoorden zijn zes modellen gespecificeerd. In het eerste model wordt uitgegaan van een algemeen gemiddelde voor een antwoord-aoi. In de volgende modellen zijn effecten één voor één toegevoegd. Allereerst moet bij het antwoorden rekening worden gehouden met het feit dat aoi's in lengte variëren; daarom wordt in het tweede model het algemene effect van de aoi-lengte geschat. In het derde model voegen we het effect van het antwoord-aoi-type toe; we bekijken de gemiddelde leestijd voor een vraag, voor een antwoord-irrelevante aoi en voor een antwoord-irrelevante aoi afzonderlijk. De lengte kan bij deze drie aoi-typen natuurlijk een wisselende invloed hebben op de leestijd; in het vierde model wordt daarom toegestaan dat het lengte-effect varieert tussen de drie aoi-typen. In het vijfde model wordt het effect van een goed antwoord verdisconteerd: is er een samenhang tussen het geven van een goed dan wel fout antwoord en de leestijden in het antwoordproces? In het zesde model ten slotte wordt bekeken of het geven van een goed dan wel fout antwoord de leestijden van de drie antwoord-aoi-typen afzonderlijk beïnvloedt.

Om meer zicht te krijgen op de kwaliteit van het antwoorden, gebruikten we aanvullend een retrospectief *stimulated recall*-interview. Participanten werd hun antwoordgedrag bij de laatste set met vier vragen getoond en verzocht hierop te reflecteren. Daardoor kon aanvullend meer inzicht worden verkregen in de rationale achter hun antwoordgedrag, die immers nauwelijks van

uit *eye-track*-data is te deduceren (Holmqvist & Andersson, 2017). Uit de toelichting die participanten tijdens het interview gaven op hun eigen antwoordgedrag werd aan de hand van een observatieschema vastgesteld in hoeverre een participant direct na lezing van een vraag blijk gaf van a) een adequate tekstbasis, en b) een adequaat situatiemodel. Daarbij werd gescoord aan de hand van een vierpuntsschaal: apert aanwezig, voldoende aanwezig, enigszins aanwezig, niet aanwezig. Bij twijfel werden antwoorden voorgelegd aan leden van het screeningspanel.

Relatie vooraf lezen-antwoord

Om de relatie tussen het vooraf lezen en het gegeven antwoord te onderzoeken, werd eerst het aantal goede antwoorden per set met vragen berekend. Omdat de hoeveelheid vragen per set varieerde, berekenden we daarna de proportie van goede antwoorden per set. Deze proportie werd vervolgens gerelateerd aan de leestijden van kernzinnen en niet-kernzinnen bij het vooraf lezen.

Daartoe zijn drie modellen gespecificeerd. In het eerste model wordt weer uitgegaan van het model met de beste *fit* bij het vooraf lezen. In het tweede model wordt het effect van de proportie goede antwoorden op de vragen geschat op de leestijden bij het vooraf lezen. In het derde model wordt het mogelijke interactie-effect tussen de proportie goede antwoorden en de leestijden van kernzinnen en (niet-)kernzinnen verdisconteerd.

3 Resultaten

3.1. Vooraf lezen

Om het vooraf lezen te beschrijven, werden twee leesmaten afzonderlijk geanalyseerd: 1) *first-pass dwell time*: de totaal-tijd van alle fixaties binnen een regio vóórdat er voor- of achteruit wordt gekeken naar een andere regio; en (2) de totale kijktijd in een bepaalde regio. Wanneer een participant tijdens het lezen kort de laatste tekstregio fixeerte om daarna de overgeslagen aoi's lineair te lezen, kregen de overgeslagen aoi's bij de eerste leesmaat automatisch de waarde 0. Dit trad in 15 % van het aantal aoi's op. Deze regio's

Tabel 1

Fit en passingsvergelijking van vijf modellen voor (de logaritme van) de first-pass dwell time en de totale kijktijd (-2LL: -2log likelihood) bij het vooraf lezen

Model	-2LL	Modelvergelijking			
		Modellen	$\Delta\chi^2$	Δdf	p
<i>First-pass dwell time</i>					
1 intercept	2318.95				
2 + aoi-lengte	2043.34	2 vs 1	275.61	2	<.001
3 + kernzin	2043.11	3 vs 2	.23	1	.63
4 + kernzin vooraan in alinea	2038.75	4 vs 3	4.36	1	.04
5 + kernzin * aoi-lengte	2035.56	5 vs 4	3.19	1	.07
<i>Totale kijktijd</i>					
1 intercept	2030.70				
2 + aoi-lengte	843.70	2 vs 1	1187	2	<.001
3 + kernzin	827.34	3 vs 2	16.36	1	<.001
4 + kernzin vooraan in alinea	814.91	4 vs 3	12.43	1	<.001
5 + kernzin * aoi-lengte	813.5	5 vs 4	1.41	1	.24

werden uitgesloten uit de analyse van de *first-pass dwell time*. Regio's met incidentele 0-waarden bij de totale kijktijd (vrijwel steeds overgeslagen teksttitels) werden eveneens uitgesloten.

Uit tabel 1 blijkt dat de passing bij beide leesmaten drastisch verbetert wanneer we de lengte van de aoi's verdisconteren (Model 2 vs Model 1: $\Delta\chi^2(2) \geq 275.61$; $p < .001$). Wanneer we kernzinnen en niet-kernzinnen onderscheiden, genereert dat bij *first-pass dwell time* geen verbetering (Model 3 vs Model 2: $\Delta\chi^2(1) = .23$; $p = .60$), maar wel bij de totale kijktijd (Model 3 vs Model 2: $\Delta\chi^2(1) = 3.96$; $p = .05$). Toevoeging van de vooropplaatsing van de kernzin levert eveneens een betere passing op (Model 4 vs Model 3: $\Delta\chi^2(1) \geq 4.36$; $p \leq .04$). Het lengte-effect, waarin enkel het lineaire effect van de hoeveelheid karakters is betrokken, verschilt echter niet tussen kernzinnen en niet-kernzinnen (Model 5 vs Model 4: $\Delta\chi^2(1) \leq 3.19$; $p \geq .07$).

