

Lezen en antwoorden bij teksten met vragen

Een cross-sectionele eye-trackstudie onder 52 vwo-leerlingen

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

Samenvatting

In Nederland traint men vaak tekstbegrip met een tekst met vragen. Daarbij bestuderen leerlingen vooraf een tekst en beantwoorden vervolgens vragen, waarbij de tekst raadpleegbaar blijft. Bij deze taak werden verschillen in leesgedrag onderzocht tussen vwo2- ($n=17$), vwo4- ($n=16$) en vwo6- leerlingen ($n = 19$). Via oogbewegingenregistratie en retrospectieve interviews werd onderzocht in hoeverre jongere en oudere vwo-leerlingen ($N = 52$) in lees- en antwoordgedrag verschillen en hoe het vooraf lezen van een tekst in deze jaarlagen de erop volgende vraagbeantwoording beïnvloedt. De resultaten: bij de tekstbestudering vooraf doorbraken deelnemers slechts zelden het lineaire leespatroon; ze schonken aan kernzinnen nipt meer leestijd. Zesdeklassers en vierdeklassers lezen beduidend sneller dan tweedeklassers, maar zesdeklassers niet sneller dan vierdeklassers. Bij de vraagbeantwoording lokaliseerden zesdeklassers beter het antwoord in de tekst en antwoordden vaker correct dan tweedeklassers en (in mindere mate) vierdeklassers. In interviews gaven deelnemers aan bij vragen doorgaans te weten waar het antwoord in de tekst stond, maar nog geen (concept)antwoord paraat te hebben. Deelnemers vertoonden ogenschijnlijk sterk pragmatisch leesgedrag. De vraag wordt opgeworpen in hoeverre deze taak oppervlakkig lezen in de hand werkt en of andere taken niet beter zelfstandig 'diep lezen' kunnen bewerkstelligen.

Kernwoorden: tekstbegrip, leesproces, vwo-leerlingen, eye-tracking, tekst met vragen

1 Inleiding

1.1 Inleiding

Er is de afgelopen twintig jaar relatief veel onderzoek verricht naar het tekstbegrip van Nederlandse scholieren. Er zijn echter nog heel wat lacunes in onze kennis. Zo is het tekstbegrip van vmbo-scholieren weliswaar relatief scherp in kaart gebracht (bijv. De Milliano, 2013; Land, 2009; Schram, 2007), maar longitudinaal onderzoek naar hun ontwikkeling in tekstbegrip ontbreekt goeddeels. En er zijn slechts enkele deelstudies naar het tekstbegrip van vwo'ers (bijv. Van Silfhout, 2014). Er is ook internationaal leesonderzoek zoals PISA (Gubbels, Van Langen, Maassen, & Meelissen, 2019), dat 15-jarige vmbo- en havo- én vwo-leerlingen qua tekstbegrip internationaal op rang ordent. Maar voor vwo'ers ontbreken longitudinale of zelfs cross-sectionele studies naar hun ontwikkeling in tekstbegrip. Bij gebrek aan empirisch onderzoek naar tekstbegrip van vwo-leerlingen noemden de voorzitters van de Expertgroep Doorlopende Leerlijnen Taal (2008), het belangrijkste document over tekstbegrip in het Nederlandse secundair onderwijs, het gestelde eindniveau voor vwo 6 zelfs een "educated guess" (Robben, 2008).

Dit gebrek aan onderzoek naar vwo'ers is opvallend. Er zijn duidelijke zorgen over het tekstbegrip van instromende leerlingen op de Nederlandse universiteiten (Bonset, 2010; Raad voor de Nederlandse Taal en Letteren, 2015). Recent signaleerde het PISA-rapport over alle opleidingsniveaus in het voortgezet onderwijs een daling in tekstbegrip (Gubbels et al., 2019). En in het bijzonder tekstbegrip is een belangrijke voorwaarde voor het succes op vervolgstudies in het hbo en wo (bijv. Bogaert, Devlieghere, & Hacquebord, 2008).

Al is er weinig onderzoek verricht naar het tekstbegrip van vwo-scholieren, de

gemiddelde leesles Nederlands op het vwo laat zich redelijk scherp uittekenen. Uit vrij recent descriptief onderzoek naar de praktijk van de leeslessen Nederlands in onder meer vwo-klassen (Linthorst & De Gloppe, 2015), een verkenning van de inrichting van de schoolexamens Nederlands lezen (Meestringa & Ravesloot, 2012) en de drie dominante huidige vwo-methodes Nederlands (Van der Leeuw et al., 2017) rijst een vrij eenduidig beeld op. In de gemiddelde vwo-leesles wordt weinig tot geen aandacht besteed aan technische leesvaardigheid of aan hardop lezen; in belangrijke mate domineert expliciete strategie-instructie de les. En mede onder invloed van het landelijk Centraal Examen Nederlands vwo speelt het individueel schriftelijk beantwoorden van vragen bij een tekst een cruciale rol in de klas (vgl. Canton, Aler, Heemskerk, Van der Westen, & Willemsen, 2013; Meestringa & Ravesloot, 2012; Pronk-Van Eunen & De Vos, 2018; enz.).

Deze taak, de tekst met vragen, bepaalt in sterke mate wat momenteel bekend is over het tekstbegrip van vwo'ers – en dat is niet zonder problemen. Weliswaar kent de tekst met vragen een lange traditie in het schoolvak Nederlands (Wesdorp, 1981), de aloude, veelgehoorde kritiek erop luidt dat een lezer erdoor niet beter wordt in het doorgronden van teksten, maar alleen beter in het beantwoorden van vragen (Hoogeveen & Bonset, 1998). Die kritiek, die de laatste jaren herhaaldelijk de media bereikte (bijv. Welgraven & De Vos, 2019), lijkt niet onzinnig. Leerlingen ontvangen bij deze taak doorgaans de algemene instructie een tekst vooraf terdege te bestuderen, en daarna dienen ze de tekstafhankelijke vragen te beantwoorden, waarbij de tekst raadpleegbaar blijft. Van oudsher wordt daarbij verondersteld dat wie vooraf goed de tekst heeft gelezen, de tekstafhankelijke vragen ook goed zal beantwoorden (vgl. Wesdorp, 1981). Maar die aanname lijkt op gespannen voet te staan met internationaal onderzoek waaruit blijkt dat een leesdoel vooraf het lezen in zeer sterke mate bepaalt (Andre, 1979; Graesser, Singer, & Trabasso, 1994; Yeari, Van den Broek, & Oudega, 2015; etc.). Bij de tekst

met vragen weten leerlingen pas bij de vragen waarom ze precies de tekst moesten lezen. De vraag is, met andere woorden, of scores bij de traditionele tekst met vragen iets zeggen over het vooraf lezen of vooral over het beantwoorden van de vragen.

Er is dus alle reden om het lees- en antwoordgedrag van vwo-scholieren, specifiek bij deze taak, nader te bestuderen. Dat doen we in deze studie op relatief non-invasieve wijze via eye-tracking, en we onderzoeken dit cross-sectioneel in drie jaarlagen (vwo 2, 4 en 6) met een in complexiteit oplopende set stimuli. Omdat eye-tracking geen inzicht biedt in het waarom achter lees- en antwoordgedrag, zetten we ondersteunend 'retrospectieve *stimulated recall*-interviews' in.

Vwo-leerlingen behoren gemiddeld tot de betere lezers in het voortgezet onderwijs (vgl. Expertgroep Doorlopende Leerlijnen Taal, 2008; Gubbels et al., 2019; Van Silfhout, 2014; etc.). Wie verschillen in tekstbegrip tussen oudere en jongere vwo-leerlingen wil aantonen, zou dus logischerwijs moeten vertrekken vanuit de kenmerken van expert-lezers. Hun leesgedrag is redelijk scherp in kaart gebracht. In de studie van Pressley en Afflerbach (1995) zijn 38 hardopdenkstudies geanalyseerd om de algemene kenmerken van een volwassen lezer op te kunnen stellen. Een goede lezer blijkt zich vooral te kenmerken door de inzet van tekstbegrip bevorderende leesstrategieën, zoals leesdoelen stellen en voorkennis activeren tijdens het lezen, kenmerken die inmiddels hun weg hebben gevonden in tal van overzichtsstudies over tekstbegrip (Duke, Pearson, Strachan, & Billman, 2011; McNamara, 2007; Perfetti, Landi, & Oakhill, 2005; etc.).

Tekstbegrip is een verschijnsel dat zich enkel indirect laat beschrijven. Om zicht te krijgen op de constructie van begrip bij lezers, gebruikten onderzoekers in de jaren tachtig en negentig van de twintigste eeuw vooral hardopdenkprotocollen, (vgl. Pressley & Afflerbach, 1995). Een belangrijk nadeel van hardopdenkprotocollen is echter dat de verbalisatie tijdens het lezen het natuurlijke

handelingsproces nogal verstoort (bijv. 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; De Leeuw, Segers, & Verhoeven, 2016; Kraal, 2020; Van der Schoot, Reijntjes, & Van Lieshout, 2012; Van Silfhout, 2014). Eye-tracking registreert waarop en hoelang de lezer diens ogen fixeert. 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 lezen minder dan hardop denken. Eye-trackdata bieden evenwel geen zicht op het waarom achter het waargenomen leesgedrag (Bax, 2013; Cozijn, 2000).

Hoe toont men met eye-tracking verschillen aan in tekstbegrip tussen groepen jongere en oudere vwo-leerlingen? Een eerste, indirecte indicator kan de leessnelheid van een lezer zijn. Een snelle lezer is doorgaans een vaardiger lezer, een verschijnsel dat zich nog versterkt voordoet bij een orthografisch tamelijk regelmatige taal als het Nederlands (bijv. Jackson & McClelland, 1979; Perfetti, 1985; Wimmer, 1993). Vaardige lezers beschikken over het algemeen over een rijke woordenschat en een ruime situationele kennis en herkennen daarmee doorgaans vrijwel geautomatiseerd woorden. Daardoor lezen ze vaak relatief snel, waarmee ze meer cognitieve capaciteit beschikbaar hebben voor het oplossen van begripsproblemen dan minder vaardige lezers (Perfetti & Stafura, 2014). Natuurlijk hangt leessnelheid niet één op één samen met tekstbegrip (bijv. Wallot et al., 2014), goede lezers lezen over het algemeen wel sneller dan minder goede lezers. Kuijpers et al. (2003) vonden bijvoorbeeld dat Nederlandse volwassenen sneller lezen dan adolescenten en dat vwo-leerlingen gemiddeld sneller lezen dan mavoleerlingen. In de studie van Van Silfhout (2014) lazen de vwo-leerlingen

sneller dan de vmbo-leerlingen. Alleen al om ontwikkelingspsychologische redenen zullen de vwo 6-lezers dus naar verwachting gemiddeld sneller lezen dan de vwo 2-lezers.

Een tweede mogelijke benadering is directer aan tekstbegrip gerelateerd: het al dan niet onderscheiden van kernzinnen. Volgens de Constructie-Integratietheorie van Kintsch (1998), uitgangspunt van vele studies naar leesvaardigheid en leesprocessen (vgl. McNamara & Magliano, 2009), is een tekst een netwerk van hiërarchisch samenhangende, betekenisvolle beweringen (proposities), waarin een lezer relaties bouwt tussen woorden en zinnen (een microstructuur) en relaties tussen alinea's (een macrostructuur). Dit constructieproces mondt uit in een 'tekstbasis', een vrij letterlijke mentale representatie van de tekst, en een 'situatiemodel', een mentaal model waarin de lezer de tekstuele wereld integreert in zijn eigen voorkennis, opvattingen, etc. Pas op dit niveau kent een lezer aan een tekst diepere betekenis toe.

Een cruciale - en voor eye-tracking bijzonder relevante - aanname in Kintsch' theorie is dat aandachtige lezers vrijwel automatisch de inhoudsstructuur van een tekst herkennen en reconstrueren, hoofd- en bijzaken meteen onderscheiden. Meerdere onderzoeken hebben aangetoond dat volwassen lezers aan kernzinnen meer tijd (aandacht) besteden dan aan detailzinnen, zelfs al bij eerste lezing ervan (Hyönä & Niemi, 1990; Kintsch, 1998; Lorch, Lorch, & Matthews, 1985; Pressley & Afflerbach, 1995; Yeari et al., 2015). 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). De Expertgroep Taal (2008) merkt op dat leerlingen op niveau 2F (grotweg vwo 2) in een tekst van dit niveau hoofd- en bijzaken kunnen scheiden; leerlingen op niveau 4F (grotweg vwo 6) zouden dit onderscheid ook bij teksten op niveau 4F moeten kunnen aanbrengen (Expertgroep Taal, 2008). Als we leerlingen confronteren met in complexiteit wisselende teksten, ligt het daarom voor de hand aan te nemen dat vwo 6-leerlingen dit herkennen

van kernzinnen duidelijker zullen laten zien dan vwo 2-leerlingen, vooral bij de eerste lezing van zinnen.

Een derde manier om met eye-tracking verschillen in tekstbegrip te tonen is verbonden aan de inzet van leesstrategieën. Zoals gezegd, kenmerken expert-lezers zich door het adequate gebruik van leesstrategieën. Veel van deze strategieën - zoals 'goede lezers vergelijken en integreren hun voorkennis met de uitspraken in de tekst' en 'goede lezers construeren, reviseren en bevragen de concepten die ze tijdens het lezen opstellen' (Duke et al., 2011) - kenmerken zich door een herlezing van eerdere zinnen en tekstdelen (bijv. Hyönä & Niemi, 1990). Vaardigere lezers teksten zullen dus minder lineair, minder zin voor zin lezen dan minder vaardige lezers (vgl. ook Schotter, Tran, & Rayner, 2014). En omdat strategie-instructie in de vwo-leesles Nederlands vaak een prominente plaats inneemt (bijv. Linthorst & De Gloppe, 2015), valt te verwachten dat oudere vwo-leerlingen strategieën eerder breed in zullen zetten bij een gedifferentieerd scala aan teksten, dan jongere vwo-leerlingen. Met andere woorden: een vwo 6-lezer zal vaker regressies naar eerdere zinnen en tekstdelen vertonen dan een vwo 2-lezer.

