

Klaar voor de toekomst?

**EEN LONGITUDINAAL ONDERZOEK
NAAR DE ONTWIKKELING VAN
21^E-EEUWSE VAARDIGHEDEN IN
HET PRIMAIR ONDERWIJS**

Ard Lazonder, Hannie Gijlers,
Noortje Janssen en Amber Walraven



Klaar voor de toekomst?

EEN LONGITUDINAAL ONDERZOEK NAAR DE ONTWIKKELING VAN 21^E-EEUWSE VAARDIGHEDEN IN HET PRIMAIR ONDERWIJS

Ard Lazonder, Hannie Gijlers, Noortje
Janssen en Amber Walraven

Nijmegen / Enschede
Augustus 2018

COLOFON

Auteurs: Ard Lazonder (Radboud Universiteit),
Hannie Gijlers, Noortje Janssen (Universiteit Twente) en
Amber Walraven (Radboud Universiteit) /
Vormgeving: Bas van der Horst, BUREAUBAS /
Productie: Rosenmullers Communicatie & Organisatie

**Dit werk valt onder een Creative Commons
Naamsvermelding-NietCommercieel-
GeenAfgeleideWerken 4.0 Internationaal licentie.**

Bezoek <http://creativecommons.org/licenses/by-nc-sa/4.0/>
om een kopie te zien van de licentie of stuur een brief naar
Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.



*Dit onderzoek is gefinancierd door het Nationaal Regieorgaan Onderwijsonderzoek
(projectnummer 405-15-546)*

NR  **NATIONAAL REGIEORGAAN
ONDERWIJSONDERZOEK**

Inhoud

VOORWOORD

4

KORT EN BONDIG

Samenvatting

7

EENVOUDIG EN COMPLEET

1.	Inleiding	10
2.	Deelnemers aan het onderzoek	13
3.	Kritisch denken	15
4.	Digitale geletterdheid	19
5.	Samenwerken	23
6.	Taal en rekenen	27
7.	De vaardigheidsgroei in samenhang	30
8.	Conclusies en aanbevelingen	33

VERDIEPING EN VERANTWOORDING

1.	Probleemoplossend vermogen	37
2.	Kritisch denken: keuze voor het soort test	39
3.	Kritisch denken: ontwerp van de test	41
4.	Statistische analyses van de scores voor kritisch denken	43
5.	Digitale geletterdheid: inhoud van de testen	45
6.	Digitale geletterdheid: keuze voor het soort test	47
7.	Digitale geletterdheid: ontwerp van de testen	49
8.	Statistische analyses van de scores voor digitale geletterdheid	54
9.	Samenwerken: ontwerp van de test	57
10.	Statistische analyses van de scores voor samenwerken	60
11.	Statistische analyses van de scores voor taal en rekenen	62
12.	De onderlinge samenhang tussen de resultaten	64
13.	Geraadpleegde literatuur	68

KLAAR VOOR DE TOEKOMST?

Een longitudinaal onderzoek naar de ontwikkeling
van 21st-eeuwse vaardigheden in het primair onderwijs

3

Voorwoord

In mei 2014 publiceerde De Volkskrant een paginagroot artikel over basisschool Het Talent uit Lent. Deze school werkt volgens een innovatief onderwijsconcept waarin veel aandacht is voor 21e-eeuwse vaardigheden zoals kritisch denken, digitale geletterdheid en samenwerken. Directeur Carla van den Bosch zou graag zien dat het effect van dit soort nieuwe initiatieven wordt gemeten, en dan vooral waar het de 21e-eeuwse vaardigheden betreft. Want: “pas als we de resultaten langdurig bijhouden, kunnen we beoordelen of dit echt een goede manier van onderwijs is.”

Deze oproep vormde de aanleiding voor het langlopend praktijkgericht NRO-project TALENTontwikkeling. Het project bestond uit twee onderdelen: een longitudinale meting van 21e-eeuwse vaardigheden bij de leerlingen van Het Talent en de ontwikkeling van een model om de leerkrachten te begeleiden bij het onderzoeken van hun eigen onderwijs. Dit rapport beschrijft de resultaten van het leerlingonderzoek; over het onderzoek met de leerkrachten is een apart verslag verschenen.¹

Bij het schrijven van dit eindrapport hebben we geprobeerd zoveel mogelijk rekening te houden met de verschillende doelgroepen waarvoor ons onderzoek relevant zou kunnen zijn. Schoolleiders, leerkrachten, onderzoekers en beleidsmakers kijken nu eenmaal ‘door een andere bril’ naar een onderzoeksverslag. Daarom heeft dit rapport een gelaagde opbouw.

