

MASTER'S THESIS

Effecten van elaborated feedback in een computerondersteunde leeromgeving op het leerresultaat van eindexamenleerlingen.

Vermeulen, Ellen

Award date:
2023

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 15. Feb. 2023

Open Universiteit
www.ou.nl



**Effecten van elaborated feedback in een computerondersteunde leeromgeving op het
leerresultaat van eindexamenleerlingen**

**Effects of elaborated feedback in a computer-based learning environment on the
learning outcome of graduating students**

Ellen Vermeulen

Master Onderwijswetenschappen, Open Universiteit

E-mailadres: vermeulen.ejh@gmail.com

Cursuscode en cursusnaam: OM9906 Masterscriptie

Naam begeleider: Prof. Dr. Desirée Joosten-ten Brinke

Woordenaantal: 9146

Datum: 04-01-2023

Samenvatting

Door het maken van oefentoetsen ter voorbereiding op een toets of eindexamen kan het *testing effect* bereikt worden. De beschikbaarheid van feedback tijdens het oefenen versterkt het *testing effect*, maar feedback is een breed begrip en er is vooralsnog weinig bekend over de effecten van verschillende feedbackvariabelen (Shute, 2008; Adesope et al., 2017). In dit exploratieve, quasi experimentele onderzoek is op basis van een *between group* ontwerp in een computerondersteunde leeromgeving het effect van *elaborated feedback* op het leereffect van wiskunde oefentoetsen onderzocht. Daarbij is *elaborated feedback* geoperationaliseerd in drie verschillende vormen en zijn achteraf frequentiegroepen bepaald. Ook zijn de modererende effecten van zelf-effectiviteit en het geslacht van de leerling onderzocht. Hiervoor hebben 54 eindexamenleerlingen van vier verschillende wiskunde niveaus een pre-test gedaan, examentrainingsoftware gebruikt, een post-test gedaan en een vragenlijst ingevuld. Non-parametrische tests wezen uit dat het oefenen met examentrainingsoftware inclusief feedback niet leidt tot betere leerresultaten, $U = 353.50$, $z = -0.156$, $p = .876$, $r = -0.02$. Ook bleken er geen significante verschillen te zijn in zelf-effectiviteitsscores en effecten van verschillende feedbackvormen. Het onderzoek toonde wel aan dat leerlingen vooral behoefte hadden aan het controleren van het tussenantwoord (EF-PRC, $M = 33,16$) en in veel mindere mate aan het controleren van het eindantwoord en hints tussendoor (respectievelijk EF-RC, $M = 13,23$ en EF-HI, $M = 7,82$). Alle resultaten moeten met terughoudendheid worden geïnterpreteerd, omdat het aantal participanten laag was.

Keywords: testing effect, feedback, computer, wiskunde

Abstract

The testing effect can be achieved through practice tests in preparation for a test or final exam. The availability of feedback during practice enhances the testing effect, but little is known about the effects of distinct feedback variables (Shute, 2008; Adesope et al., 2017). In this quasi experimental study the effects of elaborated feedback on the students' math learning outcome of practice tests were investigated in a computer-based learning environment, based on a between group design. Elaborated feedback was operationalized in three different types and afterwards frequency groups were identified. The moderating effects of self-efficacy and students' gender was also included. 54 graduating students of four different math classes took a pre-test, used the practicing software, took a post-test and completed a questionnaire. Non-parametric tests showed that software usage including feedback does not improve students' learning outcome, $U = 353.50$, $z = -0.156$, $p = .876$, $r = -0.02$. In addition, there were no differences found between self-efficacy scores and feedback types. The study did show that students prefer checking a subanswer (EF-PRC, $M = 33.16$) over checking the final answer and hints (EF-RC, $M = 13,23$ and EF-HI, $M = 7,82$ respectively). All results should be interpreted with caution due to the low number of participants.

Keywords: testing effect, feedback, computer, mathematics

Inhoud

Samenvatting	2
Abstract	3
Inhoud.....	4
1. Inleiding	5
1.1 Probleemschets en Doel.....	5
1.2 Theoretisch Kader.....	6
1.3 Huidige Studie	15
2. Methode.....	17
2.1 Ontwerp	17
2.2 Deelnemers	18
2.3 Meetinstrumenten en Materialen	19
2.4 Procedure	22
2.5 Data-Analyse	23
3. Resultaten	25
4. Discussie.....	30
4.1 Effect examentrainingssoftware op het leerresultaat	30
4.2 Verschillen tussen feedbackvorm en -frequentie.....	32
4.3 Modererende effecten van zelf-effectiviteit en geslacht van de leerling.....	34
4.4 Beperkingen van het onderzoek	35
4.5 Wetenschappelijke en maatschappelijke relevantie.....	36
Referenties.....	39
Bijlagen	44

Effecten van elaborated feedback in een computerondersteunde leeromgeving op het leerresultaat van eindexamenleerlingen

1. Inleiding

1.1 Probleemschets en Doel

Leerlingen in het voortgezet onderwijs (VO) sluiten hun middelbare schoolperiode af met de centrale eindexamens (CE's). Deze CE's zijn voor elk vak voor alle leerlingen in Nederland hetzelfde (Cito, z.d.), maar elke leerling bereidt zich er anders op voor. Gezien de grootte van de consequenties die het wel of niet slagen voor de CE's met zich meebrengt (het bepaalt tot welk vervolgonderwijs de leerling toegang heeft en zijn/haar kansen op de arbeidsmarkt), is het de vraag of leerlingen op een uniforme manier voorbereid zouden moeten worden op deze toetsen. Daartoe heeft een *startup* een softwareprogramma ontwikkeld waarmee wiskunde leerlingen CE-opgaven uit voorgaande jaren kunnen oefenen. Het originele Cito correctievoorschrift hebben zij zo bewerkt dat leerlingen inzicht krijgen in de denkstappen die gemaakt moeten worden om tot het correcte antwoord te komen. Er kunnen losse wiskunde opgaven, oefenreeksen of complete eindexamens geoefend worden, waarbij de leerling zelf het antwoord kan controleren. Als de leerling de vraag niet kan oplossen, kan er tussendoor een tip opgevraagd worden.

Het maken van vergelijkbare toetsopgaven ter voorbereiding op een examen is een effectieve leerstrategie (Adesope et al., 2017). Het ophalen van informatie uit het geheugen draagt namelijk in de meeste gevallen méér bij aan het versterken van het geheugen van diezelfde informatie dan wanneer het opnieuw bestudeerd wordt; het zogenaamde *testing effect* (Rowland, 2014). Daarnaast is in verschillende onderzoeken aangetoond dat de beschikbaarheid van feedback het *testing effect* versterkt (Adesope et al., 2017). Ook is aangetoond dat uitgebreide feedback in een online leeromgeving leidt tot hogere leerresultaten

in vergelijking met simpele feedback zoals het geven van het juiste antwoord (Van der Kleij et al., 2015). In de examentrainingssoftware is uitgebreide feedback opgenomen, maar de *startup* vraagt zich af hoe ze de feedback moeten vormgeven voor een optimaal leereffect. Helaas is er nog weinig bekend over de effecten van verschillende feedback variabelen, zoals de vorm en frequentie, op het *testing effect* (Adesope et al., 2017; Shute, 2008). Het doel van dit onderzoek is om inzicht te krijgen in de rol van feedbackgebruik in de examentrainingssoftware in relatie tot het leerresultaat. Daarmee levert het onderzoek een bijdrage aan het vergroten van inzicht van de effecten van feedbackvariabelen op het *testing effect* in een digitale leeromgeving. Daarnaast zal het onderzoek in een natuurlijke leeromgeving plaatsvinden, wat een bijdrage kan leveren aan de generaliseerbaarheid en ecologische validiteit van het *testing effect* (Adesope et al., 2017). De *startup* kan de resultaten gebruiken om de feedback in hun software op zo'n manier vorm te geven dat een optimaal leereffect bereikt wordt.

1.2 Theoretisch Kader

Volgens Bloom (1984) is één-op-één begeleiding de meest effectieve vorm van instructie. In het Nederlandse onderwijssysteem is zulke begeleiding waarbij de docent de uitleg onmiddellijk kan aanpassen aan de behoeften van een individuele leerling ondenkbaar. Technologie biedt echter veelbelovende oplossingen voor dit probleem. Net als bij individuele begeleiding door een tutor of docent, kan feedback in een computerondersteunde leeromgeving de brug vormen tussen de huidige kennis en vaardigheden van de leerling en de beoogde leerresultaten (Van der Kleij, Feskens & Eggen, 2015).

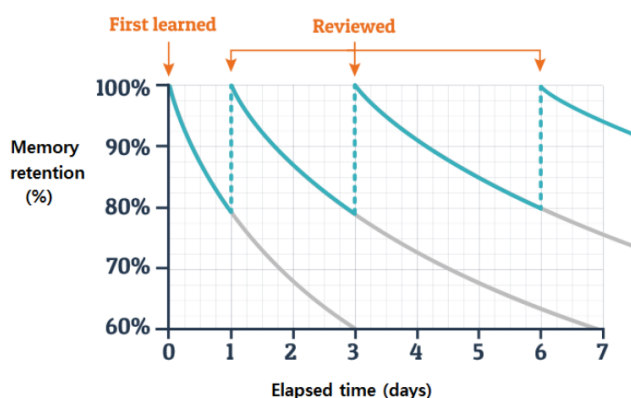
1.2.1 *Testing effect*

Wanneer iets nieuws geleerd is, wordt dat in de meeste gevallen maar voor een beperkte tijd onthouden. Deze correlatie tussen geheugen en tijd is al in 1880 aangetoond in het beroemde onderzoek van Ebbinghaus en is onlangs gerepliceerd door Murre & Dros

(2015). De correlatie houdt in dat informatie uit het geheugen verloren gaat naarmate de tijd vordert. Volgens het model wordt 40% van de geleerde kennis in drie dagen vergeten (zie de grafiek “*First learned*” in Figuur 1). Door het geleerde weer op te halen uit het geheugen wordt de vergeetcurve doorbroken en start een nieuwe, vlakkere vergeetcurve (zie de grafieken “*Reviewed*” in Figuur 1).

Figuur 1

Ebbinghaus' vergeetcurve en reviewcyclus



Noot. Herdrukt uit “The effect of flipped learning on academic performance as an innovative method for overcoming ebbinghaus’ forgetting curve,” door B. A. Chun en H. J. Heo, 2018, *Proceedings of the 6th International Conference on Information and Education Technology*, p. 56. Copyright 2018 door ACM.

Om geleerde informatie langer te onthouden, is het nodig om die informatie (meerdere keren) op te halen uit het geheugen (Chun & Heo, 2018). Hiervoor bestaan verschillende leerstrategieën die ook in leerondersteunende software vormgegeven kan worden. Zo kan de software de leerling aanzetten tot het samenvatten, herlezen of markeren van de leerstof. Alhoewel middelbare scholieren deze strategieën veel gebruiken, zijn deze ineffectief (Dunlosky et al., 2013; Kirschner & Neelen, 2015). Veel effectiever is *retrieval practice* waarbij eerder geleerde informatie wordt opgeroepen uit het geheugen. Een voorbeeld is het maken van een oefentoets. Het ophalen van informatie uit het geheugen draagt namelijk in de

meeste gevallen méér bij aan het versterken van het geheugen van diezelfde informatie dan wanneer het opnieuw bestudeerd wordt; een fenomeen dat bekend staat als het *testing effect* (Rowland, 2014). Het maken van oefentoetsen is een effectieve leerstrategie voor leerlingen van alle leeftijden en vaardigheidsniveaus en krijgt daarom ook een hoge bruikbaarheidsbeoordeling in het onderzoek van Dunlosky et al. (2013).

Aangezien het maken van vergelijkbare toetsopgaven ter voorbereiding op een examen een effectieve leerstrategie is (Adesope et al., 2017), heeft de *startup* examentrainingsoftware ontwikkeld waarin CE-opgaven wiskunde uit voorgaande jaren zijn opgenomen. De opgaven hebben zij voorzien van feedback op basis van het Cito correctievoorschrift. De feedback is facultatief.

Uit diverse onderzoeken blijkt dat de beschikbaarheid van feedback de sterkte van het *testing effect* bevordert in vergelijking met oefentoetsen zonder feedback (Adesope et al., 2017). Toch werd in de meta-analyse van Adesope et al. (2017) geen modererend effect van feedback gevonden op het *testing effect*. Hun vermoeden is dat verschillende feedbackvariabelen (waarop zij de data niet hebben kunnen uitsplitsen omdat deze variabelen niet of nauwelijks beschreven waren) verschillende effecten kunnen hebben op het *testing effect*. Het gaat dan onder andere om het vaardigheidsniveau dat behaald moet worden, de feedbackvorm die daarbij past en de feedbackfrequentie.

1.2.2 Feedbackvorm en vaardigheidsniveau

Door in leerondersteunende software voorbeeldexamenvragen of complete examens toe te voegen kan dan wel het *testing effect* gefaciliteerd worden, het is daarmee nog geen vervanger voor de zeer effectieve één-op-één begeleiding. De reden van effectiviteit is dat de docent of tutor gepersonaliseerde feedback geeft op antwoorden van de leerling, waarbij de feedback wordt aangepast op het vaardigheidsniveau van de leerling.