De beantwoording van vragen, waarbij we eveneens verschillen willen laten zien, staat wat theorievorming betreft nog in de kinderschoenen. Het antwoordproces wordt vaak omschreven als een grillig, door velerhande cognitieve processen veroorzaakt complex (vgl. Kintsch, 1998; Rouet, 2006). Bestaande studies belichten vaak de relatief eenvoudig meetbare opzoekvaardigheid in teksten naar aanleiding van vragen (Cerdán, Vidal-Abarca, Martínez, Gilabert, & Gil, 2009; Rouet, 2006; Rouet, Vidal-Abarca, Bert-Erboul, & Millogo, 2011; Vidal-Abarca et al., 2011). In een paar studies werd de cognitieve validiteit van leestoetsen onderzocht (Bax, 2013; Brunfaut & McCray, 2015). Typerend genoeg bleek van de talloze hypothesen die Brunfaut en McCray (2015) over het beantwoorden bij een leestoets opstelden, een enkele empirisch aantoonbaar: vaardige lezers besteden minder tijd aan het

maken van een vraag dan de minder vaardige lezers. Ontwikkelingspsychologisch ligt het voor de hand aan te nemen dat een vwo 6-leerling op dezelfde vraag bij dezelfde tekst sneller en adequater zal antwoorden en vaker een goed antwoord zal formuleren dan een vwo 2-leerling.

Ten slotte: het is essentieel het vooraf lezen en het vragen beantwoorden ook op hun samenhang te onderzoeken. In het schoolvak Nederlands gaat men doorgaans ervan uit dat als leerlingen de tekst vooraf goed lezen, ze vervolgens vaker vragen goed beantwoorden (bijv. Wesdorp, 1981). Wie vooraf een tekst heeft gelezen, lijkt inderdaad beter te presteren op de aansluitende tekstbegripsvragen dan wie direct met de beantwoording van de vragen begint (Cerdán et al., 2009; vgl. ook Salmerón et al., 2015): leerlingen construeren waarschijnlijk vooraf een coherent beeld van de tekst, waardoor ze bij de vragen tot een dieper begrip van de tekst kunnen komen. In hoeverre *goed* lezen het resultaat hierbij bepaalt, is bij ons weten nog nauwelijks onderzocht. In Rooijackers, Van Silfhout, Schuurs, Mulders, en Van den Bergh (2020) bleken vwo 4-leerlingen tijdens vooraf lezen vooral een redelijk scherpe tekstbasis op te bouwen, maar pas tijdens de vragen tot een rijker geschakeerd, dieper begrip van de tekst te komen. Vanuit de vigerende didactiek in methodes Nederlands ligt het voor de hand aan te nemen dat naarmate vwo-leerlingen ouder worden, ze een tekst vooraf grondiger en adequater hebben leren bestuderen: ze zouden dan beter moeten scoren bij de vragen. De vraag is dus in hoeverre we hier verschillen tussen jaarlagen kunnen aantonen: vinden we bij de oudere leerlingen een sterkere relatie tussen vooraf lezen en vragen beantwoorden dan bij de jongere leerlingen?

1.2 Onderzoekshypothesen

Het doel van deze studie is een (cross-sectionele) ontwikkeling aan te tonen bij vwo-leerlingen bij teksten met vragen, de gebruikelijke leestaak in het schoolvak Nederlands in het vo. Daarom onderzoeken we cross-sectioneel het vooraf lezen van vier teksten en het beantwoorden van bijbehorende

vragen van vwo 2-, vwo 4- en vwo 6-leerlingen. Deze vwo-jaarlagen zijn grofweg te relateren aan resp. niveau 2F, 3F en 4F van de Expertgroep Taal (2008). Daarbij instrueren we, analoog aan de praktijk in het schoolvak Nederlands, leerlingen vooraf dat ze de tekst eerst grondig hebben bestudeerd voordat ze de vragen gaan beantwoorden. Onze basis-hypothesen zijn de volgende: we verwachten dat het vooraf lezen en het vragen beantwoorden een sterke relatie op tekstbasis- en situatiemodelniveau onderhouden (A). We gaan ervan uit dat naarmate leerlingen ouder worden, ze adequatere tekstbases en situatiemodellen opbouwen (B).

Daarvan uitgaande, poneren we dan voor *het vooraf lezen* (1) de volgende drie verwachtingen: (1a) oudere vwo-leerlingen zullen in het algemeen sneller lezen dan jongere vwo-leerlingen (vwo 6 < vwo 4 < vwo 2); (1b) oudere leerlingen maken nadrukkelijker onderscheid tussen hoofd- en bijzaken: ze besteden relatief meer aandacht (kijktijd) aan kernzinnen dan aan niet-kernzinnen; (1c) oudere leerlingen lezen meer strategisch: ze herlezen vaker eerdere zinnen en tekstdelen en lezen dus minder lineair.

Voor *het vragen beantwoorden* (2) verwachten we dat oudere vwo-leerlingen adequater antwoorden dan jongere leerlingen: (2a) oudere leerlingen hebben vaker een goed antwoord; (2b) ze geven gemiddeld sneller antwoord; (2c) ze kijken vaker en langer naar de voor het antwoord relevante tekstdelen; en (2d) ze kijken korter en minder vaak naar de voor het antwoord irrelevante tekstdelen.

Voor *de relatie tussen vooraf lezen en beantwoorden* (3) verwachten we, deels bij gebrek aan vooronderzoek, het volgende: (3a) als leerlingen bij het vooraf lezen relatief veel tijd besteden aan kernzinnen, dan kijken ze bij het beantwoorden ook langer en vaker naar de voor het antwoord relevante tekstdelen; (3b) als leerlingen relatief veel tijd besteden aan kernzinnen, kijken ze ook korter en minder vaak naar de voor het antwoord irrelevante tekstdelen; (3c) als lezers kernzinnen positief onderscheiden, hebben ze vaker een goed antwoord op de vragen; (3d) bij oudere vwo-leerlingen zijn deze

voorgaande drie verwachtingen duidelijker aantoonbaar dan bij jongere vwo-leerlingen.

Eye-trackingdata staan alleen indirect uitspraken toe over de door lezers geconstrueerde tekstbases en situatiemodellen. Vooral wat de relatie tussen vooraf lezen en vragen beantwoorden betreft zijn eye-trackdata al snel meerduidelijk: heeft een leerling bijvoorbeeld snel het antwoord op een vraag geformuleerd doordat hij de tekst vooraf goed heeft begrepen óf doordat hij snel het antwoord in de tekst vond? Om de relatie tussen vooraf lezen en vragen beantwoorden scherper te kunnen beschrijven, gebruiken we naast eye-tracking daarom 'retrospectieve *stimulated recall*-interviews', een werkwijze ontleend aan redelijk recente Britse studies naar antwoordgedrag (Bax, 2013; Brunfaut & McCray, 2015). Leerlingen wordt hun antwoordgedrag tijdens een deel van het experiment getoond en wordt gevraagd te reflecteren op de mate waarin ze hun antwoord baseerden op de initiële lezing van de teksten.

2 Methode

2.1 Participanten en experimentopzet

De afname van het experiment werd om technische redenen beperkt tot één grote collegiale middelbare school in het zuiden van Nederland. Deze school kent in de vwo-afdeling een voor Nederland representatieve leesdidactiek (zoals getypeerd in bijv. Linthorst & De Glopper, 2015): de leidende methode Nederlands wordt in alle jaarlagen gebruikt, en leerlingen krijgen als taak vooral teksten met vragen. De cijfers voor het landelijk Centraal Schriftelijk Examen Nederlands tekstbegrip cirkelden de laatste jaren rondom het landelijk gemiddelde.

Twintig vwo 2-, twintig vwo 4- en twintig vwo 6-leerlingen werden aselekt gekozen. Alle deelnemers zijn moedertaalsprekers Nederlands. Vier leerlingen besloten om persoonlijke redenen niet te participeren. Achteraf bleken de data van vier leerlingen geheel onbruikbaar vanwege detectieproblemen. Bij vijf leerlingen werden data deels uitgesloten vanwege detectie- of

Tabel 1

Uitval van (data van) de aangezochte twintig participanten per jaarlaag

(partiële) uitval	vwo 2 (n = 17)	vwo 4 (n = 16)	vwo 6 (n = 19)	totaal (N = 52)
geen toestemming	1	3	0	4
data geheel onbruikbaar	2	1	1	4
data deels onbruikbaar	2	3	0	5

technische problemen; doorgaans betrof het hier de detectie van een enkele gelezen tekst. In Tabel 1 is de uitval van de beoogde participanten over de drie jaarlagen veranschouwd.

In totaal deden 22 jongens en 30 meisjes mee aan het experiment, redelijk evenwichtig verspreid over de drie jaarlagen. De gemiddelde leeftijd van de vwo 2-leerlingen was 13;10 jaar ($SD = 5$ maanden), van de vwo 4-leerlingen 16;2 jaar ($SD = 6$ maanden) en van de vwo 6-leerlingen 17;10 jaar ($SD = 7$ maanden).

Het experiment was voor alle jaarlagen identiek in opzet en inhoud. Participanten ($N = 52$) bestudeerden op het computerscherm vier teksten, variërend in lengte van 240 tot 290 woorden. Na het vooraf lezen van de tekst volgden telkens vier of vijf vragen. Bij het maken van de vragen bleef de tekst steeds naast de vraag raadpleegbaar, zoals in het schoolvak Nederlands gebruikelijk. Leerlingen gaven hardop antwoord. De aanwezige onderzoeker noteerde hun antwoorden. Na elk tekstblok met vragen volgde een mogelijkheid tot pauze. Gedurende de hele *online* sessie werden de oogbewegingen van de deelnemers met een eye-tracker gevolgd. Na afronding van het eye-trackexperiment vulden de leerlingen een controlevragenlijst in, waarin naar algemene gegevens werd gevraagd maar ook naar een inschatting van hun motivatie en concentratie gedurende het experiment. Ten slotte volgde een retrospectief *stimulated-recall*-interview, waarbij leerlingen reageerden op een *scan-path* van hun antwoordgedrag bij de laatste set stimuli.

2.2 De stimuli

De inhoud en opzet van de methodes Nederlands en van het landelijk Centraal

Schriftelijk Examen Nederlands vwo waren leidraad in het samenstellen van de stimuli. Uit landelijke dagbladen of tijdschriften selecteerden we vier korte opiniërende of uiteenzettende teksten, in moeilijkheidsgraad wisselend van niveau 2F tot 4F. Per tekst construeerden we acht tot tien vragen, eveneens in moeilijkheid wisselend. Door teksten en vragen in moeilijkheid te spreiden, kon worden verwacht dat dezelfde stimuli voor alle deelnemers inzetbaar waren.

Een screeningspanel, bestaande uit twee gepromoveerde docenten taalbeheersing en twee geroutineerde bovenbouwdocenten Nederlands, screende al het materiaal uitvoerig en stelde het correctiemodel vast. Daarna volgde een papieren pretestafname onder telkens twee vwo 2-, vwo 4- en vwo 6- klassen, waarna per tekst vier of vijf items werden 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).

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 alinea- 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

de verbinding van een tekstpassage met de achtergrondkennis van een lezer; *bridging inference questions* vereisen dat een lezer twee of meer tekstpassages inhoudelijk verbindt. Deze twee vraagtypen hebben vooral op het situatiemodel betrekking (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 meer moeite ermee zullen ervaren de passage te vinden. Leerlingen met een redelijke tekstbasis én een adequaat situatiemodel zullen hun aandacht geheel of grotendeels schenken aan de voor beantwoording noodzakelijke passages of direct een antwoord geven of zullen direct een antwoord geven, zonder de tekst verder te raadplegen (vgl. Rouet, 2006).

Omdat we wilden weten of lezers meer aandacht schenken aan kernzinnen, verzochten we het screeningspaneel de kernzinnen in de vier teksten aan te wijzen. Daartoe ontvingen ze vooraf een uitgebreide schriftelijke instructie. De overeenstemming tussen de vier leden van dit paneel hierbij bleek hoog (Cronbachs $\alpha = .82$). Daarnaast gaf het paneel per vraag aan welke tekstpassages relevant zijn om de vraag correct te beantwoorden. Hierin bleken de vier leden van het paneel zeer sterk overeen te stemmen (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. 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 tamelijk courante snelheidsdrempelwaarde van 40° per seconde (vgl. Holmqvist & Andersson, 2017).

De deelnemers ontvingen vooraf een uitgebreide toelichting op het experiment door de onderzoeker. Deze zette de werking van de eye-tracker kort uiteen en benadrukte dat het een algemeen onderzoek naar de leesontwikkeling van vwo-leerlingen betrof. Hij 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 omwille van de detectiescherpte bij elke tekst herhaald werd. Deelnemers bepaalden zelf het tempo waarin de stimuli gepresenteerd werden en schakelden naar eigen believen naar een volgend scherm. Terugkeren naar een vorig scherm was niet mogelijk. Alle deelnemers kregen dezelfde stimuli in dezelfde volgorde voorgelegd. Er werd vooraf geen maximumduur aan het experiment gesteld.

Tijdens het retrospectief *stimulated recall*-interview kregen leerlingen een *scanpath* van hun antwoordgedrag te zien waarop ze werden uitgenodigd te reageren. De aanwezige onderzoeker vroeg daarbij na eerste lezing van een vraag steeds in hoeverre leerlingen wisten waar het antwoord stond en of ze al een antwoord in hun hoofd hadden.

De 52 deelnemers spendeerden gemiddeld 29.7 minuten ($SD = 7.9$) aan het *online* deel van het experiment. Het hele experiment, met introductie, pauze, interview en afronding, duurde per leerling doorgaans een zestigtal minuten.

2.4 Analyse

Vooraf lezen. Voorafgaand aan de analyse werd de kwaliteit van de oogbewegingsdata bekeken en werden alle fixaties handmatig gecontroleerd via het programma Fixation (Cozijn, 2006). De koppeling van fixaties aan bekeken regio's leverde doorgaans weinig problemen op. De kwaliteit van de data was over het algemeen hoog.

