

# Computergestuurd Oefenen: Wie Oefent en Waarom?

Chris Van Klaveren

Ilja Cornelisz

# 1 Introductie

In september 2014 is een wetenschappelijk onderzoek uitgevoerd om de effectiviteit van adaptief oefenen op leeruitkomsten te meten. In dit onderzoek hebben leerlingen een jaar lang digitaal geoefend met hetzij een statisch digitaal oefenprogramma, danwel een adaptief digitaal oefenprogramma. Het statische oefenprogramma genereerde voor elke leerling unieke digitale oefenboekjes bestaande uit vragen die in algemene zin varieerde in moeilijkheidsniveau en kennisvaardigheid. Het adaptieve oefenprogramma hield ook rekening met moeilijkheidsniveau en kennisvaardigheid, maar registreerde bovendien tijdens het oefenen welke onderwerpen leerlingen (relatief ten opzichte van de klas) minder goed beheerste en gaf vragen over deze onderwerpen met een relatief grotere kans terug aan de leerling. In het onderzoek werden leerlingen binnen klassen willekeurig toegewezen aan een van de twee oefenprogramma's. Omdat zowel leraren als leerlingen niet wisten aan welke groep zij werden toegewezen, en doordat beide digitale oefencondities er hetzelfde uitzagen, kon het effect van adaptief oefenen ten opzichte van statisch oefenen op een zuivere en betrouwbare manier worden bepaald. De onderzoeksresultaten geven gemiddeld genomen aan dat het adaptieve oefenprogramma geen significant effect had op leeruitkomsten. De leeruitkomsten die beschouwd werden waren de behaalde toetsscores, het aantal gemaakte oefenopgaven, oefentijd en oefenprestaties. De resultaten van het onderzoek en een precieze uiteenzetting van hoe het onderzoek werd uitgevoerd wordt beschreven in Van Klaveren, Cornelisz & Vonk (2015).<sup>1</sup>

Het voorliggende rapport is een aanvulling op het wetenschappelijke onderzoeksrapport. Het beschrijft in Hoofdstuk 2 hoe leeruitkomsten (behaalde toetsscores, gemaakte oefenopgaven, oefentijd, oefenprestaties) samenhangen met achtergrondkenmerken van de leerling en verschillen per school. Ook werden er vragenlijsten afgenomen na afloop van het experiment (in juni 2015) die tot doel hadden om meer inzicht te krijgen in hoe leerlingen het oefenen hebben ervaren. Deze ervaringen worden beschreven in Hoofdstuk 3. We merken op dat de vragenlijsten werden ingevuld door 33 van de 46 klassen die deelnamen aan het onderzoek. De respons in deze 33 klassen was 69 procent (d.w.z.

---

<sup>1</sup>Het wetenschappelijk onderzoeksrapport kan gevonden worden op <https://sites.google.com/site/chrispbjvanklaveren> onder working papers.

509 van de 739 leerlingen vulden de vragenlijsten in). Tenslotte, is het van belang om te benadrukken dat de resultaten in dit document puur beschrijvend zijn en gegeven verklaringen voor bepaalde verschillen kunnen dus in werkelijkheid anders zijn.

# Samenhang tussen leeruitkomsten en achtergrondkenmerken

De samenhang tussen oefen- en leeruitkomsten met achtergrondkenmerken wordt weergegeven in Tabel 1. De onderste rij van de tabel laat de gemiddelde leeruitkomsten zien voor leerlingen die gebruik hebben gemaakt van de digitale oefenmodule. Er deden 1021 leerlingen mee aan het experiment en voor 926 leerlingen geldt dat zij gedurende het experiment hebben geoefend. Dit komt neer op een respons van 90.1 procent. Gemiddeld genomen halen deze leerlingen een 6,3 op de gemaakte toetsen. Zij maken per oefensessie 32 oefenvragen, doen daar ongeveer 15-16 minuten over en beantwoorden 73 procent van deze vragen correct.