Nieuw aan te leren vaardigheden worden meestal onderverdeeld in lagere orde vaardigheden (herkennen, begrijpen en onthouden) en hogere orde vaardigheden (analyseren en het toepassen van kennis en vaardigheden) (Attali & Van der Kleij, 2017). De feedback moet afgestemd worden op het te bereiken vaardigheidsniveau, waarbij complexere vaardigheden ook uitgebreidere feedback vereisen (Attali & Van der Kleij, 2017). Shute (2008) heeft verschillende feedbackvormen beschreven en gerangschikt op basis van het complexiteitsniveau: *Knowledge of results* (KR) informeert de leerling enkel over de correctheid van het antwoord (goed of fout). *Knowledge of correct response* (KCR) informeert de leerling enkel wat het goede antwoord is. *Try again* (TA) informeert de leerling over de correctheid van het antwoord (goed of fout), met de mogelijkheid om het opnieuw te proberen. *Elaborated feedback* (EF) is een algemene term met betrekking tot het geven van uitleg waarom het antwoord al dan niet correct was en informeert de leerling over de juiste procedure. In de examentrainingsoftware zijn alleen oefeningen voor hogere orde vaardigheden opgenomen. De facultatieve feedback die de leerling kan gebruiken valt in Shute's (2008) *elaborated feedback* categorie.

EF kan op verschillende manieren vormgegeven worden. Zo kan het bestaan uit verklaringen waarom een antwoord correct of incorrect is, informatie over waar de fout gemaakt is, het soort fout dat gemaakt is, hints over procedurele vaardigheden of probleemoplossende strategieën etc. (Narciss & Huth, 2004). In de examentrainingsoftware zijn drie typen EF aanwezig die Shute (2008) als volgt beschrijft:

- *Hints* (HI): vorm van EF die de leerling in de goede richting leidt door bijvoorbeeld de procedurele stappen te herhalen of een uitgewerkt voorbeeld te tonen. Informatie over het antwoord wordt niet weergegeven.
- *Response contingent* (RC): vorm van EF die zich richt op de reactie van de leerling. Zo kan de leerling bijvoorbeeld informatie krijgen over waarom een antwoord goed of fout is.

- *Partial response contingent* (PRC): adaptatie op de RC vorm, waarbij de “P” staat voor “*partial*”. De leerling kan bijvoorbeeld informatie krijgen over waarom een tussenantwoord goed of fout is.

Door middel van feedback kunnen leerlingen gestimuleerd worden te reflecteren op de kwaliteit van hun cognitieve strategieën en mentale modellen (Van Merriënboer & Kirschner, 2018), wat de efficiëntie van het leerproces kan verbeteren en kan leiden tot betere leerresultaten (Attali & Van der Kleij, 2017). Dit is met name van toepassing op uitgebreide feedback (*elaborated feedback*) gedefinieerd als feedback die als doel heeft om de denkwijze of het gedrag van de leerling te veranderen om het leren te verbeteren (Shute, 2008). Feedback is daarmee een van de krachtigste middelen om het leren van studenten te vergroten (Van der Kleij, Feskens & Eggen, 2015). Als leerlingen gebruikmaken van de verschillende feedbackmogelijkheden in de examentrainingssoftware zou het *testing effect* versterkt kunnen worden. Als er snel gebruik gemaakt wordt van feedback tijdens het oefenen (HI en PRC) zou dat het proces van *retrieval practice* eerder op gang kunnen brengen, waardoor geleerde kennis en vaardigheden al aan het begin van de oefenopgave opgehaald worden uit het geheugen. Dat zou kunnen leiden tot het sneller oplossen van de opgave in tegenstelling tot het opvragen van feedback aan het einde van de opgave (RC). Als opgaven sneller opgelost worden, zou in dezelfde tijd meer geoefend kunnen worden wat een positief effect kan hebben op het leerresultaat.

1.2.3 Feedbacktiming en -frequentie

Feedback geven is nuttig omdat het bijdraagt aan de reflectie op de kwaliteit van de cognitieve strategieën en mentale modellen van leerlingen (Van Merriënboer & Kirschner, 2018) en leidt het tot het herstellen of vervangen van de onjuistheid van de strategieën en modellen en het versterken van nuttige mentale verbanden (Pashler et al., 2005). Dit suggereert dat het geven van feedback altijd nuttig is en dat de feedbackfrequentie in een

computerondersteunde leeromgeving altijd hoog zou moeten zijn. Verrassend genoeg heeft een aantal onderzoeken aangetoond dat die gedachte niet altijd opgaat. Met name voor het aanleren van cognitieve vaardigheden op een hoger niveau blijkt feedback niet altijd een positief effect te hebben op de leerresultaten (Pashler et al., 2005). Een verklaring hiervoor kan zijn dat het achterhouden van feedback de leerling dwingt tot een diepere verwerking tijdens het leren, wat retentie op langere termijn verbetert (Pashler et al., 2005). Het lijkt daarom aannemelijk dat de feedbackfrequentie een grens heeft waarbij het toevoegen van meer feedbackmomenten een negatief effect begint te krijgen op de leerresultaten.

In de computerondersteunde leeromgeving bepaalt de leerling zelf de frequentie van de feedback. Ook het moment waarop feedback wordt getoond kan de leerling zelf kiezen. Zo kan de leerling direct feedback opvragen wanneer er een antwoord geformuleerd is (*immediate feedback*), of op een later moment bijvoorbeeld na het maken van een toets (*delayed feedback*) (Shute, 2008). Onderzoek waarin de frequentie en het moment van opvragen van feedback onderzocht is, tonen geen eenduidig effect op het leerresultaat (Shute, 2008). Voorstanders van *immediate feedback* argumenteren dat directe feedback een middel is om te voorkomen dat fouten in het geheugen worden opgeslagen (Shute, 2008), terwijl voorstanders van *delayed feedback* betogen dat als feedback wordt aangeboden met een vertraging dit effectiever is in het corrigeren van fouten omdat het foute antwoord niet ‘concurrereert’ met het juiste antwoord in het hoofd van de leerling volgens de zogenaamde *interference-perseveration theory* (Attali & Van der Kleij, 2017). Attali & Van der Kleij (2017) suggereren dat zelfregulatie en zelf-effectiviteit waarschijnlijk een grote rol spelen bij het effect van feedbacktiming; variabelen die in “laboratorium” omstandigheden niet worden meegenomen. Zij argumenteren dat als *immediate* en *delayed feedback* beide worden aangeboden door het leersysteem, leerlingen zelf kunnen kiezen of en in welke mate ze gebruik willen maken van de feedback. Het onderscheid tussen de “laboratorium”

omstandigheden en de praktijk zou aanzienlijke gevolgen kunnen hebben voor de verklaring van de feedbacktiming en -frequentie, maar wordt nog onvoldoende begrepen (Attali & Van der Kleij, 2017).

1.2.4 Elaborated feedback en het testing effect

Van der Kleij et al. (2015) toonden aan dat *elaborated feedback* (EF) in combinatie met *knowledge of results* (KR) en/of *knowledge of correct response* (KCR) in de vorm van uitleg, informatie met betrekking tot strategieën en hints (inclusief *try again*, TA) het meest bestudeerd is in computerondersteunde leeromgevingen en over het algemeen zeer effectief is. Ze toonden aan dat de effectgroottes van EF groter zijn (0.49) dan van KR (0.05) en KCR (0.32) en dat de effectgrootte bij wiskunde het hoogst was vergeleken met andere schoolvakken. Ze plaatsten wel de kanttekening dat het wiskunde resultaat gebaseerd is op een klein aantal studies, waardoor het effect onterecht hoog zou kunnen zijn. In hun onderzoek werd geen onderscheid gemaakt tussen de verschillende EF-vormen. Ze toonden ook aan dat wanneer het om hogere orde vaardigheden gaat (het toepassen van kennis in vaardigheden in een nieuwe situatie) EF significant effectiever is dan KR en KCR. Van den Broek et al. (2019) toonden aan dat het toevoegen van een hint niet per se een nuttige aanvulling is op feedback wanneer het om een mislukte ophaalpoging uit het geheugen gaat. De leerresultaten waren namelijk niet beter in vergelijking met het direct tonen van het antwoord na een mislukte ophaalpoging (KCR). Zij redeneren dat in dat geval tijd verloren is gegaan door het lezen van de tips. Tijd die wellicht beter besteed had kunnen worden aan het opnieuw bestuderen van de informatie (Van den Broek et al., 2019).

In de examentrainingsoftware zijn drie typen EF aanwezig die in paragraaf 1.2.2. beschreven zijn volgens de classificering van Shute (2008). De praktische betekenis is als volgt:

- EF-HI: tips die de leerling kan opvragen tijdens het maken van de oefening. Deze tips hebben geen corrigerende functie (Glover & Brown, 2006), maar bieden strategische informatie om de leerling aan te moedigen het antwoord of de strategie te heroverwegen (Attali & Van der Kleij, 2017). Tips of hints kunnen gezien worden als een extra moment om informatie uit het geheugen op te halen (Van den Broek et al., 2019).
- EF-RC: het uitgewerkte antwoord dat leerlingen zelf kunnen opvragen aan het einde van de oefening. Deze antwoorden hebben wel een corrigerende functie, maar worden gecombineerd met aanvullende instructie-informatie waarmee leerlingen gestimuleerd worden te reflecteren op hun antwoord en werkwijze (Attali & Van der Kleij, 2017). In sommige gevallen wordt ook beschreven waarom een bepaald antwoord fout is.
- EF-PRC: deze naamgeving is een adaptatie op de EF-RC, waarbij de “P” staat voor “*partial*”. Het gaat hierbij om een tussenantwoord dat leerlingen zelf kunnen opvragen tijdens het maken van de oefening wanneer de hints niet afdoende zijn. Dit tussenantwoord heeft ook als functie om de leerling aan te moedigen het antwoord of de strategie te heroverwegen en zo toch zelfstandig tot het totale antwoord te komen. Verder heeft het dezelfde eigenschappen als EF-RC.

Door gebruik te maken van de *elaborated feedback* vormen in de examentrainingsoftware zou het *testing effect* versterkt kunnen worden. De vraag welke (combinatie van) gebruikte feedbackvormen het *testing effect* het meest zal versterken is onderdeel van dit onderzoek.

1.2.5 Omgevingsfactoren en het vak wiskunde

Eerder onderzoek naar het effect van *elaborated feedback* in de vorm van hints is schaars en leverde tegenstrijdige bevindingen op (Van den Broek et al., 2019). In de literatuur is geen consistent positief effect van feedback gevonden voor wiskunde (Van der Kleij et al., 2015). Een mogelijke verklaring voor de uiteenlopende resultaten kan zijn dat naast passende inhoudelijke feedback (vorm en frequentie) ook individuele cognitieve factoren van leerlingen

relevant zijn (Narciss & Huth, 2004) zoals bekwaamheid, zelfvertrouwen en motivatie (Attali & Van der Kleij, 2017). Deze factoren staan in nauw verband met zelf-effectiviteit dat omschreven kan worden als het niveau van vertrouwen dat iemand heeft om een bepaalde taak goed uit te voeren (Alqurashi, 2016), voortkomend uit Bandura's (1977) theoretische raamwerk over zelf-effectiviteit. Overtuigingen over eigen zelf-effectiviteit bepalen hoe leerlingen zich voelen en hoe gemotiveerd ze zijn, wat belangrijke factoren zijn in hun handelen (Alqurashi, 2016). Multon et al. (1991) toonden aan dat zelf-effectiviteit inderdaad een significante predictor is van het leerresultaat. Daarnaast toonden zij aan dat de effectgrootte heterogeen is, wat wijst op modererende factoren zoals leeftijd en geslacht van de leerling. Een overvloed aan studies heeft aangetoond dat op het gebied van wiskundeprestaties jongens een hogere zelf-effectiviteit hebben dan meisjes (Meece & Agger, 2018). Gezien de positieve invloed van zelf-effectiviteit op prestaties en motivatie is een beter begrip nodig van verschillen tussen jongens en meisjes (Meece & Agger, 2018). In dit onderzoek zijn daarom zelf-effectiviteit en geslacht meegenomen. Deze variabelen zouden meer inzicht kunnen geven waarom er geen consistent positief effect van feedback gevonden kan worden omtrent wiskunde leerprestaties. In onderzoek dat gelijktijdig wordt uitgevoerd door medestudent H el ene Kommers wordt het aspect motivatie op de leerresultaten onderzocht met dezelfde dataset.

Het is belangrijk op te merken dat aangetoonde effecten gelden in een traditionele leeromgeving (klas). Onderzoek naar de rol van zelf-effectiviteit in online leeromgevingen staat nog in de kinderschoenen (Hodges, 2008). Wel is aangetoond dat de zelf-effectiviteit van leerlingen in een online leeromgeving wordt beïnvloed door extra factoren zoals eerder succes met online leren, training voorafgaand aan de cursus, feedback van docenten en angst voor online leertechnologie (Alqurashi, 2016). De rol van docenten daarin is cruciaal, omdat zij

bepalen of er instructie over de online leeromgeving wordt gegeven, of ze zelf feedback geven en hoe ze omgaan met leerlingen die minder digitaal vaardig zijn.