Om te kunnen onderzoeken in hoeverre vwo-leerlingen onderscheid maken tussen kernzinnen en detailzinnen, zijn de vier teksten opgedeeld in een sequentie van *area's of interest* (aoi's), waarbij we deze als kernzin-aoi of niet-kernzin-aoi codeerden. 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 (lettertekens 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 als een afzonderlijke categorie. Deze beide factoren zijn in de data-analyse verdisconteerd.

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

Voor de analyse wordt vooral gebruikgemaakt van multiniveau-modellen waarin observaties steeds genest zijn binnen leerlingen en binnen teksten.

Voor het vooraf lezen van de teksten zijn acht modellen gespecificeerd. In het eerste model schatten we een algemeen gemiddelde leestijd van een aoi (intercept), een variantie tussen leerlingen, een variantie tussen teksten en een residuele variantie. In de volgende modellen voegen we effecten één voor één toe 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 dan de gemiddelde leestijd voor kernzinnen en niet-kernzinnen. Omdat lezers doorgaans meer leestijd besteden aan de voorop geplaatste zinnen in een alinea, schatten we in het vierde model nog afzonderlijk het effect van deze alinea-

openende kernzinnen. In het vijfde en zesde model wordt bekeken of het effect van de lengte van aoi's wisselt tussen kernzinnen en niet-kernzinnen. En in het zevende model wordt onderzocht of de leestijden van aoi's in het algemeen per jaarlaag variëren, waarna we met het laatste model nagaan of er tussen jaarlagen verschillen bestaan in het lezen van kernzinnen en niet-kernzinnen.

Deze acht modellen worden toegepast op de (natuurlijke logaritme van de) *first-pass dwell time* en de (natuurlijke logaritme van de) totale kijktijd per aoi. De *first-pass dwell time* geeft de totaal tijd weer van alle fixaties binnen een gecodeerde regio vóórdat er achteruit of vooruit wordt gekeken naar een andere regio en is daarmee indicatief voor de eerste verwerking van een zin. De totale kijktijd geeft de totaal tijd van alle fixaties binnen een gecodeerde regio weer tijdens het hele leesproces; deze kijktijd verrekenet dus alle regressies naar deze regio en incorporeert eerder complexere cognitieve processen zoals het oplossen van begripsproblemen (Holmqvist & Andersson, 2017).

Om te bekijken in hoeverre leerlingen terugkijken naar eerdere zinnen in de tekst, kijken we op basis van het best passende model aanvullend nog naar de correlatie tussen *first-pass dwell time* en totale kijktijd: in hoeverre wijken beide maten van elkaar af? Daarnaast berekenen we het percentage regressietijd binnen de totale kijktijd: welk deel van de totale kijktijd van een aoi ontstond door herlezing, i.e. nadat een lezer vanuit een erop volgende aoi ernaar terugkeek?

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

Antwoordproces: eye-tracking. Om het oogbewegingengedrag tijdens het beantwoorden van de vragen in kaart te brengen, ontwierpen we een tweede serie van *area's 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 per vraag nog twee soorten aoi's worden onderscheiden: 1) tekst-aoi's die voor beantwoording van de vraag relevant zijn, 2) 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 berekend¹.

Voor het beantwoorden werden de achttien vragen dus ondergebracht in achttien afzonderlijke aoi's. Daarnaast ontwierpen we bij een tekst minimaal zeven, maximaal elf tekst-aoi's van minimaal een, maximaal vier tekstregels. In totaal werden er 33 tekst-aoi's onderscheiden, waarvan er naargelang de vraag doorgaans een of twee relevant waren voor het antwoord. De gemiddelde lengte van deze 51 aoi's samen was 198 karakters.

Voor het beantwoorden zijn tien modellen gespecificeerd. In het eerste model schatten we een algemene gemiddelde leestijd van een antwoord-aoi (intercept), waarna we in de volgende modellen effecten één voor één toevoegen. Allereerst moet rekening worden gehouden met het feit dat de antwoord-aoi's in lengte variëren; daarom wordt in het tweede model het lineaire effect van de aoi-lengte geschat. (Initiële analyses hebben geleerd dat hier enkel van een lineair lengte-effect sprake is, niet kwadratisch.) In het derde model voegen we het effect van het antwoord-aoi-type toe: een gemiddelde leestijd wordt geschat voor een vraag, voor een antwoord-relevante aoi en voor een antwoord-irrelevante aoi. De lengte kan bij deze drie aoi-typen natuurlijk de leestijd wisselend beïnvloeden; in het vierde model wordt daarom toegestaan dat het lengte-effect varieert tussen de drie aoi-typen.

Met de modellen 5 tot en met 7 onderzoeken we de relatie met jaarlaag. In model 5 wordt bekeken of de leestijd van een antwoord-aoi verschilt tussen jaarlagen. Model 6 staat toe dat de leestijden van de drie afzonderlijke antwoord-aoi-typen wisselen tussen jaarlagen. En in model 7 nemen we de interactie tussen jaarlaag en aoi-lengte mee.

Modellen 8 tot en met 10 bekijken de invloed die het geven van een goed antwoord op de leestijden had. In hoeverre het geven van een goed dan wel fout antwoord in het algemeen de leestijd beïnvloedt, wordt in het achtste model bekeken. Model 9 splitst dit uit voor de drie antwoord-aoi-typen afzonderlijk. In model 10 ten slotte bekijken we of het geven van een goed dan wel fout antwoord per jaarlaag een wisselend effect heeft.

In bovenstaande analyse blijft echter een belangrijke factor buiten beschouwing. Niet alleen de leestijd van de drie aoi-types kan wisselen tussen jaarlagen, maar ook het aantal bekeken aoi's: leerlingen zullen waarschijnlijk per vraag naar meerdere antwoord-relevante en antwoord-irrelevante aoi's hebben gekeken. Om dat te onderzoeken, stelden we aanvullend een multinomiaal model op, waarin we per aoi-type bekijken of de hoeveelheid bekeken antwoord-relevante en antwoord-irrelevante aoi's wisselt tussen jaarlagen, en tussen leerlingen met een goed en een fout antwoord. Daarin verdisconteren we wederom de aoi-lengte en schatten we de variantie tussen leerlingen, de variantie tussen teksten en de residuele variantie.

Antwoordproces: retrospectief interview. Om meer duidelijkheid te krijgen in het waarom achter het waargenomen antwoordgedrag, gebruiken we aanvullend een retrospectief *stimulated recall*-interview. We willen hier vooral weten in hoeverre leerlingen direct na het lezen van de vraag, op basis van de initiële tekstbestudering al een antwoord in hun hoofd hadden. Uit de reflectie van participanten op het eigen antwoorden werd na afname aan de hand van een van tevoren vastgesteld observatieschema vastgesteld in hoeverre een deelnemer meteen 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.²

De door participanten gegeven antwoorden werden gecodeerd door een van de onderzoekers aan de hand van dit observatieschema; bij twijfel werden antwoorden voorgelegd aan leden van het screeningspanel.

Tabel 2

Fit en passingsvergelijking van acht 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	ΔX^2	Δdf	p
<i>First-pass dwell time</i>					
1 Intercept	7675.83				
2 + lengte aoi ^a + lengte aoi ²	6395.83	2 vs 1	1280	2	<.001
3 + kernzin	6390.24	3 vs 2	5.59	1	.02
4 + kernzin vooraan in alinea	6385.42	4 vs 3	4.82	1	.03
5 + kernzin * lengte aoi ^a	6380.90	5 vs 4	4.52	1	.03
6 + kernzin * lengte aoi ²	6379.73	6 vs 5	1.17	1	.28
7 + jaarlaag	6355.34	7 vs 6	24.39	2	<.001
8 + jaarlaag * kernzin	6353.94	8 vs 7	1.4	2	.50
<i>Totale kijktijd</i>					
1 Intercept	6901.82				
2 + lengte aoi ^a + lengte aoi ²	2883.73	2 vs 1	4018.09	2	<.001
3 + kernzin	2854.52	3 vs 2	29.21	1	<.001
4 + kernzin vooraan in alinea	2821.70	4 vs 3	32.82	1	<.001
5 + kernzin * lengte aoi ^a	2816.51	5 vs 4	5.19	1	.02
6 + kernzin * lengte aoi ²	2815.13	6 vs 5	1.38	1	.24
7 + jaarlaag	2803.36	7 vs 6	11.77	2	<.01
8 + jaarlaag * kernzin	2801.81	8 vs 7	1.55	2	.46

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

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

Daartoe zijn vier modellen gespecificeerd. In het eerste model wordt uitgegaan van het model met de beste *fit* bij het vooraf lezen. In het tweede model wordt het effect van de proportie goede antwoorden 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. In het vierde model verkennen we of er in dit opzicht een verschil is waar te nemen tussen jaarlagen:

hebben leerlingen uit een jaarlaag meer of minder leestijd aan (niet-)kernzinnen besteed dan leerlingen uit een andere jaarlaag als ze een goed of fout antwoord hadden?

3. Resultaten

3.1 Algemeen

Allereerst: de deelnemers gaven aan dat ze zich tijdens de experimentafname gemiddeld ruim voldoende tot goed geconcentreerd voelden (op een tienpuntsschaal: $M = 7.5$, $SD = 1.3$) en goed gemotiveerd ($M = 7.9$, $SD = 1.0$).

3.2 Vooraf lezen

Passingsvergelijking. Voor de beschrijving van het vooraf lezen werden met behulp van multiniveau-modellen twee leesmaten geanalyseerd: (1) *first-pass dwell time*: de

Tabel 3

Parameterschattingen van de (logaritme van de) leestijd van zinnen per leesmaat

Parameter	First-pass dwell time		Totale kijktijd	
	Regressiegewicht	(SE)	Regressiegewicht	(SE)
niet-kernzin (vwo 6)	8.286	(.067)	8.561	(.056)
Δ kernzin	.032	(.045)	.035	(.021)
Δ kernzin vooraan in alinea	.119	(.065)	.157	(.030)
- lengte aoi ^a	.010	(.003)	.109	(.001)
- lengte aoi ²	-.003	(.001)	-.003	(<.001)
- kernzin * lengte aoi	-.012	(.005)	-.006	(.002)
<i>Jaarlaag</i>				
Δ vwo 2	.386	(.081)	.253	(.073)
Δ vwo 4	-.031	(.083)	.047	(.074)
<i>Varianties</i>				
S ² tekst	.003	(.003)	.002	(.002)
S ² leerling	.042	(.012)	.044	(.009)
S ² residu	.787	(.023)	.169	(.005)

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

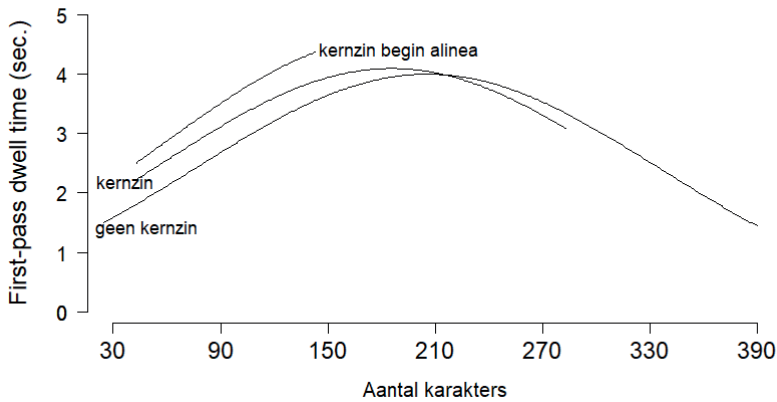
totaaltijd van alle fixaties binnen een gecodeerde regio vóórdat er voor- of achteruit wordt gekeken naar een andere regio; en (2) de totale kijktijd in een bepaalde regio. Wanneer een participant tijdens lezing kort de laatste tekstregio fixeerte om daarna de overgeslagen aoi's lineair te lezen, kregen de overgeslagen aoi's bij de eerste leesmaat automatisch de waarde 0. Deze regio's met 0-waarden (5.1 % van het aantal aoi's) werden uitgesloten uit de analyse van de *first-pass dwell time*, evenals de incidentele 0-waarden bij de totale kijktijd (1 %, vrijwel steeds overgeslagen teksttitels).

Uit Tabel 2 blijkt dat de passing van een model met één algemeen gemiddelde voor beide leesmaten sterk verbetert wanneer daarin ook de aoi-lengte wordt meegenomen (Model 2 vs Model 1: $\Delta\chi^2(2) \geq 1280$; $p \leq .001$). Toevoeging van het onderscheid kernzin/niet-kernzin levert eveneens bij beide een betere passing op (Model 3 vs Model 2: $\Delta\chi^2(1) \geq 5.59$; $p \leq .02$), evenals de toevoeging van voorop geplaatste kernzinnen (Model 4 vs Model 3: $\Delta\chi^2(1) \geq 4.82$; $p \leq .03$). Een model waarin de leestijd van kernzinnen of van niet-kernzinnen afhankelijk is van de

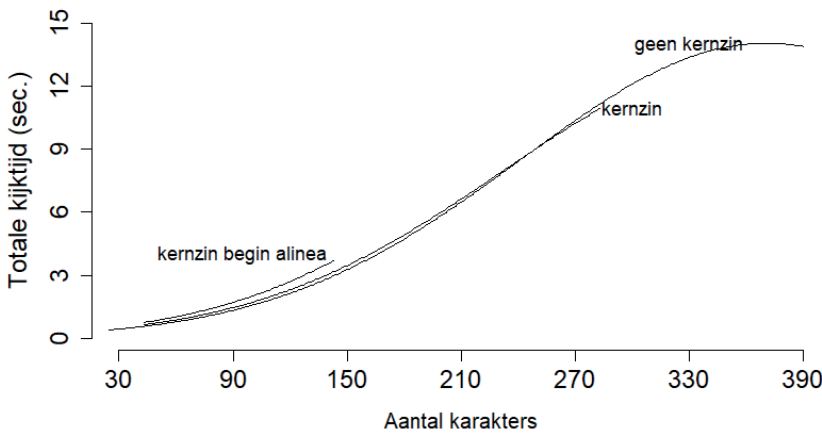
lineaire aoi-lengte, levert eveneens een verbetering op (Model 5 vs Model 4: $\Delta\chi^2(1) \geq 4.52$; $p \leq .03$); van een interactie tussen (niet-)kernzinnen en de kwadratische aoi-lengte is bij beide leesmaten evenwel geen sprake (Model 6 vs Model 5: $\Delta\chi^2(1) \leq 1.38$; $p \geq .24$). Toevoeging van de jaarlaag zorgt bij beide leesmaten voor een aanmerkelijk betere passing (Model 7 vs Model 6: $\Delta\chi^2(2) \geq 11.77$; $p \leq .001$), maar er blijkt geen interactie-effect tussen jaarlaag en (niet-)kernzinnen (Model 8 vs Model 7: $\Delta\chi^2(2) \leq 1.55$; $p \geq .46$). Hypothese 1b kan dus bij beide leesmaten niet worden aangetoond: oudere vwo-leerlingen kijken niet beduidend langer naar kernzinnen dan jongere vwo-leerlingen.