Tabel 2 geeft een schatting van de leestijden in *first-pass dwell time* en totale kijktijd voor kernzinnen en niet-kernzinnen, gebaseerd op model 4. Allereerst de *first-pass dwell time*: hier vinden we een duidelijk effect van de aoi-lengte, lineair ($\beta = .087$, $se = .006$) en kwadratisch ($\beta = -.002$, $se = .001$). Naarmate het aantal karakters toeneemt, kijken leerlingen duidelijk langer naar de zinnen

in de tekst. Er is echter geen significant algemeen onderscheid aantoonbaar tussen kernzinnen en niet-kernzinnen ($\Delta\beta = -.047$, $se = .091$; $p = .60$). De kernzinnen die vooraan in de alinea staan, krijgen wel duidelijk meer aandacht dan niet-kernzinnen ($\beta = .277$, $se = .132$).

Dan de totale kijktijd: hier vinden we weer het te verwachten effect van aoi-lengte, zowel lineair ($\beta = .105$, $se = .004$) als kwadratisch ($\beta = -.003$, $se < .001$). We vinden een (nipt) significant onderscheid tussen kernzinnen en niet-kernzinnen: kernzinnen worden lichtjes langer gelezen dan (niet-)kernzinnen ($\Delta\beta = .072$, $se = .036$; $p = .05$). Staat een kernzin daarbij voorop in een alinea, dan wordt hij duidelijk langer bekeken dan een niet-kernzin ($\beta = .187$, $se = .053$).

Een vergelijking tussen de waarden bij beide leesmaten laat zien dat het effect van de aoi-lengte vrijwel gelijk blijft. Kernzinnen worden bij de *first-pass dwell time* niet langer bekeken dan niet-kernzinnen, maar wel lichtjes langer bij de totale kijktijd. Kernzinnen die voorop staan in de alinea, worden daarentegen zowel in de *first-pass dwell time* als in de totale kijktijd duidelijk merkbaar langer bekeken.

Vergelijken we de gemiddelde (niet-)kernzinwaarden bij de beide leesmaten, dan lijken die van de totale kijktijd duidelijk hoger te

Tabel 2

Parameterschatting van de (logaritme van de) leestijd van zinnen (karakters) per leesmaat

Parameter	First-pass dwell time		Totale kijktijd	
	Regressiegewicht	(se)	Regressiegewicht	(se)
kernzin	8.020	(.095)	8.645	(.057)
niet-kernzin	8.067	(.082)	8.572	(.053)
kernzin vooraan in alinea	.277	(.132)	.187	(.053)
lengte ^a	.087	(.006)	.105	(.002)
lengte ²	-.002	(.001)	-.003	(.001)
<i>Varianties</i>				
S ² tekst	.003	(.001)	.061	(.003)
S ² leerling	.056	(.031)	.155	(.010)
S ² residu	.988	(.053)	.408	(.009)

^a gecentreerd rond gemiddelde (136.8 karakters) en gedeeld door 10

liggen: na eerste lezing lijken leerlingen relatief vaak terug te kijken naar eerdere aoi's. Hierin dient nog verdisconteerd dat de *first-pass dwell time* de (geringe) duur van de saccades meeneemt en de totale kijktijd niet. Het aandeel terugkijken naar eerdere aoi's in de oogbewegingen ligt dus nog wat hoger.

Wordt er veel teruggekeken naar eerdere aoi's? De correlatie tussen beide leesmaten is hoog ($r = .61$; $p < .001$). Wanneer we deze correlatie atteneren voor onbetrouwbaarheid, dan blijken beide leesmaten niet significant van elkaar af te wijken ($r = 1.0$). Er is geen reden om *first-pass dwell time* en de totale kijktijd als twee verschillende leesprocessen te beschouwen: de hoeveelheid regressie naar eerdere aoi's tijdens het gehele proces is te verwaarlozen.

Ten slotte: de residuele variantie bij beide leesmaten is hoog, vooral bij *first-pass dwell time*. Daarvoor zijn meerdere verklaringen. Een deel van deze variantie ontstaat door de interactie tussen leerling en tekst en door een gedeelte error. Een ander belangrijk deel wordt waarschijnlijk veroorzaakt door het tamelijk onsystematische karakter van oogbewegingen, dat zich nog het sterkst voordoet bij de *first-pass dwell time*. Men denke aan 'standaard' regressies (Rayner, 1998): lezers fixeren kort een volgende zin, om dan vanuit technische leesproblemen even terug te keren naar de zojuist gelezen eerdere zin. In dat opzicht genereert elke zin ander, individueel

bepaald leesgedrag. Het gedrag van leerlingen daarentegen is betrekkelijk stabiel te noemen, terwijl de verschillen ten gevolge van teksten relatief gering zijn.