1.3 Huidige Studie

In dit onderzoek is de praktijkvraag van de *startup* ('hoe moet de facultatieve feedback vormgegeven worden in de examentrainingsoftware om een optimaal leerresultaat te bereiken') omgezet in de centrale vraagstelling: in hoeverre is er een verschil in effect van *elaborated feedback* op het leereffect van oefentoetsen in een online leeromgeving? Om deze vraagstelling te kunnen beantwoorden wordt de feedbackvorm en -frequentie onderzocht, wat resulteert in de volgende deelvragen en hypothesen:

- Deelvraag 1: is er een verschil in leerresultaat tussen leerlingen die de examentrainingsoftware wel en niet gebruikt hebben? Voordat er ingezoomd wordt op de verschillende feedbackvariabelen wordt eerst bepaald of er een *testing effect* is opgetreden bij de groep leerlingen die de examentrainingsoftware met facultatieve feedback gebruikt hebben. Omdat de software nog niet lang gebruikt wordt door de school, zal de zelf-effectiviteit van sommige leerlingen in de online leeromgeving laag zijn omdat ze nog geen succes hebben ervaren met online leren of angst hebben voor online leertechnologie (Alqurashi, 2016). Om die reden is de verwachting dat er op natuurlijke wijze een groep leerlingen ontstaat die de software niet zullen gebruiken als leermiddel. Verwacht wordt dat het gebruiken van de examentrainingsoftware met facultatieve feedback het leerresultaat meer zal vergroten dan wanneer de software niet gebruikt wordt, omdat het oefenen met vergelijkbare toetsopgaven een effectieve leerstrategie is en omdat *elaborated feedback* zeer effectief is bevonden in online leeromgevingen (Van der Kleij et al., 2015).
- Deelvraag 2: in hoeverre is er een verschil in het gebruik van verschillende feedbackvormen (EF-HI, EF-RC en EF-PRC) in de examentrainingsoftware? Dit

onderzoek wordt in een natuurlijke omgeving uitgevoerd, waarbij de leerlingen zelf bepalen of en wanneer ze feedback krijgen. De verwachting is dat sommige leerlingen tussentijdse hints (EF-HI) en antwoorden (EF-PRC) zullen overslaan, terwijl anderen ervoor zullen kiezen om alle feedback zorgvuldig te beoordelen (Attali & Van der Kleij, 2017). Er wordt verwacht dat leerlingen al snel willen weten of ze goed op weg zijn en daarom vaker tussentijds feedback opvragen (EF-HI en EF-PRC) in vergelijking met het controleren van het eindantwoord (EF-RC).

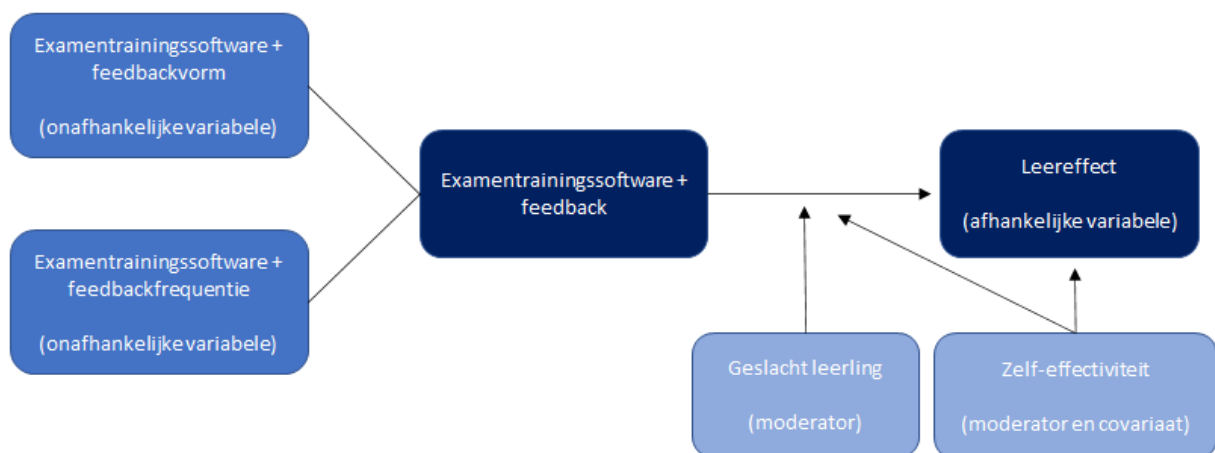
- Deelvraag 3: in hoeverre is er een verschil in effect van de feedbackfrequentie in de examentrainingsoftware op het leerresultaat? Het ontvangen van frequente feedback kan de cognitieve capaciteit van leerlingen overweldigen, wat een negatief effect kan hebben op taakspanning (Lam et al., 2011). Omdat leerlingen de feedbackfrequentie zelf kunnen reguleren in de online oefenomgeving, zal alleen feedback opgevraagd worden als daar behoefte aan is en zal waarschijnlijk niet ‘teveel’ feedback opgevraagd worden. Verwacht wordt dat leerlingen die weinig feedback opvragen in de examentrainingsoftware een lager leerresultaat zullen hebben in vergelijking met leerlingen die gemiddeld of veel feedback opvragen. De reden daarvoor is dat feedback bijdraagt aan reflectie en leidt tot het herstellen of vervangen van de onjuistheid van de strategieën en modellen en het versterken van nuttige mentale verbanden (Van Merriënboer & Kirschner, 2018).
- Deelvraag 4: in hoeverre wordt het leereffect van oefentoetsen gemodereerd door zelf-effectiviteit en het geslacht van de leerling? Van den Broek et al. (2019) beschreven dat eerder onderzoek naar *elaborated feedback* en de relatie tot het leerresultaat tegenstrijdige bevindingen opleverde. Narciss & Huth (2004) gaven als mogelijke oorzaak dat ook individuele cognitieve factoren van leerlingen relevant zijn. De modererende rol van deze variabelen worden daarom met deze deelvraag onderzocht. Een overvloed aan studies heeft aangetoond dat op het gebied van wiskundeprestaties jongens een hogere zelf-

effectiviteit hebben dan meisjes (Meece & Agger, 2018), dus dat wordt in dit onderzoek ook verwacht. Verwacht wordt dat leerlingen met een hogere zelf-effectiviteit meer vertrouwen hebben in het kunnen uitvoeren van oefenopdrachten. Dat kan leiden tot het meer of beter gebruiken van de examentrainingsoftware en feedback, dat vervolgens kan leiden tot een hoger leerresultaat.

In figuur 2 is het conceptuele model samengevat waarin alle effecten schematisch zijn weergegeven.

Figuur 2

Conceptueel model van de relaties tussen de variabelen



Noot. Het conceptueel model van het effect van feedback in de examentrainingsoftware (bestaande uit twee onafhankelijke variabelen feedbackvorm en -frequentie) op het leereffect (afhankelijke variabele) en de rol van de modererende variabelen geslacht van de leerling en zelf-effectiviteit op dit effect. Voor de volledigheid is ook het directe effect van zelf-effectiviteit op het leereffect (Meece & Agger, 2018) weergegeven.

2. Methode

2.1 Ontwerp

Het doel van dit exploratieve, quasi experimentele onderzoek is om inzicht te krijgen

in de rol van feedbackgebruik in examentrainingsoftware in relatie tot het leerresultaat, waarmee het gehoor geeft aan de oproep dat meer onderzoek naar de effecten van feedback variabelen op het *testing effect* nodig is (Adesope et al., 2017; Shute, 2008; Attali & Van der Kleij, 2017). Het gaat om een quasi experiment in de zin dat het om bestaande schoolklassen gaat, waarbij de leerlingen niet willekeurig worden toegewezen aan een experimentele conditie (Creswell, 2014). Als een groep leerlingen geen gebruik mag maken van het examentrainingproduct met feedback, zouden zij mogelijk benadeeld worden (Kember, 2003), dus is er geen controlegroep opgenomen.

Voor het onderzoek is een *between group* ontwerp gebruikt. Aangezien het opvragen van feedback in de examentrainingsoftware facultatief is, worden achteraf verschillende groepen bepaald op basis van de gebruikte feedbackvorm en -frequentie in de online leeromgeving.

2.2 Deelnemers

Op basis van de “F tests” en de statistische test “ANCOVA: Fixed effects, main effects and interactions” is de “A priori” poweranalyse uitgevoerd (Mayr et al., 2007). Voor een effectgrootte van 0.25 (*medium effect*), α -foutmarge van 5%, betrouwbaarheidsinterval van 95%, aantal vrijheidsgraden 4, aantal groepen 9 en aantal covariaten 2 is een steekproefgrootte van 303 nodig.

Om de bereidwilligheid voor deelname aan het onderzoek zo groot mogelijk te maken, is de meetperiode zo gekozen dat de examentrainingsoftware gebruikt kon worden als leermiddel ter voorbereiding op de laatste schoolexamen periode. Op basis van dit criterium bleken drie scholen de software te gaan gebruiken in die periode. Aan die scholen is gevraagd mee te doen met het onderzoek. School A is na het ontvangen van de informatiebrief afgehaakt. School B heeft alleen meegedaan met de T0 meting. School C heeft meegedaan

met het volledige onderzoek. Alleen de data van die school is daarom gebruikt voor dit onderzoek.

School C is een scholengemeenschap voor havo en vwo. Totaal hebben 86 leerlingen deelgenomen aan (een deel van) het onderzoek, verdeeld over vier verschillende wiskundevakken en niveaus. 32 van hen hebben alleen de toets van T0 of T1 gemaakt of hebben de vragenlijst onvolledig ingevuld. Deze leerlingen zijn daarom niet meegenomen in het onderzoek. Gegevens over de 54 leerlingen die wel beide oefentoetsen gemaakt hebben en de vragenlijst volledig hebben ingevuld staan in Tabel 1.

Tabel 1

Kenmerken van deelnemers

Klas	5 havo		5 havo		6 vwo		6 vwo		Totaal	
	wiskunde A		wiskunde B		wiskunde A		wiskunde B			
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Geslacht										
Jongen	11	20,4	3	5,6	6	11,1	2	3,7	22	40,7
Meisje	17	31,5	1	1,9	9	16,7	0	0	27	50
Anders	0	0	0	0	0	0	0	0	0	0
Wil ik niet zeggen	2	3,7	3	5,6	0	0	0	0	5	9,3
Totaal	30	55,6	7	13,0	15	27,8	2	3,7	54	100

2.3 Meetinstrumenten en Materialen

2.3.1 Oefentoetsen voor T0 en T1 meting

Aan het begin en eind van de meetperiode is een oefentoets afgenomen. Voor elk wiskundeniveau zijn daarvoor twee toetsen gemaakt, wat resulteert in acht verschillende toetsen. De onderwerpen van de oefentoetsen zijn samen met een wiskunde docent van school C bepaald, om er voor te zorgen dat deze aansluiten bij de laatste schoolexamenperiode.

School- en eindexamens op havo- en vwo-niveau toetsen voornamelijk hogere orde vaardigheden zoals het probleemoplossend vermogen van de leerling. Bij dit vaardigheidstype is de meest relevante test om te bepalen of studenten een probleemoplossingsprocedure hebben geleerd, het oplossen van isomorfe problemen (Attali & Van der Kleij, 2017). De toetsen van T0 en T1 bevatten daarom vragen uit de examentrainingsoftware uit dezelfde wiskunde domeinen met dezelfde structurele kenmerken, oplossingsprocedure en gemiddelde moeilijkheidsgraad. De moeilijkheidsgraad van elke opgave is in de examentrainingsoftware gecategoriseerd als 'laag', 'gemiddeld' en 'hoog' op basis van de p-waarde van Cito (<https://www.cito.nl/kennis-en-innovatie/microlearningbank/toets--en-itemanalyse/p-waarde>). De verschilscore van de twee oefentoetsen is de afhankelijke, continue variabele. De toetsen zijn met de Cito nakijkmodellen nagekeken door de onderzoeker.

2.3.2 Examentrainingsoftware

Tussen T0 en T1 hebben leerlingen gebruik gemaakt van de examentrainingsoftware als leermiddel. In de software kunnen leerlingen zelf wiskunde oefenopgaven zoeken, maar de docent kan ook oefeningen klaarzetten (zie bijlage C). Tijdens het oefenen kunnen leerlingen feedback opvragen in de software. Die feedback bestaat uit hints, oftewel EF-HI (zie bijlage D), het uitgewerkte correcte antwoord EF-RC en het uitgewerkte correcte tussenantwoorden EF-PRC (zie bijlage E) op basis van het Cito-correctievoorschrift. De *startup* logt op de achtergrond gegevens over hoe de leerlingen de software gebruikt hebben. Per participant zijn de volgende gegevens gelogd, waarmee de onafhankelijke variabelen gedefinieerd zijn:

- Hoe vaak er ingelogd is (ratio)
- Hoe vaak een opgave geopend is (ratio)
- Welke (EF-HI, EF-PRC en EF-RC) en hoe vaak feedback opgevraagd is (nominaal)
- Hoe vaak het (tussen)antwoord is nagekeken (ratio)
- Hoe vaak het (tussen)antwoord is gemarkeerd als correct is (ratio)

- Hoe vaak het antwoord is gemarkeerd correct is (ratio)

Na afloop van de interventieperiode is het databestand aangeleverd aan de onderzoeker, waarin elke leerling geïdentificeerd is op basis van het e-mailadres waarmee ingelogd wordt in de software. Met het e-mailadres wordt de data gekoppeld aan de andere variabelen.

