

De effectiviteit van een lessenserie over diep lezen

Een interventieonderzoek onder vwo 4-leerlingen

P. Rooijackers, E. Mantingh, G. van Silfhout, en H. van den Bergh

Samenvatting

Men traint in Nederland tekstbegrip vaak op traditionele wijze: leerlingen bestuderen teksten en beantwoorden vervolgens tekstafhankelijke vragen. Er loopt momenteel in Nederland een discussie of en hoe het tekstbegripsonderwijs op middelbare scholen moet worden herzien; een alternatieve benadering van de tekstbegripsles lijkt wenselijk.

Gebaseerd op recente wetenschappelijke inzichten rondom tekstbegrip, is daarom een digitale lessenserie van zes lessen ontwikkeld over het 'diep lezen' van zakelijke teksten. Tekstbegrip wordt steeds functioneel ingebed en via observerend leren procesgericht benaderd, via een breed spectrum aan taken; het zelfstandig bestuderen van teksten staat voorop.

De lessenserie werd afgenomen onder 27 vwo 4-klassen, verdeeld over twee condities: in de conditie 'Zelfstandig leren' ($n = 329$) werkten leerlingen individueel de lessen door; in de conditie 'Docent-gestuurd leren' ($n = 372$) namen leerlingen aanvullend aan klassengesprekken deel.

Resultaten: in beide condities geven leerlingen zelf aan dat ze na de lessenserie meer leesgedrag zijn gaan inzetten dat voor diep lezen kenmerkend is; dit geldt sterker voor de conditie 'Docent-gestuurd leren'. In de conditie 'Docent-gestuurd leren' vertonen leerlingen na de lessenserie ook meer (diep) tekstbegrip dan vóór de lessenserie. In beide condities blijkt de kwaliteit van de uitvoering door leerlingen bepalend voor de leerwinst. De lessenserie lijkt daarmee een effectief eigentijds alternatief voor de gebruikelijke leesles Nederlands.

Kernwoorden: diep begrip, vwo-leerlingen, zakelijke teksten, e-learning, procesgericht

1 Inleiding

De laatste jaren is er internationaal meer aandacht gekomen voor de specifieke kenmerken van diep lezen (Allen & McNamara, 2020; Wolf, 2018). Diep lezen omvat "the array of sophisticated processes that propel comprehension and that include inferential and deductive reasoning, analogical skills, critical analysis, reflection, and insight" (Wolf & Barzillai, 2009, p. 32; vgl. Baron, 2015; Piper, 2012; Hayles, 2012; Wolf, 2018). Daarmee sluit diep lezen aan bij het theoretisch gefundeerde concept situatiemodel, zoals dat met name in de Constructie-Integratietheorie is uitgewerkt (Kintsch, 1998). Volgens Wolf (2018) zouden jonge lezers in de 21^e eeuw minder met diep lezen vertrouwd zijn dan voorgaande generaties; door de exponentiële informatietoename via internet en sociale media zouden jongeren juist bedreven zijn in scannend lezen, in *hyper-reading* (vgl. Liu, 2005; Sosnoski, 1999).

In Nederland, dat een sterk gedigitaliseerde samenleving kent (bijv. DESI, 2020), is deze analyse van de actuele leesvaardigheid van jongeren inmiddels door overheidsinstanties overgenomen: de jongeren in Nederland "lezen weliswaar volop korte tekstjes – berichten op hun smartphone of samenvattende stukjes in schoolboeken – maar besteden minder tijd aan 'diep lezen': het geconcentreerd lezen van langere teksten of boeken. Mede hierdoor gaat hun leesvaardigheid achteruit" (Onderwijsraad & Raad voor Cultuur, 2019, p. 2; vgl. Ministerie van Onderwijs, Cultuur en Wetenschappen, 2020; Pereira & Nicolaas, 2019).

1.1 Leesonderwijs in Nederland

Evenals in Nederlandstalig buurland Vlaanderen zijn er in Nederland de laatste jaren zorgen gerezen over het dalende tekst-

begrip van scholieren. In de internationaal vergelijkende PISA-test vertonen Nederlandse vijftienjarigen sinds 2012 een dalende trend in hun tekstbegripsscore; in vergelijking met Europese leeftijdsgenoten blijken ze vooral minder sterk in evalueren en reflecteren (OECD, 2019). Deze laatste bevinding wordt weerspiegeld in de bekende PIRLS-test naar het tekstbegrip van basisschoolleerlingen: Nederlandse groep 6-leerlingen zijn aanmerkelijk sterker in het vinden van informatie dan in het reflecteren op de tekstinhoud (Gubbels et al., 2017). De zorgen hierover hebben inmiddels ook de politiek bereikt (Ministerie van Onderwijs, Cultuur en Wetenschappen, 2020). Het lezen van informatieve teksten is dan ook een van de meest basale vaardigheden in de huidige informatiesamenleving (Catrysse, 2017; OECD, 2019).

In Nederland wijkt de inrichting van het domein leesvaardigheid in enkele opzichten af van de internationale standaard. Waar bijvoorbeeld in leermiddelen in het Angelsaksische onderwijs het lezen van fictionele en narratieve teksten vaak nauwelijks tot niet wordt onderscheiden van dat van zakelijke teksten (bijv. Fisher & Frey, 2015), wordt dit onderscheid in Nederlandse leermiddelen en leeslessen vrijwel steeds aangehouden: het is terug te vinden in leergangen begrijpend lezen voor het basisonderwijs (bijv. Bogaerds-Hazenberg et al., 2022), in leergangen van het schoolvak Nederlands voor het voortgezet onderwijs (bijv. Frank et al., 2019), in de door de overheid vastgelegde niveaubeschrijvingen voor taalvaardigheid in het Nederlandse primaire en secundaire onderwijs (Expertgroep Doorlopende Leerlijnen Taal, 2008), en in de centrale afsluitende toetsing van het voortgezet onderwijs (CvTE, 2022).

Er loopt momenteel in de Nederlandse politiek en het Nederlandse onderwijsveld een discussie of het taalonderwijs, en in het bijzonder het leesonderwijs in de Nederlandse taal, niet aan een fundamentele herziening toe is (bijv. Inspectie van het Onderwijs, 2019; Ministerie van Onderwijs, Cultuur & Wetenschappen, 2022; Onderwijsraad & Raad voor Cultuur, 2019; Pereira & Nicolaas, 2019). Vanuit de wetenschap zijn er eveneens zorgen geuit over het Nederlandse leesonder-

wijs (Van den Broek et al., 2021; Van Steensel et al., 2020). Op dit moment wordt er een grondige herziening van het schoolvak Nederlands voorbereid, die waarschijnlijk andere accenten en aanvullingen zal opleveren in de landelijke beschrijvingen van de beheersingsniveaus en examenprogramma's voor leesvaardigheid (vgl. Curriculum.nu, 2019; Rijks-overheid, 2022). We lopen een aantal van de gesignaleerde probleempunten in het huidige Nederlandse leesonderwijs langs.

Ten eerste: zoals wel in meer landen het geval is (bijv. Fang, 2016; Wijekumar et al., 2019), is in Nederland de taakinrichting van de lessen tekstbegrip vaak nogal eenzijdig: Nederlandse basisschoolleerlingen en middelbarescholieren lezen doorgaans een relatief lange tekst, zonder een functioneel leesdoel, en beantwoorden daarna, doorgaans individueel, een set tekstafhankelijke begripsvragen, waarna ze individueel, in groepjes of klassikaal hun antwoorden bespreken of nakijken (Bogaerds-Hazenberg et al., 2022; Linthorst & De Gloppe, 2015; SLO, 2021a). Al zijn er zeker positieve afwijkingen van deze standaardpraktijk (vgl. bijv. Pronk-Van Eunen & De Vos, 2018), daarmee spiegelt de Nederlandse praktijk in grote lijnen het leesonderwijs zoals dat in de klassieke studie van Durkin (1978) werd bekritiseerd: inzet van de leesles is nog altijd vaak of leerlingen een goed of fout antwoord op een vraag hebben. Op deze eenzijdige inrichting in de leesles is in Nederland al sinds decennia kritiek: leerlingen zouden ermee in hoge mate taak-specifieke verwachtingen en vaardigheden opbouwen, toetsing van tekstbegrip wordt verward met training, en leerlingen leren ermee vooral 'een kunstje' (bijv. Hoogeveen & Bonset, 1998; Hoogeveen, 2018; Westhoff, 2012). Voor deze kritiekpunten werd in eerder onderzoek onderbouwing gevonden: leerlingen lijken er inderdaad zeer specifieke taakverwachtingen mee op te bouwen (Rooijackers et al., 2020) en ze leren met deze taak waarschijnlijk niet zelfstandig een tekst te bestuderen (Rooijackers et al., 2021c).

Een tweede probleempunt: in de les zakelijk lezen is er slechts in geringe mate aandacht voor het leesproces, en de aandacht die er is, kent duidelijke beperkingen. In de lees-

les staan, zoals gezegd, doorgaans de leerlingantwoorden op begripsvragen centraal. Docenten richten zich met enige regelmaat op het leesproces via *modelling*, ze doen voor hoe ze zelf de inhoud van een (pittige) tekst zich eigen maken; hoe leerlingen zelf een tekst lezen en begrijpen, de effectiviteit van hun leesgedrag, blijft echter buiten beschouwing (Bogaerds-Hazenbergh et al., 2020; Linthorst & De Gloppe, 2015). Daarnaast is de vraag in hoeverre docenten in het Nederlandse voortgezet onderwijs in hun lessen evidencebased leesstrategieën (bijv. Duke et al., 2011) toepassen. Actuele leergangen Nederlands voor het voortgezet onderwijs bieden namelijk geen strategieën aan maar generieke manieren van lezen: oriënterend lezen, globaal lezen, studerend lezen, etc. (Schuppert & Van den Bergh, 2009; Rooijackers et al., 2021b). Deze methodiek is gebaseerd op verouderde inzichten uit de neerlandistiek, met name die in Drop en De Vries (1974), en verschilt markant van internationale inzichten over effectieve strategie-instructie (bijv. Almasi & Fullerton, 2012).

Een derde probleempunt: de aandacht gaat in de leesles vooral uit naar enkelvoudigtekstbegrip (*single text comprehension*) en de argumentatieve aanvaardbaarheid van informatie binnen een tekst (SLO, 2021a); meer-voudig-tekstbegrip (*multiple text comprehension*) en de beoordeling van de kwaliteit, betrouwbaarheid en bruikbaarheid van informatie komen doorgaans weinig aan bod (Van den Broek et al., 2021; Gubbels, 2020; Rooijackers et al., 2021c). Deze meer 21^e-eeuwse dimensies van leesvaardigheid hebben de laatste twee decennia internationaal ruim wetenschappelijke aandacht gekregen (bijv. Braasch et al., 2018; Britt et al., 2018) en spelen inmiddels een grote rol in de theoretische raamwerken van internationale toetsen als PISA en Pirls (OECD, 2019; Mullis et al., 2017), evenals in voorstellen voor curriculumherzieningen in bijv. de VS (Pearson et al., 2020). Evaluatie en reflectie bij een of meer teksten “were arguably always part of reading literacy, but their importance has increased with the increased amount and heterogeneity of information readers are faced with today” (OECD, 2019, p. 36). In

Nederland sluit de inhoud van leesparagrafen in leergangen, centrale toetsen en beschrijvingen van beheersingsniveaus daarentegen nog sterk aan bij inzichten uit *single-text*-lees-theorieën (met name Kintsch, 1998). Er weerklinken in Nederland dan ook al geruime tijd geluiden deze meer eigentijdse begripsdimensies een prominentere plek in het onderwijs te geven (bijv. Gubbels, 2020; Kirschner, 2017; Rooijackers et al., 2021c; SLO, 2021b).

Een laatste probleempunt, samenhangend met het voorgaande: de inzet van tekstbegrip voor een schoolse taak, zoals bij een schrijfo opdracht of presentatie, blijft in leergangen Nederlands nogal eens buiten beschouwing (vgl. Van den Broek et al., 2021; Pereira & Nicolaas, 2019; Rooijackers et al., 2021c; SLO, 2021a). Onderzoek laat echter zien dat vaardige lezers zich niet alleen van minder vaardige lezers onderscheiden in tekstbegrip, maar ook in de wijze waarop ze dat begrip voor schoolse taken opbouwen en toepassen: leerlingen kunnen een opgegeven leestaak onderling bijvoorbeeld nogal afwijkend interpreteren en op grond daarvan afwijkende representaties van teksten opbouwen (Britt et al., 2018). Deze inbedding van tekstbegrip binnen een specifieke, contextgebonden leestaak ontbreekt in de lessen Nederlands doorgaans. Zonder een functionele, doelmatige context wordt lezen echter een geïsoleerde, schijnbaar ‘doelloze’ vaardigheid (bijv. Berenst et al., 2016; Van den Broek et al., 2021; Pereira & Nicolaas, 2019); bovendien leidt het integreren van leesvaardigheid in bijvoorbeeld schrijftaken vaak tot een beter begrip van teksten (bijv. Graham & Perrin, 2007; Graham & Hebert, 2011). Er is in Nederland daarom wel ervoor gepleit om tekstbegrip consequent te koppelen aan kennisdomeinen in het schoolvak Nederlands (Meesterschapsteam Nederlands, 2019). Deze functionele inbedding van tekstbegrip Nederlands is dan ook onderwerp van discussie bij de aanstaande curriculumherziening van het schoolvak (Curriculum.nu, 2019; SLO, 2021b).

1.2 Diep lezen centraal

Het concept diep lezen, vaak beschreven als de tegenhanger van *hyper reading* (Sosnoski,

1999; Wolf, 2018; vgl. Baron, 2015; Hayles, 2012), vormt sinds het laatste decennium steeds vaker onderwerp van onderzoek (vgl. Allen & McNamara, 2020). Diep lezen sluit sterk aan bij het begrip situatiemodel zoals beschreven in de meest courante theoretische modellen over de constructie van tekstbegrip (vgl. McNamara & Magliano, 2009), met name de Constructie-Integratie-Theorie (Kintsch & Van Dijk, 1978; Kintsch, 1998) en het Landschapsmodel (Van den Broek et al., 1999). In deze theorieën wordt een tekst beschouwd als een netwerk van hiërarchisch samenhangende, betekenisvolle beweringen, waarin lezers relaties leggen tussen woorden, zinnen en alinea's. Dit constructieproces mondt gaandeweg uit in een tekstbasis, een vrij letterlijke mentale representatie van de tekstinhoud, en in een situatiemodel, een mentaal model waarin lezers de tekstuele wereld integreren in hun eigen voorkennis en opvattingen. Pas op dit laatste niveau kennen lezers aan een tekst werkelijk begrip toe (Kintsch, 1998). Door de persoonlijke, actieve verwerking van de tekstinhoud, tijdens én na lezing, ontstaat bij lezers vervolgens diep begrip van een tekst, aldus Wolf (2018): doordat lezers een tekst geconcentreerd, herhaaldelijk lezen en zich op analyse, reflectie en evaluatie richten, zijn ze in staat de tekstinhoud uit te diepen: te analyseren en interpreteren, de kwaliteit en geloofwaardigheid ervan te bepalen, vanuit de tekstinhoud logische vervolgdienstapen te zetten, etc. (Wolf & Barzillai, 2009). Diep lezen is als concept sterk verbonden aan literatuurwetenschappelijke inzichten over de denkprocessen die met name literatuur van lezers kan vergen (vgl. Wolf, 2018; Hayles, 2012).

Deze geconcentreerde, convergente en herhaalde lezing van één tekst vormt in veel studies de tegenhanger van *hyper-reading* (vgl. Sosnoski, 1999; Wolf, 2018). Lezers blijken informatie op internet doorgaans eenmalig, non-lineair en scannend te lezen, vooral op zoek naar kernwoorden, springend van tekst naar tekst; het lineair, diep en langdurig geconcentreerd lezen van één tekst raakt daardoor op de achtergrond (bijv. Halamish & Elbaz, 2020; Liu, 1999; Mangen et al., 2013; Sosnoski, 1999; Wolf, 2018). Deze

aangepaste leeswijze zouden lezers hebben ontwikkeld om de overdadige hoeveelheid beschikbare informatie op internet en social media efficiënt te kunnen verwerken (bijv. Baron, 2015; Piper, 2012; Wolf, 2018). Echter, met deze aangepaste leeswijze laten ze vaak tal van diepere begripsdimensies in een tekst buiten beschouwing, zoals de betrouwbaarheid van de verstrekte informatie en de inhoudelijke consistentie ervan (Wolf, 2018). De dichotomie *hyper reading* en diep lezen wordt overigens nog weleens gerelativeerd: het is de vraag in hoeverre de moderne lezer door de hedendaagse overvloedige informatiestroom niet juist vooral nieuwe strategieën en nieuwe leeswijzen heeft ontwikkeld, waarbinnen *deep reading* een van de vele instrumenten is (Clowes, 2019; Jones & Hafner, 2012).

In hoeverre een lezer een tekst diep leest, is afhankelijk van leesdoel en leessituatie (Wolf, 2018). Een uitgebreid krantenartikel over een onderzoek kan een lezer thuis snel en *hyper-readend* tot zich nemen, maar hetzelfde krantenartikel kan een lezer diep lezen als het als basis dient voor een presentatie op werk of school. Sinds de millenniumwisseling is in internationaal leesonderzoek dan ook steeds meer deze intentionaliteit achter leesgedrag in kaart gebracht: de situationele en functionele inbedding van tekstbegrip (Britt et al., 2018; McCrudden et al., 2011; Van den Broek & Helder, 2017; etc.). Uiteenlopende leesdoelen en leessituaties blijken bij lezers te leiden tot deels afwijkende begripsprocessen en een afwijkende inzet van vaardigheden en strategieën: wie een tekst voor een schoolse opdracht bestudeert, zet andere vaardigheden en strategieën in tijdens het lezen en legt andere accenten in een tekst dan wie diezelfde tekst voor recreatieve doeleinden leest (Van den Broek & Helder, 2017; Yeari et al., 2015).

Deze intentionaliteit achter tekstbegrip wordt ook verdisconteerd in recente theoretische leesmodellen (Britt et al., 2018; vgl. Britt & Rouet, 2012; McCrudden & Schraw, 2007). Volgens het recente RESOLV-model, dat leesgedrag bij schoolse leestaken wil verklaren (Britt et al., 2018; zie ook Rouet et al., 2017), bouwen lezers bij een taak over een of

meerdere teksten vooraf eerst een statisch contextmodel op. Daarin verdisconteren ze zaken als taakinterpretatie, de status van de ‘opdrachtgever’, hulpmiddelen en belemmeringen bij de taak, evenals een zelfbeoordeling bij deze taak. Vanuit dit contextmodel stellen lezers vervolgens een dynamisch taakmodel op: taakspecifieke doelen en scenario’s, waarin ze interpreteren wat ze moeten doen om de doelen te bereiken, en welke procedures ze daarbij moeten hanteren. Deze context- en taakmodellen sturen vervolgens de constructie van begrip bij een tekst of teksten bij een lezer. Vaardige lezers onderscheiden zich bij het opbouwen en toepassen van dergelijke context- en taakmodellen sterk van minder vaardige lezers (Britt & Rouet, 2012; Gil et al., 2010).