3.2 Beantwoorden

Eye-trackresultaten

Voor de beschrijving van het beantwoorden werd maar één leesmaat in de analyse betrokken: de totale kijktijd. Uit tabel 3 blijkt dat de passing van een model waarin de lengte van de antwoord-aoi's wordt meegenomen, verschilt van het intercept (Model 2 vs Model 1: $\Delta\chi^2(2) = 11.57$; $p = .003$). Toevoeging van de drie antwoord-aoi-typen levert een duidelijk betere passing op (Model 3 vs Model 2: $\Delta\chi^2(2) = 1181.33$; $p < .001$), maar we vinden geen verbetering wanneer we het effect van de lengte laten variëren tussen de drie aoi-typen (Model 4 vs Model 3: $\Delta\chi^2(2) = 4.49$; $p = .11$). Een model dat de invloed van het geven van een goed antwoord op de leestijden in het antwoord meeneemt, heeft wel een betere passing (Model 5 vs Model 4: $\Delta\chi^2(1) = 3.71$; $p = .05$), evenals een model waarin we het effect van een goed antwoord laten verschillen tussen de drie aoi-typen (Model 6 vs Model 5: $\Delta\chi^2(2) = 15.49$; $p < .001$).

Tabel 4 geeft een schatting van de leestijden in totale kijktijd, gebaseerd op model 6, waarbij de toegevoegde parameter van model 4 (het effect van lengte verschilt tussen aoi-typen) is weggelaten. Voor het beantwoorden

Tabel 3

Fit en passingsvergelijking van zes modellen voor de (logaritme van de) totale kijktijd (-2LL: -2log likelihood) bij het beantwoorden

Model	-2LL	Modelvergelijking			
		Modellen	$\Delta\chi^2$	Δdf	p
1 intercept	7746.36				
2 + aoi-lengte	7734.79	3 vs 2	11.57	2	.003
3 + aoi-type	6553.46	2 vs 1	1181.33	2	<.001
4 + aoi-lengte * aoi-type	6548.97	4 vs 3	4.49	2	.11
5 + correct antwoord	6545.26	5 vs 4	3.71	1	.05
6 + correct antwoord * aoi-type	6529.77	6 vs 5	15.49	2	<.001

vinden we een duidelijk lengte-effect, in zowel lineair als kwadratisch opzicht (resp.: $\Delta\beta = .062$, $se = .012$; $\Delta\beta = -.001$, $se < .001$). Het effect van lengte wisselt echter niet tussen antwoord-relevante en -irrelevante aoi's. Uit de schattingen voor de drie antwoord-aoi-typen blijkt dat leerlingen tijdens het beantwoorden gemiddeld het langst naar de vraag kijken. En over het algemeen kijken leerlingen gemiddeld langer naar relevante antwoord-aoi's dan naar niet-relevante aoi's.

Er treedt een significant interactie-effect op tussen de antwoord-aoi-types en het geven van een goed antwoord. Leerlingen die een goed antwoord geven, kijken relatief korter naar de vraag dan leerlingen met een fout antwoord ($\Delta\beta = -.296$, $se = .127$; $p = .02$). Leerlingen met een goed antwoord kijken bovendien langer naar antwoord-relevante aoi's in de tekst ($\Delta\beta = .206$, $se = .101$; $p = .04$) en korter naar antwoord-irrelevante aoi's ($\Delta\beta = -.232$, $se = .078$; $p < .01$) dan leerlingen met een fout antwoord.

Aanvullend hebben we een extra model opgesteld waarbij op leerlingniveau de covarianties geschat zijn tussen de vijf factoren bij het vooraf lezen en beantwoorden: aan de ene kant kernzinnen en niet-kernzinnen bij het vooraf lezen, aan de andere kant de drie aoi-typen bij het beantwoorden.

Dan blijkt het volgende: als leerlingen bij het vooraf lezen kernzinnen snel of langzaam lezen, lezen ze ook de niet-kernzinnen snel of langzaam ($r = 1.0$). Als leerlingen bij het beantwoorden een vraag snel/langzaam lezen, lezen ze een antwoord-relevante of een antwoord-irrelevante aoi ook snel/langzaam

Tabel 4

Parameterschatting voor de (logaritme van de) totale kijktijd van antwoord-aoi's

Parameter	Coëfficiënt	se	p
vraag	9.776	.124	<.001
antw-relevante aoi	8.261	.139	<.001
antw-irrelevante aoi	7.135	.092	<.001
- aoi-lengte ^a	.062	.012	<.001
- aoi-lengte ²	-.001	.000	<.001
<i>Correct antwoord</i>			
Δ vraag * correct antw	-.296	.127	.02
Δ antw-relevante aoi * correct antw	.206	.101	.04
Δ antw-irrelevante aoi * correct antw	-.232	.078	<.01
<i>Varianties</i>			
S ² tekst	.027	.013	
S ² leerling	.026	.012	
S ² residu	1.024	.031	

^a gecentreerd rond gemiddelde (189 karakters) en gedeeld door 10

($r \geq .69$). En als leerlingen bij het vooraf lezen de kernzinnen of niet-kernzinnen snel/langzaam lezen, lezen ze bij het antwoorden ook de vraag snel/langzaam ($r \geq .79$). We vinden dus steeds relatief hoge correlaties tussen de vijf factoren wanneer de leesnelheid van de participanten een verklaring vormt.

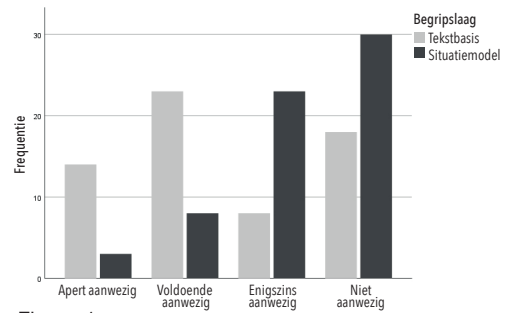
Is er ook een meer inhoudelijke samenhang aantoonbaar tussen vooraf lezen en beantwoorden? Als een leerling bij het vooraf lezen een kernzin of een niet-kernzin korter/

langer leest, dan hangt dit niet significant samen met het korter dan wel langer lezen van een antwoord-relevante of -irrelevante aoi bij het beantwoorden ($r \leq .51$; $p \geq .12$). Anders gezegd: leerlingen die vooraf kernzinnen extra aandacht geven, hebben tijdens het beantwoorden niet significant meer oog voor antwoord-relevante tekstdelen. En leerlingen die vooraf niet-kernzinnen langer aandacht geven, hebben niet significant meer oog voor antwoord-irrelevante tekstdelen. Een meer inhoudelijke samenhang tussen lezen en beantwoorden is dus niet aantoonbaar.