2.3.3 Self-Efficacy Questionnaire for Children

De Nederlandstalige SEQ-C (Muris, 2001) bestaat uit 24 items, onderverdeeld in drie domeinen van zelf-effectiviteit: 1. sociale zelf-effectiviteit (vraag 1, 5, 7, 10, 13, 16, 18 en 23), 2. academische zelf-effectiviteit (vraag 3, 6, 9, 12, 15, 17 en 20) en 3. emotionele zelf-effectiviteit (vraag 2, 4, 8, 11, 14, 19 en 21) (bijlage A). Uit de factoranalyse van Muris (2001) blijkt dat drie items (één uit elk domein) geen overtuigende lading vertoonden (vraag 1, 23 en 18). Op basis van dit resultaat is gesuggereerd om de drie items uit de vragenlijst te verwijderen (Muris, 2001), wat in andere onderzoeken ook gebeurd is (Habibi et al., 2014; Suldo & Huebner, 2006). Ook in dit onderzoek zijn de drie items verwijderd.

Elk item wordt gescoord op een vijfpunts Likertschaal, met 1 = “lukt niet”, en 5 = “lukt erg goed”. De validiteitsanalyse van Muris (2001) vertoonde een Cronbach's α van 0.88 met subschaalscores tussen 0.85 en 0.88. In dit onderzoek bleek uit de validiteitssanalyse dat de totale zelf-efficiëntiescore een Cronbach's α had van 0.89 en de academische, emotionele en sociale subschalen respectievelijk 0.75, 0.84, 0.90. Daarmee scoort de SEQ-C boven de Alpha (α) betrouwbaarheidsgrenswaarde van 0.70 (Hair et al., 2013).

2.3.4 Overige omgevingsfactoren

Om het modererende effect van het geslacht van de leerling te onderzoeken, is de vraag naar geslacht toegevoegd aan de SEQ-C vragenlijst. Ook is het e-mailadres waarmee wordt ingelogd in de examentrainingssoftware uitgevraagd om de antwoorden te kunnen koppelen aan de gelogde data.

2.4 Procedure

Aan de wiskunde docenten van de scholen die hebben aangegeven mee te willen doen aan het onderzoek is drie weken voorafgaand aan het onderzoek per e-mail de informatiebrief voor docenten en toestemmingsverklaring gestuurd. Een week later is een herinneringsmail gestuurd aan docenten die nog niet gereageerd hadden. Aan de docenten die toestemming gaven om mee te doen is gevraagd om de informatiebrief voor leerlingen door te sturen naar hun leerlingen. In de informatiebrieven zijn docenten en leerlingen geïnformeerd dat hun deelname vrijwillig was en dat ze vrij waren om hun deelname te stoppen wanneer ze maar wilden.

Fase 1: een week voorafgaand aan het onderzoek zijn vier oefentoetsen (wiskunde A/B voor havo/vwo) naar de deelnemende docenten gestuurd met instructies over hoe de eerste oefentoets (T0-meting) afgenomen moet worden en naar welk adres ze de uitwerkingen per post konden toesturen. Ook is uitgelegd dat ze in de periode daarna konden oefenen met de examentrainingsoftware.

Fase 2: na de T0-meting start de interventieperiode van twee weken. Tijdens deze periode gaan leerlingen onder begeleiding van hun docent aan de slag met de examentrainingsoftware. Docenten bepalen zelf hoe en hoe vaak ze de software inzetten als leermiddel in hun lessen. Tijdens deze periode worden alle onafhankelijke variabelen die gemeten worden via de software gelogd en beheerd door de *startup*.

Fase 3: in de week na de interventieperiode is aan de deelnemende docenten een e-mail gestuurd met vier andere oefentoetsen (wiskunde A/B voor havo/vwo) en instructies over hoe de tweede oefentoets (T1-meting) afgenomen moet worden. Daarnaast is gevraagd of ze de link naar de online Lime Survey vragenlijst willen doorsturen naar hun leerlingen. In de online vragenlijst is de toestemmingsverklaring voor leerlingen toegevoegd en is herhaald dat deelname aan het onderzoek vrijwillig was en dat ze vrij waren om hun deelname te stoppen

wanneer ze maar wilden. Nadat toestemming is verleend wordt het geslacht en het school e-mailadres uitgevraagd. Tot slot is de SEQ-C opgenomen. Na twee weken is aan docenten gevraagd om hun leerlingen te herinneren aan het invullen van de vragenlijst.

Na afloop van het onderzoek is de data bij de *startup* opgevraagd. Voor de privacy en gegevensbescherming neemt de *startup* deel aan het Privacyconvenant Digitale Onderwijsmiddelen (zie <https://www.privacyconvenant.nl> voor meer informatie). Daarvoor worden modelovereenkomsten voor de Privacybijsluiter en Beveiligingsbijlage gebruikt, die zijn opgesteld met de brancheorganisatie voor educatieve uitgevers GEU. Alle onderzoeksdata is opgeslagen op de beschikbaar gestelde veilige opslaglocatie van de Open Universiteit. Daarnaast is het onderzoeksvorstel getoetst door de commissie Ethische Toetsing Onderzoek (cETO) van de Open Universiteit.

2.5 Data-Analyse

De beschrijvende statistieken en analyses zijn met IBM Corp. SPSS Statistics versie 28.0.1.0 (IBM, Armonk, NY, USA) uitgevoerd. De minimaal benodigde steekproefgrootte voor voldoende betrouwbare resultaten is berekend met de software G*Power versie 3.1.9.7 (Faul et al., 2009).

Voorafgaand aan de analyses ter beantwoording van de onderzoeksvragen zijn data-exploraties uitgevoerd waaruit bleek dat de leerresultaten niet-normaal verdeeld zijn. Daarom zijn non-parametrische tests gebruikt voor het analyseren van de data.

Om de eerste deelvraag te kunnen beantwoorden is de verschilscore van het leerresultaat berekend middels de T0- en T1-meting. Met de Mann-Whitney U test is vervolgens onderzocht of de gemiddelden tussen de leerlingen die de examentrainingsoftware gebruikt hebben verschilt van leerlingen die de software niet gebruikt hebben. Om te kunnen onderzoeken of zelf-effectiviteit een rol speelt bij het wel of niet gebruiken van de examentrainingsoftware zijn de resultaten van de SEQ-C vragenlijst omgezet in een

totaalscore en score per domein. Met een Mann-Whitney U test is onderzocht of de totaalscore en scores per domein verschillen tussen de twee groepen.

Voor beantwoording van deelvraag twee en drie is de gelogde data omgezet in de variabelen EF-HI, EF-RC en EF-PRC. Op basis van percentielscores 33,33% en 66,67% zijn de drie variabelen opgesplitst in drie categorieën ‘weinig’, ‘gemiddeld’ en ‘veel’ gebruikt. Om de verschillen tussen de categorieën te onderzoeken zijn Kruskal-Wallis tests uitgevoerd. Leerlingen kunnen alle drie de feedbackvormen gebruiken en kunnen die weinig, gemiddeld of veel gebruiken. Daarom is bij elke leerling een binair negen-cijferig feedbackpatroon beschreven zoals ‘geen HI, geen RC, geen PRC’ (= 000000000); ‘weinig HI, geen RC, geen PRC’ (= 100000000); ‘gemiddeld HI, geen RC, geen PRC’ (= 010000000), ‘vaak HI, geen RC, weinig PRC’ (= 001000000); ‘geen HI, weinig RC, geen PRC’ (=000100000) etc. Totaal zijn er 37 patronen mogelijk. Verschillen tussen deze groepen zijn onderzocht met Kruskal-Wallis tests.

Om de laatste deelvraag te beantwoorden is met de Quade’s nonparametrische ANCOVA onderzocht of zelf-effectiviteit en het geslacht een modererende werking hebben op het leereffect van oefentoetsen. Daarvoor zijn de leerresultaat verschilcores van de feedbackgroepen die voortkomend uit deelvraag 3 significant met elkaar verschillen vergeleken, met als covariaat (de afzonderlijke domeinen van) zelf-effectiviteit en geslacht.

In de data-analyse wordt een significantieniveau van 5% gehanteerd (Creswell, 2014). Afsluitend is een datapakket samengesteld om hergebruik van de onderzoeksdata mogelijk te maken bestaande uit datadocumentatie, metadata, onbewerkte databestanden en verwerkte databestanden.

3. Resultaten

Het leerresultaat van de 54 leerlingen die met beide metingen hebben meegedaan en de vragenlijst volledig hebben ingevuld was $M = 4.8$, $SD = 2.3$ op T0. Op T1 was dat $M = 4.9$, $SD = 2.2$. De gemiddelde stijging van het leerresultaat was $M = 0.1$, $SD = 2.4$.

Voor deelvraag 1 bleek uit de beschrijvende data-analyse dat in twee klassen leerlingen de examentrainingsoftware nooit gebruikt hebben (zie tabel 2).

Tabel 2

Gebruik examentrainingsoftware per klas

Klas	<i>n</i>	havo 5	havo 5	vwo 6	vwo 6
		wiskunde A	wiskunde B	wiskunde A	wiskunde B
Software gebruik					
Niet	25	15	0	10	0
Wel	29	15	7	5	2

Noot. Het wel of niet gebruiken van de software is gebaseerd op of er wel of niet is ingelogd.

De leerresultaten op T0 en T1 van leerlingen die de examentrainingsoftware wel en niet gebruikt hebben zijn samengevat in tabel 3.

Tabel 3

Leerresultaat op T0 en T1

Meetmoment	<i>n</i>	T0		T1	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Software gebruik					
Niet	25	5.12	2.19	4.99	1.93
Wel	29	4.52	2.46	4.84	2.42

Noot. Het wel of niet gebruiken van de software is gebaseerd op of er wel of niet is ingelogd.

De leerresultaat verschilscore van leerlingen die nooit ingelogd hebben in de examentrainingsoftware ($n = 25$) was $M = -0.13$, $SD = 0.48$ en dat van leerlingen die de software wel gebruikt hebben ($n = 29$) was $M = 0.32$, $SD = 0.44$. Shapiro-Wilk-tests toonden

aan dat het leerresultaat op T0, $W = 0.95$, $p < .05$, het leerresultaat op T1, $W = 0.95$, $p < .05$ en de leerresultaat verschilscore $W = 0.94$, $p < .05$ alle drie niet normaal verdeeld waren. Voor de vervolganalyses zijn daarom niet-parametrische tests gebruikt.

De Mann-Whitney U test liet zien dat het verschil tussen het leerresultaat van leerlingen die de software wel en niet gebruikt hebben niet significant is, $U = 353.50$, $z = -0.156$, $p = .876$, $r = -0.02$.

Gezien het exploratieve karakter van dit onderzoek is op basis van de vorige resultaten een vervolganalyse gedaan. Zelf-effectiviteitscores zouden namelijk een verklaring kunnen geven waarom leerlingen de software niet gebruikt hebben. Tabel 4 geeft een overzicht van de zelf-effectiviteit per groep.

Tabel 4

Zelf-effectiviteit van leerlingen

Type zelf-effectiviteit	n	Academisch		Sociaal		Emotioneel		Totaal	
		M	SD	M	SD	M	SD	M	SD
Sotware gebruik	25	3.27	0.51	3.77	0.75	3.53	0.83	10.57	1.68
Niet	25	3.27	0.51	3.77	0.75	3.53	0.83	10.57	1.68
Wel	29	2.98	0.73	3.70	0.60	3.39	0.88	10.07	1.71

Noot. Elk item in een zelf-effectiviteitsdomein is gescoord op een vijfpunts Likertschaal, met 1 = “lukt niet”, en 5 = “lukt erg goed.”

Mann-Whitney U tests lieten zien dat de verschillen in zelf-effectiviteit tussen leerlingen die de software wel en niet gebruikt hebben niet significant zijn: academisch $\chi^2(1) = 1.40$, $p = .236$; sociaal $\chi^2(1) = 0.74$, $p = .388$; emotioneel $\chi^2(1) = 0.27$, $p = .602$; totaal $\chi^2(1) = 0.91$, $p = .339$.

Deelvraag 2 gaat over het gebruik van de verschillende feedbackvormen. De frequenties van de feedbackvormen die de 29 leerlingen gebruikt hebben, zijn samengevat in tabel 5.

Tabel 5*Frequenties van gebruikte feedbackvormen*

	<i>n</i>	<i>M</i>	<i>SD</i>	Grenswaarde percentiel 33,33%	Grenswaarde percentiel 66,66%
Feedbackvorm					
EF-HI	22	7.82	11.07	3.67	7.67
EF-RC	22	13.23	17.54	3.67	12.00
EF-PRC	19	33.16	37.16	14.67	31.67

Noot. EF = *elaborated feedback*, HI = *hints*, RC = *response contingent*, PRC = *partial response contingent*.