1.3 Strategie-instructie

Een belangrijk instrument om een tekst diep te lezen, vormen strategieën. Strategieën zijn voor een belangrijk deel persoons-, tekst- en doelaafhankelijk (Almasi & Fullerton, 2012; Pressley & Afflerbach, 1995). Wanneer twee vaardige lezers eenzelfde tekst ‘strategisch’ lezen, blijken ze nogal andere accenten in een tekst te plaatsen en daarbij (deels) andere strategieën toe te passen (vgl. Guthrie et al., 1996; Pressley & Afflerbach, 1995). Desalniettemin zijn er duidelijke gemeenschappelijke kenmerken in het leesgedrag van vaardige lezers aan te wijzen, zoals het identificeren en herlezen van kernzinnen, het bepalen van de hoofdgedachte in een tekst, het herlezen van ingewikkelde passages, etc. (vgl. Duke & Pearson, 2002; Duke et al., 2011; Pressley et al., 1995).

Een grondige herlezing van de belangrijkste elementen in een tekst lijkt een minimaal vereiste voor begrip bij een enigszins complexe tekst. Door zinnen, regels en alinea’s terug te lezen, kunnen lezers zowel falend tekstbegrip repareren als hun begrip van een tekst controleren en verdiepen (Van den Broek & Helder, 2017). Dat geldt hoogstwaarschijnlijk *a fortiori* voor situaties waarin lezers een complexe tekst moeten begrijpen. Een belangrijke voorwaarde voor diep begrip is immers de onderdompeling van de lezer in de wereld van de tekst (Wolf, 2018). Het intensief teruglezen van tekstgedeelten is dan

ook een overkoepelend kenmerk van veel van de expert-leesstrategieën (Pressley & Afflerbach, 1995).

Onderzoek naar de effectiviteit van strategie-instructie kent inmiddels een lange traditie (Duke & Pearson, 2002; Afflerbach et al., 2020). Een breed arsenaal aan onderzoek toont de effectiviteit van expliciete strategie-instructie aan (Donker et al., 2014; Graham & Harris, 1994; Hattie et al., 1996; Rogiers, 2019; etc.; vgl. ook Okkinga et al., 2018). Ouder onderzoek naar de effectiviteit van leesstrategie-instructie vertrekt nog weleens vanuit een min of meer vaste set leesstrategieën die onderwezen wordt (bijv. Oostdam & Bimmel, 1996; Palincsar & Brown, 1984; Lysynchuk et al., 1990; etc.); vanaf midden jaren negentig vertrekt interventieonderzoek naar strategie-instructie sterker vanuit de opvatting dat lezers zelfstandig en gecoördineerd een flexibel strategie-instrumentarium moeten kunnen toepassen in uiteenlopende leessituaties (bijv. Brown et al., 1996; Pressley et al., 1995). Onderzoek naar de situationele inbedding van tekstbegrip in de laatste twee decennia (zie 1.2) heeft de legitimatie van deze flexibele benadering alleen maar versterkt: niet alleen wisselen lezers in de wijze waarop ze begrip opbouwen, ze variëren ook sterk in de wijze waarop ze dit begrip in uiteenlopende leessituaties toepassen (Afflerbach et al., 2020). Goed strategie-onderwijs zou daarom niet alleen begrip, maar ook het doel en de situationele inbedding van dit begrip moeten verdisconteren (ibidem; NAEP Reading Framework, 2017): met andere woorden, strategieonderwijs moet nadrukkelijk oog hebben voor de conditionele dimensie van strategiegebruik.

Bij strategieonderwijs kan *observational learning* (Bandura, 1986) een logisch vertrekpunt vormen. Bij dit leerprincipe voeren leerlingen niet direct zelf een taak uit maar observeren ze eerst de taakuitvoering van een model, reflecteren erop en evalueren het. Omdat leerlingen de taak niet zelf uitvoeren, hebben ze meer cognitieve ruimte beschikbaar om te leren: ze verzamelen zo metacognitieve en procedurele kennis over de taakuitvoering en kunnen deze vervolgens toepassen als ze zelf een soortgelijke taak gaan verrichten. Vooral binnen het domein schrijfvaardig-

heid is de positieve werking van observerend leren redelijk uitgebreid in kaart gebracht (Braaksma et al., 2004; Couzijn, 1995; Rijlaarsdam et al., 2008; etc.), maar ook binnen het domein leesvaardigheid zijn er inmiddels studies beschikbaar die de effectiviteit ervan laten zien. Leerlingen blijken na een lessenserie met *observational learning* vaker strategisch leesgedrag in te zetten dan leerlingen die de lessenserie niet gevolgd hebben (Salmerón et al., 2020; Salmerón & Llorens, 2019), en ze geven naderhand meestal ook van een beter tekstbegrip blijk (Couzijn, 1995; Rooijackers et al., 2021a; Salmerón & Llorens, 2019). Gedragsveranderingen door observerend leren hoeven zich overigens niet steeds te vertalen in een meetbaar verhoogd resultaat op lees- of schrijfvaardigheidsproducten, aldus de *social learning theory* (Bandura, 1986; Bandura & Walters, 1977): iets leren over het proces uit zich niet steeds in een beter product. In Salmerón et al. (2020) bijvoorbeeld vertoonden de leerlingen in de eye-trackresultaten wel duidelijke gedragsveranderingen – maar dit leidde niet tot een aantoonbaar effect op hun tekstbegrip.

1.4 Doelen en onderzoeksvragen

Aansluitend op voornoemde ontwikkelingen in het Nederlandse (lees)onderwijs, construeerden wij een zelfstandig uitvoerbare, digitale module over diep lezen voor het voortgezet onderwijs, bestaande uit zes lessen. Met deze lessenserie stellen we onszelf twee algemene doelen: ten eerste willen we erin moderne inhoudelijke inzichten rondom tekstbegrip een ruime plaats geven, met name de intentionaliteit achter tekstbegrip (Britt et al., 2018; McCrudden et al., 2011), een brede conceptualisatie van evaluatie en beoordeling van teksten (OECD, 2019), en daarnaast inzichten rondom (de kenmerken van) diep begrip (Wolf, 2018).

Het tweede doel komt voort uit de uitzonderlijke educatieve omstandigheden, ontstaan door de Covid19-epidemie: we wilden de effectiviteit van de lessenserie laten zien bij een volledig zelfstandige, individuele afname ervan. Door de maandenlange nationale schoolsluiting gedurende twee lockdowns nam het digitaal lesgeven (blended learning

en e-learning) in Nederland een hoge vlucht, met name in het voortgezet onderwijs (SLO, 2021a; Smeets, 2021). Doordat een heropening van de scholen onzeker leek, ontwierpen wij daarom de lessenserie als een vorm van e-learning, zodat deze, bij de leerling thuis of op school, als een individueel, zelfstandig maakbare lesmodule kon worden afgenomen.

Doordat de Nederlandse scholen bij de start van het schooljaar 2021-2022 wederom de deuren voor hun leerlingen openden, deed de gelegenheid zich alsnog voor om de ontwikkelde digitale lessenserie niet alleen zelfstandig, maar ook onder gedeeltelijke begeleiding van een docent af te nemen. De begeleidende rol van de docent vormt immers een cruciale meerwaarde voor leren (Hattie, 2012), een belang dat tijdens de lockdowns weleens werd onderschat (Smeets, 2021). Daardoor werd besloten het experiment in twee *experimentele condities* aan te bieden, naast een controleconditie: in de ene experimentele conditie werkten de leerlingen de lessenserie (vrijwel) geheel individueel en zelfstandig door, terwijl in de andere experimentele conditie de docent in drie van de zes lessen een sterk sturende rol had, met name door middel van klassengesprekken. Om de ecologische validiteit van deze studie te verhogen, hebben we ervoor gekozen docenten zelf het experiment af te laten nemen, zonder uitgebreide training of assistentie in de klas door onderzoekers.

De *wijze van uitvoering* van deze lessenserie door leerlingen is daarmee wel een nadrukkelijk aandachtspunt in deze studie. In een eerdere pilotstudie (Rooijackers et al., 2021a) werden de lessen in een klas in een mengvorm van individuele, groeps- en klassikale lesactiviteiten afgenomen, door de docent Nederlands van de klas, samen met een onderzoeker. In deze pilotstudie bleken de leerlingen die een lessenserie van vier lessen hadden gevolgd, naderhand meer diep begrip te hebben opgebouwd dan leerlingen die deze lessenserie niet hadden gevolgd. De vraag is in hoeverre de resultaten uit deze studie in deze vervolgstudie gerepliceerd kunnen worden. Leerlingen werken het lesmateriaal immers (naargelang conditie: voor een belangrijk deel of geheel) zelfstandig en indi-

vidueel door. e-Learning blijkt echter lang niet altijd een duidelijk leerrendement op te leveren; de interactie met de docent, die tussen leerlingen en die tussen leerlingen en het computersysteem hebben bijvoorbeeld een impact op de mate waarin leerlingen van een e-learning-module leren: zonder feedback van én interactie met anderen kunnen leerlingen een isolement ervaren dat hun engagement met de module in de weg staat (bijv. Bouhnik & Marcus, 2006; Liaw, 2008; Martínez et al., 2007; Udo et al., 2011). De vraag is dus in hoeverre leerlingen zich, zonder de begeleidende, bemoedigende en sturende rol van de docent én zonder de afwisseling in werkvormen en lesactiviteiten, voldoende engageren en passend op het lesmateriaal reageren, vooral ook omdat de lessenserie in een aantal opzichten markant afwijkt van de gebruikelijke inrichting van de leesles Nederlands.

Met de lessenserie richten we ons specifiek op vwo 4-leerlingen. Er zijn wel zorgen geuit over het tekstbegrip van Nederlandstalige universitaire studenten (Van Houtven et al., 2010; Raad voor Nederlandse Taal en Letteren, 2015; Warps et al., 2021). Deze zouden over het algemeen te weinig kritisch kunnen lezen (Warps et al., 2021), terwijl juist academisch georiënteerde leerlingen later in studie en werk vaak geconcentreerd en diepgaand een tekst zullen moeten bestuderen en vanuit meerdere teksten over een onderwerp in kwaliteit wisselende informatie zullen moeten evalueren en integreren (Goldman, 1997; Hermida, 2009; Perfetti, 1997).

In deze studie onderzoeken we dus de effectiviteit van deze lessenserie over diep begrip van zakelijke teksten, onder vwo 4-leerlingen; we zullen daarbij enerzijds verkennen in hoeverre leerlingen na de lessenserie meer metacognitieve kennis over diep begrip hebben opgebouwd en anderzijds verkennen in hoeverre leerlingen ná de lessenserie blijken van meer diep begrip in begripstoetsen dan vóór de lessenserie: iets leren over het proces uit zich volgens de *observational learning theory* immers niet steeds in een beter product. Daarnaast onderzoeken we ook in hoeverre de wijze van uitvoering door leerlingen op de effectiviteit van de lessenserie van invloed is.

Daarmee staan in deze studie dus drie onderzoeksvragen centraal:

- a) In hoeverre maakt de lessenserie vwo 4-leerlingen bewust van (de voorwaarden voor) diep begrip bij zakelijke teksten, en wijkt dit af per conditie?
- b) In hoeverre geven de vwo 4-leerlingen die de lessenserie hebben doorlopen, blijk van meer diep begrip als ze daarna zelf zakelijke teksten lezen, en wijkt dit af per conditie?
- c) In hoeverre is de wijze van uitvoering (implementatie) van deze lessenserie van invloed op de effectiviteit ervan?

Vertrekkend vanuit de staande praktijk in de Nederlandse leesles, gebruikten we in de lessenserie geen narratieve of fictionele teksten – al staan deze natuurlijk bij uitstek diep lezen toe (vgl. Hayles, 2012; Wolf, 2018) – maar uiteenzettende en betogende zakelijke teksten uit kranten en (digitale) tijdschriften.

2 Methode

2.1 Participanten

Het experiment is afgenomen op zeven Nederlandse middelbare scholen, in het noorden, midden en zuiden van Nederland, in 27 vwo 4-klassen ($N = 701$). De experimentele lessenserie als geheel bestond uit zes lessen van ca. 40 minuten: drie observerend-lerenlessen, afgewisseld met drie werklessen met speciaal ontworpen leestaken. Alle lessen zijn gegeven onder toezicht van de betrokken docenten Nederlands ($N = 17$) van de school, zonder enige actieve inbreng van de onderzoekers.

Voorafgaand aan het experiment werd met de deelnemende docenten de randvoorwaarden voor deelname besproken. Hierna ontvingen de deelnemende docenten een uitgebreide handleiding bij het experiment, met daarin het volgende: een korte bespreking van de theoretische uitgangspunten achter de lessenserie, praktische informatie, uitgewerkte lesplannen en een technische handleiding voor het werken met de digitale app waarmee de lessenserie kon worden afgenomen. Docenten konden bij vragen, problemen of andere ver-

legeneidssituaties vóór en tijdens het experiment altijd contact opnemen met een van de onderzoekers.

2.2 Experimentopzet

Het experiment werd afgenomen in twee experimentele condities. Deze condities weken niet af in inhoud, maar enkel in instructiemodus. De eerste conditie, ‘Zelfstandig leren’, werd afgenomen op vier scholen, met dertien vwo 4-klassen ($n = 329$): in deze conditie werkten de leerlingen zelfstandig en individueel de zes lessen van de lessenserie door en had de docent geen noemenswaardige inhoudelijke inbreng. De tweede conditie, ‘Docent-gestuurd leren’, werd afgenomen op drie scholen, met veertien vwo 4-klassen ($n = 372$): in deze conditie ontvingen leerlingen klassikaal drie lessen (de observerend-lerenlessen – zie onder), waarin de docent les gaf en klassengesprekken leidde; de andere drie lessen (de werklessen) werkten de leerlingen zelfstandig en individueel door.

Om de effectiviteit van de lessenserie in het algemeen én het verschil in effectiviteit tussen de twee onderzoekscondities te kunnen meten, hebben we in beide condities gebruikgemaakt van een *switching-replications*-design (Shadish et al., 2002). Klassen zijn telkens per school gerandomiseerd verdeeld in twee groepen, die op wisselende momenten de interventie ontvingen. Beide groepen maakten allereerst een voortoets. Vervolgens ontving groep 1 als eerste de interventie, terwijl groep 2 regulier onderwijs volgde, onderwijs dat op een ander domein dan tekstbegrip betrekking had. Een tweede toets volgde.

Minimaal drie weken en maximaal vijf weken nadat groep 1 met de interventie was gestart, ontving vervolgens ook groep 2 de interventie, terwijl nu groep 1 regulier onderwijs volgde. Een natoets mat ten slotte het tekstbegrip van beide groepen: deze werd afgenomen vrijwel onmiddellijk na het afronden van de lessenserie door groep 2 en drie à vier weken na het afronden van de lessenserie door groep 1. In Figuur 1 is dit design grafisch weergegeven voor beide condities.

Een *switching-replications*-design heeft *a priori* een aantal voordelen. Allereerst is er een belangrijk ethisch voordeel: door de interventie in tijd te spreiden, kunnen beide onderzoeksgroepen de interventie ontvangen. Daarnaast zijn er drie belangrijke methodologische voordelen: (a) groep 2 kan als controlegroep dienen terwijl groep 1 de interventie ontvangt; (b) beide groepen ontvangen de interventie, zodat het effect van de interventie tweemaal kan worden getoetst; (c) nadat ook groep 2 de interventie heeft gevolgd, kan de natoets van groep 1 als een *delayed retention test* gelden (Shadish et al., 2002 – in onze studie wordt deze retentie overigens slechts binnen een beperkt tijdpad gemeten: al drie weken nadat groep 1 de lessenserie heeft afgenomen, kan de natoets zijn afgenomen).

Binnen deze studie, waarbij we gebruikmaken van twee condities, heeft dit design ten slotte het voordeel dat het ons toestaat niet alleen de effectiviteit binnen een conditie te onderzoeken, maar ook het eventuele verschil in effectiviteit tussen beide onderzoekscondities.

Conditie Docent-gestuurd leren

Groep 1	T ₁	X	T ₂	T ₃
Groep 2	T ₁		T ₂	X

Conditie Zelfstandig leren

Groep 1	T ₁	X	T ₂	T ₃
Groep 2	T ₁		T ₂	X

Figuur 1

Grafische weergave van het *switching-replications*-design voor beide onderzoekscondities, waarbij T's de toetsmomenten aanduiden en de X'en afnamemomenten van het experiment in de twee groepen van elke conditie.

2.3 Materiaal

Algemene opzet lessenserie. Het centrale leerdoel achter de lessenserie was dat we leerlingen kennis wilden laten opbouwen over diep begrip en dat we ze wilden leren doelgericht, bewust en zelf-regulerend (strategisch) tot diep begrip van een tekst te komen (Afflerbach et al., 2008; Almasi & Fullerton, 2012). We boden in de lessenserie daarom geen vaste set leesstrategieën aan, maar wilden leerlingen vooral laten reflecteren op de effectiviteit van hun eigen leesgedrag. Observerend leren (Bandura, 1986) vormde daarbij een schragend leerprincipe: de lessenserie wisselde steeds af tussen een observerend-leren-les en een werkles. In de observerend-leren-les observeren leerlingen twee *peer*-modellen die duidelijk variëren in het tekstbegrip dat ze opbouwen. Leerlingen reflecteren op het opgebouwde begrip van de twee modellen en op de leesmiddelen die deze hierbij inzetten, én leerlingen evalueren dit (vgl. Braaksma, 2002). In de werklessen kunnen leerlingen daarna de (meta)cognitieve kennis uit de observerend-leren-les zelf op leestaken toepassen.