Retrospectief interview

In het interview zagen leerlingen hun antwoordgedrag terug bij de vier vraagitems uit de laatste tekst. De aanwezige onderzoeker vroeg hun aan te geven in hoeverre ze onmiddellijk na vraaglezing a) wisten waar in de tekst het antwoord te vinden was (tekstbasis), b) een antwoord in gedachten hadden (situatiemodel). Over het algemeen leverden de uitspraken van leerlingen tijdens het retrospectieve *stimulated recall*-interview geen interpretatieproblemen op. Zes reflecties bij een item werden voorgelegd aan leden van het screeningspanel. Bij één leerling bood de reflectie op de tekstbasis bij één item onvoldoende aanknopingspunten. Deze werd daarom uitgesloten.

Er is een duidelijk verschil waarneembaar in de verdeling van de tekstbasisscores en situatiemodelscores bij de laatste set met vragen ($\chi^2(3) = 24.62, p = .001$). Zie Figuur 1. Leerlingen gaven over het algemeen blijk van een redelijk aanwezige tekstbasis, waarbij ‘apert aanwezig’ en ‘voldoende aanwezig’ de meerderheid vormen. Bij de situatiemodelscores overheersen juist de indicatoren ‘enigszins aanwezig’ en ‘niet aanwezig’. Het onderscheid tussen beide begripsslagen bij ‘apert aanwezig’ is sterk. Bij tekstbasis scoorden leerlingen herhaaldelijk ‘apert aanwezig’: ze waren in staat erg precies een in de vraag vermeld citaat te lokaliseren. Bij het situatiemodel trad dit slechts in drie gevallen op: slechts in uitzonderlijke gevallen gaven leerlingen na vraaglezing direct antwoord of bleken ze een antwoord in hun hoofd te hebben.



Figuur 1

Frequentie van de begripsscores voor tekstbasis en situatiemodel

3.4 Relatie vooraf lezen-antwoord

Voor het verband tussen het vooraf lezen en het antwoord (het aantal goede antwoorden) zijn beide leesmaten geanalyseerd: *first-pass dwell time* en totale kijktijd. Uit tabel 5 blijkt dat de passing van een model waarin de proportie goede antwoorden per set wordt meegenomen, bij beide leesmaten niet verschilt van het basismodel (Model 2 vs Model 1: $\Delta\chi^2(1) \leq .49$; $p \geq .48$). Toevoeging van het interactie-effect tussen (niet-)kernzinnen en de proportie goede antwoorden levert evenmin verbetering op (Model 3 vs Model 2: $\Delta\chi^2(1) \leq .55$; $p \geq .46$). Er kan dus voor beide leesmaten geen verband aangetoond worden met het gegeven antwoord.

4 Discussie

In deze studie onderzochten we het vooraf lezen en vragen beantwoorden bij teksten met vragen bij zestien vwo 4-leerlingen. We bespreken eerst het vooraf lezen, vervolgens gaan we in op het beantwoorden en de relatie tussen het vooraf lezen en het beantwoorden; we eindigen met enkele algemene slotobservaties.

4.1 Vooraf lezen

De als eerste geformuleerde onderzoekshypothese luidde: vwo bovenbouw-leerlingen bouwen tijdens het vooraf lezen over het algemeen een adequate tekstbasis en adequaat situatiemodel op, en ze zullen dus meer leestijd wijden aan kernzinnen dan aan niet-kernzinnen. Er is bij de totale kijktijd inder-

Tabel 5

Fit en passingsvergelijking van drie modellen voor de (logaritme van de) first-pass dwell time en totale kijktijd (-2LL: -2log likelihood) voor de relatie tussen vooraf lezen en het antwoord

Model	-2LL	Modelvergelijking			
		Modellen	$\Delta\chi^2$	Δdf	p
<i>First-pass dwell time</i>					
1 intercept	2038.75				
2 + proportie correct	2038.55	2 vs 1	.20	1	.65
3 + proportie correct* (niet-)kernzin	2038.38	3 vs 2	.17	1	.68
Totale kijktijd					
1 intercept	814.91				
2 + proportie correct	814.42	2 vs 1	.49	1	.48
3 + proportie correct* (niet-)kernzin	813.87	3 vs 2	.55	1	.46

daad sprake van meer leestijd voor kernzinnen dan niet-kernzinnen, en daarmee ondersteunt dit onderzoek de bevindingen uit eerdere studies dat geoefende lezers bij het lezen van teksten deze twee zinstypen onderscheiden (Kintsch, 1998; Hyönä & Niemi, 1990; Yeari et al., 2015). In vergelijking met deze eerdere studies valt het in dit onderzoek op dat de positieve kernzindiscriminatie alleen bij de totale kijktijd optreedt en niet in de *first-pass dwell time*. Dat laatste troffen Yeari et al. (2015) wel aan in hun onderzoek-conditions.

Deze bevinding kan daarmee de uitspraak onderbouwen van de Expertgroep Taal (2008) dat leerlingen op niveau 3F onderscheid maken tussen hoofd- en bijzaken: over het algemeen is er in deze studie sprake van het scheiden van hoofd- en bijzaak. Wel dienen we terughoudend te zijn met deze conclusie. Deze studie toont enkel aan dat dit effect zich over de groep van de zestien leerlingen als geheel voordeed. Bovendien troffen we dit onderscheid aanmerkelijk minder geprononceerd aan in de *first-pass*-maat.