Leestijden first-pass dwell time. Tabel 3 geeft een parameterschatting van de leestijden in *first-pass dwell time* en totale kijktijd voor kernzinnen en niet-kernzinnen, gebaseerd op model 7, waarbij niet-significante parameters buiten de analyse zijn gelaten (model 6: het interactie-effect tussen kernzinnen en kwadratische aoi-lengte).

In Figuur 1 zijn de leestijden in de *first-pass*-maat voor de zesdeklassers weergegeven,



Figuur 1
 De first-pass dwell time van niet-kernzin-aoi's, kernzin-aoi's en aoi's van voorop geplaatste kernzinnen, afhankelijk van de aoi-lengte (aantal karakters), voor vwo 6, in seconden



Figuur 2
 De totale kijktijd van kernzin-aoi's, niet-kernzin-aoi's en aoi's van voorop geplaatste kernzinnen, afhankelijk van de aoi-lengte (aantal karakters), voor vwo 6, in seconden

gebaseerd op Tabel 3. Bij de *first-pass dwell time* hebben leerlingen (de zesdeklassers) gemiddeld $e^{8.286}$ (= 3968) milliseconden nodig om een niet-kernzin te lezen van gemiddelde aoi-lengte (189 karakters).

Hiermee is de leessnelheid van zesdeklassers duidelijk hoger dan die van tweedeklassers ($e^{(8.286 + .386)} = 5831$ ms), maar ze verschilt niet van de leessnelheid van vierdeklassers ($e^{(8.286 - .031)} = 3874$ ms). Ook vierdeklassers lezen gemiddeld beduidend

sneller dan tweedeklassers. Daarmee kunnen we hypothese 1a ('oudere vwo-leerlingen lezen sneller dan jongere vwo-leerlingen') deels bevestigen.

Vanzelfsprekend is de leestijd afhankelijk van de lengte van een aoi. Per tien karakters meer neemt de leestijd in eerste instantie toe, maar naarmate aoi's langer worden, neemt de leestijd juist opvallend af. De relatie met zinslengte is $(.010 * N_{\text{lengte aoi}} - .003 * N_{\text{lengte aoi}}^2)$. Die afname in leestijd ontstond doordat

bij langere aoi's vaker regressies optraden naar eerdere aoi's, waardoor de *first-pass*-waarde relatief laag is.

Wat kernzinnen betreft: aan betrekkelijk korte kernzinnen wordt meer leestijd besteed dan aan dito niet-kernzinnen, maar bij langere kernzinnen verdwijnt het onderscheid. Voor kernzinnen is de relatie $(.010 - .012) * N_{\text{ lengte aoi}} - .003 * N_{\text{ lengte aoi}}^2$. De kernzinnen die voorop in de alinea zijn geplaatst, worden duidelijk langer gelezen dan niet-kernzinnen.

Leestijden totale kijktijd. In Figuur 2 zijn de leestijden van niet-kernzinnen en beide typen kernzinnen weergegeven voor de totale kijktijd. Zesdeklassers gebruiken gemiddeld $e^{8.561}$ (= 5224) milliseconden voor het lezen van een niet-kernzin van gemiddelde lengte (189 karakters). Daarmee lezen vwo 6-leerlingen beduidend sneller dan vwo 2-leerlingen ($e^{(8.561 + .253)}$ = 6721 ms) maar niet sneller of langzamer dan vwo 4-leerlingen.

De leestijd is weer duidelijk afhankelijk van de aoi-lengte: naarmate een aoi langer wordt, stijgt de leestijd sterk. Deze toename vlakkt echter af bij heel lange niet-kernzinnen. Zie Figuur 2. De relatie met zinslengte is $(.109 * N_{\text{ lengte aoi}} - .003 * N_{\text{ lengte aoi}}^2)$.

Een kernzin van gemiddelde lengte wordt over het algemeen nauwelijks langer ($e^{(8.561 + .035)}$ = 5410 ms) bekeken dan niet-kernzinnen van gemiddelde lengte ($e^{8.561}$ = 5223 ms), al is het onderscheid tussen beide wel significant. Naarmate een kernzin langer dan gemiddeld wordt, verdwijnt het onderscheid met een niet-kernzin. Kernzinnen die vooraan in de alinea zijn geplaatst, worden wel duidelijk consequent langer bekeken ($e^{(8.561 + .035 + .157)}$ = 6330 ms, bij een gemiddelde lengte) dan niet-kernzinnen. Voor kernzinnen is de relatie $((.109 - 0.006) * N_{\text{ lengte aoi}} - .003 * N_{\text{ lengte aoi}}^2)$.

Regressies en varianties. In hoeverre keken leerlingen vaak terug naar eerdere zinnen of tekstdelen? Opvallend in Tabel 3 is dat de waarden in de totale kijktijd van de kortere aoi's relatief dicht bij de waarden in de *first-pass*-maat liggen. Wanneer voor alle jaarlagen de correlatie tussen beide leesmaten wordt bekeken, dan blijkt deze opvallend hoog ($r = .71$; $p < .001$): van niet-lineair leesgedrag

tussen aoi's lijkt over het algemeen beperkt sprake. Beide leesmaten lijken erg sterk geassocieerd. Als deze correlatie wordt geattenuëerd voor onbetrouwbaarheid, dan wijken de twee leesmaten zelfs niet van elkaar af ($r = 1.0$). Als daarnaast het percentage regressietijd binnen de totale kijktijd per aoi wordt berekend – dat deel van de totale kijktijd van een aoi dat ontstond door terugkijken –, dan blijkt 9 % van de totale kijktijd van een aoi te zijn besteed aan terugkijken: een laag percentage. Leerlingen blikten ogenschijnlijk dus over het algemeen weinig terug.

Als hierbij vervolgens de jaarlagen nader worden vergeleken, dan blijkt de correlatie tussen de *first-pass*-maat en de totale kijktijd voor vwo 2-leerlingen opvallend hoog ($r = .81$; $p < .001$). De correlatie voor vwo 4-leerlingen ligt in vergelijking daarmee lager ($r = .65$; $p < .001$), evenals die voor vwo 6-leerlingen ($r = .69$; $p < .001$). Het verschil van vwo 2 met vwo 6 en met vwo 4 is daarmee significant ($z \geq 5.09$; $p < .001$); dat tussen vwo 4 en 6 echter niet ($z = -1.04$; $p = .30$).

Het aandeel regressietijd binnen de totale kijktijd verschilt eveneens tussen jaarlagen ($F(2, 2515) = 4.96$, $p < .01$), waarbij enkel het onderscheid tussen vwo 2 en vwo 4 significant is ($p < .01$). Vwo 4-leerlingen kijken 11 % van de totale leestijd terug naar aoi's, 4 % meer dan vwo 2-leerlingen (die 7 % van de totale kijktijd eraan besteden). Vwo 6-leerlingen besteden 9 % van hun tijd hieraan. Hypothese 1c), 'oudere vwo-leerlingen lezen minder lineair dan jongere vwo-leerlingen', kunnen we dus bevestigen, al lijkt de omvang van dit niet-lineaire lezen gering.

Kijkt men dan ten slotte onder in Tabel 3 nog naar de varianties, dan blijkt vooral bij *first-pass dwell time* de residuele variantie zeer groot. Deze variantie, die bestaat uit de interactie tussen leerling en tekst, de interactie tussen aoi's en leerlingen en een gedeelte error, neemt beduidend af bij de totale kijktijd. Deze afname hangt hoogstwaarschijnlijk voor een belangrijk deel samen met het deels onsystematische karakter van oogbewegingen, dat zich vooral voordoet in de *first-pass*-maat: regelmatig keerden leerlingen na twee, drie fixaties in een nieuwe aoi weer terug naar de laatste woorden van de vorige aoi. De twee

Tabel 4

Fit en passingsvergelijking van tien 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	25563.41				
2 + lengte aoi	25527.79	2 vs 1	35.62	1	<.001
3 + aoi-type	22200.50	3 vs 2	3327.29	2	<.001
4 + aoi-type * lengte aoi	22187.60	4 vs 3	12.93	2	.001
5 + jaarlaag	22180.70	5 vs 4	6.84	2	.03
6 + jaarlaag * aoi-type	22141.34	6 vs 5	39.36	4	<.001
7 + jaarlaag * lengte aoi	22141.05	7 vs 6	.29	2	.87
8 + antwoord correct	22119.13	8 vs 7	21.92	1	<.001
9 + antwoord correct * aoi-type	22038.36	9 vs 8	80.77	2	<.001
10 + antwoord correct * jaarlaag	22037.23	10 vs 9	1.13	2	.57

andere varianties gedragen zich bij beide leesmaten vergelijkbaar: het aandeel variantie binnen leerlingen is behoorlijk, terwijl het aandeel variantie dat toegeschreven kan worden aan de teksten, opvallend beperkt is.

3.3 Beantwoorden

Het antwoord. Als het gemiddelde totaalaantal goede antwoorden per jaarlaag wordt bekeken, dan blijken vwo 2-leerlingen gemiddeld 7 van de 18 vragen goed te hebben ($SD = 2.5$), vwo 4-leerlingen 9.6 vragen ($SD = 2.2$) en vwo 6-leerlingen 11.3 vragen ($SD = 2.2$). Er is daarmee een duidelijk onderscheid tussen jaarlagen ($F(2, 51) = 15.95, p < .001$), waarbij zowel het onderscheid tussen vwo 2 en vwo 4/6 significant is ($p < .001$) als dat tussen vwo 4 en 6 ($p = .05$). Daarmee kunnen we dus hypothese 2a ('oudere leerlingen hebben vaker een goed antwoord') bevestigen.

Passingsvergelijking. Voor de beschrijving van het beantwoorden werd alleen de totale kijktijd in de analyse betrokken. Uit Tabel 4 blijkt dat het verdisconteren van de lineaire lengte van de antwoord-aoi's leidt tot passingsverbetering (Model 2 vs Model 1: $\Delta\chi^2(1) = 35.62; p < .001$). Als de drie verschillende aoi-typen (vraag-aoi, antwoord-relevante aoi en antwoord-irrelevante aoi) in leestijd verschillen, leidt dit tot verbetering (Model 3 vs Model 2: $\Delta\chi^2(2) = 3327.29; p < .001$), evenals wanneer de aoi-lengte van de

drie verschillende aoi-typen wordt meegenomen (Model 4 vs Model 3: $\Delta\chi^2(2) = 12.93; p = .001$). Een model dat bekijkt in hoeverre leestijden verschillen tussen jaarlagen, heeft een betere passing (Model 5 vs Model 4: $\Delta\chi^2(2) = 6.84; p = .03$). Bij nadere bestudering van de schattingen bij dit model blijken vwo 6-leerlingen (net) niet significant sneller te antwoorden dan vwo 2-leerlingen ($\beta = -.147; p = .06$), noch sneller of langzamer dan vwo 4-leerlingen ($p = .44$), maar vwo 4-leerlingen gemiddeld sneller te antwoorden dan vwo 2-leerlingen ($\beta = -.207; p = .01$), waarmee we hypothese 2b ('oudere lezers antwoorden sneller') deels kunnen bevestigen. Als we de leestijden van jaarlagen vervolgens laten verschillen tussen de drie aoi-typen, bewerkstelligt dit evenwel een zeer duidelijke verbetering (Model 6 vs Model 5: $\Delta\chi^2(4) = 39.29; p < .001$). Een model dat de interactie tussen jaarlagen en aoi-lengte meeneemt, doet dit niet (Model 7 vs Model 6: $\Delta\chi^2(2) = .29; p = .87$). Bij de vervolgmogelijkheden, die de invloed van het geven van een goed of fout antwoord op de leestijd onderzoeken, blijkt telkens een significante verbetering, behalve bij het interactie-effect met jaarlaag (Model 10 vs Model 9: $\Delta\chi^2(2) = 1.13; p = .57$).³

Leestijden aoi's antwoordproces. Tabel 5 geeft de parameterschattingen van de leestijden in totale kijktijd, gebaseerd op

Tabel 5
Parameterschattingen voor de (logaritme van de) totale kijktijd naar aoi's bij het beantwoorden, per jaarlaag

Parameter	coëfficiënt	SE
vraag-aoi (vwo 6)	9.646	.107
antwoord-relevante aoi (vwo 6)	8.320	.087
antwoord-irrelevante aoi (vwo 6)	7.356	.105
- vraag-aoi * aoi-lengte ^a	.057	.007
- antwoord-rel. aoi * aoi-lengte ^a	-.023	.008
- antwoord-irrel. aoi * aoi-lengte ^a	-.020	.008
<i>Jaarlaag</i>		
Δvwo 2 * vraag-aoi	.336	.108
Δvwo 2 * antwoord-relevante aoi	-.326	.108
Δvwo 2 * antwoord-irrelev. aoi	-.260	.099
Δvwo 4 * vraag-aoi	.051	.051
Δvwo 4 * antwoord-relevante aoi	.041	.108
Δvwo 4 * antwoord-irrelev. aoi	-.267	.099
<i>Antwoord</i>		
Δantwoord correct * vraag-aoi	-.204	.073
Δantwoord correct * antwoord-relevante aoi	.478	.092
Δantwoord correct * antwoord-irrelevante aoi	-.162	.083
<i>Varianties</i>		
S ² tekst	.049	.016
S ² leerling	.040	.009
S ² residu	1.087	.018

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

model 9, waarbij niet-significante parameters uit de analyse zijn gelaten (model 7: het interactie-effect tussen jaarlaag en aoi-lengte). Wanneer allereerst het beantwoorden in het algemeen wordt bekeken, dan blijken zesdeklassers van de drie aoi-typen relatief het langst naar de vraag te kijken (bij een gem. aoi-lengte: $e^{9.646} = 15460$ ms). Ze kijken vervolgens het langst naar een antwoord-relevante aoi (bij gem. lengte: $e^{8.320} = 4105$ ms). Naar een niet-relevante aoi kijken deze leerlingen het kortst (bij gem. lengte: $e^{7.356} =$

1566 ms). Vwo 6 kijkt dus duidelijk langer (bij gem. lengte: $e^{(8.320)} - e^{(7.356)} = 2540$ ms) naar een antwoord-relevante aoi dan naar een antwoord-irrelevante aoi.