Daarbij wilden we dat leerlingen doelbewust en zelfstandig de in de lessenserie opgenomen teksten zouden lezen, met als doel diep begrip voor een specifieke leestaak. Dat diepe leesgedrag probeerden we op twee manieren af te dwingen. Eerste manier: de leestaken werden in de werklessen zo geconstrueerd dat leerlingen vooraf de tekst intensief moesten bestuderen, waarna ze bij de eigenlijke verwerkingstaak (bijv. begripsvragen beantwoorden, een sorteertaak maken) niet langer over de tekst zouden beschikken. In eerder onderzoek naar het lezen van teksten en het beantwoorden van begripsvragen (Rooijackers et al., 2020) bleken vwo-leerlingen bij het vooraf-lezen van de teksten vaak geneigd tot oppervlakkig leesgedrag, waarna ze zich vooral bij het beantwoorden van de begripsvragen zetten tot een intensievere tekstraadpleging. Het feit dat leerlingen weten dat ze tijdens de verwerkingstaak de tekst nog kunnen raadplegen, lijkt een belangrijke verklaring voor dit oppervlakkige vooraf-leesgedrag (vgl. Rooijackers et al., 2021c). Wanneer lezers weten dat ze bij het beant-

woorden van begripsvragen juist niet meer de tekst kunnen raadplegen, zetten ze zich vaak wél tot zo'n diepgaande vooraf-bestudering: lezers blijken de tekst dan vooraf langer te bestuderen, met meer oog voor diepe betekenis, en hun antwoordscore op begripsvragen hangt dan vooral samen met de mate waarin ze tijdens vooraf-lezen een adequate coherente mentale representatie van de tekstinhoud hebben gemaakt (Schroeder, 2011; vgl. Ferrer et al., 2017; Higgs et al., 2017; Ozuru et al., 2007). Omdat we wilden dat leerlingen zelfstandig zo'n coherente mentale representatie van de tekstinhoud opbouwden, benadrukten we dus dat leerlingen de verwerkingstaak moesten kunnen maken op basis van hun vooraf-bestudering van de tekst – en dat ze daarom niet langer de tekst bij de verwerkingstaak konden raadplegen.¹

Tweede manier: we wisselden bij elke werkles sterk in de vorm van de uiteindelijke leestaak. Door veelvuldige blootstelling aan één taak (de tekst met vragen) lijken Nederlandse leerlingen in de leesles zich vooral te richten op de taak, namelijk het beantwoorden van vragen: een zelfstandige bestudering van een tekst schiet er dan gemakkelijk bij in (Bogaerds-Hazenberg et al., 2022; Rooijackers et al., 2020). Daarom gebruikten we in de lessenserie naast tekstafhankelijke vragen (Fisher & Frey, 2015), ook sorteertaken (McNamara & Kintsch, 1996) en samenvattings- en argumentatietaken (Gil et al., 2010).

Daarnaast wilden we leerlingen via de lessenserie drie typen kennis over diep lezen laten opbouwen. Effectieve, kwaliteitsrijke strategie-instructie heeft idealiter op drie typen kennis betrekking: declaratieve, procedurele en conditionele kennis (bijv. Lorch et al., 1993; Paris et al., 1983). Vooral de conditionele inbedding van strategieën en vaardigheden ontbreekt in het Nederlandse leesonderwijs nogal eens (Bogaerds-Hazenberg et al., 2022; Rooijackers et al., 2021b). Daarom werd de lessenserie opgedeeld in drie sets van twee lessen, waarbij elke set vooral een van deze drie kennistypen centraal stelde.

De conditionele inbedding van begrip vertaalde zich ook in twee andere ontwerpprincipes. Ten eerste werden alle leestaken in

Tabel 1
 Overzicht van leerdoelen en leeractiviteiten in lessenserie met bijbehorende lesnummers

Leerdoel	Les	Leeractiviteit				
		observerend leren		individueel werken	voortgang monitoren	
		via EMME	via animatie	zelfstandig	quiz-vragen	open-vragen
declaratieve kennis van diep lezen	1-6	1, 3, 5	-	2	2	2
procedurele kennis van diep lezen	3-6	3, 5	-	2, 4	4	4
conditionele kennis van diep lezen	5-6	-	5	2, 4, 6	6	6
reflectie op taakmodel bij leesopdracht	5-6	-	5	6	5, 6	6
zelfstandig begrip opbouwen	1-6	-	-	2, 4, 6	-	-
scenario-based leestaak	1-6	-	-	-	-	-
meervoudig-tekstbegrip	6	-	-	6	-	-

de lessenserie binnen een functionele context aangeboden (McCrudden et al., 2011). Leestaken werden steeds voorafgegaan door een voor leerlingen herkenbaar scenario, met een duidelijk leesdoel, waarvoor diep lezen noodzakelijk was. Men denke aan een voordracht over een artikel, een voorbereiding via een artikel op schrijftoetsen, deelname aan een discussie naar aanleiding van een artikel, etc. Ten tweede wilden we leerlingen niet alleen diep begrip laten opbouwen en/of hierop reflecteren, maar we wilden ze ook confronteren met de vraag hoe ze een adequaat taakmodel bij een leestaak opbouwen: bouwen modellen, en leerlingen zelf, wel het juiste begrip *voor de taak* op? We gebruikten daarom (in de laatste observerend-leren-les) een tekst die ogenschijnlijk vooral niet bruikbaar leek voor een opdracht en (in de laatste werkles) een tekst die juist ogenschijnlijk zeer bruikbaar leek, maar, vanuit diep begrip, dat nauwelijks of niet bleek te zijn. Daarbij werd in deze twee lessen aanvullend meervoudig-tekstbegrip (*multiple text comprehension*) ingezet, zodat de bruikbaarheid van een besproken tekst beter kon worden bepaald.

Ten slotte: we hanteerden het principe van de *gradual release of responsibility* (Fisher & Frey, 2013). Omdat we met name de gemid-

deld vaardige en minder vaardige vwo 4-lezer met deze lessenserie wilden bedienen – lezers die waarschijnlijk minder geverseerd zijn in diep begrip –, werden in de eerste lessen teksten gekozen en opdrachten geconstrueerd met een gemiddelde moeilijkheidsgraad. In de laatste twee lessen werd de complexiteit opgevoerd en werd het concept diep begrip rijk uitgewerkt. In Tabel 1 is voor elk van de zes lessen de leerdoelen en -activiteiten aangegeven.

Materiaal experimentele lessenserie: de eerste twee observerend-leren-lessen. In les 1 en 3, de eerste twee observerend-leren-lessen, reflecteerden leerlingen via modellen op de vragen: (1) ‘Wat is diep begrip?’ (declaratieve kennis) en (2) ‘Hoe bouw je het op?’ (procedurele kennis). Omdat begripsconstructie door lezers doorgaans een vrij fijnmazig proces is, gebruikten we voor deze modellen EMME’s: *Eye-Movements Modeling Examples* (vgl. Jarodzka et al., 2013). Een proefpersoon heeft eerst een taak uitgevoerd op een beeldscherm, terwijl een eye-tracker diens oogbewegingen (fixaties) op het beeldscherm registreerde. Een onderzoeker construeert vervolgens van deze oogbewegingen op het scherm een filmpje en monteert het

audiocommentaar van de proefpersoon eronder. Zo'n EMME kunnen docenten vervolgens in de les gebruiken om met leerlingen op leeractiviteiten te reflecteren. Een EMME biedt duidelijke didactische voordelen: leerlingen krijgen een vrij directe inkijk in de complexe perceptuele en cognitieve processen van een lezer; bovendien blijkt het bekijken van EMME's vaak motiverend (Salmerón & Llorens, 2019). Veel internationale didactische studies vonden dan ook positieve leer-effecten bij EMME-inzet in de klas (zie voor een overzicht: Van Marlen, 2019).

Als basis voor de EMME's maakten we gebruik van de eye-track-data uit een eerder *mixed-method*-onderzoek (Rooijackers et al., 2020). In elke les boden we via EMME's twee modellen aan die sterk contrasteerden in leesvaardigheid (Salmerón & Llorens, 2019). De vaardige lezers in de filmpjes onderscheidden zich duidelijk van de minder vaardige lezers a) in leestempo, want vaardige lezers zijn doorgaans snellere lezers dan minder vaardige lezers (Perfetti & Stafura, 2014); en b) in strategiegebruik, want vaardige lezers zetten doorgaans vaker en efficiënter strategieën in die het opgebouwde begrip bij een tekst verdiepen (Van den Broek & Helder, 2017). Per model werden kenmerkende fragmenten gebruikt van in totaal doorgaans een anderhalve minuut.

In het audiocommentaar reflecteerden de model-leerlingen achteraf op de eye-track-beelden van hun leesgedrag: waarom keken ze terug naar eerdere zinnen of tekstdelen? Begrepen ze naar hun eigen idee wat ze lazen? De modellen varieerden hier sterk in a) de mate waarin ze in eigen woorden de tekstinhoud al dan niet adequaat weergaven, en b) de kwaliteit waarmee ze reflecteerden: vaardige lezers kunnen namelijk doorgaans scherper de rationale achter eigen leesgedrag verwoorden dan minder vaardige lezers (McNamara & Magliano, 2009).

Bij de leerlingtaken voor deze twee EMME-lessen analyseerden leerlingen niet alleen de EMME's afzonderlijk, maar moesten ze ook beargumenteren welk model op basis van de vertoonde beelden en het hoorbare commentaar naar hun idee het duidelijkst bij de gelezen tekst diep begrip opbouw-

de, en welk model het minst (vgl. Braaksma, 2002). Tijdens verschillende stadia van de constructie werd dit materiaal gescreend door een groep screeners en er vonden meerdere try-outs plaats.

Materiaal experimentele lessenserie: de derde observerend-leren-les. Les 5, de derde observerend-leren-les richtte zich op de vraag 'Wanneer en waarom bouw je diep begrip op?' (conditionele kennis). Omdat in deze les vooral de kwaliteit van het taakmodel bij een leestaak centraal stond (Britt et al., 2018), gebruikten we voor de modellen in deze les geen EMME's, maar video-animaties waarin de modellen hardop nadachten tijdens het verrichten van een complexe leestaak.

Hiertoe ontwierpen we allereerst een leestaak waarbij leerlingen een korte tekst zich eigen moesten maken voor een beleidswerkgroep voor een schoolkantine. Een zestal vwo 4-leerlingen verrichtte daarna hardop de taak, terwijl we audio-opnames van hun hardopdenkproces maakten. Hieruit werden twee opnames geselecteerd die contrasteerden in de kwaliteit en scherpheid a) waarmee een leerling de inhoud van de tekst kon weergeven, b) waarmee deze reflecteerde op de bruikbaarheid van de tekst, en c) waarmee deze de tekstuele gedachtegang toepaste binnen het scenario.

In de animaties stond het audiocommentaar van de modellen centraal, terwijl screenshots van alinea's in de tekst werden getoond waarop het commentaar betrekking had. Leerlingen hoorden alleen het audiocommentaar tijdens de belangrijkste stappen in het denkproces van de modellen. De animaties werden zo gecomprimeerd van ca. 10 minuten naar een duur van 2-3 minuten.

Constructie materiaal experimentele lessenserie: de werklessen. Voor les 2, 4 en 6, de drie werklessen, zijn vier korte uiteenzettende of betogende teksten uit Nederlandse dagbladen of tijdschriften geselecteerd. Bij elke leestaak is een scenario ontwikkeld waarvoor diep begrip noodzakelijk was. Daarna werd daarbij een passende verwerkingstaak ontwikkeld. Bij elke werkles werd sterk gewisseld in de vorm van de uiteindelijke leestaak.

Elke tekst werd voorafgegaan door de instructie dat de tekst ná bestudering, bij het specificeren van de verwerkingstaak, niet meer mocht worden ingezien, zodat leerlingen aan de hand van de verwerkingstaak konden controleren in hoeverre ze de tekst vooraf voldoende diep hadden gelezen.

Digitale implementatie van de lessenserie.

Voor de digitale afname van de lessenserie werd een Nederlandse online onderwijsapp gebruikt (LessonUp, lessonup.com). Deze werkwijze had als belangrijk voordeel dat alle antwoorden van leerlingen centraal konden worden vastgelegd en inzichtelijk waren voor de onderzoekers. Tijdens het afnemen van de lessenserie bood de app daarnaast docenten de mogelijkheid de voortgang van hun leerlingen te volgen, de antwoorden van leerlingen in te zien en deze eventueel van feedback te voorzien.

Tijdens de verwerking van de inhoud van de lessenserie in het format van de app bleek dat een van de ontwerpprincipes van de lessenserie technisch niet realiseerbaar was: leerlingen kregen in de lessenserie steeds de nadrukkelijke instructie dat ze een tekst na bestudering niet meer mochten inzien bij de verwerkingstaak; in de uiteindelijke versie was dat niet het geval. Om te controleren of leerlingen zich aan onze instructie hielden, voegden we in twee lessen hierover een controlevraag in, na het afronden van de leestaak.

Onderscheid tussen condities. Om het onderscheid tussen condities te kunnen waarborgen, werden er twee varianten van de lessenserie ontwikkeld. Als eerste werd het materiaal voor de conditie ‘Zelfstandig leren’ geconstrueerd, waarbij er zorg voor werd gedragen dat de leerdoelen en instructies helder en volledig waren, taken zelfstandig maakbaar waren én dat er na elke leestaak duidelijke feedback werd gegeven.

Het materiaal voor de conditie ‘Docent-gestuurd leren’ werd hiervan afgeleid. De drie werklessen werden in hun geheel overgenomen. De drie observerend-leren-lessen daarentegen werden voor klassikale afname aangepast: de uitleg bij de

leerdoelen en de instructie bij taken werden ingekort, individuele taken werden vrijwel steeds vervangen door een sheet die met een prikkelende vraag uitnodigde tot een klassengesprek, en de feedback op antwoorden was beknopt, om bespreking ervan door de docent mogelijk te maken.

2.4 Controle-instrumenten implementatie

Om meer zicht te krijgen op hoe de lessenserie per conditie in de les werd geïmplementeerd, ontwikkelden we een aantal controle-instrumenten, waarbij we gebruikmaakten van een *mixed-methods*-aanpak: we verzamelden parallel aan elkaar tijdens en na afname van het experiment zowel kwantitatieve als kwalitatieve data (Creswell & Plano Clark, 2018). Na het verzamelen van de data werden ze afzonderlijk van elkaar gecodeerd en geanalyseerd, om de interpretatie van de data aan te kunnen scherpen.

Logboeken docenten. Deelnemende docenten ontvingen een logboek waarin hen werd verzocht notities te maken over het verloop van de lessenserie en van de afname van de toetsen. Na de les moesten ze aangeven of leerlingen de les volledig hadden kunnen afronden, of leerlingen geconcentreerd en gemotiveerd werkten, in hoeverre leerlingen gemakkelijk dan wel moeilijk vonden en welke lesactiviteiten de docent zelf had verricht. In het logboek voor de conditie ‘Docent-gestuurd leren’ werd aanvullend gevraagd of de leerlingparticipatie voldoende was tijdens de klassengesprekken; in de conditie ‘Zelfstandig leren’ werd gevraagd of leerlingen (zoals geïnstrueerd) zelfstandig en individueel werkten.

Interviews docenten. Na afloop van het experiment werd per school de coördinerende docent Nederlands ondervraagd door een van de onderzoekers, via een voor het experiment ontwikkeld semi-gestructureerd interview. Vooraf was de coördinerende docent gevraagd de indrukken rondom het experiment bij deelnemende docenten Nederlands te inventariseren.

Doel van het interview was tweeledig: ten eerste om te controleren in hoeverre de expe-

rimentele lessenserie was geïmplementeerd zoals beoogd; ten tweede om te achterhalen hoe docenten de lessenserie inhoudelijk beoordeelden.

Speciaal punt van aandacht bij de conditie ‘Docent-gestuurd leren’ was de vraag in hoeverre docenten het voeren van klassengesprekken tijdens de observerend-leren-lessen geslaagd vonden: in hoeverre slaagden ze erin met de leerlingen adequaat op het leesgedrag van de modellen te reflecteren? In hoeverre lukte het de modellen onderling te vergelijken? In hoeverre gaven de leerlingen de gewenste respons?

Punt van aandacht bij de conditie ‘Zelfstandig leren’ was de vraag in hoeverre leerlingen gedurende de hele lessenserie individueel en zelfstandig en daarbij geconcentreerd en gemotiveerd werkten.

Lesobservaties. Om te controleren op het onderscheid in lesactiviteit tussen de twee condities, maakten we gebruik van lesobservaties. Per conditie observeerden we op één school telkens de lesactiviteiten van twee klassen en hun docenten; per klas werden in totaal vier lessen geobserveerd, waarvan steeds twee observerend-leren-lessen en twee werklessen.

Voor de lesobservatie werd gekozen voor een werkwijze geïnspireerd op die van Elving-Heida (2019). Daarbij werden zowel de docent als drie leerlingen geobserveerd via een schema op basis van systematische directe observatie (vgl. Briesch et al., 2010). Bij de constructie van het observatieschema werden enkele uitgangspunten uit het State-Event Classroom Observation System toegepast (SECOS: Saudargas, 1997).

Voor de docentobservatie gold de volgende handelwijze: de onderzoeker observeerde de docent per minuut telkens tien seconden en legde diens leshandeling vast gedurende dit korte tijdvak. De lijst mogelijke handelingen omvatte enerzijds klassikale activiteiten – instructie geven, met de klas bronnen bekijken, klassengesprekken voeren – en anderzijds individuele activiteiten: een leerling hulp bieden. Handelingen waarbij de docent zich niet actief met de klas bezighield – zoals het observeren van leerlingen, het voorberei-

den van lesactiviteiten, het corrigeren van leerlingwerken –, werden eveneens genoteerd.

Een vergelijkbare handelwijze werd gevolgd bij het observeren van de leerlingen. Eerst werden drie leerlingen gekozen die goed zichtbaar waren voor de onderzoeker; geselecteerde leerlingen werden niet geïnformeerd dat ze werden geobserveerd. Vervolgens observeerde de onderzoeker elke leerling per minuut telkens tien seconden en legde met een observatieschema de lesactiviteit van de leerling vast gedurende dit tijdvak. Het observatieschema omvatte enerzijds aan de lesinhoud gerelateerde activiteiten – schoolwerk, overleggen met docent – en anderzijds niet aan de lesinhoud gerelateerde activiteiten – om zich heen kijken, kletsen met klasgenoot en andere dingen doen.

Lesmateriaal. Aan de hand van de digitale app was eenvoudig meetbaar in hoeverre leerlingen ook daadwerkelijk alle onderdelen van de lessenserie hadden doorgewerkt. De app registreerde per les a) het percentage dia's van een les die een leerling had bekeken, b) het percentage gecontroleerde (goede) antwoorden op de vragen in een les: leerlingen hadden bij elke gesloten vraag de mogelijkheid meerdere malen antwoord te geven, totdat ze het goede antwoord hadden.

Om te kunnen achterhalen of leerlingen zich aan de instructie hielden om *niet* terug te kijken naar een bestudeerde tekst bij de verwerkingstaak, stelden we bij twee werklessen aan het einde van de lessenserie de vraag of leerlingen zich aan deze instructie hadden gehouden.

Evaluatie lessenserie. Om te kunnen bezien hoe leerlingen de lessenserie evalueerden, voegden we aan het einde van de lessenserie enkele evaluerende vragen toe: we vroegen aan leerlingen aan te geven in hoeverre ze de lessenserie als geheel leerzaam en op niveau vonden en we vroegen om eventuele suggesties voor verbetering.

2.5 Meetinstrumenten effect lessenserie

Post-then-pre-test. Om te bezien in hoeverre leerlingen ander strategisch leesgedrag ver-

toonden ná de lessenserie, construeerden we een zogenaamde *post-then-pre-test*. Daarin konden leerlingen na de lessenserie op een vijfpuntsschaal bij een twintigtal stellingen aangeven hoe zij zelf hun leesgedrag en tekstbegrip in het algemeen beoordeelden, vóór en ná de lessenserie. Een dergelijke test meet vooral de door deelnemers ervaren verandering door een interventie (Hill & Betz, 2005). Daarbij vroegen we leerlingen aan te geven hoe vaak ze bepaald leesgedrag inzetten zowel vóór als ná de lessenserie (Lam & Bengo, 2003): bijvoorbeeld ‘Als ik een tekst diep moet lezen, dan breng ik in mezelf de belangrijkste denkstappen van de tekst in kaart.’