Andere uitspraken van de Expertgroep Taal lijken we vanuit deze studie eerder te kunnen relativiseren. Binnen de subdomeinen 'begrijpen' en 'interpreteren' van zakelijk lezen wordt in niveaubeschrijving 3F verondersteld dat leerlingen bij het lezen van lastige teksten enkele niet-lineaire leesstrategieën inzetten, zoals lastige tekstdelen herlezen, opgestelde betekenissen herzien en

de tekst scannen om alineafuncties te bepalen – bekende strategieën van expert-lezers (Pressley & Afflerbach, 1995; Duke & Pearson, 2002), strategieën ook die in bovenbouwmethodes Nederlands nadruk krijgen. De twee leesmaten *first-pass dwell time* en totale kijktijd blijken echter in leestijd niet duidelijk van elkaar af te wijken. Daardoor is het onaannemelijk dat de onderzochte groep vwo 4-leerlingen deze strategische middelen bij het vooraf lezen actief en veelvuldig heeft ingezet. Ook een ander in de marge ontwikkeld meetinstrument liet zien dat er nauwelijks sprake was van strategisch leesgedrag: leerlingen vertoonden bij het vooraf lezen zelden 'grote regressies', ze sprongen, met andere woorden, zelden terug naar eerdere tekstdelen (Hyönä, Lorch, & Kaakinen, 2002)¹. We kunnen wel stellen dat leerlingen vooral lineair lezen.

We moeten voorzichtig zijn hiermee harde conclusies te trekken over de kwaliteit van opgebouwde tekstbases en situatiemodellen bij het vooraf lezen van de vwo 4-leerlingen: immers, vanuit de Kintschiaanse basisaanname van dit onderzoek – lezers maken onderscheid tussen kernzinnen en niet-kernzinnen – mogen we aannemen dat er een zekere tekstbasis en situatiemodel in het algemeen aanwezig zijn. Maar de overige bevindingen wijzen vooral op een tamelijk oppervlakkig leesproces, zoals dat in enkele andere studies bij lezende vwo-leerlingen is aangetroffen (Stevens, 2018; Breukink, [in voorbereiding]). De

vraag kan dan ook worden opgeworpen of het hier beschreven leesgedrag, zeker bij de moeilijker teksten, volstaat om de teksten adequaat te doorgronden.

4.2 Beantwoorden

Voor het *beantwoorden* vertrokken we in deze studie vanuit twee hypothesen. Allereerst: de leerlingen bouwen tijdens het vooraf lezen ten minste een adequate tekstbasis op en besteden dus aanzienlijk meer tijd aan antwoord-relevante tekstdelen dan aan antwoord-irrelevante tekstdelen. Inderdaad, de deelnemers schenken bij de vragen meer aandacht aan antwoord-relevante dan aan antwoord-irrelevante tekstdelen. Tegelijk is de gemiddelde tijd die aan niet-relevante tekstdelen wordt besteed aanzienlijk: leerlingen kunnen dus eventueel de vereiste tekstbasis pas tijdens het beantwoorden hebben opgebouwd.

Nadere duiding biedt hier het retrospectieve interview, dat inzicht geeft in de rationale achter het antwoordgedrag bij de laatste set met vragen. Direct na de vraaglezing blijkt er over het algemeen sprake van een redelijk aanwezige tekstbasis: participanten geven bij de vier bevraagde items aan dat ze na vraaglezing doorgaans weten in welke alinea zich het antwoord bevindt. We lijken dus te mogen stellen dat leerlingen tijdens vooraf lezen inderdaad een redelijke tekstbasis construeren.

De tweede hypothese – dat leerlingen tijdens vooraf lezen een adequaat situatiemodel hebben opgebouwd en daarmee snel en adequaat antwoorden – kunnen we vanuit de *eye-track*-data niet duidelijk beantwoorden. In het interview echter geven leerlingen bij de vier bevraagde vraagitems meestal aan dat ze het voor het antwoord noodzakelijke situatiemodel pas construeren tijdens het beantwoorden van de vraag. Dat versterkt de eerdere conclusie bij het vooraf lezen dat leerlingen tamelijk oppervlakkig lezen.

We vonden verder in deze studie, zoals valt te verwachten, een duidelijke relatie tussen het beantwoorden en het uiteindelijke antwoord: leerlingen met een goed antwoord kijken beduidend langer naar relevante tekstdelen dan leerlingen met een onjuist ant-

woord; leerlingen met een goed antwoord kijken ook beduidend korter naar irrelevante tekstdelen. Ze kijken bovendien korter naar de vraag dan leerlingen met een onjuist antwoord. De verklaring lijkt eenvoudig: leerlingen die de vraag snel doorgronden, begrijpen beter waar het antwoord te vinden is en formuleren gemakkelijker een goed antwoord, terwijl leerlingen die veel moeite hebben met de vraag, vaker de vraag lezen, meer moeite hebben het antwoord in de tekst te vinden en ook minder snel een goed antwoord hebben.