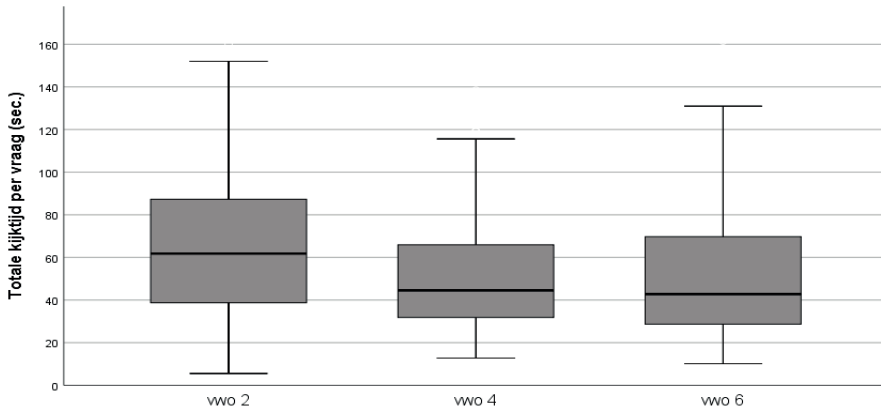
Naarmate de vraagstelling in lengte toeneemt, kijken leerlingen duidelijk langer naar de vraag ($.057 * N_{\text{lengte aoi}}$). Naarmate antwoord-relevante aoi's in lengte toenemen, besteden leerlingen er iets minder totale kijktijd aan ($-.023 * N_{\text{lengte aoi}}$); hetzelfde geldt voor de antwoord-irrelevante aoi's ($-.020 * N_{\text{lengte aoi}}$). Het verschil in leestijd tussen vraag en antwoord-(ir)relevante aoi's wordt dus groter naarmate hun lengte toeneemt.

Tweedeklassers kijken opvallend langer naar de vraag (bij gem. lengte: $e^{(9.646 + .336)} = 21634$ ms) dan zesdeklassers. Ze kijken duidelijk korter naar een antwoord-relevante aoi (bij gem. lengte: $e^{(8.320 - .326)} = 2963$ ms) en ook korter naar een antwoord-irrelevante aoi (bij gem. lengte: $e^{(7.356 - .260)} = 1207$ ms) dan zesdeklassers. Vwo 2 besteedt dus meer aandacht aan de vraag en minder aan de tekst.

Vierdeklassers kijken niet significant korter of langer naar een vraag dan zesdeklassers (bij gem. lengte: $e^{(9.646 + .051)} = 16269$ ms). Ze wijden niet significant meer of minder kijktijd aan een antwoord-relevante aoi (bij gem. lengte: $e^{(8.320 + .041)} = 4277$ ms), maar kijken wel korter dan zesdeklassers naar een niet-relevante tekst-aoi (bij gem. lengte: $e^{(7.356 - .267)} = 1199$ ms).

Wat de relatie met het uiteindelijke antwoord betreft: leerlingen met een goed antwoord kijken over het algemeen duidelijk korter naar een vraag dan leerlingen met een onjuist antwoord (bij gem. lengte: voor vwo 6: $e^{9.646} - e^{(9.646 - .204)} = 2853$ ms korter). Leerlingen die een correct antwoord formuleren, bekijken duidelijk langer een antwoord-relevante aoi (voor vwo 6, bij gem. lengte: $e^{(8.320 + .478)} - e^{8.320} = 2516$ ms langer), en wat korter een antwoord-irrelevante aoi (voor vwo 6, bij gem. lengte: $e^{7.356} - e^{(7.356 - .162)} = 234$ ms korter).

Proportie bekeken aoi's antwoordproces. In bovenstaande werd echter niet verdisconteerd dat leerlingen per vraag vaker naar een antwoord-irrelevante of antwoord-relevante aoi hebben kunnen kijken. Om te bekijken of



Figuur 3

Frequentieverdeling van de totale kijktijd per vraagitem, per jaarlaag, uitgedrukt in seconden

de proportie bekeken aoi's varieert tussen jaarlagen, pasten we een multinomiaal model toe, waarin we de lengte van aoi's verdisconteerden en jaarlaag als hoofdeffect meenamen.

Dan blijkt dat er weliswaar over het algemeen tussen vwo 2 en vwo 6 een verschil aantoonbaar is ($F(2, 4962) = 3.65, p = .03$), maar niet tussen vwo 2 en vwo 4 ($F(2, 4962) = 1.78, p = .17$), en evenmin tussen vwo 4 en vwo 6 ($F(2, 4962) = 0.30, p = .74$). Als nader wordt ingezoomd op de afzonderlijke aoi-typen, dan zijn er voor de proportie antwoord-relevante aoi's geen significante verschillen tussen jaarlagen te vinden, maar wel voor de proportie antwoord-irrelevante aoi's. De kans dat een vwo 6'er naar een voor de vraag irrelevante tekst-aoi kijkt, is heel wat kleiner ($\beta = -.341, p = .01$) dan de kans dat een vwo 2'er dat doet (bij een gem. aoi-lengte: $\beta = .399$). Een vierdeklasser heeft echter geen kleinere kans naar een irrelevante aoi te kijken dan een tweedeklasser ($\beta = -.259; p = .07$). Vwo 6 kijkt dus gemiddeld weliswaar langer naar een irrelevant tekstdeel dan vwo 2, ze kijkt wel duidelijk minder vaak naar deze delen. Uitgaande van een gemiddeld lange antwoord-irrelevante aoi (195 karakters), dan bestaat 32 % van de door een zesdeklasser bekeken aoi's uit voor beantwoording irrelevante tekstdelen, terwijl dat voor een tweedeklasser 61 % is. Gevolg

is dat de totale tijd die vwo 6 aan antwoord-irrelevante tekstdelen besteedt, aanzienlijk korter is, en dat vwo 6 daarmee het kortst naar de antwoord-irrelevante tekstdelen kijkt.

Nemen we in dit model ook mee of een leerling een goed dan wel fout antwoord had, dan blijkt dit duidelijk significant ($F(2, 4966) = 9.65, p < .001$). Als leerlingen een goed antwoord hebben, is de kans veel kleiner dat ze naar antwoord-irrelevante aoi's kijken ($\beta = -.444; p < .001$) dan wanneer ze een fout antwoord hebben. Door een vrij sterk interactie-effect met de aoi-lengte ($\beta = .137; p = .001$) hebben vooral de kortere antwoord-irrelevante aoi's aanmerkelijk minder kans bekeken te worden. Uitgaande van een gemiddeld lange antwoord-irrelevante aoi (195 karakters), bestaat 55 % van de aoi's die een leerling met een goed antwoord bekijkt, uit voor beantwoording irrelevante tekstdelen terwijl dat voor een leerling met een fout antwoord 60 % is.

Wanneer ten slotte de totaal bestede tijd aan een vraag wordt bekeken, dan blijken leerlingen ruim de tijd te nemen voor het lezen en beantwoorden van vragen. De gemiddeld bestede totale kijktijd voor het lezen en beantwoorden van een vraag was 60.5 seconden ($SD = 41.5$): ter vergelijking, de gemiddelde totale kijktijd voor het vooraf lezen van een tekst bedroeg 86 seconden ($SD = 25.9$). In Figuur 3 is de frequentieverdeling

van de totaal bestede tijd per vraagitem weergegeven per jaarlaag. Zoals al eerder vastgesteld, is enkel het onderscheid tussen vwo 2 en vwo 4 hier significant ($p = .01$); dat tussen vwo 2 en vwo 6 (nipt) niet ($p = .06$). In het leeuwendeel van de observaties blijken leerlingen aanzienlijk tijd nodig te hebben om te antwoorden, zeker als men zich realiseert dat leerlingen het antwoord niet op hoefden te schrijven en mondeling konden formuleren.

Varianties. Onderaan in Tabel 5 staan, ten slotte, de varianties. De residuele variantie, allereerst, is opmerkelijk hoog. Dat is niet bevreemdend. De hoge residuele variantie neemt het uiterst wisselende gedrag per vraag per persoon mee. De ene vraag is moeilijker dan de andere en veroorzaakt onder andere daardoor telkens wisselend leesgedrag voor de vraag en voor de antwoord-relevante aoi's en antwoord-irrelevante aoi's. Daarmee vergeleken is de variantie die toegeschreven kan worden aan leerlingen gering, en dat geldt eveneens voor de variantie veroorzaakt door de vier sets van tekst en vragen.

Samenvattend: de verschillen tussen jaarlagen zijn hier voor een belangrijk deel conform verwachting. We zien allereerst hypothese 2c bevestigd: vwo 4 en vwo 6 kijken langer naar de antwoord-relevante aoi's dan vwo 2. Hypothese 2d zien we eveneens bevestigd, zij het via een omweg: zesdeklassers kijken weliswaar langer naar een niet-relevante aoi dan tweedeklassers, maar ze kijken er wel duidelijk minder vaak naar dan tweedeklassers, waardoor ze in totaal juist korter naar irrelevante tekstdelen kijken. Voorts: vwo 4 en vwo 6 kijken korter naar de vraag dan vwo 2. Wat vwo 6-leerlingen bij de leestijden van antwoord-aoi's onderscheidt van vwo 2-leerlingen, onderscheidt *grosso modo* ook de leerlingen met een goed antwoord van de leerlingen met een fout antwoord. En leerlingen besteden over het algemeen veel tijd aan een vraag.

3.4 Relatie tussen het vooraf lezen en het beantwoorden

Om de relatie tussen het vooraf lezen en het beantwoorden te verkennen, is een extra model opgesteld waarbij naast de gemiddelde

varianties op leerlingniveau de covarianties tussen de vijf aoi-typen zijn geschat: aan de ene kant de kernzinnen en niet-kernzinnen bij het vooraf lezen, en aan de andere kant de drie aoi-typen bij het beantwoorden. Daarbij hebben we wederom gecontroleerd voor aoi-lengte.

Leerlingen die bij het vooraf lezen kernzinnen snel dan wel langzaam lezen, lezen ook de niet-kernzinnen snel dan wel langzaam ($r = 1.0$). Als leerlingen bij het beantwoorden een vraag snel/langzaam lezen, lezen ze bij het beantwoorden ook een antwoord-relevante of een antwoord-irrelevant tekstdeel snel/langzaam ($r \geq .66$). En wie een antwoord-relevant tekstdeel snel/langzaam leest, leest ook een antwoord-irrelevant tekstdeel snel/langzaam ($r = .80$). We vinden een minder hoog, matig verband tussen het lezen van de vraag en het vooraf lezen van kernzinnen ($r = .65$) en niet-kernzinnen ($r = .51$): wie kernzinnen en niet-kernzinnen snel leest, leest ook de vraag snel. We vinden dus vanuit de leessnelheid van een leerling over het algemeen redelijk hoge correlaties tussen deze vijf factoren.

Is er ook een meer inhoudelijke samenhang aantoonbaar tussen vooraf lezen en beantwoorden? Een leerling die bij het beantwoorden korter of langer naar een antwoord-irrelevante tekstdeel kijkt, heeft eerder ook enigszins korter of langer naar een kernzin gekeken ($r = .45$) én naar een niet-kernzin ($r = .37$) – een zwak verband. De relatie tussen het korter of langer kijken naar een antwoord-relevant tekstdeel bij het beantwoorden en het lezen van een kernzin en van een niet-kernzin bij het vooraf lezen is echter zeer zwak, zelfs verwaarloosbaar (resp. $r = .17$ & $r = .02$). Door de relatief zwakke relaties blijken deze effecten niet te variëren tussen jaarlagen.

We kunnen hier dus niet of nauwelijks aantonen dat leerlingen die bij het vooraf lezen langer kijken naar kernzinnen, ook langer naar de antwoord-relevante tekstdelen kijken bij het antwoorden (hypothese 3a); wel kunnen we voorzichtig stellen dat leerlingen die bij het vooraf lezen langer kijken naar kernzinnen, bij het beantwoorden wat korter naar antwoord-irrelevante tekstdelen kijken (hypothese 3b): al is het

Tabel 6

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 basismodel	6356.54				
2 + proportie correct	6356.46	2 vs 1	.08	1	.78
3 + proportie correct * (niet-)kernzin	6355.11	3 vs 2	1.35	1	.25
4 + proportie correct * (niet-)kernzin * jaarlaag	6354.67	4 vs 3	.44	2	.80
<i>Totale kijktijd</i>					
1 basismodel	2804.73				
2 + proportie correct	2802.96	2 vs 1	1.77	1	.18
3 + proportie correct * (niet-)kernzin	2802.09	3 vs 2	.87	1	.35
4 + proportie correct * (niet-)kernzin * jaarlaag	2799.96	4 vs 3	2.13	2	.35

verband zwak en de hoeveelheid verklaarde variantie hierbij niet hoog (20 %). Deze effecten blijken niet te variëren tussen jaarlagen (hypothese 3d).