We onderscheidden vier sets stellingen – zie voor alle stellingen de Tabel 1.1 in bijlage 1. Vijf stellingen betroffen de inzet van terugkijkgedrag, conform het uitgangspunt van flexibele strategie-inzet (intensief teruglezen van tekstdelen); twee stellingen bevroegen hier aspecten die niet expliciet aan bod kwamen (terugkijken naar afzonderlijke / moeilijke woorden). Vijf stellingen betroffen de mate waarin leerlingen tekstelementen in eigen woorden omzetten tijdens lezen, een basisvoorwaarde voor het bereiken van diep begrip; daarbij bevroeg de laatste van deze vijf de inschatting van leerlingen in hoeverre ze een tekst voldoende denken te begrijpen. Vijf stellingen betroffen dimensies van diep begrip: de betrouwbaarheid en bruikbaarheid bepalen van bronnen, vervolgdienstapen zetten vanuit de tekstinhoud. Omdat we ook geïnteresseerd waren in eventuele neven-effecten, waren vijf stellingen toegevoegd over de inzet van andere leesstrategieën. Twee daarvan kwamen expliciet in de EMME-lessen aan bod, de andere impliciet.

Voor-, tussen- en natoetsen. Om te kunnen controleren op verschillen in tekstbegrip vóór en na de interventie bij beide experimentele groepen, gebruikten we een voortoets, tussen-toets en twee natoetsen. Als voor-, tussen- en (eerste) natoets werd een cloze-toets gebruikt, een gatentekst, waarin woorden strategisch zijn weggelaten (de ‘*rational deletion*’-variant: vgl. Cain et al., 2005; Oller & Jonz, 1994; Gellert & Elbro, 2013). Er loopt sinds

de jaren tachtig van de vorige eeuw een discussie over de vraag in hoeverre gatenteksten tekstbegrip op een hoger niveau kunnen meten dan op zinsniveau. Voor de ‘traditionele’ vorm van de gatentekst, waarin woorden in een vaste interval (elk N^e woord) uit de tekst worden weggelaten, lijkt deze kritiek deels valide: slechts een minderheid van de items doet in studies dan aantoonbaar een appel op begripsvaardigheden over zinsgrenzen heen (voor deze discussie: Pearson & Hamm, 2005; zie ook Kleijn, 2018). Voor de ‘*rational deletion*-variant’ daarentegen, waarin onderzoekers een strategische keuze maken in de weggelaten woorden (bijv. connectieven, modale uitdrukkingen en bepalingen, verwijzingen en antecedenten), lijkt deze kritiek niet op te gaan: in deze vorm weerspiegelt de testscore vooral de vaardigheid van lezers om op macroniveau coherentie in een tekst aan te brengen (vgl. Cain et al., 2005; Gellert & Elbro, 2013; Kleijn, 2018; Oller & Jonz, 1994). De cloze-test heeft als belangrijke voordelen dat een relatief groot aantal items binnen een kort tijdsbestek kan worden afgenomen, doorgaans eenduidig te beoordelen is en dat deze tekstbegrip doorgaans relatief betrouwbaar meet (ibidem).

Als tweede natoets gebruikten we een reguliere tekst-met-vragen-test, met zowel open als gesloten vragen. Wij gebruiken deze tweede natoets vooral, omdat de begripsvragentest in het schoolvak Nederlands zeer gebruikelijk is, al staat deze situatie de laatste tien jaar in Nederland scherp ter discussie (bijv. Hoogeveen, 2018; Kamalski, 2007). In hoeverre, in welke mate en onder welke condities een begripsvragentest valide tekstbegrip kan meten, lijkt af te hangen van een complexe set interacterende factoren van tekst- en vraagkenmerken, waardoor begripsvragentesten zeer wisselende begripsconstructen lijken te bevragen (bijv. Cutting & Scarborough, 2006; Keenan et al., 2008; vgl. Kleijn, 2018, p. 15-16). Omdat we het effect van onze lessenserie ook wilden meten via een in het onderwijs vertrouwde toetsvorm - gatenteksten worden in het schoolvak Nederlands immers weinig toegepast –, hebben we desalniettemin deze toets als tweede natoets opgenomen.

De vier gebruikte toetsen werden eerder gepretest en bleken toen afdoende betrouwbaar (zie Rooijackers et al., 2021a). Deelnemende docenten Nederlands corrigeerden de drie cloze-toetsen. Een van de betrokken onderzoekers corrigeerde ter controle daarnaast per school de toetsen van één klas. De overeenstemming in correctie tussen deze twee beoordelaars bleek bij deze steekproef bij de drie gatenteksten (voor- en tussentoets, eerste deel natoets) steeds zeer hoog ($\alpha \geq .97$), waarschijnlijk doordat na uitgebreid pretesten een zeer nauwsluitend correctie-model was geconstrueerd.

Wat de natoets met tekstafhankelijke begripsvragen betreft: deze werd beoordeeld door een bovenbouwdocent Nederlands en een van de onderzoekers. Deze scoorden onafhankelijk van elkaar alle natoetsen. Ook hier bleek de overeenstemming in beoordeling hoog ($\alpha = .95$). Afwijkingen in score tussen deze twee beoordelaars werden in onderling overleg in overeenstemming gebracht.

2.6 Data-analyse

Post-then-pre-test. Voor het meten van het effect van de interventie op deze toets is gebruikgemaakt van multiniveau-modellen, waarbij als vaste effecten conditie en de interactie tussen conditie en interventie is meegenomen. Bij het toetsen van de eventuele verschillen in scores op de post-then-pre-test zijn observaties steeds genest binnen leerlingen en klassen.

Voor-, tussen- en natoetsen. Voor het meten van het effect van de interventie is hier eveneens gebruikgemaakt van multiniveau-modellen, waarbij observaties steeds genest zijn binnen leerlingen en binnen klassen. Daarbij vertrekken we vanuit een vaste modellenopbouw: als vaste effecten gelden a) tijdstip van meting (de vier afgenomen toetsen), b) interventie, c) tijdstip van interventie (*switching replications*) en d) conditie. Daarnaast nemen we e) de interactie tussen conditie en interventie mee: we zijn dan geïnteresseerd in de verschillen in gemiddelde scores tussen condities, veroorzaakt door de interventie.

Uitgaande van deze vaste modellenopbouw, controleren we vervolgens in hoeverre

de twee implementatievariabelen een effect hebben op de tekstbegripscore, zowel als interactie-effect met interventie als interactie-effect met interventie en conditie. Nu voegen we telkens een effect toe, zodat we van elk toegevoegd model de *fit* en passingsverbetering kunnen controleren. In Bijlage 2 is de modellenopbouw in Tabel 2.1 veraanschouwelijkt.

Een belangrijk voordeel van het werken met multiniveau-modellen is dat ermee alle beschikbare data kan worden meegenomen in de parameterschattingen; deels ontbrekende data bij een proefpersoon leiden niet ertoe dat alle data van deze proefpersoon buiten beschouwing moeten worden gelaten. Dat is in onze studie extra van belang: doordat afname tijdens de Covid19-epidemie plaatsvond, vielen er immers meer ontbrekende data bij leerlingen te verwachten in het experiment dan onder normale omstandigheden.

3 Resultaten

3.1 Implementatie van de lessenserie

Om zicht te krijgen op de implementatie van de lessenserie, worden de logboekgegevens, interviews, lesobservaties en de gegevens uit de digitale app in onderstaande getrianguleerd. Eerst wordt de deelname van leerlingen aan de lessenserie langsgelopen, vervolgens het onderscheid tussen condities, en ten slotte worden enkele implementatievariabelen nader bekeken.

Participatie lessenserie. Het experiment werd afgenomen in de eerste maanden van het schooljaar 2021-2022. De Covid19-epidemie in Nederland zorgde voor meer lesuitval in het experiment dan gebruikelijk en leidde ertoe dat enkele docenten gaandeweg het experiment gedwongen waren een enkele (werk)les als huiswerk op te geven.

Volgens de logboeken van de docenten werd de complete lessenserie door vrijwel alle 27 klassen doorlopen. Drie klassen vormden daarop volgens de logboeken een uitzondering: door ziekte van één docent maakte één klas de twee laatste lessen niet; door onvoorziene lesuitval en bijzonder verlof van een docent maakten twee klassen op één

school de twee laatste lessen als huiswerk, zonder dat de docent hierop later controleerde. Vier klassen maakten één werkles als huiswerk, één klas maakte twee werklessen als huiswerk – de docenten van deze klassen controleerden wel dit huiswerk.

Vanuit de digitale data, beschikbaar via de app, zien we dit beeld bevestigd. Per les namen gemiddeld 630 van de 701 leerlingen deel. Vooral bij de eerste vier lessen is de aanwezigheid relatief hoog (per les: *gem.* = 646, *sd* = 14), terwijl deze bij de twee laatste lessen aanmerkelijk lager ligt (per les: *gem.* = 598, *sd* = 2.5).

Bij aanwezigheid blijken leerlingen de dia's van een les doorgaans vrijwel in hun geheel door te werken.² Voor de drie werklessen, voor de twee condities samen, geldt dat een leerling bij deelname gemiddeld 97.5% (*sd* = 10.9) bekeek van de dia's van de lessen. Dit percentage verschilt tussen condities ($F(2, 634.34) = 5.40, p = .02$), al is dit verschil beperkt ('Docent-gestuurd leren': *gem.* = 97%, *sd* = 11.8; 'Zelfstandig leren': *gem.* = 98.2%, *sd* = 9.7). Dit verschil ontstond vooral doordat docenten in de conditie 'Docent-gestuurd' lessen vaak kort klassikaal openden en daarmee de eerste dia's van een les bespraken. Voor de drie observerend-leren-lessen vinden we eveneens hoge percentages, die daarbij wel duidelijk verschillen per conditie ($F(2, 73571.10) = 133.47, p < .001$): in de conditie 'Zelfstandig leren' bekijkt een leerling bij aanwezigheid gemiddeld 94.3% (*sd* = 20.3) van de dia's van een les door. Voor de conditie 'Docent-gestuurd leren' is dat percentage aanmerkelijk lager: gemiddeld 81.5% (*sd* = 26.6). Dit grote verschil is eenvoudig verklaarbaar: de docent gaf in deze conditie immers vooral klassikaal les: een leerling kon de voortgang van de les zowel volgen op het eigen scherm als op het digitale klassenbord.

Overigens dient hier gesignaleerd dat sommige docenten via de digitale app erop controleerden dat leerlingen de gehele les doorwerkten. Op twee scholen in de conditie 'Zelfstandig leren' werd de werkwijze gehanteerd dat een leerling in principe elke les volledig doorwerkte en bij absentie de gemiste lessen inhaalde.

Onderscheid tussen condities: docentactiviteiten. Bekijken we allereerst de lesactiviteiten van de docenten. Bij de drie *observerend-leren-lessen* geven docenten uit de conditie 'docent-gestuurd leren' in logboeken en interviews steeds aan een groot deel van hun tijd te hebben besteed aan instructie en klassengesprekken, in tegenstelling tot docenten uit de conditie 'Zelfstandig leren', die aangeven hieraan geen tijd te hebben besteed.

In de acht geobserveerde lessen vinden we deze bevindingen bevestigd. De hoeveelheid geregistreerde instructietijd van docenten in een les verschilt van conditie tot conditie ($F(2, 861.13) = 15.53; p < .01$): in de conditie 'Zelfstandig leren' is de hoeveelheid instructietijd 11% (*sd* = 9.27), terwijl deze in de conditie 'Docent-gestuurd leren' 31.8% (*sd* = 5) is. Ook de hoeveelheid tijd besteed aan het klassikaal bekijken van bronnen wijkt af ($F(2, 496.13) = 90.90; p < .001$): in de conditie 'Zelfstandig leren' beslaat dit 1.3% van de lestijd (*sd* = 1.5), terwijl in de conditie 'Docent-gestuurd leren' dit 17% is (*sd* = 2.9). Er is daarnaast een zeer duidelijk verschil aantoonbaar in de hoeveelheid bestede tijd aan klassengesprekken ($F(2, 2278.13) = 475.44; p < .001$): deze vinden in de conditie 'Zelfstandig leren' in het geheel niet plaats, terwijl het percentage in de andere conditie 33.8% (*sd* = 3.1) is.

Wat de drie *werklessen* betreft: docenten geven in logboeken en interviews aan zich hier niet actief opgesteld te hebben: ze voerden geen klassengesprekken, noch gaven ze uitgebreide instructies, zoals de bedoeling was. Een tweetal docenten gaf in de interviews aan dat de hoeveelheid hulpvragen van individuele leerlingen tijdens de les wel sterk kon variëren. Deze opmerkingen stroken met de bevindingen in de acht geobserveerde lessen: docenten voerden geen klassengesprekken, noch bekeken ze met de klas bronnen.

Alle resultaten wijzen erop dat docenten zich dus inderdaad aan de lesinstructie in beide condities hebben gehouden.

Onderscheid tussen condities: leerlingactiviteiten. Bekijken we dan de handelingen van leerlingen. Docenten geven in logboeken en

interviews aan dat leerlingen doorgaans het grootste deel van hun tijd besteden aan de lessenserie. Wel gaven enkele docenten aan dat het orde houden de nodige tijd vergde. In vier klassen werkten leerlingen daardoor regelmatig samen of was de concentratie laag.

Binnen de acht geobserveerde lessen is er een duidelijk verschil tussen condities waar te nemen. Zo is er allereerst een verschil aantoonbaar tussen de hoeveelheid tijd die de leerling gemiddeld zichtbaar aan de lesinhoud besteedt ('schoolse taak') ($F(2, 936.33) = 26.35, p < .001$): leerlingen in de conditie 'Docent-gestuurd leren' besteden gemiddeld 87.8% ($sd = 7.6$) hieraan, terwijl leerlingen in de conditie 'Zelfstandig leren' 96.7% ($sd = 3.8$) hieraan besteden. De tijd dat een leerling in (klassikaal) gesprek met een docent is, varieert eveneens per conditie ($F(2, 14.08) = 11.60, p = .001$): leerlingen in de conditie 'Docent-gestuurd leren' besteden gemiddeld 4.1% ($sd = 1.6$) hieraan, terwijl leerlingen in de conditie 'Zelfstandig leren' geen tijd hieraan besteden. Ten slotte is er ook een verschil in tijd dat een leerling met een medeleerling praat ($F(2, 500.52) = 31.42, p < .001$): leerlingen in de conditie 'Docent-gestuurd leren' besteden gemiddeld 8.8% ($sd = 4.9$) hieraan, terwijl leerlingen in de conditie 'Zelfstandig leren' 2.4% ($sd = 2.9$) hieraan besteden. Waarschijnlijk doordat docenten in de eerste conditie veelvuldig met klassen in gesprek gingen, praatten leerlingen vaker met elkaar, al dan niet over de lesinhoud.

Leerlingen besteden dus in de conditie 'Zelfstandig leren' hun lestijd grotendeels aan het doorwerken van de lessenserie. In de conditie 'Docent-gestuurd leren' is er meer interactie tussen docent en leerlingen én meer interactie tussen leerlingen onderling.

Implementatie. Om te bezien in hoeverre leerlingen zich aan de instructies in de lessenserie hielden en zich engageerden met de lessenserie, bouwden we enkele controle-instrumenten in.

Ten eerste werd bekeken in hoeverre leerlingen bij gesloten vragen in de lessen hun *antwoord controleerden*. Omdat er in de conditie 'Docent-gestuurd leren' in de drie observerend-leren-lessen nauwelijks gesloten vra-

gen aanwezig waren, is het zinnig onderscheid te maken tussen werk- en observerend-lerenlessen. Bekeken over de drie werklessen, is het percentage gecontroleerd goede antwoorden hoog (87.9%, $sd = 11.94$). Dit percentage wijkt duidelijk af tussen condities ($F(2, 3348.60) = 24.53, p < .001$). In de conditie 'Docent-gestuurd leren' is dit percentage 85.4% ($sd = 13.5$), terwijl het in de conditie 'Zelfstandig leren' hoger ligt: 90.4% ($sd = 9.5$). In de drie observerend-lerenlessen in de conditie 'Zelfstandig leren' ligt het percentage eveneens hoog: 95.4% ($sd = 9.9$).

Ten tweede werd leerlingen twee maal aan het einde van een werkles gevraagd of ze, zoals geïnstrueerd, *niet terug hadden gekeken* naar de tekst bij de vervolgoopdracht. Bij de eerste werkles gaf 27% van de leerlingen aan zich niet aan deze instructie te hebben gehouden; bij de tweede werkles gaf 28% dit aan – percentages die overigens niet per conditie verschilden ($p \geq .51$). Het aandeel leerlingen dat aangeeft terug te kijken, overlapt tussen beide lessen slechts gedeeltelijk ($r = .39$): 38.5% van de leerlingen blijkt hiermee zich minimaal één keer niet aan de instructie te hebben gehouden.

In de logboeken en interviews gaven docenten in de conditie 'Docent-gestuurd leren' aan dat ze soms moeite hadden om een goed lopend klassengesprek te voeren over de geobserveerde modellen, omdat zowel de docenten als de leerlingen zelf nogal onervaren waren met een dergelijke procesgerichte benadering. Ook wezen vier docenten uit deze conditie erop dat leerlingen niet altijd het gedrag van de modellen doorgrondde: "leerlingen vinden het redelijk gemakkelijk, maar uit de klassengesprekken blijkt juist dat ze echt nog niet alles doorzien."

Samenvattend. Ondanks het feit dat het experiment tijdens de Covid19-epidemie werd uitgevoerd, was de deelname door leerlingen ruim voldoende en verliep de afname zonder noemenswaardige incidenten. Leerlingen die aan een les deelnamen, lijken doorgaans ook de lesinhoud vrijwel geheel te hebben doorgewerkt.

Vanuit de logboeken, interviews en lesobservaties over docent- en leerlinggedrag

rijst het beeld op dat docenten en leerlingen grotendeels conform de opgestelde afnamecondities hebben gehandeld; in de conditie 'Zelfstandig leren' lijken docenten (nauwelijks tot) niet actief te hebben deelgenomen aan de lessen en werkten leerlingen veelal zelfstandig; docenten in de conditie 'Docent-gestuurd leren' gaven vaak klassikale instructie, bekeken vaak met leerlingen bronnen en voerden vaak klassikale gesprekken. De instructies in de lessenserie zelf lijken leerlingen redelijk te hebben opgevolgd: ze controleerden over het algemeen zeer vaak hun antwoorden in de observatie- en werklessen; echter, aan de instructie om bij de verwerkingsopdracht *niet* terug te kijken naar de bestudeerde tekst, hielden lang niet alle leerlingen zich: 38.5% van de leerlingen gaf aan hiervan te zijn afgeweken. Ook gaven sommige docenten uit de conditie 'Docent-gestuurd leren' aan dat ze het moeilijk vonden goede klassengesprekken te voeren over de leerlingmodellen in de EMME's.