De statistieken die weergegeven worden onder toetscores geven de gemiddelde behaalde toetsscores weer, maar nu opgesplitst naar achtergrondkenmerken. De behaalde cijfers per leerjaar verschillen in de zin dat leerlingen in jaar 1 gemiddeld lagere cijfers halen dan leerlingen in jaar 2 en 3. Ook is er veel variatie in de behaalde cijfers voor de verschillende vakken, maar er moet opgemerkt worden dat leerlingen in de verschillende vakken niet gelijkmatig verdeeld zijn over de leerjaren. Leerlingen scoren dus, bijvoorbeeld, relatief hoge cijfers voor het vak economie maar dit zijn ook alle leerlingen in het derde leerjaar. Zodoende kan niet ontrafeld worden in hoeverre dit een “ vak-effect” is en/of een “leerjaar-effect” en blijft het onduidelijk hoe deze waargenomen verschillen in behaalde cijfers tussen vakken precies geïnterpreteerd moeten worden.

De overige achtergrondkenmerken die beschouwd worden in de tabel zijn opleidingsniveau van de moeder, of leerlingen oefenen met het statische of het dynamische programma en het geslacht van de leerling. De verschillen tussen deze kenmerken zijn beter interpreteerbaar omdat (1) we weten dat opleidingsniveau van de ouders positief zou moeten correleren met het behaalde cijfers, en (2) omdat de leerlingverdeling over het type oefenprogramma bij constructie gelijk verdeeld is binnen klassen (door willekeurige toewijzing) en er gemiddeld genomen evenveel meisjes als jongens zitten in een klas. De tabel laat inderdaad zien dat leerlingen hogere cijfers halen naarmate moeders hoger opgeleid zijn. Alleen het opleidingsniveau van de moeder wordt beschouwd omdat het opleidingsniveau van

beide ouders sterk samenhangt. We merken op dat een laag opleidingsniveau correspondeert met basis/vmbo als hoogste opleidingsniveau, de middencategorie correspondeert met havo/vwo/mbo als hoogst behaalde opleidingsniveau en de hoogste opleidingscategorie correspondeert met minimaal hbo of wo als hoogst behaalde opleidingsniveau. Tabel 1 laat zien dat het gemiddelde cijfer van leerlingen met middelbaar- en laag opgeleide moeders gelijk is en dat leerlingen met hoger opgeleide moeders iets meer dan een half punt hoger scoren. Leerlingen die oefenen met het adaptieve en statische oefenprogramma behalen gemiddeld genomen dezelfde cijfers. Er is weliswaar een klein verschil, maar dit verschil is niet significant. Vanzelfsprekend komt dit resultaat overeen met de bevindingen in de uitgevoerde effectstudie, namelijk, dat behaalde toetsscores niet verbeteren doordat er met een adaptief oefenprogramma geoefend wordt. Tenslotte zijn de behaalde cijfers van meisjes iets hoger dan de behaalde cijfers van jongens.

Omdat zowel het aantal gemaakte vragen per sessie, de tijd die daaraan besteed wordt en de oefenprestatie inzicht geven in het oefengedrag van leerlingen, bespreken we deze uitkomsten gezamenlijk in relatie tot de hierboven beschouwde achtergrondkenmerken. Leerlingen in het eerste leerjaar maken de meeste oefenopgaven (39) per sessie en doen daar gemiddeld genomen ongeveer 19 minuten over. Zij beantwoorden ongeveer 70 procent van de vragen correct. Leerlingen in het tweede jaar maken per oefensessie significant minder vragen (28) en doen daar vanzelfsprekend ook minder lang over (12 minuten). Hun oefenprestatie is met 71 procent ongeveer gelijk aan de oefenprestatie van leerlingen die in het eerste jaar zitten. Leerlingen in het derde jaar, dus voor het vak economie, oefenen het minst en maken ongeveer 22 vragen per oefensessie in 12.5 minuut. Gemiddeld genomen presteren zij wel beter tijdens de oefensessies door ongeveer 7 procent van de gemaakte vragen meer goed te beantwoorden (78 procent). De oefenverschillen tussen vakken zijn (opnieuw) moeilijk te duiden, maar wat opvalt is dat leerlingen veruit de meeste oefenvragen maken per oefensessie voor het vak Nederlands.