4.3 Relatie vooraf lezen - beantwoorden

Voor de *relatie tussen vooraf lezen en beantwoorden* formuleerden we op situatiemodelniveau twee verwachtingen. Ten eerste: lezers die tijdens het vooraf lezen langer naar kernzinnen hebben gekeken dan naar niet-kernzinnen, zullen bij het beantwoorden van vragen langer kijken naar de voor beantwoording noodzakelijke tekstdelen, dan lezers die kernzinnen *niet* positief onderscheiden. We hebben deze hypothese niet kunnen bevestigen. Waar we in de analyse van de covarianties wel hoge correlaties binnen het vooraf lezen aantreffen én binnen het beantwoorden, vonden we tussen het lezen en beantwoorden enkel een correlatie tussen het vooraf lezen en het lezen van de vraag, en daarvoor vormt leessnelheid een voor de hand liggende verklaring. Maar voor het overige gedragen vooraf lezen en beantwoorden zich hier als twee op zichzelf staande gehelen.

De tweede verwachting luidde: lezers die langer naar kernzinnen en/of korter naar niet-kernzinnen kijken, zullen vaker een goed antwoord op een vraag geven. Deze verwachting konden we evenmin bevestigen. Leerlingen met een goed antwoord kijken bij het vooraf lezen over het algemeen niet korter of langer naar de tekst, noch kijken ze korter of langer naar kernzinnen en niet-kernzinnen.

We konden al met al op situatiemodelniveau geen relatie aantonen tussen het vooraf lezen en het beantwoorden.

4.4 Slotobservaties

Enkele algemene voorbehouden bij deze studie zijn op hun plaats. Het betreft allereerst een *eye-track*studie: de aanname dat kijken

naar een aoi een directe weerspiegeling is van de aandacht die een participant aan deze aoi schenkt, staat nog altijd ter discussie (vgl. Holmqvist & Andersson, 2017); eye-tracking biedt daarnaast alleen zicht op het *hoe* en niet op het *waarom* achter lezen. Weliswaar vullen we deze lacune deels door een retrospectief interview in te zetten, dit had enkel betrekking op de laatste vier vragen bij een tekst op niveau 3F. Daarnaast betreft het een onderzoek naar het lezen en antwoorden op één Nederlandse school, bij slechts vier relatief korte teksten en een beperkt aantal vragen, onder zestien proefpersonen.

Overigens, vanwege de arbeidsintensiviteit van de onderzoeksmethode zijn de steekproeven bij vergelijkbare *mixed method*-eye-trackstudies niet veel groter (vgl. Bax, 2013; Brunfaut & McCray, 2015). De technische uitval van participanten is bovendien voor een *eye-tracking*-studie laag: vanwege de voor *eye-tracking* bijzondere lengte van het experiment kozen we voor een meermaals herhaalde kalibratie- en validatieprocedure die zeker aan de hoge kwaliteit van de data heeft bijgedragen. Deelnemers schatten ook hun motivatie en concentratie tijdens het experiment meestal in als ruim voldoende tot goed; incidentele opgaven van onvoldoende motivatie en/of concentratie lijken de data niet significant te hebben beïnvloed, zo bleek uit aanvullende analyses.

Inhoudelijk is het in deze studie dan allereerst opvallend dat we niet bij beide leesmaten meer leestijd voor kernzinnen vonden dan voor niet-kernzinnen, maar enkel bij de totale kijktijd. Daarvoor kunnen verschillende verklaringen zijn, zoals de beperktere leeservaring van de steekproef en vooral ook de heterogeniteit binnen de groep. Maar het kan ook samenhangen met de afwijkende experimentinrichting. Eerdere *eye-track*-studies naar kernzindetectie vertrokken vanuit *self-paced reading* (Mross, 1989; Hyönä & Niemi, 1990), óf vergeleken de leestijd van kernzinnen met die van perifere zinnen (Yeari et al., 2015). In deze studie is het natuurlijk lezen dichter benaderd dan bij *self-paced reading* en wordt de leestijd van kernzinnen vergeleken met de leestijd van de rest van de tekst. Dat we onder deze omstandigheden

geen extra leestijd voor kernzinnen vinden bij het vooraf lezen, kan komen doordat een lezer pas bij het opnieuw lezen prioriteit aan kernzinnen toekent. Kintsch' aanname van een vrijwel geautomatiseerde kernzindetectie zou dan kunnen worden betwijfeld. Dergelijke kritiek ligt in lijn met het vaak geuite bezwaar dat in Kintsch' tekstbegripstheorie de ideale lezer te zeer centraal staat (McNamara & Magliano, 2009).

Ten tweede: bij de tekst met vragen wordt doorgaans verondersteld dat het vooraf lezen sterk het antwoord bepaalt. We konden op situatiemodel echter nauwelijks tot niet een relatie tussen het vooraf lezen en het beantwoorden vaststellen. Wat we niet hebben kunnen aantonen, kan er natuurlijk wel degelijk zijn. Maar er tekent zich in de resultaten een duidelijk patroon af. We vonden enkel een voorzichtige 'zwakke' relatie tussen het vooraf lezen en het beantwoorden op tekstbasisniveau. De participanten lazen vooraf vooral lineair, tamelijk oppervlakkig en bouwden waarschijnlijk pas tijdens het beantwoorden een meer dan rudimentair situatiemodel van de teksten op, ondanks de zeer nadrukkelijke instructie vooraf om de teksten terdege te bestuderen en het soms pittige karakter van de teksten. Het antwoordproces lijkt daarmee inhoudelijk vooral op zichzelf te staan.

Werkt de tekst met vragen dus oppervlakkig lezen in de hand, zoals vaak is beweerd (bv. Hoogeveen & Bonset, 1998; SLO, 2016, 2017)? Die kritiek lijkt vanuit deze studie terecht. Wij vonden in deze studie vrijwel geen diepere samenhang tussen het vooraf lezen en het beantwoorden van de vraag. En dat is misschien ook niet zo verwonderlijk: leerlingen weten bij deze taak nu eenmaal niet wat ze na het lezen zal worden gevraagd. In internationale leestaken en -toetsen wordt voorafgaande aan het lezen vaak een leesdoel en specifieke taakcontext gegeven, maar in methodes en toetsen in het schoolvak Nederlands ontbreekt deze meestal. Dat het beantwoorden zelf centraal komt te staan en niet het lezen, lijkt een vanzelfsprekend gevolg van de huidige inrichting van de tekst met vragen in het schoolvak Nederlands. De suggestie om voorafgaande aan leestoetsen

steeds een context en doelstelling te formuleren lijkt daarom zinnig (vgl. Nederlands Nu! & Sectiebestuur Nederlands Levende Talen, 2018).