3.5 Relatie tussen het vooraf lezen en het antwoord

Voor het verband tussen het vooraf lezen en het antwoord (het aantal goede antwoorden) zijn weer beide leesmaten geanalyseerd: *first-pass dwell time* en totale kijktijd. Uit Tabel 6 blijkt dat de passing van een model waarin de proportie goede antwoorden wordt meegenomen, bij beide leesmaten niet verschilt van het basismodel (Model 2 vs Model 1: $\Delta\chi^2(1) \leq 1.77$; $p \geq .18$). Toevoeging van het interactie-effect tussen (niet-)kernzinnen en de proportie goede antwoorden levert evenmin verbetering op (Model 3 vs Model 2: $\Delta\chi^2(1) \leq 1.35$; $p \geq .25$). En een model dat het effect tussen (niet-)kernzinnen en de proportie correcte antwoorden laat variëren tussen jaarlagen, genereert evenmin een betere *fit* (Model 4 vs Model 3: $\Delta\chi^2(2) \leq 2.13$; $p \geq .35$).

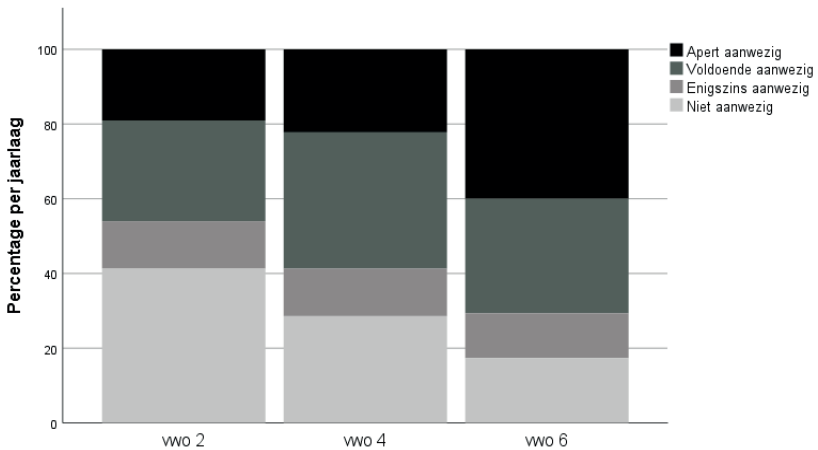
Hypothese 3c ('Leerlingen die langer bij het vooraf lezen naar kernzinnen kijken, hebben bij de vragen vaker een goed antwoord') kunnen we dus vanuit beide leesmaten niet aantonen, en evenmin hypothese 3d ('Bij oudere leerlingen is dit duidelijker aantoonbaar').

3.6 Retrospectief interview

In het interview zagen deelnemers hun oogbewegingengedrag terug bij de beantwoording van de vier vraagitems bij de laatste tekst. We vroegen hun aan te geven in hoeverre ze meteen na het lezen van de vraag a) wisten in welke tekstpassages het antwoord te vinden was (tekstbasis), en b) een antwoord in gedachten hadden (situatiemodel). Bij vier leerlingen (twee tweedeklassers, een vierdeklasser, een zesdeklasser) bleek bij één item de reflectie afwezig of om technische redenen onbruikbaar; bij één tweedeklasser gold dit voor drie items.

Over het algemeen was de audioreflectie, mede door de gelijktijdig afgespeelde scanpath, eenduidig en helder. Een beperkt aantal reflecties (twaalf gevallen, bij zes leerlingen) bleek onvoldoende eenduidig; deze gevallen werden voorgelegd aan twee leden van het screeningspanel. De drie codeurs stemden in tien gevallen overeen in hun codering. De twee overige gevallen werden buiten de analyse gelaten.

Worden dan allereerst de tekstbasis-scores bekeken, dan blijkt er een aantoonbaar verschil tussen de drie jaarlagen ($\chi^2(6) = 14.31$, $p = .03$). Het onderscheid tussen vwo 2 en vwo 6 is significant ($\chi^2(3) = 12.50$, $p \leq .01$) maar niet dat tussen vwo 2 en vwo 4, noch dat tussen vwo 4 en vwo 6 ($\chi^2(3) < 5.68$, $p \geq .07$).



Figuur 4
 Percentuele verdeling van de vier antwoordcategorieën in de tekstbasisscores, per jaarlaag

In Figuur 4 is per jaarlaag de percentuele spreiding in tekstbasisscores weergegeven. Waar tweedeklassers in ruim 50% van de observaties aangeven niet/nauwelijks over een adequate tekstbasis bij beantwoording te beschikken, geven zesdeklassers hiervan blijk in een kleine 30% van de observaties. Tweedeklassers geven in 19% van de observaties aan over een apert aanwezige tekstbasis te beschikken, waar het bij zesdeklassers 40% betreft. De vwo 4-deelnemers positioneren zich tussen de scores van deze beide groepen in, al valt ook bij hen de geringere hoeveelheid ‘apert aanwezig’-scores op in vergelijking met vwo 6. Vwo 6-leerlingen rapporteerden dus in het algemeen een aanzienlijk adequatere tekstbasis dan vwo 2-leerlingen.

Wanneer vervolgens de situatiemodelscores per jaarlaag worden vergeleken, dan blijkt er wederom een apert verschil tussen jaarlagen ($\chi^2(6) = 29.55, p < .001$). Het verschil tussen vwo 2 en vwo 6 is significant, evenals dat tussen vwo 4 en vwo 6 ($\chi^2(3) \geq 15.50, p \leq .001$), maar niet dat tussen vwo 2 en 4 ($\chi^2(3) = 1.92, p = .30$).

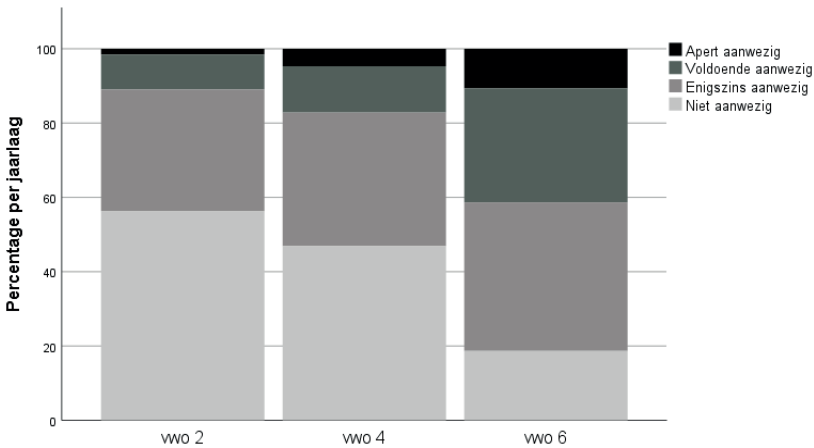
In Figuur 5 is per jaarlaag de percentuele spreiding in situatiemodelscores weergegeven. Hier is te zien dat vwo 2 na vraaglezing in 56% van de observaties en vwo 4 in 47% van de observaties niet/

nauwelijks van een adequaat situatiemodel blijk gaf terwijl vwo 6 dit in 19% van de observaties aangaf. Er is ook een duidelijk verschil bij ‘voldoende aanwezig’: bij tweedeklassers betreft dit 9% van de scores en bij vierdeklassers 13%; bij zesdeklassers daarentegen betreft het 31%. Bij ‘apert aanwezig’ liggen de (lage) scores van de drie jaarlagen aanmerkelijk dicht bij elkaar. In veel opzichten geven vwo 6-leerlingen al met al blijk van een aanmerkelijk sterker opgebouwd situatiemodel dan vwo 2- en vwo 4-leerlingen, maar ook zij hebben zelden meteen een scherp antwoord in hun hoofd.

4. Discussie

In deze studie onderzochten we verschillen in het vooraf lezen van teksten en het beantwoorden van bijbehorende vragen tussen vwo 2-, vwo 4- en vwo 6-leerlingen. Daarbij generaliseren we over leerlingen én over teksten. We vatten allereerst de bevindingen samen voor het vooraf lezen, het beantwoorden en voor de retrospectieve interviews.

We beginnen bij het leesproces. Vwo 2-leerlingen lezen duidelijk langzamer dan vwo 6-leerlingen; vwo 4-leerlingen lezen sneller dan vwo 2-leerlingen, maar niet sneller of



Figuur 5
Percentuele verdeling van de vier antwoordcategorieën in de situatiemodellscores, per jaarlaag

langzamer dan vwo 6-leerlingen (hypothese 1a). Deelnemers maken bij eerste lezing over het algemeen een zeker onderscheid tussen kernzinnen en niet-kernzinnen, maar bij het herlezen verdwijnt dit onderscheid vrijwel. We konden hierbij niet aantonen dat oudere leerlingen meer aandacht hebben voor kernzinnen dan jongere leerlingen (hypothese 1b). Over het algemeen lezen deelnemers sterk lineair en ogenschijnlijk weinig strategisch. Tweedeklassers lezen een tekst meer lineair (minder strategisch) dan vierde- en zesdeklassers; vierdeklassers lezen echter niet meer of minder lineair (strategisch) dan zesdeklassers (hypothese 1c).

Betreffende het vragen beantwoorden: tweedeklassers hebben conform verwachting minder vragen goed dan de vierdeklassers, en de vierdeklassers weer minder vragen goed dan de zesdeklassers (hypothese 2a). Gemiddeld geldt: het beantwoorden van een vraag neemt veel tijd in beslag, en slechts relatief zelden geeft men snel en adequaat antwoord op een vraag. Alleen voor vwo 4-leerlingen konden we aantonen dat ze over het algemeen sneller antwoord geven dan vwo 2-leerlingen (hypothese 2b). Leerlingen hebben bij het beantwoorden gemiddeld meer oog voor een antwoord-relevant tekstdeel dan voor een antwoord-irrelevant tekstdeel; de vraagstelling zelf krijgt duidelijk

de meeste aandacht. Vierde- en zesdeklassers kijken langer naar een antwoord-relevant tekstdeel dan tweedeklassers. Maar zesdeklassers kijken niet langer naar een antwoord-relevant tekstdeel dan vierdeklassers; noch kijken zesdeklassers vaker ernaar dan tweede- en vierdeklassers (hypothese 2c). Wat de irrelevante tekstdelen betreft: hiernaar kijken zesdeklassers beduidend minder vaak dan tweedeklassers, maar ze kijken er niet minder vaak naar dan vierdeklassers; zesdeklassers kijken, onverwacht, ook juist langer naar een antwoord-irrelevant tekstdeel dan tweede- en vierdeklassers (hypothese 2d). Voor het overige: zesdeklassers kijken gemiddeld beduidend korter naar de vraag dan tweedeklassers.

Over het algemeen komt het onderscheid tussen oudere en jongere leerlingen overeen met het onderscheid tussen leerlingen met een goed of fout antwoord: zowel zesdeklassers als leerlingen met een goed antwoord kijken gemiddeld korter naar de vraag en gebruiken de tekst selectiever.

Wat de relatie tussen het lees- en het antwoordproces betreft: leerlingen die bij het vooraf lezen langer kijken naar kernzinnen, kijken bij het antwoorden niet of nauwelijks langer naar de antwoord-relevante tekstdelen (hypothese 3a). Wel kunnen we voorzichtig

stellen dat deze leerlingen bij het beantwoorden wat korter naar antwoord-irrelevante tekst delen kijken (hypothese 3b), al is het aangetoonde verband zwak. We konden tussen jaarlagen bij het voorgaande geen onderscheid aantonen (hypothese 3d). Voorts: leerlingen die langer naar kernzinnen kijken, hebben bij het beantwoorden, onverwacht, niet aantoonbaar vaker een goed antwoord (hypothese 3c), waarbij jaarlaag – wederom – niet aantoonbaar een rol speelt (hypothese 3d). We kunnen vanuit de eye-trackdata daarmee opvallend weinig relaties aantonen tussen vooraf lezen en vragen beantwoorden (basishypothese A).

Dan het retrospectieve interview: leerlingen gaven daarin aan direct na vraaglezing over het algemeen over een redelijke tekstbasis te beschikken, maar in het grote merendeel van de gevallen niet over een redelijk tot goed situatiemodel (basishypothese A). De zesdeklassers beschikten daarbij over een beduidend betere tekstbasis dan de tweedeklassers; een onderscheid tussen vierdeklassers en twee- dan wel zesdeklassers konden we hier niet aantonen. Vierdeklassers en zesdeklassers gaven vaker blijf van een adequaat situatiemodel dan de tweedeklassers, en de zesdeklassers weer vaker dan de vierdeklassers (basishypothese B).

Ten slotte: bij het vooraf lezen valt het op dat de variantie tussen leerlingen aanmerkelijk hoger ligt dan de variantie tussen teksten: vooral de individuele leesaanpak van een leerling verklaart de variatie in kijkgedrag, de teksten zelf aanmerkelijk minder. Daarnaast: de residuele variantie is vooral bij het beantwoorden uiterst omvangrijk. Elke vraag lijkt vooral telkens wisselend, van persoon tot persoon sterk afwijkend antwoordgedrag te genereren.

Het doel van deze studie is verschillen aan te tonen tussen jongere en oudere vwo-leerlingen bij de tekst met vragen. In zeer algemene lijnen tekenen in de resultaten de verschillen zich dan af zoals verwacht: zesdeklassers onderscheiden zich van tweedeklassers door een hogere leessnelheid, meer terugkijken (meer strategisch lezen) bij het doornemen van de teksten, en door een

effectiever antwoordproces en adequater tekstbegrip bij het beantwoorden van de vragen. Tegelijkertijd treffen we ook onverwachte resultaten aan. De verschillen tussen jaarlagen bij vooraf lezen én beantwoorden zijn vaak minder markant dan verwacht. Vierdeklassers gedragen zich vooral in de eye-trackresultaten vaak als zesdeklassers, met name bij het vooraf lezen. Bovendien blijkt tussen het vooraf lezen en het vragen beantwoorden een aanzienlijk minder sterke relatie dan verwacht. We lopen de opvallende bevindingen langs.