Ook voor de leeruitkomsten die direct verwant zijn aan het oefengedrag van leerlingen is het zo dat de achtergrondkenmerken gerelateerd aan het opleidingsniveau van de moeder, het type oefenprogramma en het geslacht van de leerling het beste geïnterpreteerd kunnen worden. De tabel laat zien dat oefengedrag niet samenhangt met het opleidingsniveau van de moeder: leerlingen maken ongeveer evenveel vragen, leerlingen van moeders met een middelbare opleiding doen er gemiddeld genomen iets langer over (maar dit verschil is niet significant) en ook de oefenprestaties van leerlingen is gelijk.

Het oefengedrag van leerlingen hangt wel samen met of de leerling werd toegewezen aan het statische of adaptieve oefenprogramma. Er lijkt een groot verschil te bestaan tussen het aantal gemaakte opgaven, maar dit verschil wordt voornamelijk veroorzaakt doordat

er een klein aantal leerlingen in de statische groep buitensporig veel vragen maakten. Het gevolg is dat er wel een gemiddeld verschil waarneembaar is in het aantal gemaakte oefenvragen, maar dit verschil is niet significant.

Leerlingen die oefenen met het adaptieve oefenprogramma presteren gemiddeld genomen minder goed tijdens het oefenen (69% t.o.v. 77%). Dit is niet verwonderlijk omdat zij juist ook relatief meer vragen ontvangen op onderwerpen waarop eerder is gebleken dat zij deze lastig vinden. Dat is ook de reden waarom leerlingen die adaptief oefenen meer tijd besteden aan het oefenen dan leerlingen die statisch oefenen. In Figuur 1 wordt de moeilijkheidsgraad van de geoefende opgaven die leerlingen ontvangen afgezet tegen de proportie leerlingen die adaptief hebben geoefend. De intuïtie van deze grafiek is eenvoudig: als leerlingen in beide oefenprogramma's dezelfde vragen ontvangen qua moeilijkheidsniveau dan is de verdeling 50-50 en zou de lijn horizontaal moeten lopen op 0.5 (zie de stippellijn). Er is echter sprake van een stijgende lijn wat aangeeft dat vooral leerlingen die adaptief oefenen moeilijkere vragen ontvangen. Het omslagpunt ligt bij de moeilijkheidsgraad van 0.28. Voor vragen van dit niveau geldt dat 28% van de overige vragen minder (of net zo) moeilijk zijn en 72% moeilijker. Links van deze grens is de proportie leerlingen die adaptief oefenen relatief laag en rechts van deze moeilijkheidsgraad is deze proportie juist relatief hoog. Gegeven dat de moeilijkheidsgraad van leerlingen die adaptief oefenen significant hoger is lijkt het verschil in oefenprestatie relatief klein te zijn. Met andere woorden, ondanks dat leerlingen die adaptief oefenen echt moeilijkere vragen krijgen, leidt dit niet zozeer tot een sterk verschil in goed/fout gemaakte vragen. Dat kan aangeven dat het adaptieve programma inderdaad oefenvragen retourneert die beter passen bij het niveau wat de leerling op dat moment aan kan.