Bij eerdere studies bleek de tekst met vragen al een weinig valide toetsvorm om tekstbegrip op situatiemodelniveau te meten (Kamalski, 2007; Land, 2009). Vanuit deze studie is de vraag of leerlingen met deze taak wel leren lezen met dieper tekstbegrip: de vragen lijken de meeste aandacht te krijgen. De vraag is dus ook of deze taak wel zo geschikt is om tekstbegrip te trainen en of de huidige dominantie van de tekst met vragen in het schoolvak Nederlands als leestaak wel zo wenselijk is. Vervolgonderzoek zou zich moeten richten op de vraag of afwijkende inrichtingen van de tekst met vragen of alternatieve leestaken zoals de sorteertaak (Kamalski, 2007) geschikter zijn om valide tekstbegrip te oefenen en te trainen.

Noot

1. Een regressie werd als een 'grote regressie' beschouwd wanneer er meer dan twee tekstregels op het scherm (> 100 karakters) werd teruggekeken. In 64 % van de geobserveerde tekstlezingen trad er geen of slechts één grote regressie op; slechts drie tekstlezingen kenden vijf of meer grote regressies.

Literatuur

- Andre, T. (1979). Does answering higher level questions while reading facilitate productive reading? *Review of Educational Research*, 49, 280-318.
- Bax, S. (2013). The cognitive processing of candidates during reading tests: Evidence from eye-tracking. *Language Testing*, 30(4), 441-465.
- Breukink, C. (in voorbereiding). Secondary school students' text comprehension for poetry and prose: Missing cues and coherence.
- Brunfaut, T., & McCray, G. (2015). *Looking into test-takers' cognitive processes whilst completing reading tasks: a mixed-method eye-tracking and stimulated recall study*. (ARAGs Research Reports). British Council: s.l.

- Canton, J., Aler, T., Heemskerk, K., van der Westen, W., & Willemsen, K. (red.). (2013). *Effecten van sturing op discrepantie tussen de cijfers van het centraalexamen: Onderzoek naar de sturing door schoolleiders en de effecten daarvan op het taalonderwijs op havo en vwo en het school-examen bij de talen*. Utrecht: VLLT.
- Cerdán, R., Vidal-Abarca, E., Martínez, T., Gilabert, R., & Gil, L. (2009). Impact of question-answering tasks on search processes and reading comprehension. *Learning and Instruction*, 19(1), 13-27.
- Cozijn, R. (2000). *Integration and inference in understanding causal inferences*. Dissertatie. Tilburg: Faculteit der Letteren, KUB.
- Cozijn, R. (2006). Het gebruik van oogbewegingen in leesonderzoek. *Ephemerides Theologicae Lovanienses*, 28, 220-232.
- De Bruin, E. (2018, 26 oktober). Nederlands, wat is dat eigenlijk? [Interview met Marc van Oostendorp.] *NRC Handelsblad*.
- Duchowski, A. T. (2007). *Eye tracking methodology: Theory and practice*. Springer: London.
- Duke, N. K., & Pearson, P. D. (2002). *What research has to say about reading instruction*. Newark, DE: International Reading Association.
- Expertgroep Doorlopende Leerlijnen Taal. (2008). *Over de drempels met taal*. Enschede: SLO.
- Gerjets, P., Kammerer, Y., & Werner, B. (2011). Measuring spontaneous and instructed evaluation processes during Web search: Integrating concurrent thinking-aloud protocols and eye-tracking data. *Learning and Instruction*, 21(2), 220-231.
- Hamilton, R.J. (1985). A framework for the evaluation of the effectiveness of adjunct questions and objectives. *Review of Educational Research*, 55, 47-85.
- Holmqvist, K., & Andersson, R. (2017). *Eye Tracking: A comprehensive guide to methods, paradigms and measures*. CreateSpace Independent Publishing Platform.
- Hoogeveen, M., & Bonset, H. (1998). *Het schoolvak Nederlands onderzocht*. Leuven/Apeldoorn: Garant.
- Hulshof, H., Kwakernaak, E., & Wilhelm, F. (2015). *Geschiedenis van het talenonderwijs in Nederland: Onderwijs in de moderne talen van 1500 tot heden*. Passage: Groningen.
- Hyönä, J., & Niemi, P. (1990). Eye movements during repeated reading of a text. *Acta psy-*