Kijkt men allereerst naar het vooraf lezen, dan valt op dat de deelnemers, onafhankelijk van jaarlaag, tijdens de eerste doorlezing aan kernzinnen wat meer aandacht schenken. Dit bevestigt voorzichtig de aanname van de Expertgroep Taal (2008) dat lezers vanaf niveau 2F hoofd- en bijzaak scheiden, én de belangrijke aanname in de Constructie-Integratie-theorie (Kintsch, 1998) dat geoefende lezers in eerste instantie onderscheid maken tussen hoofd- en bijzaken. Bij herlezing van zinnen verdwijnt echter onverwacht het onderscheid tussen kernzinnen en niet-kernzinnen. Vanuit de olopende moeilijkheidsgraad van de stimuli verwachtten we bovendien dat de oudere leerlingen kernzinnen gemiddeld langer bekijken dan de jongere leerlingen, maar ook dit konden we niet aantonen. Is kernzindetectie bij jongere expert-lezers misschien niet afhankelijk van tekstcomplexiteit? Een andere, verleidelijke verklaring volgt uit de studie van Yeari et al. (2015). Zij vonden dat leerlingen zich bij een leestaak met open vragen vooral bij de herlezing op de hele tekst richten, niet op de hoofdzaken: ze kunnen immers over alle tekstelementen vragen krijgen. Speciaal aandacht schenken aan kernzinnen wordt niet door de taak beloond.

Daarnaast valt het op dat deelnemers ondanks de deels pittige teksten over het algemeen sterk lineair lezen, ook al vinden we wel voorzichtig de verwachte verschillen tussen jaarlagen. Frequent inzet van leesstrategieën lijkt over het algemeen onaannemelijk. De Expertgroep Taal (2008) veronderstelt in de beschrijving van het domein zakelijk lezen 3F en 4F de adequate

inzet van enkele leesstrategieën, evenals de Nederlandse eindtermen voor het vwo. Dit strategische leesgedrag hebben wij bij het vooraf lezen nauwelijks kunnen aantonen. Daarmee is natuurlijk niet gezegd dat vwo-leerlingen deze strategieën niet kennen en kunnen toepassen. Eerder lijken ze intensieve inzet ervan bij dit experiment niet noodzakelijk te hebben geacht. Uit enkele recente verkennende studies naar het leesgedrag van vwo'ers bleek tamelijk oppervlakkig, sterk pragmatisch leesgedrag (Breukink, [in voorbereiding]; Stevens, 2018; Rooijackers et al., 2020). In deze studie lijkt zich een soortgelijk oppervlakkig leesproces af te tekenen.

Ten derde valt op dat de zesdeklassers zich in de eye-trackresultaten vaak nauwelijks onderscheiden van de vierdeklassers. Dit kan meerdere verklaringen hebben, bijvoorbeeld de samenstelling van de steekproef (veel goede lezers onder de vwo 4-leerlingen?). Eén verklaring ligt voor de hand: een zesdeklasser weet tijdens het lezen beter te infereren en expliciteren, zonder dat hij daarbij duidelijk anders op zinnen en woorden fixeert. Daarmee wordt op een bekende begrenzing van eye-tracking gestuit (Holmqvist & Andersson, 2017). De betere situatiemodelscores van de zesdeklassers bij het retrospectieve interview kunnen hiervoor als bewijs dienen.

Wat het vragen beantwoorden betreft, valt op dat zesdeklassers vaak adequater antwoordgedrag vertonen dan tweedeklassers en, in mindere mate, vierdeklassers. Zesdeklassers lijken het meest verfijnde antwoordgedrag te vertonen: ze doen niet korter over de vraagbeantwoording dan tweedeklassers, maar lezen de vraag veel korter en bekijken een antwoord-irrelevant tekstdeel minder vaak maar wel langer dan tweedeklassers. Daarmee hebben ze ook vaker vragen goed. Deze resultaten zijn logisch te verklaren: wie vooraf een redelijk beeld van de tekst heeft gekregen, begrijpt de vraagstelling sneller en weet beter waar het antwoord zich bevindt, waardoor hij zijn aandacht gericht en relatief ruimer kan wijden aan de tekstbestudering. En andersom: wie de vraag minder makkelijk of niet

begrijpt, zal deze vaker moeten herlezen en richt daarmee minder de aandacht op de tekst, waardoor zij of hij al met al gemakkelijker een fout antwoord geeft. Ook Brunfaut en McCray (2015) vonden in hun eye-trackdata dat leerlingen met een goed antwoord sneller antwoordgedrag vertoonden, met minder aandacht voor de vraag, evenals Bax (2013) dat vond bij onderzochte testitems. Betere lezers antwoorden gericht en effectiever.

Wat de relatie tussen vooraf lezen en beantwoorden betreft: opvallend genoeg konden we onze twee basishypothesen maar zeer ten dele bevestigen. We konden een redelijk sterk verband tussen het vooraf lezen en het vragen beantwoorden aantonen op tekstbasisoniveau, maar niet op situatiemodelniveau. Desalniettemin rijst er een tamelijk eenduidig beeld op uit de resultaten: leerlingen bouwen tijdens het vooraf lezen een nogal globaal, weinig specifiek situatiemodel op, dat wel wat in scherpte toeneemt naarmate de jaarlaag hoger is. Ondanks onze zeer nadrukkelijke instructie vooraf om de teksten intensief te bestuderen en de relatief hoge concentratie- en motivatiecijfers, volstonen leerlingen tijdens vooraf lezen, lijkt het, met *shallow processing* (Sanford & Sturt, 2002). Vooral bij het beantwoorden van de vragen bouwden ze een accuraat en adequaat situatiemodel op. Dit correspondeert met eerdere bevindingen over het effect van het stellen van open vragen bij een tekst (bijv. Cerdán et al., 2009; Rouet et al., 2001), waarin de participanten door het beantwoorden van open vragen een diepgaand situatiemodel bij de tekst construeerden. Pas toen leerlingen door de vragen te lezen in detail duidelijk werd wat ze met de gelezen tekst moesten doen, zetten ze zich tot diepgaande bestudering. Daarmee lijkt vooral de beantwoordingfase te bepalen of een leerling een goed dan wel fout antwoord geeft (vgl. hierover: Embretson & Wetzel, 1987; Rupp, Ferne & Choi, 2006).

Meer dan vooraf verwacht, wijzen de aangetoonde verschillen tussen jaarlagen vaak op sterk taakafhankelijk lees- en antwoordgedrag bij de tekst met vragen. De vraag is in hoeverre dit sterk taakafhankelijke

gedrag onverwacht is. Leesgedrag is immers altijd sterk doelfafhankelijk (Graesser et al., 1994), en het wordt vanuit onderzoek steeds duidelijker dat lezers vaak niet met volle, rijke aandacht lezen, maar hun leesgedrag zeer pragmatisch aanpassen aan de vereisten voor een taak (vgl. de ‘*Good Enough*’ *approach* - Ferreira & Patson, 2007). Bij de tekst met vragen wordt het doel voor leerlingen pas bij het lezen van de vragen duidelijk en dan is het niet zinvol tijdens het vooraf lezen een accuraat, rijk gespecificeerd situatiemodel op te bouwen. Globaal de tekst lezen is echter voor een lezer wel zinnig: hij krijgt zo een beeld van de hele tekst en voorkomt daarmee de valkuil dat hij op basis van louter lokale informatie een vraag onjuist interpreteert of beantwoordt (vgl. Cerdán et al., 2009; Ozuru et al., 2007). De soms weinig markante verschillen tussen jaarlagen bij het vooraf lezen in deze studie kunnen hiermee worden verklaard.

Hoezeer taakverwachting en -vertrouwdheid van leerlingen bij de resultaten in deze studie inderdaad een rol spelen, kan blijken uit een tot nog toe onbenoemd detail. Acht van de 52 participanten gaven bij het retrospectief interview ongevraagd (!) aan dat ze zich minstens eenmaal in hun antwoordgedrag volledig lieten leiden door conventies rondom de taak tekst met vragen, niet door enig tekstbegrip: “Ik zocht het antwoord in alinea 2, omdat dit de tweede vraag bij de tekst was.”

Dit pragmatische leesgedrag wordt waarschijnlijk nog in de hand gespeeld door de huidige inrichting van de tekst met vragen in het vak Nederlands. Men stelt meestal geen duidelijk leesdoel vooraf en een leerling kan sterk in aard variërende vragen verwachten (vragen op lokaal of globaal niveau; analytische, stilistische, interpreterende vragen), waarbij hogere-orde-denkvaardigheden als reflecteren en evalueren enkel beperkt aan bod komen. Vaak hebben de vragen ook betrekking op één tekst, en de taak dient doorgaans individueel uitgevoerd. Met deze conventies rond de tekst met vragen is een leerling al vertrouwd vanaf de lagere school en in bovenbouwmethodes Nederlands als *Nieuw Nederlands* en *Op niveau* domineert ze de leesvaardigheidsparagrafen (vgl. Van

der Leeuw et al., 2017).

Enkele algemene voorbehouden bij deze studie zijn echter wel op hun plaats. Het betreft primair een eye-trackstudie. De basisaanname achter deze methode, namelijk dat kijken naar een aoi een directe weerspiegeling is van de aandacht die een participant eraan schenkt, staat nog altijd ter discussie, evenals de praktische implicaties van deze aanname (vgl. Holmqvist & Andersson, 2017). Eye-tracking biedt daarnaast alleen zicht op het hoe, niet op het waarom achter lezen. En weliswaar vulden we deze lacune deels door een retrospectief interview in te zetten, hierin werden alleen de laatste vier vragen bij één tekst besproken. Daarnaast betrof het een cross-sectioneel, dus niet longitudinaal onderzoek, afgenomen op één Nederlandse school, bij slechts vier relatief korte teksten en een beperkt aantal vragen. Ten slotte: de verschillen tussen de drie jaarlagen onderzochten we door telkens groepen van minder dan twintig leerlingen met elkaar te vergelijken. De steekproeven bij vergelijkbare *mixed method*-eye-trackstudies zijn overigens vergelijkbaar of zelfs aanmerkelijk kleiner (vgl. Bax, 2013; Brunfaut & McCray, 2015).

Tegelijk wijzen we graag op het methodologisch vernieuwende karakter van deze studie. Hoe de mentale ‘complexen’ bij een tekst met vragen zich tot elkaar verhouden – initieel leesproces, antwoordproces en het uiteindelijke antwoord –, is methodologisch ingewikkeld te ontrafelen. Gebruik van hardop-denkprotocollen werd wel ingezet, maar grijpt sterk in in het natuurlijke lezen (bijv. Norris, 1990; vgl. Pressley & Afflerbach, 1995). Om de relatie tussen vooraf lezen en vragen beantwoorden te onderzoeken, gebruikten Cerdán et al. (2009) daarom sterk op eye-tracking gelijkende software om het vooraf lezen te beschrijven, maar zij stelden in hun studie tekstbegrip gelijk aan de score op de vragen. Enkele redelijk recente Britse studies combineerden eye-tracking met hardop-denken via retrospectieve interviews (Bax, 2013; Brunfaut & McCray, 2015), maar lieten de initiële tekstlezing buiten beeld. In onze studie wordt deze *mixed-method* benadering overgenomen om initieel

leesgedrag en antwoordgedrag, afzonderlijk en in samenhang, te kunnen beschrijven. Door te generaliseren over leerlingen en teksten, konden we zo ‘holistisch’ verschillen in tekstbegrip tussen jongere en oudere vwo-leerlingen aantonen bij de taak tekst met vragen.

Al met al lijkt deze studie daarmee de vaak geuite kritiek bij de tekst met vragen te ondersteunen dat leerlingen met de tekst met vragen vooral vragen leren beantwoorden (bijv. Hoogeveen, & Bonset, 1998). Eerdere Nederlandse studies vonden dat taken zoals de sorteertaak en de mental model-taak beter op het niveau van het situatiemodel tekstbegrip meten dan de tekst met vragen (Kamalski, 2007; Land, 2009). Vanuit deze studie is de vraag in hoeverre de tekst met vragen in zijn huidige inrichting leerlingen leert om zelfstandig, met dieper tekstbegrip teksten te lezen. Het is bepaald niet onwaarschijnlijk dat met de veelvuldige inzet van deze taak eerder taakvaardigheid wordt bevorderd dan leesvaardigheid.

De sterke dominantie van de tekst met vragen als individuele taak in de leeslessen Nederlands in zijn huidige vorm lijkt zo ongewenst. Recent klonken er in Nederland geluiden dat de motivatie van leerlingen om ‘diep te lezen’ afneemt (Raad voor Cultuur & Onderwijsraad, 2019) en er zijn tekenen dat het tekstbegrip in Nederland en Vlaanderen daalt (Gubbels et al., 2019; Van den Branden, Ghesquière & Van Keer, 2019). De vraag is in hoeverre de huidige didactische dominantie van de tekst met vragen daaraan bijdraagt, een samenhang die in kranten nogal eens wordt verondersteld (bijv. Welgraven & De Vos, 2019). De kritiek op het Nederlandse leesonderwijs luidt van oudsher dat het sterk toetsgericht is (bijv. Hoogeveen & Bonset, 1998): leraren trainen hun leerlingen vooral die kennis en vaardigheden die nodig zijn om ‘teksten met vragen’ goed te kunnen maken. Daarmee ontstaat al snel het gevaar dat andere cognitieve vaardigheden, betrokken bij ‘diep lezen’, het onderspit delven: er ontstaat “survival of content that fittest best the high-stakes test” (Popham, 2001).