Op basis van de grenswaarden van de percentielen 33,33% en 66,66% uit tabel 5, zijn de leerlingen verdeeld in de groepen ‘weinig’, ‘gemiddeld’ en ‘vaak’ gebruik zodat deelvraag 3 beantwoord kan worden. De resultaten staan in tabel 6.

Tabel 6*Indeling van gebruiksfrequentie per feedbackvorm*

Feedbackvorm	EF-HI			EF-RC			EF-PRC		
	leerresultaat			leerresultaat			leerresultaat		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Weinig gebruik	7	0.96	0.54	7	0.24	3.70	6	1.63	3.25
Gemiddeld gebruik	8	-0.24	3.59	9	-0.36	2.05	7	-1.43	2.15
Vaak gebruik	7	1.03	2.20	6	0.91	2.40	6	0.88	2.38

Noot. EF = *elaborated feedback*, HI = *hints*, RC = *response contingent*, PRC = *partial response contingent*.

Kruskal-Wallis H tests lieten zien dat er geen significante verschillen waren tussen het weinig, gemiddeld en vaak gebruiken van de *elaborated feedback* vormen. In tabel 7 een paarsgewijze vergelijking.

Tabel 7*Paarsgewijze vergelijking gebruiksfrequentie per feedbackvorm*

Feedbackvorm	EF-HI		EF-RC		EF-PRC	
	χ^2	<i>p</i>	χ^2	<i>p</i>	χ^2	<i>p</i>
Weinig – gemiddeld	2.04	.544	0.29	.930	0.75	.817
Weinig - veel	-2.82	.400	-3.05	.398	5.92	.058
Gemiddeld - veel	-4.86	.161	-3.33	.329	-5.17	.098

Noot. EF = *elaborated feedback*, HI = *hints*, RC = *response contingent*, PRC = *partial response contingent*.

Leerlingen konden alle drie de feedbackvormen gebruiken. Alle mogelijke feedbackpatronen (37 mogelijkheden) zijn paarsgewijs met elkaar vergeleken. Zie bijlage B voor alle resultaten. De Kruskal-Wallis H tests lieten zien dat het leerresultaat tussen “geen HI, weinig RC, gemiddeld PRC” en “gemiddeld HI, weinig RC, weinig PRC” significant verschilt $\chi^2(1) = -22.50, p = .024$ alsmede tussen “gemiddeld HI, gemiddeld RC, gemiddeld PRC” en “gemiddeld HI, weinig RC, weinig PRC” $\chi^2(1) = -16.13, p = .041$.

Modererende factoren rondom feedbackvariabelen en het leereffect kunnen zelf-effectiviteit en het geslacht van de leerling zijn. Om deelvraag 4 te beantwoorden is de Quade nonparametrische ANCOVA uitgevoerd voor de twee groepen die bij deelvraag 3 significant van elkaar verschillen:

- De Quade nonparametrische ANCOVA kon niet uitgevoerd worden tussen de groepen “geen HI, weinig RC, gemiddeld PRC” en “gemiddeld HI, weinig RC, weinig PRC” omdat elke groep uit een enkele leerling bestaat.

Om dezelfde reden kon ook de modererende rol van geslacht niet onderzocht worden.

- De Quade nonparametrische ANCOVA toonde tussen de groepen “gemiddeld HI, gemiddeld RC, gemiddeld PRC” en “gemiddeld HI, weinig RC, weinig PRC” geen significant verschil met de totale zelf-effectiviteitsscore als covariaat $F = 0.879, p =$

.447. Ook met de afzonderlijke zelf-effectiviteit domeinen als covariaat werd er geen significant verschil gevonden: academische zelf-effectiviteit $F = 3.000$, $p = .225$, sociale zelf-effectiviteit $F = 2.168$, $p = .279$ en emotionele zelf-effectiviteit $F = 0.250$, $p = .667$.

De modererende rol van geslacht kon niet onderzocht worden, omdat alle leerlingen in de twee groepen meisjes waren.

Gezien het exploratieve karakter van dit onderzoek zijn de volgende vervolganalyses uitgevoerd. Kruskal-Wallis H tests toonden geen significante verschillen tussen het leerresultaat van jongens, meisjes en “wil ik niet zeggen” $\chi^2(2) = 1.32$, $p = .516$ maar wel tussen zelf-effectiviteit van jongens, meisjes en “wil ik niet zeggen”: academisch $\chi^2(2) = 6.84$, $p = .033$; sociaal $\chi^2(2) = 1.26$, $p = .532$; emotioneel $\chi^2(2) = 9.89$, $p = .007$; totaal $\chi^2(2) = 5.97$, $p = .051$. In tabel 8 paarsgewijze vergelijking.

Tabel 8

Paarsgewijze vergelijking zelf-effectiviteit en geslacht van leerlingen

Type zelf-effectiviteit	Academisch		Sociaal		Emotioneel		Totaal	
	χ^2	p	χ^2	p	χ^2	p	χ^2	p
Wil ik niet zeggen – jongen	18.21	.019	4.89	.529	13.79	.076	16.23	.037
Wil ik niet zeggen – meisje	19.76	.010	0.02	.998	0.16	.983	7.78	.309
Jongen - meisje	-1.55	.730	4.87	.838	13.63	.002	8.45	.061

Noot. Elk item in een zelf-effectiviteitsdomein is gescoord op een vijfpunts Likertschaal, met 1 = “lukt niet”, en 5 = “lukt erg goed.”

Jongens scoorden een gemiddelde emotionele zelf-effectiviteit van $M = 3.92$, $SD = 0.69$ en meisjes $M = 3.16$, $SD = 0.77$.

4. Discussie

Dit exploratieve, quasi experimentele onderzoek staat in het teken van de vraag ‘in hoeverre is er een verschil in effect van *elaborated feedback* op het leereffect van oefentoetsen in een online leeromgeving’. Hierna worden de conclusies van de vier deelvragen stapsgewijs besproken.

4.1 Effect examentrainingsoftware op het leerresultaat

Vooraf konden leerlingen niet toegewezen worden aan verschillende experimentele condities omdat leerlingen uit de controlegroep mogelijk benadeeld zouden worden (Kember, 2003). Maar zoals verwacht bleken op natuurlijke wijze toch twee groepen te zijn ontstaan: leerlingen die de examentrainingsoftware nooit gebruikt hebben in de meetperiode ($n = 25$) en leerlingen die de software wel gebruikt hebben ($n = 29$). Opvallend was dat alle leerlingen uit de wiskunde B klassen de software gebruikt hadden en in de havo wiskunde A klas maar de helft. In de vwo wiskunde A klas heeft slechts een derde van de leerlingen de software gebruikt. Vermoed wordt dat deze verschillen te maken hebben met verschillen tussen docenten. Omdat de examentrainingsoftware nog niet zo lang gebruikt wordt op de school, zou het kunnen dat sommige docenten de effectiviteit van het oefenen met voorbeeldopgaven niet kennen en de software daarom niet inzetten als leermiddel. Dunlosky et al. (2013) suggereren dat docenten het oefenen met vergelijkbare opgaven niet veel inzetten, omdat ze niet geïnformeerd worden over de effectiviteit van deze en andere leermethodes. Andere factoren kunnen zijn dat elke docent op een andere manier instructie heeft gegeven over het gebruik van de online leeromgeving of dat sommige docenten minder digitaal vaardig zijn (Alqurashi, 2016). In toekomstig onderzoek zouden deze factoren meegenomen kunnen worden. Een andere mogelijkheid is dat leerlingen van wiskunde B meer behoefte hebben aan het oefenen met vergelijkbare examenopgaven dan wiskunde A leerlingen. Wellicht hebben wiskunde A leerlingen alternatieve leermiddelen om zich voor te bereiden op een toets. In

vervolgonderzoek zou uitgevraagd kunnen worden of er gebruik gemaakt is van andere leermiddelen.

Het gemiddelde leerresultaat van leerlingen die de examentrainingsoftware niet gebruikt hebben was op T0 hoger dan van leerlingen die de examentrainingsoftware wel gebruikt hebben. Dat zou er op kunnen wijzen dat de leerlingen die de software wel gebruikt hebben meer gemotiveerd zijn om hun leerresultaat te verbeteren en de software te gebruiken. De leerresultaat verschillscore van die groep was positief ($M = 0.32$) terwijl het leerresultaat van leerlingen die de software niet gebruikt hebben achteruit is gegaan ($M = -0.13$), maar het verschil tussen de twee groepen bleek niet significant te zijn. Er kan dus niet geconcludeerd worden dat met het gebruik van de examentrainingsoftware en feedback een *testing effect* heeft plaatsgevonden. Dat resultaat is onverwacht aangezien *elaborated feedback* zeer effectief is bevonden in online leeromgevingen (Van der Kleij et al., 2015). Toch worden in de literatuur geen consistente positieve effecten gevonden van feedback op het leerresultaat wiskunde (Van der Kleij et al., 2015). Een verklaring voor het niet kunnen vinden van een effect in dit onderzoek zou de interventieperiode van twee weken kunnen zijn. Misschien is deze periode te kort geweest om een effect te hebben op het leerresultaat. Kang et al. (2007) toonden wel significante effecten aan met een interventieperiode van een aantal dagen, maar in hun onderzoek ging het om correctieve feedback. In dit onderzoek ging het om complexe, hogere orde vaardigheden en *elaborated feedback*. Een andere verklaring kan zijn dat leerlingen die de software niet gebruikt hebben andere leermethoden gebruikt hebben. In vervolgonderzoek zou die factor meegenomen kunnen worden om deze reden uit te sluiten.

Voorspelling voor het niet gebruiken van de examentrainingsoftware was dat leerlingen in die groep een lagere zelf-effectiviteitsscore zouden hebben. Opvallend is dat die leerlingen juist hogere zelf-effectiviteitsscores hadden, al bleek dit verschil niet significant te zijn: academisch $\chi^2(1) = 1.40, p = .236$; sociaal $\chi^2(1) = 0.74, p = .388$; emotioneel $\chi^2(1) =$

0.27, $p = .602$; totaal $\chi^2(1) = 0.91$, $p = .339$. Een mogelijke verklaring voor het niet vinden van een effect is dat de SEQ-C is gebruikt als vragenlijst. De gemeten academische, sociale en emotionele zelf-effectiviteit zijn misschien niet de variabelen die een effect hebben. In vervolgstudies zou de *math self-efficacy* gemeten kunnen worden, die een negatieve correlatie heeft met verveling. Verveling is een deactiverende emotie en is in het algemeen schadelijk voor het maken van huiswerk (Luo et al., 2016). Een andere verklaring zou kunnen zijn dat leerlingen geen behoefte hadden aan het oefenen in de online leeromgeving omdat ze andere leermethoden gebruikt hebben.

4.2 Verschillen tussen feedbackvorm en -frequentie

Uit frequentietabel 5 blijkt dat gedurende de gehele interventieperiode vooral EF-PRC is opgevraagd ($M = 33.16$) en in veel mindere mate EF-RC ($M = 13.23$) en EF-HI ($M = 7.82$). Verwacht werd dat er met name behoefte zou zijn aan EF-PRC én EF-HI, maar blijktbaar hebben leerlingen vooral behoefte om tussentijds te toetsen of ze goed op weg zijn door het tussenantwoord te controleren. Hints bevatten namelijk geen (tussen)antwoord. De gedachte dat leerlingen waarschijnlijk al snel willen weten of ze goed op weg zijn in het beantwoorden van de vraag lijkt daarom te kloppen. Wel moet opgemerkt worden dat de feedbackbehoefte van de leerling afhangt van de leerfase waarin de leerling zit. In de beginfase van het bestuderen van een nieuw onderwerp is veel ondersteuning nodig (waarschijnlijk dus ook meer EF-HI en EF-PRC behoefte), wat afneemt naarmate de kennis over het onderwerp toeneemt, het zogenaamde *scaffolding* (Jumaat & Tasir, 2014). Of het gebruik van EF-HI en EF-PRC afneemt en EF-RC toeneemt naarmate de onderzoeksperiode vordert, zou meegenomen kunnen worden in toekomstig onderzoek.

Wel moeten de resultaten met voorzichtigheid geïnterpreteerd worden gezien de grootte van de spreiding in de resultaten (zie tabel 5). De oorzaak van de grote spreiding ligt vermoedelijk in het kleine aantal respondenten.

In de examentrainingsoftware zijn drie vormen van *elaborated feedback* aanwezig. Aangezien leerlingen alle feedbackvormen konden gebruiken, zijn de verschillende vormen met elkaar vergeleken door elk mogelijk feedbackpatroon samen te stellen. Omdat er veel patronen mogelijk zijn (37) en het aantal gebruikers van de software laag (29), gaat het in de meeste gevallen om een vergelijking tussen twee leerlingen. De resultaten van deze paarsgewijze vergelijking van alle feedbackpatronen moeten dus met terughoudendheid geïnterpreteerd worden. Er werden twee significante verschillen aangetoond. Bij beide effecten was dezelfde leerling betrokken die op T0 een 1.3 scoorde en op T1 een 8.7, waarmee bediscussieerd kan worden of dit geen outlier is.