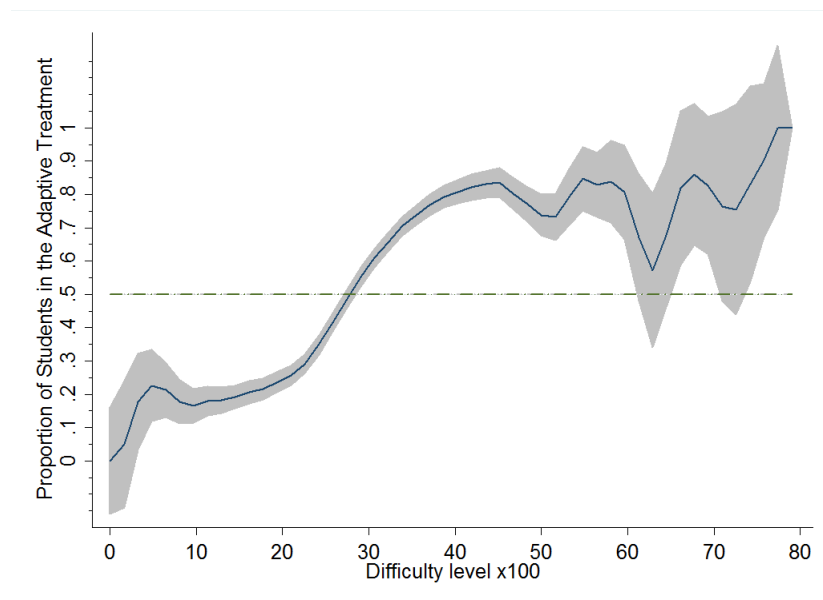
Om volledig te zijn merken we op dat het grijze vlak wat de geplotte lijn omringt de betrouwbaarheid weergeeft. Wanneer het grijze vlak smaller is geeft dit aan dat vragen met dit specifieke moeilijkheidsniveau relatief vaak werden aangeboden (en dus dat Figuur 1 voor deze gebieden betrouwbaarder is). De figuur laat zien dat de moeilijkheidsgraag van een geretourneerde vraag gemiddeld genomen ligt tussen de 0,1 en 0,5, en dat de betrouwbaarheid van de plot op deze gebieden relatief hoog is (d.w.z. het grijze omringende vlak relatief smal is).

TABEL I

	Toetsscores		# vragen	Oefentijd	Oefenprestatie
	N	Gem.	(per sessie)	(minuten)	Gem.
Leerjaar 1	491	5,87	39,00	18,65	0,70
Leerjaar 2	135	6,62	27,74	11,55	0,71
Leerjaar 3	300	6,78	22,32	12,48	0,78
Biologie	178	6,65	11,42	6,92	0,72
Economie	300	6,78	22,24	12,48	0,78
Geschiedenis	222	5,22	29,62	14,23	0,71
Nederlands	226	6,35	59,19	27,99	0,69
Opl. Moeder Laag	80	6,07	15,54	15,35	0,75
Opl. Moeder Middel	187	6,07	17,75	21,64	0,72
Opl. Moeder Hoog	265	6,62	15,10	14,37	0,75
Opl. Moeder Onbekend	315	6,00	16,05	16,47	0,71
Statisch	461	6,33	36,29*	15,60	0,77
Adaptief	465	6,21	27,66	14,29	0,69
Meisje	422	6,34	30,32	15,86	0,73
Jongen	433	6,11	36,17	15,05	0,73
Medianen	926	6,27	31,95	15,62	0,73

\*Verschil tussen statische en adaptieve groep is niet significant is.

Figuur 1: Proportie leerlingen die adaptief oefenen ten opzichte van moeilijkheidsniveau van de geoefende vragen



# Leerlingervaringen met digitaal oefenen

Dit hoofdstuk presenteert informatie die verkregen werd door leerlingen aan het eind van het schooljaar een vragenlijst te laten invullen over hoe zij het digitaal oefenen hebben ervaren. Deze vragenlijst werd afgenomen in 33 van de 46 deelnemende klassen. Een totaal van 509 van de 739 leerlingen vulden de vragenlijsten in zodat de respons in de 33 klassen ongeveer 69 procent was. Een totaal van 13 klassen vulden de vragenlijsten niet in en dit kwam waarschijnlijk doordat de onderzoekers de vragenlijsten (te) kort voor de zomervakantie aanboden.<sup>2</sup> Er zijn twee verschillende digitale oefenprogramma's aangeboden (adaptief en statisch), die er voor leerlingen in principe hetzelfde uitzien. Omdat de ervaringen met beide oefenprogramma's kunnen verschillen, tonen we in dit hoofdstuk telkens de ervaringen voor beide groepen apart. Ook geven we aan wat het verschil is in oefenervaring tussen beide groepen en laten we zien of het verschil in ervaring significant is. Met significant bedoelen we dat we met grote zekerheid kunnen vaststellen dat er daadwerkelijk een verschil is, en dit verschil dus niet berust op toeval.