Het lijkt in die zin raadzaam om de dominantie van de ‘reguliere’ tekst met

vragen te doorbreken. Dat kan door een veelheid aan andere leestaken aan te bieden: men denke aan samenvattingstaken of synthesesetaken (Van Ockenburg, Van Weijen, & Rijlaarsdam, 2018), argumentatietaken (Anmarkrud, Bråten, & Strømsø, 2014) of sorteertaken (Land, 2009). Ook een andere inrichting van de tekst met vragen is zeer te overwegen: tekstafhankelijke vragen kunnen steeds in groeps- of klassikaal verband worden behandeld (vgl. Fisher & Frey, 2015); hogere-orde-denkvaaardigheden zoals evalueren en reflecteren kunnen meer een plaats vinden, en de taak kan voorzien worden van een duidelijk leesdoel, met aan dat leesdoel gekoppelde vragen (Nederlands nu! & Sectiebestuur Nederlands van Levende Talen, 2018; Schuurs, 2019). Bovenal lijkt het zinnig het proces achter het ‘diep lezen’ zelf centraal te blijven stellen, bijvoorbeeld door modellen (Fisher, Frey & Lapp, 2017) of de inzet van observerend leren (Keehnen, Braaksma & De Boer, 2015).

Het in deze studie beschreven experiment vond plaats met toestemming van de Facultaire Ethische Toetsingscommissie Geesteswetenschappen (FETC-GW) van de Universiteit Utrecht.

Noten

- ¹ De *first-pass dwell time* is hier niet bruikbaar: ze veronderstelt een lineaire leeswijze en het beantwoorden kenmerkt zich juist door scannend zoekgedrag en een interactie tussen tekst en vraag (Brunfaut, & McCray, 2015; Rouet, 2006).
- ² Als concretisaties voor de scores bij de *tekstbasis* gold hierbij resp.: 1) ‘zeer duidelijk aanwezig’: weet een antwoord-relevant of een in de vraag geciteerd tekstdeel direct met een eventuele initiële afwijking van één regel te lokaliseren; 2) ‘voldoende aanwezig’: weet in welke alinea een antwoord-relevant tekstdeel of een in de vraag geciteerd tekstdeel zich bevindt; 3) ‘enigszins aanwezig’: weet of een antwoord-relevant tekstdeel of een in de vraag geciteerd tekstdeel zich in begin, midden of eind van de tekst bevindt; 4) ‘niet aanwezig’: weet niet waar een antwoord-relevant tekstdeel of een in de vraag

geciteerd tekstdeel zich in de tekst bevindt.

Als concretisaties voor de scores bij het *situatiemodel* gold resp.: 1) 'zeer duidelijk aanwezig': formuleert direct het antwoord of geeft aan direct het antwoord te weten; 2) 'voldoende aanwezig': geeft aan een idee of begin van een antwoord in het hoofd te hebben; 3) 'enigszins aanwezig': geeft weliswaar aan het antwoord nog niet te weten, maar weet wel direct waar het antwoord in de tekst te vinden is – we gaan hier uit van het idee dat een adequate tekstbasis doorgaans een rudimentair situatiemodel met zich meebrengt (vgl. Kintsch, 1998); 4) 'niet aanwezig': heeft geen idee wat het antwoord is, noch hoe het antwoord gevonden kan worden.

- ³ Ook verdere interactie-effecten met jaarlaag leiden niet tot een significant betere passing. Nemen we bijvoorbeeld, in aansluiting op de modellen in Tabel 4, als model 11 het interactie-effect van het geven van een goed antwoord met jaarlaag en de drie antwoord-aoi-typen, voor de totale kijktijd, dan vinden we: Model 11 vs Model 10: $\Delta\chi^2(4) = 3.89$; $p = .42$. Met andere woorden: het geven van een goed dan wel fout antwoord staat ons hier niet verder toe verschillen aan te tonen tussen de jongere en oudere vwo-leerlingen.

Literatuur

- Andre, T. (1979). Does answering higher level questions while reading facilitate productive reading? *Review of Educational Research*, 49, 280-318.
- Anmarkrud, Ø., Bråten, I., & Strømso, H. I. (2014). Multiple-documents literacy: Strategic processing, source awareness, and argumentation when reading multiple conflicting documents. *Learning and Individual Differences*, 30, 64-76.
- Bax, S. (2013). The cognitive processing of candidates during reading tests: Evidence from eye-tracking. *Language Testing*, 30(4), 441-465.
- Bogaert, N., Devlieghere, J., Hacquebord, H., et al. (2008). *Aan het werk: adviezen ter verbetering van functionele leesvaardigheid in het onderwijs*. Den Haag: Nederlandse Taalunie.
- Bonset, H. (2010). Nederlands in het voortgezet en hoger onderwijs: Hoe sluit dat aan? Deel 1.

Levende Talen Magazine, 97(3), 16-20.

- 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 schoolexamen 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 Leeuw, L., Segers, E., & Verhoeven, L. (2016). The effect of student-related and text-related characteristics on student's reading behaviour and text comprehension: An eye movement study. *Scientific Studies of Reading*, 20(3), 248-263.
- De Milliano, I.I.C.M. (2013). *Literacy development of low-achieving adolescents. The role of engagement in academic reading and writing*. (Proefschrift). Amsterdam: UvA.
- Duchowski, A. T. (2007). *Eye tracking methodology: Theory and practice*. London: Springer.
- Duke, N. K., & Pearson, P. D. (2002). *What research has to say about reading instruction*. Newark, DE: International Reading Association.
- Duke, N.K., Pearson, D., Strachan, S.L., & Billman, A.K. (2011). Essential elements of fostering and teaching reading comprehension. In: J. Samuels, & A.E. Farstrup (Eds.), *What research has to say about reading instruction*. Newark: International Reading Association, 51-93.
- Newark: International Reading Association, p. 51-93.
- Embretson, S. E., & Wetzel, C. D. (1987).

- Component latent trait models for paragraph comprehension tests. *Applied psychological measurement*, 11(2), 175-193.
- Expertgroep Doorlopende Leerlijnen Taal. (2008). *Over de drempels met taal*. Enschede: SLO.
- Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, 1(1-2), 71-83.
- Fisher, D., & Frey, N. (2015). *Text-dependent questions, grades K-5: Pathways to close and critical reading*. Thousand Oaks: Corwin.
- Fisher, D., Frey, N., & Lapp, D. (2008). Shared readings: Modeling comprehension, vocabulary, text structures, and text features for older readers. *The Reading Teacher*, 61(7), 548-556.
- 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.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371-395.
- Gubbels, J., van Langen, A., Maassen, N., & Meelissen, M. (2019). *Resultaten PISA-2018 in vogelvlucht*. Enschede: Universiteit Twente.
- 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.
- Hyönä, J., & Niemi, P. (1990). Eye movements during repeated reading of a text. *Acta psychologica*, 73(3), 259-280.
- Jackson, M.D., & McClelland, J. L. (1979). Processing determinants of reading speed. *Journal of Experimental Psychology*, 108(2), 151-181.
- Kamalski, J. (2007). *Coherence marking, comprehension and persuasion: On the processing and representation of discourse*. Utrecht: LOT.
- Keehnen, T., Braaksma, M., & de Boer, M. (2015). Leren door zien lezen. Observerend leren bij leesvaardigheid in 3 vwo. *Levende Talen Tijdschrift*, 16(1), 34-41.
- Kintsch, W. (1998). *Comprehension: a paradigm of cognition*. Cambridge: Cambridge University Press.
- Kraal, A. (2020). *Different readers, different texts, different processes : the effects of reader and text properties on text processing*. PhD thesis. Leiden University.
- Kuijpers, C., van der Leij, A., Been, P., van Leeuwen, T., ter Keurs, M., Schreuder, R., & van den Bos, K.P. (2003). Leesproblemen in het voortgezet onderwijs en de volwassenheid. *Pedagogische Studiën*, 80(4), 272-287.
- Land, J. (2009). *Zwakke lezers, sterke teksten? Effecten van tekst- en lezerskenmerken op het tekstbegrip en de tekstwaardering van vmbo-leerlingen*. Delft: Eburon.
- Linthorst, T.R., & de Gloppe, K. (2015). De didactiek van begrijpend lezen in het voortgezet onderwijs: lesobservaties bij Nederlands en zaakvakken. *Pedagogische Studiën*, 92, p. 150-166.
- 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.
- 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 schoolexamen Nederlands op havo en vwo; verslag van een digitale enquête*. Enschede: SLO.
- Nederlands Nu! & Sectiebestuur Nederlands Levende Talen. (2018). *Advies Examens Nederlands [aangeboden aan curriculum.nu]*. S.I.
- Norris, S. P. (1990). Effect of eliciting verbal reports of thinking on critical thinking test performance. *Journal of Educational Measurement*, 27(1), 41-58.
- 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., Best, R., Bell, C., Witherspoon, A., & McNamara, D. S. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction*, 25(4), 399-438.
- Perfetti, C. A. (1985). *Reading ability*. Oxford University Press.
- 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). Malden: Blackwell.
- Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific studies of Reading*, 18(1), 22-37.
- Popham, W. J. (2001). *The truth about testing: An educator's call to action*. Beauregard St. Alexandria: ASCD.
- Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale: Erlbaum.
- Pronk-van Eunen, M., & de Vos, B. (2018). Lezen met de leessandwich: Een kansrijke didactiek bij Nederlands en mv. *Levende Talen Magazine*, 105(2), 10-14.
- Raad voor Cultuur & Onderwijsraad. (2019). *Lees! Een oproep tot een leesoffensief*. Den Haag: Raad voor Cultuur & Onderwijsraad.
- Raad voor Nederlandse taal en letteren. (2015). *Vaart met taalvaardigheid: Nederlands in het hoger onderwijs*. Adviesrapport. Den Haag: Taalunie.
- 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.
- Rooijackers, P., van Silfhout, G., Schuurs, U., Mulders, I., & van den Bergh, H. (2020). Lezen en antwoorden bij de tekst met vragen geobserveerd; een eye-trackstudie onder vwo 4-leerlingen. *Pedagogische Studiën*, 97(1), 42-58.
- Rouet, J. F. (2006). *The skills of document use: From text comprehension to Web-based learning*. Mahwah: Lawrence Erlbaum Associates Publishers.
- Rouet, J. F., Vidal-Abarca, E., Erbou, A. B., & Millogo, V. (2001). Effects of information search tasks on the comprehension of instructional text. *Discourse Processes*, 31(2), 163-186.
- Rupp, A. A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language testing*, 23(4), 441-474.
- Salmerón, L., Vidal-Abarca, E., Martínez, T., Mañá, A., Gil, L., & Naumann, J. (2015). Strategic decisions in task-oriented reading. *The Spanish journal of psychology*, 18, E102.
- Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in cognitive sciences*, 6(9), 382-386.
- Schotter, E. R., Tran, R., & Rayner, K. (2014). Don't believe what you read (only once) comprehension is supported by regressions during reading. *Psychological science*, 25(6), 1218-1226.
- Schram, D. (red.). (2007). *Lezen in het VMBO*. Eburon Uitgeverij.
- Schuurs, U. (2019). De constructie van een (alternatieve) leestoets: verslag en opbrengsten. (Intern verslag.) Arnhem: Cito.
- 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 den Branden, K., van Keer, H., & Ghesquière, P. (2019). *Sleutels voor effectief begrijpend lezen; Inspiratie voor een eigentijds didactiek in het basisonderwijs*. Brussel: Vlaamse Onderwijsraad.
- Van der Leeuw, B., Meestringa, T., van Silfhout, G., Smit, J., Hoogeveen, M., Prenger, J., Langberg, M., & Jansma, N. (2017). *Nederlands; Vakspecifieke trendanalyse 2017*. Enschede: SLO.
- Van der Schoot, M., Reijntjes, A., & van Lieshout, E. C. (2012). How do children deal with inconsistencies in text? An eye fixation and self-paced reading study in good and poor reading comprehenders. *Reading and Writing*,

25(7), 1665-1690.

- Van Ockenburg, L., van Weijen, D., & Rijlaarsdam, G. (2019). Learning to Write Synthesis Texts: A Review of Intervention Studies. *Journal of Writing Research, 10*(3), 401-425.
- 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.
- Vidal-Abarca, E., Martínez, T., Salmerón, L., Cerdán, R., Gilabert, R., Gil, L., Mana, A., Llorens, A.C., & Ferris, R. (2011). Recording online processes in task-oriented reading with Read&Answer. *Behavior research methods, 43*(1), 179-192.
- Wallot, S., O'Brien, B. A., Haussmann, A., Kloos, H., & Lyby, M. S. (2014). The role of reading time complexity and reading speed in text comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(6), 17-45.
- Welgraven, E. & de Vos, B. (2019, 27 augustus). Laat leerlingen actief met een tekst bezig zijn, in plaats van die eindeloze vragen. *Trouw*.
- Wesdorp, H. (1981). *Evaluatietechnieken voor het moedertaalonderwijs: een inventarisatie van beoordelingsmethoden voor de stelvaardigheid, het begrijpend lezen, de spreek-, luister- en discussievaardigheid*. 's-Gravenhage: Staatsuitgeverij.
- Wimmer, H. (1993). Characteristics of developmental dyslexia in a regular writing system. *Applied Psycholinguistics, 14*, 1-33.
- 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 aan het Sint-Janslyceum in 's-Hertogenbosch 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; a cross-sectional eye-tracking study among preacademic students

In Dutch education one often uses 'texts with questions' to train reading skills: students study a text in advance and then answer questions, while the text is still available. In this eye-tracking study reading processes of eighth grade ($n = 17$), tenth grade ($n = 16$) and twelfth grade ($n = 19$) preacademic students within this task were observed. It was investigated how younger and older preacademic students ($N = 52$) differ in their reading and answering processes and how initial text reading influences answering. Results: twelfth graders read texts initially faster than eighth graders, but not faster than tenth graders. Participants generally showed few signs of strategic reading. In the answering process older students knew better where to locate answers in the text and answered more questions correctly than younger students. Participants initially showed highly pragmatic reading behavior. The question is raised to what extent this task triggers superficial reading.

Keywords: text comprehension, reading process, answering process, preacademic students, eye-tracking.