Uit de resultaten bleek tegen de verwachting in dat niet alle feedbackfrequenties het leereffect vergroten. Leerlingen die één of meerdere feedbackvormen weinig of vaak gebruikt hebben, vertoonden wel een verbetering in leerresultaat, maar gemiddelde gebruikers niet. Ook al zijn de verschillen tussen alle feedbackfrequenties niet significant, is het toch opvallend dat leerlingen die één of meerdere feedbackvormen gemiddeld gebruikt hebben, een verslechtering in leerresultaat vertoonden. Het opvragen van feedback zou de leerling kunnen weerhouden van een diepere verwerking tijdens het leren, wat retentie op langere termijn verslechterd (Pashler et al., 2005). Maar als dat het geval zou zijn, zou de verwachting zijn dat ook de groep ‘vaak’ een vergelijkbaar of nog slechter leerresultaat zou vertonen. Bij deze uitkomsten zou het helpen om te weten hoeveel opgaven de betreffende groepen geoefend hebben, maar bij nadere bestudering van het databestand bleek dat het aantal geoefende opgaven niet vastgesteld kon worden. Een andere verklaring kan zijn dat leerlingen de hint of het (tussen)antwoord wel aangeklikt hebben, maar niet gelezen hebben. Vervolgonderzoek met *eye tracking* zou hier uitsluitsel over kunnen geven.

Als de *startup* de diepere verwerking tijdens het leren zou willen bevorderen (om retentie op langere termijn te verbeteren) (Pashler et al., 2005)

4.3 Modererende effecten van zelf-effectiviteit en geslacht van de leerling

De laatste onderzoeksvraag was in hoeverre het leereffect van oefentoetsen gemodereerd wordt door zelf-effectiviteit en het geslacht van de leerling. Wanneer zelf-effectiviteit wordt meegenomen als covariaat bij de analyse van het effect van de feedbackfrequentie in de examentrainingsoftware op het leerresultaat, blijkt het eerder gevonden significante effect te zijn verdwenen. Hetzelfde geldt als de academische, sociale en emotionele zelf-effectiviteitsscore als covariaat worden meegenomen in de analyse. Dat betekent dat zelf-effectiviteit een modererend effect heeft. Maar ook dit resultaat moet met terughoudendheid geïnterpreteerd worden, omdat het significante resultaat bij deelvraag 3 al ter discussie staat. Door het lage aantal respondenten was het niet mogelijk om het modererende effect van geslacht van de leerling te onderzoeken.

Uit de verdere data-analyse kan geconcludeerd dat er volgens Kruskal-Wallis H tests significante verschillen zijn tussen jongens, meisjes en “wil ik niet zeggen” in zelf-effectiviteit. De significante verschillen waarbij “wil ik niet zeggen” betrokken is, kunnen niet verklaard worden. De betreffende leerlingen hebben namelijk wel een geslacht en het vergelijken met de groep jongens of meisjes zou betekenen dat ze worden vergeleken met hun eigen geslacht. Dat zou een onzinnige vergelijking zijn.

Verwacht werd dat jongens een hogere zelf-effectiviteit zouden hebben dan meisjes, omdat een overvloed aan studies dat heeft aangetoond (Meece & Agger, 2018). Voor de academische, sociale en totale zelf-effectiviteitsscore werd geen verschil gevonden, maar voor emotionele zelf-effectiviteit wel. Een verklaring voor het niet kunnen vinden van significante verschillen is waarschijnlijk het kleine aantal respondenten. Maar het zou ook kunnen liggen aan de gebruikte SEQ-C vragenlijst. Een veel gebruikte andere maat voor zelf-effectiviteit is de *math self-efficacy* (Luo et al., 2016).

Het gevonden significante verschil was een hogere emotionele zelf-effectiviteit van jongens ($M = 3.92$, $SD = 0.69$) ten opzichte van meisjes ($M = 3.16$, $SD = 0.77$). De mate van controle over eigen emoties, gedachten en gedrag (Muris, 2001) is dus groter voor jongens dan voor meisjes. Dit resultaat is ook in eerder onderzoek gevonden (Calandri et al., 2021) en vindt zijn oorsprong waarschijnlijk in stereotypering van genderrollen op het gebied van wiskunde (Meece & Agger, 2018).

4.4 Beperkingen van het onderzoek

Uit de poweranalyse bleek dat een steekproefgrootte van 303 deelnemers nodig was voor een betrouwbaar resultaat en het kunnen detecteren van verschillen tussen drie verschillende feedbackvormen, drie verschillende frequentiegroepen en vier verschillende wiskundeniveaus. Omdat het klantenbestand van de *startup* nog niet toereikend was om dat aantal te halen, is het onderzoek zes maanden uitgesteld. Het aantal benodigde deelnemers leek na die zes maanden haalbaar, maar uiteindelijk is de data van 54 participanten geanalyseerd wat betekent dat er een veel kleinere kans is dat effecten gevonden worden en dat generalisatie niet mogelijk is. Het onderzoek heeft dus ook niet kunnen bijdragen aan de ecologische validiteit van het *testing effect*. Ook was het door het beperkte aantal respondenten niet mogelijk om de effectiviteit van de verschillende feedbackvormen per wiskundeniveau te onderzoeken en was het modererende effect van het geslacht niet te bepalen.

Aangezien het een quasi-experiment betreft, zijn de bedreigingen op de interne betrouwbaarheid (*assignment bias*) groter dan bij een zuiver experiment (Creswell, 2014). De bedoeling was om dit effect te verkleinen door data van leerlingen van verschillende scholen te combineren. Helaas bleek maar één school mee te hebben gedaan, waardoor de interne betrouwbaarheid bedreigd is. Ook beperkt het de generaliseerbaarheid van de resultaten.

Het quasi-experimentele aspect was nodig omdat leerlingen die de examentrainingsoftware niet gebruiken mogelijk benadeeld zouden worden. Hierdoor was het effect tussen de verschillende feedbackvormen moeilijk te meten. In vervolgonderzoek zouden verschillende versies van de trainingssoftware gebruikt kunnen worden, waarbij in één versie alleen EF-HI beschikbaar is, in de andere alleen EF-RC etc. Het is dan ook mogelijk om effecten van omgevingsfactoren te controleren omdat randomisering toegepast kan worden (Creswell, 2014).

Het was de bedoeling om naast de feedbackfrequentie ook de feedbacktiming mee te nemen in het onderzoek. Bij nadere bestudering van het databestand bleek dat het referentiepunt van de timing niet eenduidig vastgesteld kon worden. In vervolgonderzoek zou meer gebruikersdata gelogd moeten worden in de examentrainingsoftware, waaronder de specifieke vraag die aangeklikt is en inactiviteit van de leerling. Het zou namelijk kunnen dat de leerling even wegloopt bij de computer en op een later moment weer verder gaat. Volgens het logbestand duurt het opvragen van de feedback dan erg lang, wat een misrepresentatie van de werkelijkheid is. Een andere oplossing kan zijn het timing effect te onderzoeken met *eye-tracking* methodes.

In dit onderzoek is de SEQ-C vragenlijst gebruikt om de zelf-effectiviteit te meten. Toekomstige studies kunnen andere manieren gebruiken om zelf-effectiviteit te operationaliseren, bijvoorbeeld middels *math self-efficacy* (Luo et al., 2016).

4.5 Wetenschappelijke en maatschappelijke relevantie

Dit exploratieve onderzoek heeft ondanks de onbetrouwbare resultaten op het gebied van effecten van *elaborated feedback* op het leerresultaat toch een aantal inzichten opgeleverd.

In eerder onderzoek naar de rol van feedback op het leerresultaat in een computerondersteunde leeromgeving gaat het in de meeste gevallen om feedback in de

breedste zin van het woord. In dit onderzoek is getracht meer inzicht te geven in verschillende feedbackvariabelen door onderscheid te maken tussen verschillende feedbackvormen volgens de rangschikking van Shute (2008). Interessant is dat leerlingen vooral behoefte hadden aan het controleren van het tussenantwoord (EF-PRC, $M = 33.16$) en in veel mindere mate aan het controleren van het eindantwoord en hints tussendoor (respectievelijk EF-RC, $M = 13.23$ en EF-HI, $M = 7.82$). Waarom deze verschillen bestaan en welke rol de leerfase van de leerling hierin speelt zou verder onderzocht kunnen worden.

Docenten die de examentrainingsoftware onlangs hadden aangeschaft waren benaderd mee te doen met het onderzoek. Het feit dat ze zelf tot aankoop van de software zijn overgegaan, suggereert dat ze de software ook willen gebruiken tijdens hun lessen. Maar uit de resultaten bleek dat een relatief grote groep leerlingen uit de wiskunde A klassen de examentrainingsoftware nooit gebruikt hebben. Dit zou er op kunnen wijzen dat de rol van de docent rondom het gebruik van de software zoals beschreven door Alqurashi (2016) wellicht groter is dan op voorhand gedacht. Een andere mogelijkheid is dat leerlingen van wiskunde B meer behoefte hebben aan het oefenen met vergelijkbare examenopgaven dan wiskunde A leerlingen. Vervolgonderzoek zou daar meer duidelijkheid over kunnen geven.

De resultaten van dit onderzoek zijn niet betrouwbaar genoeg om conclusies te trekken over het (niet) vinden van effecten. Wel kan de opzet van dit onderzoek gebruikt worden voor vervolgonderzoek. Belangrijk daarbij zijn het gebruik van verschillende versies van examentrainingsoftware en de werving van meer scholen die mee willen doen met het onderzoek. Effecten van feedbackvariabelen en modererende factoren zoals docenten, zelf-effectiviteit en geslacht kunnen dan met grotere betrouwbaarheid gepresenteerd worden en tegelijkertijd kunnen omgevingsfactoren gecontroleerd worden door randomisering van de participanten.

Tot slot kan dit onderzoek de bekendheid van het *testing effect* vergroten. Uit onderzoek blijkt keer op keer dat veel docenten nog onbekend zijn met de effectiviteit van hun gebruikte leermethodes (Dunlosky et al., 2013; Surma et al., 2018).

Referenties

- Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the Use of Tests: A Meta-Analysis of Practice Testing. *Review of Educational Research*, 87(3), 659–701. <https://doi.org/10.3102/0034654316689306>
- Alqurashi, E. (2016). Self-efficacy in online learning environments: A literature review. *Contemporary Issues in Education Research (CIER)*, 9(1), 45-52.
- Attali, Y., & Van der Kleij, F. (2017). Effects of feedback elaboration and feedback timing during computer-based practice in mathematics problem solving. *Computers & Education*, 110, 154-169.
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, 84(2), 191.
- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational researcher*, 13(6), 4-16.
- Calandri, E., Graziano, F., & Rollé, L. (2021). Social Media, Depressive Symptoms and Well-Being in Early Adolescence. The Moderating Role of Emotional Self-Efficacy and Gender. *Frontiers in Psychology*, 1871.
- Caprara, G. V., Fida, R., Vecchione, M., Del Bove, G., Vecchio, G. M., Barbaranelli, C., & Bandura, A. (2008). Longitudinal analysis of the role of perceived self-efficacy for self-regulated learning in academic continuance and achievement. *Journal of educational psychology*, 100(3), 525.
- Chun, B. A., & Heo, H. J. (2018). The effect of flipped learning on academic performance as an innovative method for overcoming ebbinghaus' forgetting curve. In *Proceedings of the 6th International Conference on Information and Education Technology* (pp. 56-60). DOI: <https://doi.org/10.1145/3178158.3178206>
- Cito, (z.d.). *Centrale examens vo. Cito*. Geraadpleegd op 15 oktober 2022, van

<https://www.cito.nl/onderwijs/voortgezet-onderwijs/centrale-examens-voortgezet-onderwijs>

- Creswell, J.W. (2014). *Educational research: planning, conducting, and evaluating quantitative and qualitative research (4e ed.)*. Harlow, Verenigd Koninkrijk: Pearson Education
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the public interest, 14*(1), 4-58.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods, 41*(4), 1149-1160.
- Glover, C., & Brown, E. (2006). Written feedback for students: too much, too detailed or too incomprehensible to be effective?. *Bioscience education, 7*(1), 1-16.
- Habibi, M., Tahmasian, K., & Ferrer-Wreder, L. (2014). Self-efficacy in Persian adolescents: Psychometric properties of a Persian version of the Self-Efficacy Questionnaire for Children (SEQ-C). *International Perspectives in Psychology, 3*(2), 93-105.
- Hair, J. F., Ringle, C. M., & Sarstedt, M. (2013). Partial least squares structural equation modeling: Rigorous applications, better results and higher acceptance. *Long range planning, 46*(1-2), 1-12.
- Hodges, C. B. (2008). Self-efficacy in the context of online learning environments: A review of the literature and directions for research. *Performance Improvement Quarterly, 20*(3-4), 7-25.
- Hwang, M. H., Choi, H. C., Lee, A., Culver, J. D., & Hutchison, B. (2016). The relationship

- between self-efficacy and academic achievement: A 5-year panel analysis. *The Asia-Pacific Education Researcher*, 25(1), 89-98.
- Jumaat, N. F., & Tasir, Z. (2014). Instructional scaffolding in online learning environment: A meta-analysis. In *2014 international conference on teaching and learning in computing and engineering* (pp. 74-77). IEEE.
- Kang, S. H., McDermott, K. B., & Roediger III, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European journal of cognitive psychology*, 19(4-5), 528-558.
- Kember, D. (2003). To Control or Not to Control: The question of whether experimental designs are appropriate for evaluating teaching innovations in higher education. *Assessment & Evaluation in Higher Education*, 28(1), 89–101.
<http://doi.org/10.1080/02602930301684>
- Kirschner, P.A., Neelen, M. (2015, 14 juli). Learning the smart way [Blog post]. Gevonden via <https://3starlearningexperiences.wordpress.com/2015/07/14/learning-the-smart-way-2/>
- Lam, C. F., DeRue, D. S., Karam, E. P., & Hollenbeck, J. R. (2011). The impact of feedback frequency on learning and task performance: Challenging the “more is better” assumption. *Organizational Behavior and Human Decision Processes*, 116(2), 217-228.
- Luo, W., Ng, P. T., Lee, K., & Aye, K. M. (2016). Self-efficacy, value, and achievement emotions as mediators between parenting practice and homework behavior: A control-value theory perspective. *Learning and Individual Differences*, 50, 275-282.
- Mayr, S., Erdfelder, E., Buchner, A., & Faul, F. (2007). A short tutorial of GPower. *Tutorials in quantitative methods for psychology*, 3(2), 51-59.
- Meece, J., & Agger, C. (2018). Achievement motivation in education. In *Oxford Research*

Encyclopedia of Education.

- Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of counseling psychology, 38*(1), 30.
- Muris, P. (2001). A brief questionnaire for measuring self-efficacy in youths. *Journal of Psychopathology and behavioral Assessment, 23*(3), 145-149.
- Murre, J. M., & Dros, J. (2015). Replication and analysis of Ebbinghaus' forgetting curve. *PloS one, 10*(7), <https://doi.org/10.1371/journal.pone.0120644>.
- Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multimedia learning. *Instructional design for multimedia learning, 181195*.
- Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 3.
- Rawson, K. A., Dunlosky, J., & Sciartelli, S. M. (2013). The power of successive relearning: Improving performance on course exams and long-term retention. *Educational Psychology Review, 25*(4), 523-548.
- Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140*(6), 1432–1463. <https://doi.org.ezproxy.elib11.ub.unimaas.nl/10.1037/a0037559>
- Shute, V. J. (2008). Focus on formative feedback. *Review of educational research, 78*(1), 153-189.
- Soderstrom, N. C., & Bjork, R. A. (2014). Testing facilitates the regulation of subsequent study time. *Journal of Memory and Language, 73*, 99-115.
- Suldo, S. M., & Huebner, E. S. (2006). Is extremely high life satisfaction during adolescence advantageous?. *Social indicators research, 78*(2), 179-203.

- Surma, T., Vanhoyweghen, K., Camp, G., & Kirschner, P. A. (2018). The coverage of distributed practice and retrieval practice in Flemish and Dutch teacher education textbooks. *Teaching and Teacher Education, 74*, 229-237.
- Van den Broek, G. S., Segers, E., Van Rijn, H., Takashima, A., & Verhoeven, L. (2019). Effects of elaborate feedback during practice tests: Costs and benefits of retrieval prompts. *Journal of Experimental Psychology: Applied, 25*(4), 588.
- Van der Kleij, F. M., Feskens, R. C., & Eggen, T. J. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of educational research, 85*(4), 475-511.
- Van Merriënboer, J. J., & Kirschner, P. A. (2018). Ten steps to complex learning: A systematic approach to four-component instructional design. New York, NY: Routledge.

Bijlage A**Korte vragenlijst voor het meten van zelf-effectiviteit van jongeren**

SEQ

Hoe goed lukt het je om

1. Je leraar te vragen je te helpen wanneer je je schoolwerk moeilijk vindt?
 lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed
2. Je mening te uiten als klasgenoten het niet met je eens zijn?
 lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed
3. Jezelf weer op te vrolijken als er iets vervelends gebeurd is?
 lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed
4. Te studeren als je ook nog andere leuke dingen te doen hebt?
 lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed
5. Weer rustig te worden als je heel erg bang bent?
 lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed
6. Vrienden te worden met andere leerlingen?
 lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed
7. Voor een proefwerk een hoofdstuk uit je hoofd te leren?
 lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed
8. Een praatje te maken met iemand die je niet goed kent?
 lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed
9. Te zorgen dat je niet zenuwachtig wordt?
 lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed
10. Elke dag je huiswerk helemaal af te hebben?
 lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed
11. Prettig samen te werken met andere leerlingen?
 lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed
12. Je gevoelens te beheersen?
 lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed
13. Elk lesuur goed op te letten?
 lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed
14. Er iets van te zeggen wanneer andere leerlingen iets doen dat je niet bevalt?

lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed

15. Jezelf op te peppen als je in de put zit?

lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed

16. Voor elk vak minimaal een voldoende te halen?

lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed

17. Een grappige gebeurtenis te vertellen aan een groepje leerlingen?

lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed

18. Tegen je vriend of vriendin te zeggen dat je je niet zo goed voelt?

lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed

19. Ervoor te zorgen dat je ouders tevreden zijn over je schoolprestaties?

lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed

20. Vrienden te blijven met andere leerlingen?

lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed

21. Niet te denken aan vervelende dingen?

lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed

22. Een proefwerk goed te maken?

lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed

23. Ervoor te zorgen dat je geen ruzie met andere kinderen krijgt?

lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed

24. Je geen zorgen te maken over dingen die kunnen gebeuren?

lukt niet lukt een beetje lukt nogal lukt goed lukt erg goed

Bijlage B**Non-parametrische paarsgewijze vergelijking van alle EF soorten**

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig. ^a
geen_HI_weinig_RC_gemiddeld_PRC- geen_HI_weinig_RC_weinig_PRC	-4.500	9.983	-.451	.652	1.000
geen_HI_weinig_RC_gemiddeld_PRC- gemiddeld_HI_weinig_RC_geen_PRC	-5.500	9.983	-.551	.582	1.000
geen_HI_weinig_RC_gemiddeld_PRC- gemiddeld_HI_gemiddeld_RC_gemiddeld_PRC	-6.375	7.892	-.808	.419	1.000
geen_HI_weinig_RC_gemiddeld_PRC- gemiddeld_HI_vaak_RC_gemiddeld_PRC	-7.000	9.983	-.701	.483	1.000
geen_HI_weinig_RC_gemiddeld_PRC- weinig_HI_weinig_RC_geen_PRC	-7.000	9.983	-.701	.483	1.000
geen_HI_weinig_RC_gemiddeld_PRC- gemiddeld_HI_gemiddeld_RC_weinig_PRC	-9.000	9.983	-.902	.367	1.000
geen_HI_weinig_RC_gemiddeld_PRC- vaak_HI_gemiddeld_RC_gemiddeld_PRC	-12.000	9.983	-1.202	.229	1.000
geen_HI_weinig_RC_gemiddeld_PRC- weinig_HI_gemiddeld) RC_vaak_PRC	-12.000	9.983	-1.202	.229	1.000
geen_HI_weinig_RC_gemiddeld_PRC- weinig_HI_geen_RC_geen_PRC	-13.250	8.645	-1.533	.125	1.000
geen_HI_weinig_RC_gemiddeld_PRC- weinig_HI_vaak_RC_geen_PRC	-14.000	9.983	-1.402	.161	1.000
geen_HI_weinig_RC_gemiddeld_PRC- vaak_HI_gemiddeld_RC_weinig_PRC	-14.250	8.645	-1.648	.099	1.000
geen_HI_weinig_RC_gemiddeld_PRC- vaak_HI_vaak_RC_vaak_PRC	-15.000	7.892	-1.901	.057	1.000
geen_HI_weinig_RC_gemiddeld_PRC- weinig_HI_weinig_RC_weinig_PRC	-15.000	8.645	-1.735	.083	1.000
geen_HI_weinig_RC_gemiddeld_PRC- gemiddeld_HI_weinig_RC_weinig_PRC	-22.500	9.983	-2.254	.024	1.000
geen_HI_weinig_RC_weinig_PRC- gemiddeld_HI_weinig_RC_geen_PRC	-1.000	9.983	-.100	.920	1.000
geen_HI_weinig_RC_weinig_PRC- gemiddeld_HI_gemiddeld_RC_gemiddeld_PRC	-1.875	7.892	-.238	.812	1.000
geen_HI_weinig_RC_weinig_PRC- gemiddeld_HI_vaak_RC_gemiddeld_PRC	-2.500	9.983	-.250	.802	1.000
geen_HI_weinig_RC_weinig_PRC- weinig_HI_weinig_RC_geen_PRC	-2.500	9.983	-.250	.802	1.000
geen_HI_weinig_RC_weinig_PRC- gemiddeld_HI_gemiddeld_RC_weinig_PRC	-4.500	9.983	-.451	.652	1.000
geen_HI_weinig_RC_weinig_PRC- vaak_HI_gemiddeld_RC_gemiddeld_PRC	-7.500	9.983	-.751	.452	1.000
geen_HI_weinig_RC_weinig_PRC- weinig_HI_gemiddeld) RC_vaak_PRC	-7.500	9.983	-.751	.452	1.000

geen_HI_weinig_RC_weinig_PRC- weinig_HI_geen_RC_geen_PRC	-8.750	8.645	-1.012	.311	1.000
geen_HI_weinig_RC_weinig_PRC- weinig_HI_vaak_RC_geen_PRC	-9.500	9.983	-.952	.341	1.000
geen_HI_weinig_RC_weinig_PRC- vaak_HI_gemiddeld_RC_weinig_PRC	-9.750	8.645	-1.128	.259	1.000
geen_HI_weinig_RC_weinig_PRC- vaak_HI_vaak_RC_vaak_PRC	-10.500	7.892	-1.330	.183	1.000
geen_HI_weinig_RC_weinig_PRC- weinig_HI_weinig_RC_weinig_PRC	-10.500	8.645	-1.215	.225	1.000
geen_HI_weinig_RC_weinig_PRC- gemiddeld_HI_weinig_RC_weinig_PRC	-18.000	9.983	-1.803	.071	1.000
gemiddeld_HI_weinig_RC_geen_PRC- gemiddeld_HI_gemiddeld_RC_gemiddeld_PRC	.875	7.892	.111	.912	1.000
gemiddeld_HI_weinig_RC_geen_PRC- gemiddeld_HI_vaak_RC_gemiddeld_PRC	1.500	9.983	.150	.881	1.000
gemiddeld_HI_weinig_RC_geen_PRC- weinig_HI_weinig_RC_geen_PRC	-1.500	9.983	-.150	.881	1.000
gemiddeld_HI_weinig_RC_geen_PRC- gemiddeld_HI_gemiddeld_RC_weinig_PRC	3.500	9.983	.351	.726	1.000
gemiddeld_HI_weinig_RC_geen_PRC- vaak_HI_gemiddeld_RC_gemiddeld_PRC	6.500	9.983	.651	.515	1.000
gemiddeld_HI_weinig_RC_geen_PRC- weinig_HI_gemiddeld) RC_vaak_PRC	-6.500	9.983	-.651	.515	1.000
gemiddeld_HI_weinig_RC_geen_PRC- weinig_HI_geen_RC_geen_PRC	-7.750	8.645	-.896	.370	1.000
gemiddeld_HI_weinig_RC_geen_PRC- weinig_HI_vaak_RC_geen_PRC	-8.500	9.983	-.851	.395	1.000
gemiddeld_HI_weinig_RC_geen_PRC- vaak_HI_gemiddeld_RC_weinig_PRC	8.750	8.645	1.012	.311	1.000
gemiddeld_HI_weinig_RC_geen_PRC- vaak_HI_vaak_RC_vaak_PRC	9.500	7.892	1.204	.229	1.000
gemiddeld_HI_weinig_RC_geen_PRC- weinig_HI_weinig_RC_weinig_PRC	-9.500	8.645	-1.099	.272	1.000
gemiddeld_HI_weinig_RC_geen_PRC- gemiddeld_HI_weinig_RC_weinig_PRC	-17.000	9.983	-1.703	.089	1.000
gemiddeld_HI_gemiddeld_RC_gemiddeld_PRC- weinig_HI_weinig_RC_geen_PRC	-.625	7.892	-.079	.937	1.000
gemiddeld_HI_gemiddeld_RC_gemiddeld_PRC- gemiddeld_HI_vaak_RC_gemiddeld_PRC	.625	7.892	.079	.937	1.000
gemiddeld_HI_gemiddeld_RC_gemiddeld_PRC- gemiddeld_HI_gemiddeld_RC_weinig_PRC	-2.625	7.892	-.333	.739	1.000
gemiddeld_HI_gemiddeld_RC_gemiddeld_PRC- weinig_HI_gemiddeld) RC_vaak_PRC	-5.625	7.892	-.713	.476	1.000
gemiddeld_HI_gemiddeld_RC_gemiddeld_PRC- vaak_HI_gemiddeld_RC_gemiddeld_PRC	5.625	7.892	.713	.476	1.000
gemiddeld_HI_gemiddeld_RC_gemiddeld_PRC- weinig_HI_geen_RC_geen_PRC	-6.875	6.113	-1.125	.261	1.000