3.2 Evaluatie van de lessenserie

Evaluatie door leerlingen. Wanneer leerlingen zelf gevraagd wordt de lessenserie te evalueren na de laatste les, geeft 83.5% aan dat ze de lessenserie 'goed op niveau' vindt; 10.3% vindt de lessenserie 'gemakkelijk'; 6.3% vindt de lessenserie 'moeilijk'. 86.4% van de leerlingen geeft na de lessenserie aan deze leerzaam te hebben gevonden.

Op het verzoek om suggesties te geven voor verbetering van de lessenserie gingen leerlingen vaak niet in. Er werden vooral korte complimenten uitgedeeld ("erg tof", "leuk om te doen"). Sommige leerlingen gaven aan dat ze de lessenserie wat te lang vonden en adviseerden inkorting. Het vaakst geopperde advies was dat één type vraag in de lessenserie (zogenaamde 'sleepvragen') technisch anders zou moeten worden ingericht.

Docenten gaven aan dat leerlingen over het algemeen redelijk enthousiast waren over de lessenserie, al vonden ze de lessenserie wel vaak aan de lange kant. Ook zouden leerlingen de lessenserie over het algemeen vrij gemakkelijk hebben gevonden. Verder gaven docenten van drie scholen aan dat leerlingen

moeite hadden met de inhoudelijke opzet van de lessenserie. Ze misten 'leerbare inhoud', zoals lijsten met functietermen en argumentatieschema's en vonden het accent op de opbouw van begrip verwarrend. Twee scholen boden de leerlingen daarom voorafgaande aan de natoetsen pdf's aan met de dia's van drie van de zes lessen uit de lessenserie.

Evaluatie door docenten. Docenten evalueerden de lessenserie wisselend. De docenten van vijf van de zeven scholen waren redelijk tot zeer enthousiast: ze gaven aan de gekozen benadering 'verfrissend' te vinden ("een eye-opener") en het idee te hebben dat ze leerlingen nieuwe inzichten in tekstbegrip konden bieden ("Het helpt leerlingen om zicht te krijgen in het leesproces en dat te spiegelen aan wat ze zelf kunnen" en: "ik vind het heel mooi om te zien hoe ze in zes lessen opbouwen naar vragen over het ontkrachten van argumenten en afleiden van tegenargumenten; (...) op zo'n niveau zijn we in klas 4 eerder niet aan de slag geweest.")

Vooraf op één school waren er zeer duidelijke twijfels over de lessenserie. De twee docenten van deze school hadden twijfels over de moeilijkheid van de lessenserie ("het niveau is voor vwo te laag") en gaven aan dat ze niet inzagen wat de meerwaarde van de lessenserie vormde ("geen goede voorbereiding op het Centraal Examen Nederlands") en dat leerlingen naar hun idee vooral plichtmatig in de lessenserie hadden geparticipeerd.

3.3 Effect van de lessenserie

Ervaren leereffect: post-then-pre-test. 600 leerlingen (86% van de deelnemers) maakten deze test. De uitval op deze test in het experiment was daarmee relatief hoog, vooral doordat twee docenten elk in één klas de test niet bleken afgenomen te hebben. Elke set van vijf samenhangende stellingen bleek afzonderlijk ruim voldoende betrouwbaar ($GLB \geq .76$). Alleen de set stellingen die op niet expliciet besproken strategieën betrekking had, scoorde duidelijk lager in betrouwbaarheid, zij het nog altijd voldoende (vóór de lessenserie: $GLB = .65$; na de lessenserie: $GLB = .73$).

Tabel 2

Parameterschatting van de gemiddelde score op de post-then-pre-test per set stellingen, per conditie (S^2_k : variantie tussen klassen; S^2_l : variantie tussen leerlingen; S^2_r : residuele variantie)

Set stellingen	Conditie Docent-gestuurd leren		Conditie Zelfstandig leren			S^2_k ; S^2_l ; S^2_r
	vóór (gem./se)	na (gem./se)	vóór (gem./se)	na (gem./se)		
1. strategisch terugkijkgedrag	2.76 (.04)	3.46 (.04)	2.77 (.04)	3.28 (.04)	a c	(.00; .32; .71)
2. opbouwen van begrip	2.73 (.04)	3.47 (.04)	2.79 (.05)	3.30 (.05)	a c	(.01; .31; .86)
3. dimensies van diep begrip	2.14 (.05)	2.83 (.05)	2.28 (.06)	2.84 (.06)	a c	(.01; .37; .91)
4. ander strategisch leesgedrag	2.23 (.05)	2.92 (.05)	2.24 (.06)	2.78 (.06)	a c	(.01; .29; 1.17)

Noot: a: significant hoofdeffect voor/na de lessenserie; b: significant hoofdeffect conditie; c: significant interactie-effect voor/na lessenserie met conditie

In Tabel 2 zijn voor elke set stellingen de schattingen vóór en na de lessenserie per conditie weergegeven. (In Tabel 1.1 in bijlage 1 zijn deze schattingen per stelling te vinden.) In beide condities geven leerlingen voor alle sets stellingen aan dat ze ná de lessenserie vaker diep-leesgedrag zijn gaan vertonen, dan dat ze dat vóór de lessenserie vertoonden. Leerlingen in de conditie 'Docent-gestuurd' rapporteren daarbij een duidelijk grotere gedragswijziging dan leerlingen in de conditie 'Zelfstandig leren'.

Leerlingen geven dus over het algemeen aan dat ze na de lessenserie vaker strategisch terug zijn gaan kijken, vaker (diep) begrip zijn gaan opbouwen en vaker andere strategieën zijn gaan inzetten, dan dat ze dat vóór de lessenserie deden. Leerlingen in de conditie 'Docent-gestuurd leren' rapporteren hierbij steeds een grotere gedragsverandering dan de leerlingen in de conditie 'Zelfstandig leren'.

Effect op tekstbegrip van de lessenserie. De voortoets, die uit 27 items bestond, bleek bij het experiment voldoende betrouwbaar om te kunnen differentiëren tussen groepen ($N = 668$; $GLB = .76$), evenals de tussentoets, die uit 23 items bestond ($N = 628$; $GLB = .74$). De twee natoetsen ($N = 656$) presteerden wisselend: de gatentekst, die uit 20 items bestond, bleek voldoende betrouwbaar ($GLB = .72$), de

tekst met vragen, die uit 9 items bestond, was dat duidelijk niet ($GLB = .48$) – een probleem dat door het gebruik van twee natoetsen grotendeels wordt opgevangen. In Tabel 3 zijn de correlaties tussen de vier toetsen weergegeven: we zien dan dat de correlaties tussen de drie cloze-toetsen ($r \geq .34$) steeds hoger liggen dan de correlatie van de drie cloze-toetsen met de begripsvragentoets ($r \leq .32$).

Tabel 3

Correlaties (Pearson's r) tussen de vier gebruikte toetsen in het experiment

correlatie	voor- toets	tussen- toets	natoets: cloze	natoets: vragen
voortoets	-	.38	.44	.26
tussentoets	.38	-	.34	.25
natoets: cloze	.44	.34	-	.32

In Bijlage 2 is de *fit* en passingsvergelijking van elk toegevoegd model nader beschreven; hier rapporteren we enkel in hoeverre een toegevoegd effect, per toegevoegd model, significantie bereikt. Uitgaande van de vooraf opgestelde modellenopbouw, vinden we allereerst een hoofdeffect voor tijdstip van meting ($F(3, 702.46) = 812.17, p < .001$), wat gezien de afwijkende maximumscore tussen de gebruikte toetsen een evidente uitkomst is. We vinden evenwel geen hoofdeffect voor interventie ($F(1, 407.86) = .05, p = .82$), noch

Tabel 4
Parameterschattingen voor de score op de vier toetsen in de lessenserie

Parameter	coëfficiënt	SE	p
<i>Meetmomenten</i>			
voortoets (conditie 'Zelfstandig')	17.10	(.32)	<.001
tussentoets	13.91	(.38)	<.001
natoets: cloze	10.60	(.42)	<.001
natoets: vragen	5.96	(.41)	<.001
 Δconditie 'Docent-gestuurd'	 -.37	 (.32)	 ,26
Δtijdstip interventie	.20	(.21)	.34
 Δinterventie : conditie 'Zelfstandig'	 -.40	 (.31)	 ,20
Δinterventie: conditie 'Docent-gestuurd'	.66	(.24)	<.01
<i>Implementatie</i>			
percentage gecontroleerde antwoorden ^a	.016	(.008)	.045
<i>Varianties</i>			
S ² klas	.31	(.13)	
S ² leerlingen: voortoets	12.71	(.74)	
S ² leerlingen: tussentoets	10.45	(.64)	
S ² leerlingen: natoets cloze	9.13	(.59)	
S ² leerlingen: natoets begripsvragen	3.75	(.25)	

Noot: a) gecentreerd rond gemiddelde (88%)

vinden we een onderscheid tussen de twee momenten (*switching replications*) waarop de interventie plaatsvond ($F(1, 132.04) = .87, p = .35$). We vinden evenmin een hoofdeffect voor conditie ($F(1, 55.27) = 1.26, p = .27$). Een interactie-effect tussen conditie en interventie is wel aantoonbaar ($F(1, 919.02) = 6.46, p = .01$): de tekstbegripsscore wijkt na de interventie af tussen beide condities.

Nemen we vervolgens de twee *implementatievariabelen* van de lessenserie mee. Kijken we allereerst naar de eerste controlevariabele: het percentage gecontroleerde antwoorden tijdens de drie werklessen. Deze blijkt als interactie-effect met interventie aanwezig ($F(1, 545.53) = 4.03, p = .045$). Een interactie-effect tussen interventie, conditie en deze implementatievariabele is echter niet aantoonbaar ($F(1, 598.82) = .19, p = .67$).

Kijken we dan naar de tweede controlevariabele: de vraag of leerlingen zich aan de instructie hielden om niet terug te kijken naar

de tekst bij de vervolgopdracht. Deze blijkt noch aantoonbaar als interactie-effect met interventie ($F(1, 524.29) = 1.12, p = .29$), noch aantoonbaar als interactie-effect met interventie en conditie ($F(1, 523.54) = .45, p = .50$).

In Tabel 4 zijn de parameterschattingen gegeven voor de scores op de vier toetsen in het experiment. Daarin is de eerste implementatievariabele als interactie-effect met interventie verdisconteerd, die immers aantoonbaar was; de tweede implementatievariabele, die geen aantoonbaar interactie-effect met interventie kende, is daarin buiten beschouwing gelaten. We zien dan dat een effect voor de conditie ‘Zelfstandig leren’ niet kan worden aangetoond ($\beta = -.40, se = .31$), maar wel voor de conditie ‘Docent-gestuurd leren’: leerlingen in deze conditie scoorden na de interventie gemiddeld hoger op de toetsen dan dat ze dat vóór de interventie deden ($\beta =$

.66, $se = .24$). De grootte van dit effect (Cohen's $d = .22$) wordt traditioneel als 'klein' beschouwd (Cohen, 1969); recent werd echter betoogd dat een effectgrootte van $> .20$ binnen onderwijsonderzoek juist als een groot effect dient te worden gezien (Kraft, 2020). Dit effect per conditie blijkt niet afhankelijk van het tijdstip van de interventie: leerlingen die op het eerste tijdstip de interventie ontvingen, scoren niet aantoonbaar hoger of lager dan de leerlingen in de tweede conditie ($\beta = .20$, $se = .21$, $p = .34$).

Daarnaast: naarmate leerlingen hun antwoorden percentueel vaker controleren, scoren ze onafhankelijk van conditie gemiddeld hoger op de toetsen ná de interventie, dan dat ze dat vóór de interventie deden ($\beta = .016$, $se = .01$, $p = .045$). Gemiddeld controleerden leerlingen 88% van hun antwoorden. Leerlingen die 100% van hun antwoorden controleren tijdens de lessenserie, scoren dus $.19 (= (100-88) * .016)$ hoger op de toetsen ná de interventie dan de gemiddelde leerlingen. Andersom geldt: leerlingen die 50% van hun antwoorden controleren tijdens de lessenserie, scoren ná de interventie $.61 (= (88-50) * .016)$ lager dan gemiddelde leerlingen, en $.80 (= (100-50) * .016)$ lager dan leerlingen die 100% controleren. Drie procent van de variantie tussen leerlingen kan hiermee worden verklaard ($ES = .03$, $sd = .01$). In Bijlage 3 is de gemiddelde score van leerlingen op de vier toetsmomenten gevisualiseerd weergegeven in drie figuren.

Samenvattend. We kunnen dus aantonen dat leerlingen in de conditie 'Docent-gestuurd' leren in het algemeen van meer tekstbegrip blijkgeven ná de interventie op basis van zowel een cloze- als een begripsvragentoets. Daarnaast geldt, onafhankelijk van conditie: naarmate leerlingen hun antwoorden vaker controleren, bouwen ze meer tekstbegrip op ná de interventie, dan dat ze vóór de lessenserie lieten zien; naarmate ze minder vaak hun antwoorden controleerden, bouwen ze minder tekstbegrip op ná de interventie.

3.4. Samenvatting resultaten per onderzoeksvraag

Vooropgesteld: blijkens de beschikbare data

hebben docenten en leerlingen zich gehouden aan de instructie bij beide condities. De waardering voor de lessenserie door leerlingen en docenten was over het algemeen duidelijk positief.

De eerste onderzoeksvraag luidde: 'In hoeverre maakt de lessenserie vwo 4-leerlingen bewust van (de voorwaarden voor) diep begrip bij zakelijke teksten, en wijkt dit af per conditie?' Uit de evaluatie van leerlingen en docenten en vooral uit de resultaten op de *post-then-pre-test* kunnen we concluderen dat de lessenserie leerlingen hiervan inderdaad bewustmaakt. Deelnemers zelf geven ná de lessenserie aan dat ze vaker diep-leesgedrag zijn gaan inzetten dan vóór de lessenserie: naar eigen zeggen herlezen ze vaker tekstelementen, bouwen ze vaker (diep) begrip op en zetten ze vaker strategieën in. Dit geldt sterker voor de conditie 'Docent-gestuurd leren' dan voor de conditie 'Zelfstandig leren': leerlingen die de observerend-lernen-lessen klas-sikaal hebben doorgewerkt, ervaren dus een grotere verandering in hun leesgedrag dan leerlingen die deze lessen individueel hebben doorgewerkt.

De tweede onderzoeksvraag luidde: 'In hoeverre geven vwo 4-leerlingen die de lessenserie hebben gemaakt, blijk van meer diep begrip als ze daarna zelf zakelijke teksten lezen, en wijkt dit af per conditie?' Leerlingen in de conditie 'Docent-gestuurd leren' blijken aantoonbaar over het algemeen baat te hebben gehad van de interventie: zij blijken dan een wat hogere tekstbegripsscore te behalen dan dat ze dat vóór de lessenserie deden; voor de conditie 'Zelfstandig leren' kunnen we dit niet in het algemeen aantonen.

Kijken we naar de derde onderzoeksvraag: 'In hoeverre is de wijze van uitvoering van deze lessenserie van invloed op de effectiviteit ervan?' Er is duidelijk een effect aantoonbaar voor zowel de clozetest als de test met vragen, waarbij de wijze van uitvoering door de leerling wel een sterke invloed heeft. Voor beide condities geldt: naarmate leerlingen vaker hun gegeven antwoorden controleerden tijdens de lessenserie, leren ze meer erdoor, onafhankelijk van conditie. Leerlingen die meer dan gemiddeld hun antwoord controleren, hebben wél baat bij de lessense-

rie – terwijl leerlingen die minder dan gemiddeld hun antwoord controleren, er *niet* op vooruitgaan.

4 Discussie

In deze studie onderzochten we onder 27 vwo 4-klassen, in twee condities, de effectiviteit van een lessenserie over diep begrip op a) het leesgedrag van leerlingen, zoals ervaren door henzelf, en op b) hun tekstbegrip, zoals gemeten in een viertal toetsen. Ontwerpprincipes bij de lessenserie vormden het diep lezen van zakelijke teksten, een procesgerichte benadering van tekstbegrip via observerend leren, flexibele strategie-instructie en een functionele inbedding van tekstbegrip.

In het algemeen geldt dat leerlingen na de lessenserie zelf een duidelijke verandering in hun eigen leesgedrag door de lessenserie rapporteren: ze herlezen naar eigen zeggen naderhand vaker tekstelementen, bouwen vaker (diep) begrip op, en zetten vaker strategieën in. Dit alles geldt echter aanmerkelijk sterker voor de conditie ‘Docent-gestuurd leren’ dan voor de conditie ‘Zelfstandig leren’. Bovendien: als we niet verdisconteren hoe leerlingen de lessenserie hebben doorgevoerd, geven alleen leerlingen uit de conditie ‘Docent-gestuurd leren’ ná de lessenserie gemiddeld blijk van meer tekstbegrip dan vóór de lessenserie; in de conditie ‘Zelfstandig leren’ kunnen we dit effect niet aantonen. Niet geheel onverwacht blijkt de docent daarmee een cruciale rol te spelen in het creëren van een leereffect bij deze lessenserie (vgl. Hattie, 2012).

In de conditie waarin leerlingen zelfstandig, zonder begeleidende docent, de lessenserie doorwerken, wordt juist de wijze doorslaggevend waarop de leerling de lessen uitvoert. Voor beide condities geldt: leerlingen die consequent de tijd namen hun antwoord op gesloten vragen te controleren – leerlingen waarschijnlijk die over het algemeen wat meer gemotiveerd of geconcentreerd of gedisciplineerd te werk zijn gegaan –, blijken meer geleerd te hebben van de lessenserie dan de leerlingen die doorgaans niet de tijd hiervoor namen – dan leer-

lingen waarschijnlijk die over het algemeen wat minder gemotiveerd of geconcentreerd of gedisciplineerd te werk zijn gegaan. Deze bevinding kan aansluiten bij een bekende beperking van e-learning: in situaties waarin de intrinsieke motivatie van leerling om te leren niet vanzelfsprekend is, worden leerkenmerken (concentratie en motivatie, zelfdiscipline en *self-efficacy beliefs*, etc.) vaak doorslaggevend om te leren (Bouhnik & Marcus, 2006). Opvallend genoeg zien we hierin geen verschil optreden tussen condities: of de docent wel of niet het leerproces begeleidt, de wijze van uitvoering door een leerling lijkt in beide condities hetzelfde effect te hebben op de leerwinst.