Alvorens de ervaringen van leerlingen uiteengezet worden is het belangrijk om na te gaan in hoeverre de non-respons van leerlingen die statisch oefenen overeenkomt met de non-respons van leerlingen die adaptief oefenen. Wanneer leerlingen beide oefenprogramma's anders ervaren kan dit namelijk zijn weerslag hebben op de mate van respons op de enquête, waardoor de ervaringsuitingen tussen beide leerlinggroepen niet meer goed vergelijkbaar zijn. Tabel II laat zien dat de non-respons van leerlingen in beide oefencondities niet significant verschilt. Dat betekent dat selectieve respons geen versturende rol speelt, wanneer ervaringen van leerlingen tussen beide oefencondities worden vergeleken.

In de post-experimentele vragenlijst werd aan studenten verteld dat zij gedurende het experiment hebben geoefend met hetzij een statisch, dan wel een adaptief oefenprogramma. De tekst in de vragenlijst luidde letterlijk: "*Wat je misschien niet wist is dat er twee digitale oefenprogramma's waren en dat jij telkens met een van de twee oe-*

---

<sup>2</sup>We merken op dat de gevonden effecten van adaptief versus statisch oefenen voor deze 33 klassen gelijk waren aan die van de 46 klassen.



Tabel II

	Statisch (N = 371)		Adaptief (N = 368)		Verschil
	Gem.	Std. dev.	Gem.	Std. dev.	
Non-respons	0.313	0.024	0.310	0.024	0.003

*fenprogramma's hebt gewerkt. Beide programma's bepaalden op een andere manier welke vraag voor jou het beste was om op dat moment te maken".*

- **Versie 1** van het programma geeft vragen van alle onderwerpen en zowel makkelijke als moeilijke vragen, waarbij het niet uitmaakt of je eerdere vragen over dat onderwerp goed hebt gemaakt.
- **Versie 2** van het programma geeft vooral vragen over onderwerpen die je minder goed begrijpt en die qua moeilijkheid aansluiten bij het niveau wat je op dat moment hebt.

De eerste rij in Tabel III geeft de proportie leerlingen weer die aangeven dat zij geoefend hebben met het statische oefenprogramma. Belangrijk is dat ongeveer 65 procent van de leerlingen in beide oefencondities aangeven dat zij denken geoefend te hebben met het statische oefenprogramma. Dit is een interessant resultaat, omdat het duidelijk weergeeft (1) dat studenten niet wisten aan welke oefenconditie zij waren toegewezen bij het begin van het experiment, en (2) dat zij hier gedurende het experiment ook niet achter zijn gekomen. Dat de meerderheid van de leerlingen denkt dat zij aan de statische oefenconditie zijn toegewezen geeft aan dat beide oefencondities niet als adaptief werden ervaren.

De overige statistieken in de tabel geven informatie over of leerlingen digitaal oefenen leuk vinden, of zij oefende in de week waarin de stof behandeld werd en of zij oefenen ter voorbereiding op de toets. Een eerste opvallend gegeven is dat de ervaringen van leerlingen in de statische groep wederom vergelijkbaar zijn met de ervaringen die leerlingen hebben in de adaptieve groep. Dat betekent dat de statistieken die in de tabel weergegeven worden informatie geven over hoe leerlingen digitaal oefenen in het algemeen ervaren. Een tweede opvallend gegeven is dat de antwoorden die leerlingen geven sterk verdeeld zijn over de verschillende antwoordcategorieën, wat aangeeft dat ervaringen van leerlingen sterk uiteenlopen. Er kan niet gesteld worden dat leerlingen gemiddeld genomen digitaal oefenen leuk vinden, maar wel is het zo dat het percentage leerlingen die digitaal oefenen leuk vindt veel groter is dan het percentage leerlingen die digitaal oefenen niet leuk vindt. Iets minder dan 60 procent van de leerlingen geeft aan vaak of altijd te oefenen in de