- chologica*, 73(3), 259-280.
- Hyönä, J., Lorch Jr, R. F., & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology*, 94(1), 44-55.
- Kamalski, J. (2007). *Coherence marking, comprehension and persuasion: On the processing and representation of discourse*. Utrecht: LOT.
- Kintsch, W. (1998). *Comprehension: a paradigm of cognition*. Cambridge: Cambridge University Press.
- Land, J. (2009). *Zwakke lezers, sterke teksten? Effecten van tekst- en lezerskenmerken op het tekstbegrip en de tekstwaardering van vmbo-leerlingen*. Delft: Eburon.
- Lorch, R. F., Lorch, E. P., & Mogan, A. M. (1987). Task effects and individual differences in on-line processing of the topic structure of a text. *Discourse Processes*, 10(1), 63-80.
- Lorch Jr, R. F., Lorch, E. P., & Matthews, P. D. (1985). On-line processing of the topic structure of a text. *Journal of memory and language*, 24(3), 350-362.
- Magliano, J. P., Millis, K., Ozuru, Y., & McNamara, D. S. (2007). A multidimensional framework to evaluate reading assessment tools. In D. McNamara (ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (p. 107-136). Mahwah, NJ: Erlbaum.
- McNamara, D. S. (ed.) (2007). *Reading comprehension strategies: Theory, interventions, and technologies*. Mahwah, NJ: Erlbaum.
- McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of learning and motivation*, 51, 297-384.
- Meestringa, T., & Ravesloot, C. (2012). *Het school-examen Nederlands op havo en vwo: Verslag van een digitale enquête*. Enschede: SLO.
- Mross, E.F. (1989). *Macroprocessing in expository text comprehension*. Dissertatie. University of Colorado: Boulder.
- Narvaez, D., Van den Broek, P., & Barron-Ruiz, A. (1999). The influence of reading purpose on inference generation and comprehension in reading. *Journal of Educational Psychology*, 91, 488-496.
- Nederlands Nu! & Sectiebestuur Nederlands Levende Talen. (2018). *Advies Examens Nederlands [aangeboden aan curriculum.nu]*. S.I.
- Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior research methods*, 42(1), 188-204.
- O'Reilly, T., & McNamara, D. S. (2007). The impact of science knowledge, reading skill, and reading strategy knowledge on more traditional 'high-stakes' measures of high school students' science achievement. *American Educational Research Journal*, 44(1), 161-196.
- Ozuru, Y., Briner, S., Kurby, C. A., & McNamara, D. S. (2013). Comparing comprehension measured by multiple-choice and open-ended questions. *Canadian Journal of Experimental Psychology*, 67, 215-22.
- Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M.J. Snowling & C. Hulme (eds.), *The science of reading: A handbook* (p. 227-247). Blackwell Publishing: Malden.
- Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale: Erlbaum.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422.
- Robben, T. (2008). 'Over de drempels met taal en rekenen[: Interview met G. Rijlaarsdam & H. van den Bergh]'. *Levende Talen Magazine*, 95(5), 5-8.
- Rouet, J. F. (2006). *The skills of document use: From text comprehension to Web-based learning*. Mahwah: Lawrence Erlbaum Associates Publishers.
- SLO. (2016). Position paper SLO t.b.v. hoorzitting/ rondetafelgesprek examinering van het vak Nederlands in het voortgezet onderwijs d.d. 10 november 2016. Kamerstuk Tweede Kamer der Staten Generaal. Geraadpleegd via: <https://www.tweedekamer.nl/kamerstukken/detail?id=2016Z20696&did=2016D42622>
- SLO. (2017). *Leerplankundige analyse PISA 2015*. Arnhem: SLO.
- Stevens, N. (2018). Grip op tekstbegrip: Een onderzoek naar het tekstbegrip van proza en poëzie en de ontwikkeling ervan onder vwo-leerlingen in leerjaar 3 en 5. Masterscriptie. Utrecht: Universiteit Utrecht.
- Van Onna, M. Straat, H., & Alberts, R. (2014). *Verslag ijkingsonderzoek centrale examens Nederlands VO 2014 aan de referentiesets en*

referentieniveaus taal. Cito: Arnhem.

Van Silfhout, G. (2014). *Fun to read or easy to understand? Establishing effective text features for educational texts on the basis of processing and comprehension research*. Dissertatie. Utrecht: LOT.

Yeari, M., Van den Broek, P., & Oudega, M. (2015). Processing and memory of central versus peripheral information as a function of reading goals: Evidence from eye-movements. *Reading and writing*, 28(8), 1071-1097.

Auteurs

Patrick Rooijackers is docent Nederlands in het voortgezet onderwijs en promovendus aan de Universiteit Utrecht. **Gerdineke van Silfhout** is leerplanontwikkelaar bij de afdeling voortgezet onderwijs van Stichting Leerplanontwikkeling (SLO). **Uriël Schuurs** is toetsdeskundige Nederlands bij Cito. **Iris Mulders** is als senior onderzoeker verbonden aan het Utrechts Laboratorium voor Linguïstiek (UiL OTS) van de Universiteit Utrecht. **Huub van den Bergh** is als hoogleraar Didactiek en toetsing van het taalvaardigheidsonderwijs werkzaam aan de Universiteit Utrecht.

Correspondentieadres: Patrick Rooijackers, Universiteit Utrecht, Trans 10, 3512 JK Utrecht; E-mail: p.j.h.rooijackers@uu.nl

Abstract

Reading texts and answering questions; an eye-tracking study among Dutch tenth grade preacademic pupils

To measure and train the reading skills of students often the traditional 'text with questions' is used in Dutch secondary education. With this task it is often assumed that students who read the text well will answer more questions correctly. In this study the relationship between reading and answering process in this particular task is investigated by means of the analysis of eye-movements. Starting point is an assumption in the Construction-Integration theory of Kintsch (1998): to build an adequate comprehension

model readers have to give more attention to topical sentences than to detail sentences. It was investigated whether Dutch tenth grade preacademic students (N = 16) spent more reading time on topical sentences, and to what extent this is related to the answering process. Results show that participants paid more attention (longer reading times) to topical sentences, even if we control for differences in sentence length. However, on a deeper comprehension level a relationship between reading and answering process could not be established. The analysis indicates that participants initially read linearly and only in the answering process constructed a deeper comprehension of the text. The question is therefore to what extent the 'text with questions' in school subject Dutch fosters better text comprehension.

Keywords: text comprehension, preacademic, school subject Dutch, eye-tracking, text with questions.