4.1 Kanttekeningen

Vooropgesteld: deze interventiestudie is afgenomen in de eerste maanden van het schooljaar 2021-2022. Deze periode was door de aanhoudende Covid19-epidemie voor docenten en leerlingen vaak onoverzichtelijk en chaotisch: sommige docenten gaven in logboeken en interviews aan dat ze hun lessen als relatief rommelig ervoeren en dat ze meer dan in andere schooljaren leerroutine en -discipline bij hun vwo 4-leerlingen misten. De uitzonderlijke afnameomstandigheden leidden ook tot meer uitval dan gebruikelijk, op vrijwel alle onderdelen van het experiment.³ In hoeverre de omstandigheden daarmee een impact hebben gehad op de resultaten in deze studie, is een open vraag. Wel denken we de impact van de uitval op de resultaten te hebben kunnen minimaliseren door de gekozen statistische benaderingswijze en door de relatief omvangrijke steekproef.

Los van de afnameomstandigheden, zijn er de nodige voorbehouden bij deze studie te plaatsen. Zo hebben we in deze studie enkel via een vragenlijst gemeten in hoeverre leerlingen door de lessenserie een verandering in hun leesgedrag hebben ervaren. De vraag in hoeverre leerlingen hun leesgedrag ook in de praktijk hebben bijgesteld door de interventie, hebben we niet onderzocht via bijvoorbeeld oogbewegingsopnamen (vgl. bijv. Salmerón et al., 2020).

Verder is het goed hier te signaleren dat we tal van leer- en leerlingkenmerken niet in

het onderzoek verdisconteerd hebben, kenmerken die in andere studies wel een interessant perspectief op de resultaten boden, zoals het onderscheid tussen vaardige en minder vaardige lezers (bijv. Van Silfhout, 2014), de *self-efficacy beliefs* van leerlingen (bijv. Elving-Heida, 2019) en hun metacognitieve vaardigheden en kennis van tekstconventies, -kenmerken en -genres (bijv. Bouwer & Koster, 2016). De vraag in hoeverre sommige leerlingen meer van de interventie geprofi-teerd dan andere leerlingen, blijft in deze studie onderbelicht. Verder geldt dit onderzoek slechts vwo 4-leerlingen; in hoeverre de lessenserie ook effectief is voor andere doelgroepen (havo- en mavo-leerlingen bijvoorbeeld), hebben we niet onderzocht.

Bovendien moet hier worden opgemerkt dat we niet nader hebben onderzocht welke componenten in de lessenserie de effectiviteit ervan kunnen verklaren. In hoeverre onze studie direct de effectiviteit van observerend leren onderbouwt, in het bijzonder voor de leesles (vgl. Couzijn, 1995), valt in strikte zin niet te zeggen. We hebben immers niet onderzocht in hoeverre specifiek de component ‘observerend leren’ in de lessenserie dit positieve resultaat heeft veroorzaakt; mogelijk kunnen andere componenten in de lessenserie (bijv. de functionele inbedding van tekstbegrip of het weghalen van de tekst na bestudering) evenzeer de werkzaamheid van onze lessenserie verklaren.

4.2 Wetenschappelijke implicaties

Met de experimentele lessenserie in onze studie ontwikkelden we een grotendeels procesgerichte benadering van de les tekstbegrip, waarbij we aansloten bij observerend leren (bijv. Bandura & Walters, 1977; Bandura, 1986), bij recente inzichten rond de intentionaliteit van tekstbegrip (bijv. Britt et al., 2018; McCrudden et al., 2011), bij een brede conceptualisatie van evaluatie en beoordeling van teksten (OECD, 2019), en bij inzichten rondom (de kenmerken van) diep begrip (Wolf, 2018).

In een aantal opzichten lijkt deze studie interessante wetenschappelijke implicaties te hebben. Allereerst zou ze – zoals gezegd – als een mogelijke onderbouwing kunnen gelden

voor de effectiviteit van observerend leren voor het leesonderwijs, waarnaar nog relatief beperkt onderzoek bestaat, nationaal (Couzijn, 1995; Keehnen et al., 2015; Rooijackers et al., 2021a) en internationaal (Salmerón & Llorens, 2019; Salmerón et al., 2020). Ten tweede: in onze experimentele lessenserie worden recente theoretische inzichten geïntegreerd over intentioneel tekstbegrip, over het belang van evaluatie en beoordeling en over diep begrip. Naar de mogelijkheden om deze inzichten afzonderlijk, in specifieke interventies in het klaslokaal te vertalen, bestaat al het nodige onderzoek: eerder werd bijvoorbeeld aangetoond dat leerlingen via observerend leren effectief leren reflecteren en evalueren aangaande de kwaliteit van bronnen (bijv. Braasch et al., 2013; Wiley et al., 2013). Onze studie verkent hoe leerlingen vanuit een intentioneel kader (kenmerken van) diep begrip kan worden onderwezen: daarmee lijkt het een van de eerste interventiestudies te zijn die deze moderne inzichten in tekstbegrip *in samenhang* tot een effectieve, sterk procesgerichte lessenserie voor scholieren heeft geïntegreerd. Ten derde lijkt deze studie andermaal de belangrijke rol van interactie bij e-learning te bevestigen (vgl. Bouhnik & Marcus, 2006; Martínez et al., 2007; Udo et al., 2011): zonder de begeleidende rol van de docent en de interactie met klasgenoten lijken vooral leerlingkenmerken te bepalen in hoeverre leerlingen tot leren komen bij onze lessenserie.

4.3 Implicaties voor het onderwijs

Daarnaast heeft deze studie interessante implicaties voor het leesonderwijs binnen het schoolvak Nederlands. Er loopt momenteel een discussie in Nederland of het onderwijs in tekstbegrip in het voortgezet onderwijs niet aan herziening toe is (bijv. SLO, 2021b). In deze studie wordt tekstbegripsonderwijs duidelijk anders benaderd dan op veel Nederlandse voortgezet-onderwijsscholen in het schoolvak Nederlands momenteel de standaard lijkt: de traditionele tekst-met-vragen-aanpak (vgl. Frank et al., 2019). Zowel deelnemende docenten als leerlingen gaven aan deze afwijkende benadering te herkennen én doorgaans ook te waarderen: het leeuwendeel

van docenten vond de benadering interessant en leerzaam (“erg tof”, “een eye-opener”). Deze afwijkende benadering lijkt zich ook te hebben vertaald in een andere leesaanpak: leerlingen geven na de lessenserie aan dat ze vaker diep-leesgedrag zijn gaan inzetten dan vóór de lessenserie; bovendien lieten leerlingen na de lessenserie meer tekstbegrip zien.

In veel opzichten contrasteert de lessenserie in deze studie met de reguliere tekst-met-vragen-didactiek. Deze traditionele didactiek wordt weliswaar allang scherp bekritiseerd (Durkin, 1978), maar heeft nog altijd breed ingang in het voortgezet onderwijs (bijv. Bogaerds-Hazenberg et al., 2022; Hoogeveen & Bonset, 1998; Westhoff, 2012; etc.). In de reguliere benadering staat vooral het product centraal, namelijk de leerlingantwoorden op tekstafhankelijke vragen, en blijft het leesproces van leerlingen doorgaans onbesproken; in de observerend-leren-lesserie in onze lessenserie richten leerlingen zich daarentegen geheel op het leesproces, vanuit de vraag in hoeverre het leesgedrag van leeftijdsgenoten in filmpjes tot diep begrip leidt. In de reguliere aanpak worden vaak strategieën context- en doelloos aangeboden, als generieke ‘manieren van lezen’, en worden evidencebased leesstrategieën enkel indirect aangeboden (Rooijackers et al., 2021b); in onze lessenserie staat daarentegen het doel centraal, namelijk het opbouwen van diep begrip voor een schoolse taak, en wordt een flexibele, doelmatige inzet van evidence-based strategieën voorgestaan. In de reguliere aanpak ontbreekt vrijwel steeds een functionele context en worden leerlingen vrijwel steeds sets tekstafhankelijke vragen aangeboden, die elkaar in een schijnbaar willekeurige volgorde afwisselen; in onze lessenserie zijn daarentegen alle leestaken ingebed binnen een functioneel schoolse doel, wisselen de leestaken sterk in vorm, en in de werklessen sluiten de vervolgtaken logisch aan op dit doel.

Met deze afwijkende benadering van de les tekstbegrip kan de lessenserie aansluiten bij een aantal actuele thema’s binnen het schoolvak Nederlands. Allereerst: andermaal toont deze studie dat een meer procesgerichte benadering via observerend leren bruikbaar

en effectief kan zijn in het domein leesvaardigheid in het schoolvak Nederlands (vgl. Couzijn, 1995; Keehnen et al., 2015; Rooijackers et al., 2021a). De werkwijze in de leesles Nederlands lijkt vaak bij uitstek productgericht; de vraag is of het Nederlandse leesonderwijs niet sterker op de metacognitieve en zelfregulerende vaardigheden van leerlingen in zou moeten spelen door het leesgedrag van leerlingen zelf centraal te stellen.

De lessenserie sluit daarnaast aan bij de roep om meer diep lezen in het Nederlandse onderwijs en verdisconteert evalueren en reflecteren in tekstbegripsonderwijs, dimensies van tekstbegrip waaraan het Nederlandse leesonderwijs nog onvoldoende aandacht lijkt te schenken (Gubbels, 2020; OECD, 2019; Onderwijsraad & Raad voor Cultuur, 2019). De lessenserie kan voorts aansluiten bij de roep om meer ‘bewuste geletterdheid’ (Meesterschapsteam, 2019) en meer taalbewustzijn bij Nederlandse leerlingen in het voortgezet onderwijs (curriculum.nu, 2019): leerlingen blijken na de lessenserie immers een duidelijke bewustzijnsverandering in hun leesgedrag te ervaren. Ten slotte: doordat de lessenserie tekstbegrip steeds inkadert binnen schoolse taken en daarmee tekstbegrip inbedt binnen andere taalvaardigheden (schrijfvaardigheid, debatteren, etc.), sluit ze aan bij de roep om ontkaveling binnen het schoolvak Nederlands (Van Gelderen & Van Schooten, 2011; SLO, 2021b).

4.4 Aandachtspunten

Op een drietal punten biedt deze studie stof ter overweging. Allereerst: in eerdere studies werd al getoond hoezeer Nederlandse leerlingen grote taakvertrouwdheid en taakafhankelijkheid bij de traditionele tekst-met-vragen hebben opgebouwd (Bogaerds-Hazenberg et al., 2022; Rooijackers et al., 2020); ook in deze studie lijkt vertrouwdheid met deze standaardaanpak haar sporen na te laten. In de werklessen van de lessenserie werd bij leestaken erop gehamerd dat leerlingen vooraf een tekst diepgaand moesten bestuderen en deze daarna, bij de verwerkingstaak, niet meer moesten raadplegen, zodat ze tijdens de verwerkingstaak konden controleren of ze tij-

dens bestuderen voldoende diep begrip hadden opgebouwd. Daarmee wilden we bereiken dat leerlingen vooraf zelfstandig een tekst leerden bestuderen – iets waartoe de reguliere tekst-met-vragen-aanpak niet lijkt aan te zetten (Rooijackers et al., 2021c). Echter, een kleine 40% van de deelnemende leerlingen in het experiment gaf aan dat ze deze instructie hadden genegeerd. Daarmee vielen ze terug op het hun vertrouwde scenario ‘lees de taak, zoek daarna het antwoord in de tekst’. Ook in de evaluatie door docenten en leerlingen klonken er met enige regelmaat minder positieve geluiden die aan de standaardaanpak zijn te verbinden en onwennigheid met de afwijkende benadering suggereerden. (Vergelijk het leerlingantwoord: “Dit was niet interessant voor mij, [het] helpt me niet met de gewone les, we hebben voor mijn gevoel de helft van deze lessen naar mensen die lezen gekeken.”)

Tweede overweging: deze studie lijkt weliswaar de effectiviteit van observerend leren aan te tonen – al hebben we dit niet nader onderzocht: zie onze opmerking in § 4.1 –, een open vraag is wel in hoeverre erin de potentie ervan afdoende is benut. Om voldoende diepgang in hun reflectie te kunnen opbouwen, waren leerlingen en docenten geïnstrueerd elk modelfilmpje vooral twee keer te bekijken; deze instructie was gebaseerd op ervaringen uit eerdere try-outs. In het experiment lijkt ze echter niet breed te zijn nageleefd: docenten en leerlingen hielden zich in de geobserveerde lessen er in elk geval doorgaans niet aan. Omdat we de leerlingantwoorden in de observerend-leren-lessen niet uitgebreid kwalitatief hebben onderzocht, valt niet te zeggen in hoeverre deze en andere afwijkingen van de instructie de evaluatie en reflectie op de modellen hebben beïnvloed, en daarmee ook het leereffect van de lessenserie. Kwalitatief vervolgonderzoek naar de voorwaarden voor effectiviteit van observerend leren lijkt hoe dan ook in het algemeen raadzaam (vgl. Van der Loo, 2022).

Derde overweging, samenhangend met het vorige punt: de vraag is in hoeverre docenten niet nader geschoold en getraind zouden moeten worden in een meer procesgerichte benadering van tekstbegrip. Mede doordat docen-

ten vooraf geen uitgebreide inhoudelijke of didactische training hadden ontvangen door de onderzoekers, ontstonden er waarschijnlijk grote verschillen in de aanpak van docenten. Sommige docenten in de conditie ‘Docent-gestuurd leren’ gaven in hun logboeken aan dat ze levendige, inhoudelijke klassengesprekken hielden, waar andere docenten aangaven dit proces moeizamer te vinden verlopen. Een enkele docent uit deze conditie bleek in het geheel niet overtuigd van het nut van deze werkwijze en achtte de lessenserie “geen goede voorbereiding op het Centraal Examen Nederlands” (wat inderdaad geen doel ervan was). Een zekere mate van training of scholing lijkt al met al wenselijk, al heeft de gevarieerde aanpak van de docent de algemene effectiviteit van de lessenserie in de conditie ‘Docent-gestuurd leren’ niet in de weg gestaan.

4.5 Conclusie

In een tijd waarin leerlingen hun informatie vooral via internet en sociale media tot zich nemen, lijkt er alle reden om leerlingen in het voortgezet onderwijs nadrukkelijk te confronteren met een afwijkende diepgaande, aan close-reading verwante leesaanpak en de voordelen die zo’n aanpak voor het leerproces kan hebben (vgl. Onderwijsraad & Raad voor Cultuur, 2019; zie ook Lapp et al., 2018). Dat geldt niet in het minst voor Nederlandse vwo-leerlingen die in studie en werk vaak te maken zullen krijgen met diep-lezen en de intensieve, herhaalde leeswijze die dit vergt. Huidige Nederlandse universitaire studenten zouden nu over het algemeen te weinig kritisch kunnen lezen (Warps et al., 2021). Met deze uitgebreid geconstrueerde en vooraf uitgetroefde lessenserie kan vwo-leerlingen een ‘introductiecursus’ op het concept diep lezen worden aangeboden.

Enkele van de ontwerpprincipes achter deze lessenserie zijn eenvoudig door docenten in het klaslokaal te vertalen. Te denken valt aan: a) het functioneel inkaderen van leestaken, bijvoorbeeld door leestaken te integreren in presentatie-, discussie-, en schrijftaken, b) het procesgericht benaderen van tekstbegrip, bijvoorbeeld door leerlingen onderling hun leesproces te laten bespreken en evalueren, c) het afwisselend aanbieden van soorten

leestaken, zoals sorteertaken, samenvattaken, argumentatietaken, etc., en d) het benadrukken van het belang van de zelfstandige tekstbestudering door een leerling, bijvoorbeeld door de bestudeerde tekst in te nemen bij het uitreiken van de verwerkingstaak.

We hopen te gelegener tijd deze lessenserie openbaar toegankelijk te maken voor gebruik in de les; mogelijk kan ze ook door educatieve uitgevers verder worden uitgewerkt. Het zelfstandig construeren van een lessenserie als deze is immers voor docenten nogal complex: in onze lessenserie spelen oogbewegingsfilmpjes bijvoorbeeld een belangrijke rol; Nederlands middelbare scholen beschikken momenteel niet over een eye-track-computer waarmee zulke EMME's kunnen worden geconstrueerd. In een aansluitende, alternatieve lessenserie zouden uitgevers dan aspecten van tekstbegrip kunnen uitdiepen die vooralsnog redelijk beperkt in de lessenserie zijn uitgewerkt, zoals meervoudig-tekstbegrip (Britt & Rouet, 2012) en de bruikbaarheid en betrouwbaarheid van informatie (Kirschner, 2017).

Kortom: we denken met deze lessenserie een alternatieve, effectieve en zinvolle blik op tekstbegripsonderwijs te hebben laten zien, met meer oog voor het leesgedrag van leerlingen en de effectiviteit ervan, voor reflectie en evaluatie, en voor functioneel, betekenisvol lezen. De lessenserie in onze studie kan zo een inspiratiebron vormen voor een effectieve, meer eigentijdse vakinhoudelijke en -didactische benadering van het domein leesvaardigheid.

Met de resultaten ervan denken we bovendien andermaal de cruciale toegevoegde waarde van de docent in de les te hebben laten zien (vgl. Hattie, 2012). In de twee lockdowns tijdens de Covid19-epidemie werd deze meerwaarde in het voortgezet onderwijs wel eens ter discussie gesteld doordat het thuiswerken van leerlingen tijdelijk de norm was (Smeets, 2021). Leergangen Nederlands voor het voortgezet onderwijs zetten in hun leesmodules inmiddels al enige tijd deels in op e-learning (vgl. bijv. Frank et al., 2019). Het is echter de vraag of dergelijke vormen van e-learning bij middelbarescholieren over het algemeen zinvol zijn. Deze lessenserie laat

juist zien dat in leersituaties *zonder* een docent vooral leerkenmerken van leerlingen bepalend worden of ze tot leren komen. In een samenleving zoals de Nederlandse, die sterk waarde hecht aan kansgelijkheid (bijv. Onderwijsraad, 2019), kan beter niet overmatig worden vertrouwd op onderwijs zonder de sturende rol van de docent.

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

Graag danken we de deelnemende docenten Nederlands voor hun deelname aan dit onderzoek. Speciale dank geldt de coördinerende docenten: mevr. L. von Berg (Stedelijk Gymnasium Den Bosch), dhr. H. Das (Grijdanus Zwolle), mevr. L. van Harmelen (Scholengemeenschap Johan de Witt, Den Haag), mevr. M. Megens (Maaslandcollege, Oss), mevr. J. Piët (Revis Doorn), dhr. T. Robben (Kandinsky College, Nijmegen) en mevr. H. Stinstra (Ds Pierson College, 's-Hertogenbosch).

Noten

- 1 Men zou de vraag kunnen stellen in hoeverre het valide is dat leerlingen bij de verwerkingstaak niet langer de beschikking hebben over de bestudeerde tekst, zoals in onze studie het geval is. In veel school-, studie- of werksituaties hebben lezers immers juist wél de beschikking over een tekst, wanneer ze om informatie verlegen zitten – hoewel zeker niet in alle situaties: denk aan discussies en presentaties, en leertoetsen, waarin leerlingen doorgaans informatie hebben gememoriseerd. Deze vraag lijkt ons binnen onze studie echter niet relevant: de aandacht gaat in deze studie uit naar de zelfstandige diepe bestudering van teksten, vanuit een vooraf gesteld, specifiek leesdoel. De verwerkingstaak ná de bestudering is in de lessenserie vooral een controlemiddel voor leerlingen om te bezien in hoeverre ze de tekst tijdens deze bestudering voldoende hebben doorgrond – waardoor de verhouding tussen tekstbestudering en verwerkingstaak anders ligt dan gebruikelijk.