TABEL III

		Statisch (N=255)		Adaptief (N=254)		
		Gem.	Std. dev.	Gem.	Std. dev.	Verschil
Toegewezen aan statisch?		0.639	0.030	0.685	0.029	0.046
Digitaal oefenen is leuk	Nee	0.278	0.028	0.244	0.027	0.034
	Neutraal	0.361	0.030	0.406	0.031	0.045
	Ja	0.361	0.030	0.350	0.030	0.010
Oefent in lesstof-week	Soms	0.431	0.031	0.504	0.031	0.073
	Vaak	0.255	0.027	0.240	0.027	0.015
	Altijd	0.314	0.029	0.256	0.027	0.058
Oefent ter voorbereiding toets	Soms	0.612	0.031	0.610	0.031	0.002
	Vaak	0.239	0.027	0.248	0.027	0.009
	Altijd	0.145	0.022	0.138	0.022	0.007

week waarin de lesstof werd aangeboden. Een groot percentage leerlingen (43.1%) stelt eveneens dat zij slechts soms oefenen in de week waarin de lesstof aangeboden werd. Een scenario zou kunnen zijn dat leerlingen vooral gericht oefenen ter voorbereiding op de toets een juist vlak voor een toets oefenen. Echter, de laatste vraag in Tabel III biedt slechts gedeeltelijke ondersteuning voor deze stelling, daar iets minder dan 40 procent van de leerlingen aangeeft vaak of altijd te oefenen ter voorbereiding op een toets.

Table IV geeft informatie over of digitaal oefenen door leerlingen wordt ervaren als een middel om de stof beter te begrijpen (kwaliteit), als een middel om de stof sneller te snappen (efficiëntie) en daarbij of de vragen die aangeboden worden als te moeilijk worden ervaren. De verwachting was dat adaptief oefenen beter aan zou sluiten bij de leerbehoeften van leerlingen en dat, zodoende, adaptief oefenen leerlingen helpt om de stof beter en sneller te begrijpen. Echter, ervaringen van leerlingen in de statische groep blijken wederom niet significant te verschillen van de ervaringen van leerlingen in de adaptieve groep, en dus komen de statistieken niet overeen met de voorafgestelde verwachtingen. Een andere interessante uitkomst is dat studenten in zowel de statische als de adaptieve oefenconditie de vragen als even moeilijk lijken te ervaren, ondanks de eerder besproken resultaten waaruit blijkt dat leerlingen in de adaptieve oefenconditie gemiddeld genomen moeilijkere vragen ontvangen. In dat opzicht lijkt de adaptieve oefenconditie wel goed te werken, omdat het vragen aan zou moeten bieden die beter aansluiten bij het moeilijkheidsniveau wat de leerling aankan. Iets minder dan 30 procent van de leerlingen geeft aan dat digitaal oefenen leerlingen helpt om de stof sneller en/of beter te begrijpen, terwijl rond de 40 procent van de leerlingen juist aangeeft dat digitaal oefenen niet helpt om de stof sneller of beter te begrijpen. De enqueteresultaten lijken er verder op te duiden

TABEL IV

		Statisch (N=255)		Adaptief (N=254)		
		Gem.	Std. dev.	Gem.	Std. dev.	Verschil
DO*: ik snap het beter.	Oneens	0.282	0.028	0.299	0.029	0.017
	Neutraal	0.412	0.031	0.402	0.031	0.010
	Eens	0.298	0.029	0.299	0.029	0.001
DO: ik snap het sneller	Oneens	0.322	0.029	0.331	0.030	0.009
	Neutraal	0.392	0.031	0.417	0.031	0.025
	Eens	0.278	0.028	0.248	0.027	0.030
DO: is te moeilijk	Oneens	0.435	0.031	0.390	0.031	0.046
	Neutraal	0.420	0.031	0.472	0.031	0.053
	Eens	0.133	0.021	0.134	0.021	0.001

\*DO: Digitaal Oefenen

dat een deel van de leerlingen aangeeft baat te hebben bij het oefenen. Wanneer we echter achtergrondkenmerken van de leerlingen en ouders koppelen aan de gegeven antwoorden, dan blijkt dat de achtergrondkenmerken van leerlingen die aangeven baat te hebben bij digitaal oefenen niet afwijken van de achtergrondkenmerken van leerlingen die aangeven geen baat te hebben bij het digitaal oefenen. De groep die aangeeft geholpen te worden door de digitale oefenmodule kan op grond hiervan dus niet nader gespecificeerd worden.