gemiddeld_HI_gemiddeld_RC_gemiddeld_PRC- weinig_HI_vaak_RC_geen_PRC	-7.625	7.892	-.966	.334	1.000
gemiddeld_HI_gemiddeld_RC_gemiddeld_PRC- vaak_HI_gemiddeld_RC_weinig_PRC	7.875	6.113	1.288	.198	1.000
gemiddeld_HI_gemiddeld_RC_gemiddeld_PRC- weinig_HI_weinig_RC_weinig_PRC	-8.625	6.113	-1.411	.158	1.000
gemiddeld_HI_gemiddeld_RC_gemiddeld_PRC- vaak_HI_vaak_RC_vaak_PRC	8.625	4.991	1.728	.084	1.000
gemiddeld_HI_gemiddeld_RC_gemiddeld_PRC- gemiddeld_HI_weinig_RC_weinig_PRC	-16.125	7.892	-2.043	.041	1.000
weinig_HI_weinig_RC_geen_PRC- vaak_HI_vaak_RC_vaak_PRC	8.000	7.892	1.014	.311	1.000
gemiddeld_HI_vaak_RC_gemiddeld_PRC- weinig_HI_weinig_RC_geen_PRC	.000	9.983	.000	1.000	1.000
gemiddeld_HI_vaak_RC_gemiddeld_PRC- vaak_HI_gemiddeld_RC_gemiddeld_PRC	5.000	9.983	.501	.616	1.000
gemiddeld_HI_vaak_RC_gemiddeld_PRC- weinig_HI_gemiddeld) RC_vaak_PRC	-5.000	9.983	-.501	.616	1.000
gemiddeld_HI_vaak_RC_gemiddeld_PRC- weinig_HI_vaak_RC_geen_PRC	-7.000	9.983	-.701	.483	1.000
gemiddeld_HI_vaak_RC_gemiddeld_PRC- weinig_HI_weinig_RC_weinig_PRC	-8.000	8.645	-.925	.355	1.000
gemiddeld_HI_vaak_RC_gemiddeld_PRC- gemiddeld_HI_weinig_RC_weinig_PRC	-15.500	9.983	-1.553	.120	1.000
weinig_HI_weinig_RC_geen_PRC- gemiddeld_HI_gemiddeld_RC_weinig_PRC	2.000	9.983	.200	.841	1.000
weinig_HI_weinig_RC_geen_PRC- weinig_HI_weinig_RC_weinig_PRC	-8.000	8.645	-.925	.355	1.000
gemiddeld_HI_vaak_RC_gemiddeld_PRC- gemiddeld_HI_gemiddeld_RC_weinig_PRC	-2.000	9.983	-.200	.841	1.000
gemiddeld_HI_vaak_RC_gemiddeld_PRC- weinig_HI_geen_RC_geen_PRC	-6.250	8.645	-.723	.470	1.000
gemiddeld_HI_vaak_RC_gemiddeld_PRC- vaak_HI_gemiddeld_RC_weinig_PRC	7.250	8.645	.839	.402	1.000
weinig_HI_weinig_RC_geen_PRC- vaak_HI_gemiddeld_RC_gemiddeld_PRC	5.000	9.983	.501	.616	1.000
weinig_HI_weinig_RC_geen_PRC- vaak_HI_gemiddeld_RC_weinig_PRC	7.250	8.645	.839	.402	1.000
gemiddeld_HI_vaak_RC_gemiddeld_PRC- vaak_HI_vaak_RC_vaak_PRC	8.000	7.892	1.014	.311	1.000
weinig_HI_weinig_RC_geen_PRC- weinig_HI_gemiddeld) RC_vaak_PRC	5.000	9.983	.501	.616	1.000
weinig_HI_weinig_RC_geen_PRC- weinig_HI_geen_RC_geen_PRC	6.250	8.645	.723	.470	1.000
weinig_HI_weinig_RC_geen_PRC- weinig_HI_vaak_RC_geen_PRC	7.000	9.983	.701	.483	1.000
weinig_HI_weinig_RC_geen_PRC- gemiddeld_HI_weinig_RC_weinig_PRC	15.500	9.983	1.553	.120	1.000

gemiddeld_HI_gemiddeld_RC_weinig_PRC- weinig_HI_gemiddeld)_RC_vaak_PRC	-3.000	9.983	-.301	.764	1.000
gemiddeld_HI_gemiddeld_RC_weinig_PRC- vaak_HI_gemiddeld_RC_gemiddeld_PRC	3.000	9.983	.301	.764	1.000
gemiddeld_HI_gemiddeld_RC_weinig_PRC- weinig_HI_geen_RC_geen_PRC	-4.250	8.645	-.492	.623	1.000
gemiddeld_HI_gemiddeld_RC_weinig_PRC- weinig_HI_vaak_RC_geen_PRC	-5.000	9.983	-.501	.616	1.000
gemiddeld_HI_gemiddeld_RC_weinig_PRC- vaak_HI_gemiddeld_RC_weinig_PRC	5.250	8.645	.607	.544	1.000
gemiddeld_HI_gemiddeld_RC_weinig_PRC- weinig_HI_weinig_RC_weinig_PRC	-6.000	8.645	-.694	.488	1.000
gemiddeld_HI_gemiddeld_RC_weinig_PRC- vaak_HI_vaak_RC_vaak_PRC	6.000	7.892	.760	.447	1.000
gemiddeld_HI_gemiddeld_RC_weinig_PRC- gemiddeld_HI_weinig_RC_weinig_PRC	-13.500	9.983	-1.352	.176	1.000
vaak_HI_gemiddeld_RC_gemiddeld_PRC- weinig_HI_geen_RC_geen_PRC	-1.250	8.645	-.145	.885	1.000
vaak_HI_gemiddeld_RC_gemiddeld_PRC- weinig_HI_vaak_RC_geen_PRC	-2.000	9.983	-.200	.841	1.000
vaak_HI_gemiddeld_RC_gemiddeld_PRC- gemiddeld_HI_weinig_RC_weinig_PRC	-10.500	9.983	-1.052	.293	1.000
weinig_HI_gemiddeld)_RC_vaak_PRC- vaak_HI_vaak_RC_vaak_PRC	3.000	7.892	.380	.704	1.000
vaak_HI_gemiddeld_RC_gemiddeld_PRC- vaak_HI_gemiddeld_RC_weinig_PRC	-2.250	8.645	-.260	.795	1.000
vaak_HI_gemiddeld_RC_gemiddeld_PRC- vaak_HI_vaak_RC_vaak_PRC	3.000	7.892	.380	.704	1.000
vaak_HI_gemiddeld_RC_gemiddeld_PRC- weinig_HI_weinig_RC_weinig_PRC	-3.000	8.645	-.347	.729	1.000
weinig_HI_gemiddeld)_RC_vaak_PRC- weinig_HI_geen_RC_geen_PRC	1.250	8.645	.145	.885	1.000
weinig_HI_gemiddeld)_RC_vaak_PRC- weinig_HI_vaak_RC_geen_PRC	2.000	9.983	.200	.841	1.000
weinig_HI_gemiddeld)_RC_vaak_PRC- vaak_HI_gemiddeld_RC_weinig_PRC	2.250	8.645	.260	.795	1.000
weinig_HI_gemiddeld)_RC_vaak_PRC- gemiddeld_HI_weinig_RC_weinig_PRC	10.500	9.983	1.052	.293	1.000
vaak_HI_gemiddeld_RC_gemiddeld_PRC- weinig_HI_gemiddeld)_RC_vaak_PRC	.000	9.983	.000	1.000	1.000
weinig_HI_gemiddeld)_RC_vaak_PRC- weinig_HI_weinig_RC_weinig_PRC	-3.000	8.645	-.347	.729	1.000
weinig_HI_geen_RC_geen_PRC- weinig_HI_vaak_RC_geen_PRC	-.750	8.645	-.087	.931	1.000
weinig_HI_geen_RC_geen_PRC- vaak_HI_gemiddeld_RC_weinig_PRC	1.000	7.059	.142	.887	1.000
weinig_HI_geen_RC_geen_PRC- vaak_HI_vaak_RC_vaak_PRC	1.750	6.113	.286	.775	1.000

weinig_HI_geen_RC_geen_PRC- weinig_HI_weinig_RC_weinig_PRC	-1.750	7.059	-.248	.804	1.000
weinig_HI_geen_RC_geen_PRC- gemiddeld_HI_weinig_RC_weinig_PRC	9.250	8.645	1.070	.285	1.000
weinig_HI_vaak_RC_geen_PRC- vaak_HI_gemiddeld_RC_weinig_PRC	.250	8.645	.029	.977	1.000
weinig_HI_vaak_RC_geen_PRC- vaak_HI_vaak_RC_vaak_PRC	1.000	7.892	.127	.899	1.000
weinig_HI_vaak_RC_geen_PRC- weinig_HI_weinig_RC_weinig_PRC	-1.000	8.645	-.116	.908	1.000
weinig_HI_vaak_RC_geen_PRC- gemiddeld_HI_weinig_RC_weinig_PRC	8.500	9.983	.851	.395	1.000
vaak_HI_gemiddeld_RC_weinig_PRC- weinig_HI_weinig_RC_weinig_PRC	-.750	7.059	-.106	.915	1.000
vaak_HI_gemiddeld_RC_weinig_PRC- vaak_HI_vaak_RC_vaak_PRC	.750	6.113	.123	.902	1.000
vaak_HI_gemiddeld_RC_weinig_PRC- gemiddeld_HI_weinig_RC_weinig_PRC	-8.250	8.645	-.954	.340	1.000
vaak_HI_vaak_RC_vaak_PRC- weinig_HI_weinig_RC_weinig_PRC	.000	6.113	.000	1.000	1.000
vaak_HI_vaak_RC_vaak_PRC- gemiddeld_HI_weinig_RC_weinig_PRC	-7.500	7.892	-.950	.342	1.000
weinig_HI_weinig_RC_weinig_PRC- gemiddeld_HI_weinig_RC_weinig_PRC	7.500	8.645	.868	.386	1.000

Bijlage C

Voorbeeld vraagselectie door docent en voorbeeldvraag in examentrainingsoftware

Selectie vanuit docent

Je docent heeft de volgende opgaven voor je geselecteerd:

WAH5 T1

Geselecteerd op 5 april 2022 om 17:09

[Rookgedrag van leerlingen](#)

Wiskunde A Havo 2018-II

3/3 vragen | 10 punten | ⌚ 24 min.

[Veiligheidsmonitor](#)

Wiskunde A Havo 2021-II

1/1 vraag | 3 punten | ⌚ 7 min.

9

Vraag 9 3 punten

⌚ 7 min. | 🧠 Hoog

In tabel 1 zie je dat van de leerlingen in de steekproef 23%, bijna een kwart, rookt of ooit gerookt heeft. Op basis van bovenstaande gegevens kun je het 95%-betrouwbaarheidsinterval voor de lifetime-prevalentie van roken berekenen,

Bereken dit 95%-betrouwbaarheidsinterval. Rond de percentages in je antwoord af op gehele getallen.

💡 Werk je antwoord uit op papier. Heb je de vraag gemaakt? Kijk dan je antwoord na.
Kom je er niet uit? Ontvang dan een tip.

Kijk antwoord na

Ontvang een tip

Bijlage D**Voorbeeld EF-HI feedback in examentrainingsoftware**

Kom je door de tip wel tot het juiste antwoord? Dan krijg je alsnog de punten.

Tip: hoe pak je deze vraag aan?**Gegeven:**

- tabel 1: lifetime-prevalentie van roken met prevalentie gegevens en steekproefomvang
- steekproefomvang is 6714
- lifetime-prevalentie roken is 23%
- formuleblad met formules voor betrouwbaarheidsintervallen

Gevraagd:

Wat: 95%-betrouwbaarheidsinterval voor de lifetime-prevalentie van roken.

Hoe: rond de percentages af op gehele getallen.

Aanpak:

- bepalen van de steekproefproportie
- twee grenswaarden van het 95%-betrouwbaarheidsinterval



Kom je met deze tip wel tot het juiste antwoord? Dan krijg je alsnog de punten.

Ben je tot een antwoord gekomen?

Ja

Nee

Als het niet is gelukt, dan krijg je nog een tip.

Bijlage E**Voorbeeld EF-PRC feedback in examentrainingsoftware**

Rookgedrag van leerlingen

X

Kijk vraag 9 na

Controleer je tussenantwoorden en bereken je punten.

De tussenstappen**Tussenantwoord 1**

Het 95%-betrouwbaarheidsinterval voor de populatieproportie is:

$$0,23 \pm 2 \cdot \sqrt{\frac{0,23 \cdot (1-0,23)}{6714}}$$

Heb je dit tussenantwoord goed? Ja Nee**Tussenantwoord 2**Dit geeft het interval $[0,220; 0,240]$.Heb je dit tussenantwoord goed? Ja Nee**Tussenantwoord 3**Vermenigvuldigen met 100 voor de lifetime-prevalentie geeft $[22\%; 24\%]$.Heb je dit tussenantwoord goed? Ja Nee**De opgave**

Sinds de jaren tachtig meet het Trimbos-instituut regelmatig via een enquête het gebruik van alcohol, drugs en tabak in aselecte, representatieve steekproeven onder alle leerlingen van het

[Hele vraag weergeven](#)