- 2 Voor de goede orde: we nemen hier aan dat leerlingen een dia hebben bekeken, als ze deze minimaal vier seconden op het scherm zichtbaar hebben gehad. In hoeverre leerlingen dan met aandacht deze dia's hebben bekeken, valt daarmee evenwel niet te zeggen.
- 3 Een ander veelzeggend detail in dit opzicht: drie docenten van een achtste deelnemende school haakten enkele dagen voor afname van het experiment af, omdat ze door lesuitval en tijdsdruk meenden zich niet afdoende ervoor te kunnen inzetten.

Literatuur

- Afflerbach, P., Hurt, M., & Cho, B.Y. (2020). Reading comprehension strategy instruction. In *Handbook of Strategies and Strategic Processing* (pp. 98-118). New York/Londen: Routledge.
- Allen, L.K., & McNamara, D.S. (2020). Defining deep reading comprehension for diverse readers. In *Handbook of Reading Research, Volume V* (p. 261-276). New York/Londen: Routledge.
- Almasi, J.F., & Fullerton, S.K. (2012). *Teaching Strategic Processes in Reading*. New York: Guilford.
- Anmarkrud, Ø., Bråten, I., & Strømsø, H. I. (2014). Multiple-documents literacy: Strategic processing, source awareness, and argumentation when reading multiple conflicting documents. *Learning and Individual Differences*, 30, 64-76.
- Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of Social and Clinical Psychology*, 4(3), 359-373.
- Bandura, A., & Walters, R.H. (1977). *Social learning theory. Vol. 1*. Prentice Hall: Englewood cliffs.
- Baron, N S. (2015). *Words onscreen: The fate of reading in a digital world*. Oxford University Press, USA.
- Bogaerds-Hazenbergh, S.T.M., & Evers-Vermeul, J. (2020). Modeling onder de loep, op zoek naar effectieve ingrediënten. In A. Kraal (Eds.), *Voor u gelezen in 2020: Samenvattingen van actueel internationaal wetenschappelijk onderzoek* (pp. 15-23). Kenniscentrum Begrijpend Lezen.
- Bogaerds-Hazenbergh, S., Evers-Vermeul, J., & Bergh, H. van den. (2022). What textbooks offer and what teachers teach: an analysis of the Dutch reading comprehension curriculum. *Reading and Writing*, 1-27.
- Bouhnik, D., & Marcus, T. (2006). Interaction in distance-learning courses. *Journal of the American Society for Information Science and Technology*, 57(3), 299-305.
- Braaksma, M. (2002). *Observational learning in argumentative writing*. Dissertatie UvA. UvA.
- Braaksma, M., Rijlaarsdam, G., Bergh, H. van den, & Hout-Wolters, B. van (2004). Observational learning and its effects on the orchestration of writing processes. *Cognition and Instruction*, 22(1), 1-36.
- Braasch, J.L., Bråten, I., & McCrudden, M.T. (Eds.). (2018). *Handbook of multiple source use*. New York/Londen: Routledge.
- Briesch, A., Chafouleas, S., & Riley-Tillman, T. (2010). Generalizability and dependability of behavior assessment methods to estimate academic engagement: A comparison of systematic direct observation and direct behavior rating. *School Psychology Review*, 39(3), 408-421.
- Britt, M.A., & Rouet, J.-F. (2012). Learning with multiple documents: Component skills and their acquisition. *Enhancing the quality of learning: Dispositions, instruction, and learning processes*, 276-314.
- Britt, M.A., Rouet, J.-F., & Durik, A.M. (2018). *Literacy beyond text comprehension: A theory of purposeful reading*. New York/Londen: Routledge.
- Broek, P. van den, & Helder, A. (2017). Cognitive processes in discourse comprehension: Passive processes, reader-initiated processes, and evolving mental representations. *Discourse Processes*, 54(5-6), 360-372.
- Broek, P. van den, Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading: Inferences and the on-line construction of a memory representation. In H. van Oosten-dorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (p. 71-98). Erlbaum.
- Broek, P. van den, Helder, A., Espin, C. & Liende, M. van der. (2021). *Sturen op begrip: Effectief leesonderwijs in Nederland*. Universiteit Leiden.
- Brown, R., Pressley, M., Van Meter, P., & Schuder, T. (1996). A quasi-experimental validation of transactional strategies instruction with low-achieving second-grade readers. *Journal of Educational Psychology*, 88, 18-37.

- Cain, K., Patson, N., & Andrews, L. (2005). Age- and ability-related differences in young readers' use of conjunctions. *Journal of child language*, 32(4), 877-892.
- Catrysse, L., Gijbels, D., Donche, V., De Maeyer, S., Lesterhuis, M., & Van den Bossche, P. (2017). How are learning strategies reflected in the eyes? Combining results from self-reports and eye-tracking. *British Journal of Educational Psychology*, 88(1), 118-137. doi:10.1111/bjep.12181
- Clowes, R. W. (2019). Screen reading and the creation of new cognitive ecologies. *AI & Society*, 34(4), 705-720.
- Cohen, J. (1969). *Statistical Power Analysis for the Behavioral Sciences* (1st ed.). New York: Academic Press.
- Couzijn, M.J. (1995). *Observation of writing and reading activities: Effects on learning and transfer*. Dissertatie UvA. UvA.
- Curriculum.nu. (2019). *Toelichting Nederlands; toelichting op het voorstel voor de basis van de herziening van de kerndoelen en eindtermen van de leraren en schoolleiders uit het ontwikkelteam Nederlands*. Geraadpleegd op 24 februari 2022 via curriculum.nu/voorstellen/nederlands/
- Cutting, L.E., & Scarborough, H.S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific studies of reading*, 10(3), 277-299.
- CvTE [College voor Toetsen en Examens]. (2022). *Nederlands (4F) vwo: syllabus Centraal Examen*. Utrecht: CvTE.
- DESI. (2020). *Digital Economy and Society Index (DESI) 2021*. European Commission.
- Donker, A. S., Boer, H. de, Kostons, D., Ewijk, C.D. van, & Werf, M.P. van der. (2014). Effectiveness of learning strategy instruction on academic performance: A meta-analysis. *Educational Research Review*, 11, 1-26.
- Drop, W., & Vries, J.H.L. de. (1977). *Taalbeheersing: handboek voor taalhantering*. Groningen: Wolters-Noordhoff.
- 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 (pp. 51-93). Newark: International Reading Association.
- Durkin, D. (1978). What classroom observation reveals about reading comprehension instruction. *Reading Research Quarterly* 14, 481-533.
- Elving-Heida, K. (2019). *Effectieve leeractiviteiten voor het schrijfonderwijs in havo 4*. (Proefschrift UU). LOT.
- Fang, Z. (2016). Teaching close reading with complex texts across content areas. *Research in the Teaching of English*, 51(1), 106-116.
- College voor Toetsen en Examens. (2021). *Nederlands (4F) vwo: syllabus Centraal Examen*. Utrecht: CvTE.
- Expertgroep Doorlopende Leerlijnen Taal. (2008). *Over de drempels met taal*. Enschede: SLO.
- Ferrer, A., Vidal-Abarca, E., Serrano, M.Á., & Gilbert, R. (2017). Impact of text availability and question format on reading comprehension processes. *Contemporary Educational Psychology*, 51, 404-415.
- Fisher, D., & Frey, N. (2013). *Gradual release of responsibility instructional framework*. IRA Essentials, 1-8.
- Fisher, D., & Frey, N. (2015). *Text-dependent questions, grades K-5: Pathways to close and critical reading*. Thousand Oaks: Corwin.
- Frank, H., Grezel, J.E., Kooiman, M., Steenbergen, W., Vos, M., & Zetstra, H. *Nieuw Nederlands: 4-5-6 vwo*. Noordhoff.
- Gelderen, A. van, & Schooten, E. van. (2011). *Taalonderwijs; een kwestie van ontkavelen: Openbare les in duplo*. Rotterdam: Rotterdam University Press.
- Gellert, A. S., & Elbro, C. (2013). Cloze tests may be quick, but are they dirty? Development and preliminary validation of a cloze test of reading comprehension. *Journal of Psychoeducational Assessment*, 31(1), 16-28.
- Gil, L., Bråten, I., Vidal-Abarca, E., & Strømso, H.I. (2010). Understanding and integrating multiple science texts: Summary tasks are sometimes better than argument tasks. *Reading Psychology*, 31, 30-68.
- Goldman, S.R. (1997). Learning from text: Reflections on the past and suggestions for the future. *Discourse Processes*, 23(3), 357-398.
- Graham, S., & Harris, K.R. (1994). Cognitive str-

- tegy instruction: Issues and guidelines in conducting research. In Vaughn, S., and Bos, C. (Eds.), *Research issues in learning disabilities: Theory, methodology, assessment, and ethics* (pp. 146–160). New York: Springer-Verlag.
- Graham, S., & Hebert, M. (2011). Writing-to-read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review*, 81, 710–744.
- Graham, S., & Perin, D. (2007). *Writing next: Effective strategies to improve writing of adolescent middle and high school*. New York, NY: Carnegie Corporation of New York.
- Gubbels, J., Langen, A. van, Maassen, N., & Meelissen, M. (2019). *Resultaten PISA-2018 in vogelvlucht*. Universiteit Twente.
- Gubbels, J., Netten, A., & Verhoeven, L. (2017). *Vijftien jaar leesprestaties in Nederland: PIRLS-2016*. Nijmegen: Expertisecentrum Nederlands/ Radboud Universiteit.
- Gubbels, J. (2020). Moeite met evalueren en reflecteren: wat betekent dat? Didactiek Nederlands – Handboek. Geraadpleegd op 5 juli 2021 via didactieknederlands.nl/handboek/2020/12/moeite-met-evalueren-en-reflecteren-wat-betekent-dat/
- Guthrie, J.T., Meter, P. van, McCann, A.D., Wigfield, A., Bennett, L., Poundstone, C.C., Rice, M.E., Faibisch, F.M., Hunt, B., & Mitchell, A.M., (1996). Growth of literacy engagement: Changes in motivations and strategies during concept-oriented reading instruction. *Reading Research Quarterly*, 31, 306–332.
- Halamish, V., & Elbaz, E. (2020). Children's reading comprehension and metacomprehension on screen versus on paper. *Computers & Education*, 145, 103737.
- Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research*, 66(2), 99–136.
- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. New York/ Londen: Routledge.
- Hayles, N K. (2012). *How we think: Digital media and contemporary technogenesis*. Chicago: University of Chicago Press.
- Hermida, D. (2009). The importance of teaching academic reading skills in first-year university courses. *International Journal of Research and Review*, (3), 20–30.
- Higgs, K., Magliano, J.P., Vidal-Abarca, E., Martínez, T., & McNamara, D.S. (2017). Bridging skill and task-oriented reading. *Discourse Processes*, 54(1), 19–39.
- Hill, L.G., & Betz, D.L. (2005). Revisiting the retrospective pretest. *American Journal of Evaluation*, 26(4), 501–517.
- Hoogeveen, M. (2018). *Het leesonderwijs in primair en voortgezet onderwijs: een stand van zaken en curriculaire aanbevelingen*. SLO.
- Hoogeveen, M., & Bonset, H. (1998). *Het schoolvak Nederlands onderzocht*. Leuven/Apeldoorn: Garant.
- Houtven, T. van, Peters, E., & El Morabit, Z. (2010). Hoe staat het met de taal van studenten? Exploratieve studie naar begrijpend lezen en samenvatten bij instromende studenten in het Vlaamse hoger onderwijs. *Levende Talen Tijdschrift*, 11(3), 29–44.
- Inspectie van het Onderwijs. (2019). *De staat van het onderwijs 2018. Onderwijsverslag over 2017/2018*. Utrecht: Inspectie van het Onderwijs.
- Jarodzka, H., Gog, T. van, Dorr, M., Scheiter, K., & Gerjets, P. (2013). Learning to see: Guiding students' attention via a model's eye movements fosters learning. *Learning and Instruction*, 25, 62–70.
- Jones, R., & Hafner, C. (2012). *Understanding digital literacies: A practical introduction*. New York/Londen: Routledge.
- Kamalski, J. (2007). *Coherence marking, comprehension and persuasion: On the processing and representation of discourse*. Dissertatie UU. LOT.
- Keehnen, T., Braaksma, M., & Boer, M. de. (2015). Leren door zien lezen. Observerend leren bij leesvaardigheid in 3 vwo. *Levende Talen Tijdschrift*, 16(1), 34–41.
- Keenan, J.M., Betjemann, R.S., & Olson, R.K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, 12(3), 281–300.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.
- Kintsch, W., & Van Dijk, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363–394.
- Kirschner, P.A. (2017). *Het voorbereiden van leer-*

- lingen op (nog) niet bestaande banen. Arnhem: NVSP Innovatief in Werk.
- Kleijn, S. (2018). *Clozing in on readability: How linguistic features affect and predict text comprehension and on-line processing*. Utrecht: LOT. Dissertatie UU.
- Kraft, M. A. (2020). Interpreting Effect Sizes of Education Interventions. *Educational Researcher*, 49(4), 241–253. <https://doi.org/10.3102/0013189X20912798>
- Lam, T.C., & Bengo, P. (2003). A comparison of three retrospective self-reporting methods of measuring change in instructional practice. *American Journal of Evaluation*, 24(1), 65–80.
- Lapp, D., Moss, B., Grant, B., & Johnson, K. (2018). *Close reading: werken aan dieper tekstbegrip in het basisonderwijs*. Soest: Pica.
- Liaw, S.S. (2008). Investigating students' perceived satisfaction, behavioral intention, and effectiveness of e-learning: A case study of the Blackboard system. *Computers & Education*, 51(2), 864–873.
- Linthorst, R., & Gloppe, K. de. (2015). De didactiek van begrijpend lezen in het voortgezet onderwijs. *Pedagogische Studiën*, 92(2), 150–166.
- Liu, Z. (2005). Reading behavior in the digital environment. *Journal of Documentation*, 61(6), 700–712.
- Loo, J. van der. (2022). *Mastering The Art of Academic Writing: Comparing the Effectiveness of Observational Learning and Learning by Doing*. Dissertatie. Universiteit van Tilburg.
- Lorch, R.F., Lorch, E.P., & Klusewitz, M.A. (1993). College students' conditional knowledge about reading. *Journal of Educational Psychology*, 85(2), 239–252.
- Lysynchuk, L.M., Pressley, M., & Vye, N.J. (1990). Reciprocal teaching improves standardized reading-comprehension performance in poor comprehenders. *The Elementary School Journal*, 90, 469–484.
- Mangen, A., Walgermo, B.R., & Brønnick, K. (2013). Reading linear texts on paper versus computer screen: Effects on reading comprehension. *International Journal of Educational Research*, 58, 61–68.
- Marlen, T. van. (2019). *Looking Through the Teacher's Eyes: Effects of Eye Movement Modeling Examples on Learning to Solve Procedural Problems*. Dissertatie UU. Universiteit Utrecht.
- Martínez, R.A., Del Bosch, M.M., Herrero, M.H.P., & Nuño, A.S. (2007). Psychopedagogical components and processes in e-learning. Lessons from an unsuccessful on-line course. *Computers in human behavior*, 23(1), 146–161.
- McCrudden, M. T., Magliano, J.P., & Schraw, G. (2011). Toward an integrated view of relevance in text comprehension. In M. McCrudden, J. Magliano, & G. Schraw (eds.), *Text relevance and learning from text* (p. 395–414). Information Age Publishing.
- McCrudden, M.T., & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational Psychology Review*, 19(2), 113–139.
- McNamara, D.S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14(1), 1–43.
- McNamara, D.S., & Magliano, J.P. (2009). Self-explanation and metacognition. In D.J. Hacker et al. (Eds.), *Handbook of metacognition in education* (p. 60–81). New York/Londen: Routledge.
- Meesterschapsteam. (2019). *Perspectieven op taal, taalgebruik en literatuur als basis voor nieuwe eindtermen*. Geraadpleegd op 24 februari 2022 via nederlands.vakdidactiek.nl/wp-content/uploads/sites/4/2019/11/Perspectieven_als_basis_voor_eindtermen_NL_LINE-ALReindversie181219.pdf
- Ministerie van Onderwijs, Cultuur & Wetenschappen. (2020, 13 oktober). Kamerbrief 'Stand van zaken leesoffensief'. Geraadpleegd op 24 februari 2022 via rijksoverheid.nl/documenten/kamerstukken/2020/10/13/kamerbrief-stand-van-zaken-leesoffensief
- Ministerie van Onderwijs, Cultuur & Wetenschappen. (2022, 12 mei). Kamerbrief over Masterplan basisvaardigheden. Geraadpleegd op 7 september 2022 via rijksoverheid.nl/documenten/kamerstukken/2022/05/12/kamerbrief-masterplan-basisvaardigheden
- Mullis, I.V.S., Martin, M.O., Foy, P., & Hooper, M. (2017). *PIRLS 2016 international results in reading*. TIMSS & PIRLS International Study Center.
- National Academies of Sciences, Engineering, & Medicine. (2017). *Evaluation of the achievement levels for mathematics and reading on the*

- National Assessment of Educational Progress. National Academies Press.
- OECD. (2019). *PISA 2018: Assessment and Analytical Framework*. Paris: PISA, OECD. <https://doi.org/10.1787/b25efab8-en>
- Oller, J. W., & Jonz, J. (1994). *Cloze and Coherence*. Associated University Press.
- Onderwijsraad. (2019). *Doorgeschoten differentiatie in het onderwijsstelsel; stand van educatief Nederland 2019*. Den Haag: Onderwijsraad.
- Onderwijsraad & Raad voor Cultuur. (2019). *Lees! Een oproep tot een leesoffensief*. Raad voor Cultuur/Onderwijsraad.
- Oostdam, R., & Bimmel, P. (1996). De kloof tussen theorie en praktijk; Experimenteel onderzoek naar training en transfer van leesstrategieën. *Levende Talen Magazine*, 83(511), 332-337.
- 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.
- Palincsar, A.S., & Brown, A.L. (1984). Reciprocal teaching of comprehension-fostering and monitoring activities. *Cognition and Instruction*, 1(2), 117-175.
- Paris, S.G., Lipson, M.Y., & Wixson, K.K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*, 8(3), 293-316.
- Pearson, P.D., & Hamm, D.N. (2005). The assessment of reading comprehension: A review of practices-past, present, and future. In S.G. Paris & S.A. Stahl (red.), *Children's reading comprehension and assessment* (p. 13-69). Routledge.
- Pearson, P.D., Palincsar, A. S., Biancarosa, G., & Berman, A.I. (2020). Reaping the Rewards of the Reading for Understanding Initiative. *National Academy of Education*.
- Pereira, C., & Nicolaas, M. (2019). *Effectief onderwijs in begrijpend lezen*. Nederlandse Taalunie.
- Perfetti, C. (1997). Sentences, individual differences, and multiple texts: Three issues in text comprehension. *Discourse Processes*, 23(3), 337-355.
- Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific studies of Reading*, 18(1), 22-37.
- Piper, A. (2012). *Book was there: Reading in electronic times*. Chicago: University of Chicago Press.
- Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale: Erlbaum.
- Pressley, M., Brown, R., El-Dinary, P.B., & Afflerbach, P. (1995). The comprehension instruction that students need: instruction fostering constructively responsive reading. *Learning Disabilities Research & Practice*, 10, 215-224.
- Pronk-van Eunen, M., & de Vos, B. (2018). Lezen met de leessandwich: Een kansrijke didactiek bij Nederlands en mvt. *Levende Talen Magazine*, 105(2), 10-14.
- Raad voor Nederlandse taal en letteren. (2015). *Vaart met taalvaardigheid: Nederlands in het hoger onderwijs*. Adviesrapport. Den Haag: Taalunie.
- Rijlaarsdam, G., Braaksma, M., Couzijn, M., Janssens, T., Kieft, M., Raedts, M., Steendam, E. van, Toorenaar, A., & Bergh, H. van den. (2008). Observation of peers in learning to write; Practice and Research. *Journal of Writing Research*, 1(1), 53-83.
- Rogiers, A. (2019). *Illuminating text-learning in secondary education: measuring and fostering strategy use and knowledge acquisition*. Dissertatie, Universiteit Gent.
- Rooijackers, P., Silfhout, G. van, & Bergh, H. van den. (2021a). Observerend leren lezen op het vwo; reflecteren op tekstbegrip via oogbewegingsfilmpjes. *Levende Talen Tijdschrift*, 22(3), 3-16.
- Rooijackers, P., Silfhout, G. van, & Bergh, H. van den. (2021b). Timmeren met oude spijkers en zonder hout: Waarom leergangen Nederlands nu niet echt aan strategie-instructie doen. *Levende Talen Magazine*, 108(1), 18-23.
- Rooijackers, P., Silfhout, G. van, Schuurs, U., & Bergh, H. van den. (2021c). De relatie tussen het vooraf lezen van teksten en het beantwoorden van begripsvragen: twee experimenten onder 5-vwo-leerlingen. *Pedagogische Studiën*, 98(5), 320-349.
- Rooijackers, P., Silfhout, G. van, Schuurs, U., Mulders, I., & Bergh, H. van den. (2020). Lezen en antwoorden bij teksten met vragen; een cross-sectionele eye-trackstudie onder 52 vwo-leerlingen. *Pedagogische Studiën* 97(4), 281-308.
- Rouet, J.-F., Britt, M.A., & Durik, A.M. (2017). RESOLV: Readers' Representation of Reading Contexts and Tasks. *Educatio-*

- nal Psychologist*, 52(3), 200-215. DOI: 10.1080/00461520.2017.1329015
- Salmerón, L., Delgado, P., & Mason, L. (2020). Using eye-movement modelling examples to improve critical reading of multiple webpages on a conflicting topic. *Journal of Computer Assisted Learning*, 36(6), 1038-1051.
- Salmerón, L., & Llorens, A. (2019). Instruction of digital reading strategies based on eye-movements modeling examples. *Journal of Educational Computing Research*, 57(2), 343-359.
- Saudargas, R. (1997). *State-Event Classroom Observation System (SECOS). Observation manual*. University of Tennessee, Knoxville.
- Schroeder, S. (2011). What readers have and do: Effects of students' verbal ability and reading time components on comprehension with and without text availability. *Journal of Educational Psychology*, 103(4), 877.
- Schuppert, L., & Bergh, H. van den. (2017). Leesstrategieën: Evidence-based of Nieuw Nederlands? Een experimentele studie in 4 vwo. *Levende Talen Tijdschrift*, 18(1), 14-23.
- Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Mifflin.
- Silfhout, G. van. (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.
- SLO. (2021a). *Leesvaardigheid onderbouw vo; Domeinbeschrijving ten behoeve van peilingsonderzoek*. Amersfoort: SLO.
- SLO. (2021b). *Startnotitie Nederlands bovenbouw voortgezet onderwijs*. Amersfoort: SLO.
- Smeets, E. (2021). *Monitor hybride onderwijs: rapportage over het voortgezet onderwijs*. Nijmegen: KBA.
- Sosnoski, J. (1999). Hyper-readers and their reading engines. In G. Hawisher & C. Selfe (eds.), *Passions, Pedagogies, and 21st Century Technologies* (pp. 161-177). Eds. Utah: Logan.
- Steensel, R. van, & Houtveen, T. (2020). De vele kanten van leesbegrip; Verslag van een literatuurstudie naar kernelementen van effectief onderwijs in begrijpend lezen. *Levende Talen Tijdschrift*, 21(3), 3-12.
- Udo, G.J., Bagchi, K.K., & Kirs, P.J. (2011). Using SERVQUAL to assess the quality of e-learning experience. *Computers in Human Behavior*, 27(3), 1272-1283.
- Warps, J., Visser, M. de, Lodewick, J., & Termorshuizen, T. (2021). *Verkenning vervolgonderwijs*. Nijmegen: ResearchNed.
- Westhoff, G. (2012). Mesten en meten in leesvaardigheidstraining: Leesvaardigheidsonderwijs en examentraining zijn twee verschillende dingen. *Levende Talen Magazine*, 99(4), 16-20.
- Wijekumar, K.K., Beerwinkle, A.L., Harris, K.R., & Graham, S. (2019). Etiology of teacher knowledge and instructional skills for literacy at the upper elementary grades. *Annals of dyslexia*, 69(1), 5-20.
- Wolf, M. (2018). *Reader, come home: The reading brain in a digital world*. New York: Harper.
- Wolf, M., & Barzillai, M. (2009). The importance of deep reading. *Educational Leadership, Literacy 2.0*, 6(66), 32-37.
- Yeari, M., Broek, P. van den, & 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 promovendus aan de Universiteit Utrecht, via een Dudoc Alfa-beurs. Daarnaast is hij werkzaam als toetsdeskundige bij stichting Cito en is hij als docent Nederlands verbonden aan het Sint-Janslyceum in 's-Hertogenbosch. **Erwin Mantingh** is als lerarenopleider en vakdidacticus verbonden aan de Graduate School of Teaching en het Departement voor Talen, Literatuur & Communicatie van de Universiteit Utrecht. **Gerdineke van Silfhout** is werkzaam bij SLO als curriculumexpert en is als programmamanager Actualisatie examenprogramma's verantwoordelijk voor het coördineren van de bijstelling van de examenprogramma's bovenbouw vo. **Huib 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

The effectiveness of a lesson series on deep reading

An intervention study among Dutch tenth grade pre-university pupils

In the Netherlands, text comprehension instruction often takes a traditional form: students study texts and subsequently answer text-dependent questions. Currently, it is debated whether and how text comprehension instruction in secondary education should be revised. In this study, we therefore tested an alternative approach to reading lessons.

Based on recent scientific insights, we developed a digital lesson series consisting of six lessons focused on deep reading of expository texts. It features a goal-directed, process-oriented approach to text comprehension, by means of observational learning. Independent study of the texts also takes a prominent place.

The lesson series was tested in 27 10th-grade pre-university classes, distributed over two conditions: in the independent learning condition ($n = 329$) students worked individually, and in the teacher-guided learning condition ($n = 372$) students additionally participated in class discussions.

Self-reports show that students in both conditions displayed more reading behavior characteristic of deep reading. Furthermore, students in the teacher-guided learning condition obtained better text comprehension results. In both conditions, the students' quality of execution determined learning gains. The lesson series therefore seems an effective contemporary alternative to current Dutch reading lessons.

Keywords: deep comprehension, pre-university students, expository texts, e-learning, process-oriented

Bijlage 1

Tabel 1.1

Parameterschatting van de gemiddelde score per stelling per conditie op de post-then-pre-test (S^2_l : variantie tussen leerlingen; S^2_k : variantie tussen klassen; S^2_r : residuele variantie)

Stelling	Conditie Docent-gestuurd		Conditie Zelfstandig		S ² _l ; S ² _k ; S ² _r
	gem. vóór	gem. na	gem. vóór	gem. na	
Terugkijkgedrag					
1. lees ik enkele woorden of woordgroepen terug.	2.96	3.64	3.02	3.53	^{a c} (.57; .00; .28)
2. lees ik hele zinnen terug.	2.84	3.58	2.91	3.39	^{a c} (.51; .01; .28)
3. lees ik een of meerdere alinea's terug.	2.23	3.22	2.35	2.99	^{a c} (.66; .02; .30)
4. stop ik bij een moeilijk woord dat ik niet ken.	3.21	3.69	3.18	3.48	^{a c} (.94; .00; .25)
5. lees ik veel elementen in de tekst opnieuw.	2.46	3.15	2.41	3.00	^a (.63; .01; .28)
Inzet andere strategieën					
6. scan ik vóór of ná het lezen de tekst.	2.73	3.50	2.55	3.25	^a (1.05; .06; .40)
7. let ik tijdens lezen op de hoofdzaken in de tekst.	2.50	3.40	2.66	3.25	^{a c} (.88; .04; .35)
8. pauzeer ik tijdens lezen om even na te denken.	2.20	2.78	2.21	2.69	^a (.85; .02; .30)
9. maak ik tijdens of na lezen aantekeningen, bijv. in de kantlijn.	1.30	1.73	1.35	1.66	^{a c} (.49; .00; .24)
10. formuleer ik hetgeen ik lees, in mijn hoofd om in eigen woorden.	2.45	3.18	2.44	3.07	^a (.81; .03; .32)
Opbouwen begrip					
11. lees ik de tekst zo dat ik hem naderhand in mijn hoofd kan samenvatten.	2.58	3.41	2.56	3.24	^{a c} (.76; .03; .38)
12. breng ik in mezelf de belangrijkste denkstappen van de tekst in kaart.	2.01	2.67	2.02	2.55	^{a c} (.69; .01; .28)
13. kan ik na lezen zelfstandig de hoofdgedachte ervan formuleren.	2.88	3.59	3.01	3.37	^{a c} (.66; .01; .28)
14. controleer ik na lezen af of ik de tekst goed genoeg begrepen heb.	2.94	3.84	2.97	3.65	^{a c} (.65; .01; .37)
15. denk ik dat ik na lezen de tekst voldoende begrijp.	3.25	3.85	3.33	3.67	^{a c} (.53; .01; .31)
Dimensies van diep begrip					
16. denk ik erover na of en hoe de tekst met bronnen is onderbouwd.	1.65	2.41	1.97	2.56	^{a c} (.64; .01; .33)
17. denk ik erover na of de schrijver ervan (of een bron) gezaghebbend is.	1.62	2.28	1.88	2.44	^a (.81; .02; .33)
18. denk ik erover na of en hoe ik een tekst voor een opdracht kan gebruiken.	2.61	3.27	2.55	3.12	^a (.86; .03; .29)
19. denk ik erover na of ik vanuit de gedachtegang zelf vervolgsdenkstappen kan zetten.	2.10	2.74	2.07	2.58	^a (.71; .01; .29)
20. denk ik na over wat de tekst me wil zeggen.	2.72	3.46	2.89	3.48	^{a c} (.84; .02; .33)

Noot: ^a: significant hoofdeffect voor/na de lessenserie; ^b: significant hoofdeffect conditie;
^c: significant interactie-effect voor/na lessenserie met conditie

Bijlage 2

Fit en passingsvergelijking van de modellen

In Tabel 2.1 is de fit en passingsverbetering weergegeven van tien modellen voor de score op de begripstoetsen. Als model 1 geldt een intercept: een gemiddelde score op de vier toetsen tezamen, met een schatting van de klassenvariantie en een variantie op leerling-niveau en error. Omdat de toetsen in maximumscore afwijken en op verschillende momenten zijn afgenomen, voegen we in model 2 het tijdstip van meting toe. Dit levert een passingsverbetering op – logischerwijs, gezien de wisselende maximumscore bij de vier toetsen (Model 2 vs Model 1: $\Delta\chi^2(3) = 1685.45, p \leq .001$). Zie Tabel 2.1.

Interventie: Vervolgens verdisconteren we met model 3 het algemene effect van de interventie op de scores van de vier toetsen, wat niet tot een betere passing leidt (Model 3 vs Model 2: $\Delta\chi^2(1) = .04, p = .84$). Daarna verdisconteren we hierin het tijdstip van de interventie: in hoeverre is er een verschil in score waar te nemen tussen de eerste groep die de interventie ontving en de tweede groep die de interventie ontving binnen het *switching-replications*-design (model 4)? Dit leidt evenmin tot een betere passing (Model 4 vs Model 3: $\Delta\chi^2(1) = 2.51, p = .11$). Voor deze effectmeting van de scores binnen het *switching-*

replications-design maken we hier gebruik van het *fixed-occasion*-model (Goldstein, 1987; Quené & Van den Bergh, 2004).

Conditie: Met model 5 bekijken we in hoeverre er een algemeen verschil in score aantoonbaar is tussen condities – ongeacht of leerlingen de interventie hebben gevolgd: hier vinden we geen passingsverbetering (Model 5 vs Model 4: $\Delta\chi^2(1) = .01, p = .92$). Hierna bekijken we met model 6 in hoeverre dit verschil in score per conditie afhankelijk is van het feit of leerlingen de interventie hebben gevolgd: dit blijkt wel tot verbetering te leiden (Model 6 vs Model 5: $\Delta\chi^2(1) = 3.97, p < .045$).

Wijze van uitvoering: Ten slotte testen we in hoeverre de wijze van uitvoering door leerlingen van de experimentele lessenserie van invloed is op de scores. Daarvoor stonden ons twee controlevariabelen ter beschikking: het percentage gecontroleerde goede antwoorden in de experimentele lessenserie en het ‘niet-terugkijken’ naar de bestudeerde tekst. Toevoeging van een model waarin het percentage gecontroleerd goede antwoorden wordt verdisconteerd, leidt tot een zeer duidelijke passingsverbetering (Model 7 vs Model 6: $\Delta\chi^2(1) = 939.41, p < .001$); een interactie-effect met conditie leidt echter niet tot aantoonbare verbetering (Model 8 vs Model 7: $\Delta\chi^2(1) = .53, p = .47$). Toevoeging van twee modellen waarin het al dan niet terugkijken

Tabel 2.1

Fit en passingsvergelijking van tien modellen voor de score op de begripstoetsen

Model	-2LL	Modelvergelijking			
		Modellen	ΔX^2	Δdf	p
1 intercept	13193.44				
2 + tijdstip van meting	11534.99	2 vs 1	1658.45	3	< .001
3 + interventie	11534.95	3 vs 2	.04	1	.84
4 + tijdstip interventie	11532.44	4 vs 3	2.51	1	.11
5 + conditie	11532.43	5 vs 4	.01	1	.92
6 + interventie * conditie	11528.46	6 vs 5	3.97	1	<.05
7 + percentage gecontroleerde antwoorden * interventie	10589.05	7 vs 6	939.41	1	< .001
8 + percentage gecontroleerde antwoorden * interventie * conditie	10588.52	8 vs 7	.53	1	.47
9 + niet terugkijken * interventie	10588.33	9 vs 8	.19	1	.66
10 + niet terugkijken * interventie * conditie	10587.82	10 vs 9	.51	1	.48

wordt meegenomen (model 9 en 10), heeft geen verbetering tot gevolg, niet als interactie-effect met interventie (Model 9 vs Model 8: $\Delta\chi^2(1) = .19, p = .66$), noch als interactie-effect met interventie en conditie (Model 10 vs Model 9: $\Delta\chi^2(1) = .51, p = .48$).

We rapporteren derhalve vanuit model 7 en laten model 8 tot en met 10 buiten beschouwing.

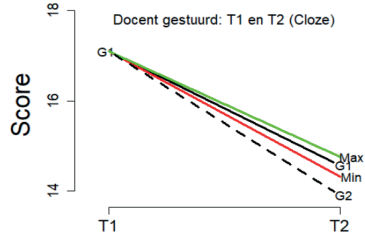
Bijlage 3

Scores op de vier toetsmomenten grafisch weergegeven

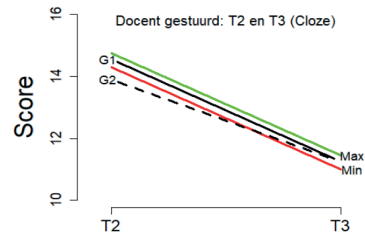
Opmerking vooraf. Een grafische visualisatie van de scores op de vier meetmomenten in één figuur, uitgesplitst voor conditie, voor moment van implementatie én voor wijze van uitvoering door de leerling, blijkt overladen en nauwelijks leesbaar.

Daarom zijn in deze bijlage drie figuren te vinden waarin de gemiddelde score op de vier meetmomenten zijn opgesplitst in een figuur voor het eerste en tweede meetmoment (figuur 3.1: voor- en tussentoets), voor het tweede en derde meetmoment (figuur 3.2: tussentoets en cloze-natoets) en voor het derde en vierde meetmoment (figuur 3.3: beide natoetsen).

Daarbij is ervoor gekozen alleen de conditie ‘Docent-gestuurd leren’ te veraanschouwelijken, omdat hier sprake is zowel van een hoofdeffect van interventie, als van een interactie-effect tussen interventie en de wijze van uitvoering door leerlingen; in de conditie ‘Zelfstandig leren’ ontbreekt dit hoofdeffect.



Figuur 3.1
Gemiddelde score voor de voor- en tussentoets, voor de conditie ‘Docent-gestuurd leren’, voor zowel de groep die de interventie na de voortoets ontving (G1) als de groep die de interventie pas na de tussentoets ontving (G2); voor G1 is daarbij de gemiddelde score veraanschouwelijkt van leerlingen die tijdens de lessenserie hun gegeven antwoorden minimaal controleerden (‘Min’) en maximaal controleerden (‘Max’).



Figuur 3.2
Gemiddelde score voor de tussentoets en cloze-natoets, voor de conditie ‘Docent-gestuurd leren’, voor zowel de groep die de interventie na de voortoets ontving (G1) als de groep die de interventie na de tussentoets ontving (G2); daarbij is de gemiddelde score veraanschouwelijkt van leerlingen die tijdens de lessenserie hun gegeven antwoorden minimaal controleerden (‘Min’) en maximaal controleerden (‘Max’).



Figuur 3.3
Gemiddelde score voor de cloze-natoets en vragen-natoets, voor de conditie ‘Docent-gestuurd leren’, voor beide groepen (G1 en G2); daarbij is de gemiddelde score veraanschouwelijkt van leerlingen die tijdens de lessenserie hun gegeven antwoorden minimaal controleerden (‘Min’) en maximaal controleerden (‘Max’).