- Het deel **kort en bondig** is bedoeld voor lezers die snel een indruk willen krijgen van de belangrijkste uitkomsten van het onderzoek.
- Het deel **eenvoudig en compleet** is vooral geschikt voor lezers die iets meer uitleg willen maar voor wie een volledige wetenschappelijke verantwoording net

iets te ver gaat. Voor lezers die hier wel in geïnteresseerd zijn, staat op diverse plaatsen in de tekst een verwijzing naar het derde deel.

- Het deel **verdieping en verantwoording** bevat een gedetailleerde beschrijving van de belangrijkste onderdelen van het onderzoek. Hier vindt u bijvoorbeeld een verslag van het ontwerp van de meetinstrumenten en de volledige statistische analyses.

Dit project had niet uitgevoerd kunnen worden zonder de medewerking van de leerkrachten en leerlingen van Het Talent. Drie jaar lang hebben wij het onderwijsprogramma in de maanden januari tot april een beetje ontwricht door leerlingen uit de les te halen voor onze longitudinale meting. Dit was nooit een probleem: de leerkrachten waren flexibel en de leerlingen vonden het erg leuk om aan onze ‘opdrachtjes’ te werken.

Een speciaal woord van dank gaat uit naar onze vaste aanspreekpunten Carla van den Bosch, Karin Kruitwagen, Moniek Minkhorst en Elleke Muller, die met ons meedachten over de praktische uitvoering van het onderzoek en ons veel regelwerk uit handen hebben genomen. Merel Boon en Marloes Warnar hebben geholpen bij het afnemen en scoren van de testen; dames, heel veel dank hiervoor! Tot slot willen we Erika Schlatter, promovendus aan de Radboud Universiteit, bedanken voor het verzamelen van de gegevens met de Toren van Hanoi taak.

**Ard Lazonder, Hannie Gijlers,
Noortje Janssen en Amber Walraven**

Augustus 2018

¹ Janssen, N., Walraven, A., Lazonder, A., & Gijlers, H. (2018). Meesterschap: Een vier-fasen model voor de begeleiding van praktijkgericht onderzoek. Verkrijgbaar www.nro.nl/nro-projecten-vinden/?projectid=405-15-546-de-effectiviteit-van-het-pdks-onderwijsconcept

Over de auteurs

Ard Lazonder is hoogleraar onderwijswetenschappen aan de Radboud Universiteit. In de eerste twee jaar van het project was hij als adjunct-hoogleraar verbonden aan de vakgroep Instructietechnologie van de Universiteit Twente. Twee van zijn medeauteurs komen eveneens uit deze vakgroep. **Hannie Gijlers** werkt daar als universitair docent en doet onderzoek naar samenwerkend leren met behulp van ICT. **Noortje Janssen** promoveerde bij de vakgroep en was er voor dit project aangesteld als medewerker onderzoek. Zij heeft ruime ervaring met praktijkgericht onderzoek en bestudeert daarin voornamelijk de professionele ontwikkeling van docenten. **Amber Walraven** tenslotte, promoveerde op een onderzoek naar digitale geletterdheid in het voortgezet onderwijs en werkt tegenwoordig als universitair docent bij de Radboud Docenten Academie.

Klaar
voor
de
toekomst

Kort en
bondig

Samenvatting

1.1 Doel van het onderzoek

Basisschool Het Talent is een integrale ontwikkelschool die werkt vanuit een Pedagogische Didactische Kern Structuur (PDKS). Vanaf de oprichting in 2001 wil deze school leerlingen op hun eigen tempo en niveau laten werken én extra aandacht besteden aan 21^e-eeuwse vaardigheden. De schoolleiding en leerkrachten wilden graag weten 'of hun aanpak werkt' en dan vooral wat betreft de ontwikkeling van 21^e-eeuwse vaardigheden. Dit verzoek is vertaald in de volgende twee onderzoeksvragen:

- In welke mate realiseert het onderwijs-concept van Het Talent de ontwikkeling van 21^e-eeuwse vaardigheden én basisvaardigheden op het gebied van taal en rekenen?
- Hoe wordt deze ontwikkeling beïnvloed door leerlingkenmerken?

1.2 Onderzoeksopzet en onderzoeksgroep

Om deze onderzoeksvragen te beantwoorden zijn de bovenbouwleerlingen van Het Talent twee jaar lang gevolgd om hun ontwikkeling van 21^e-eeuwse vaardigheden in kaart te brengen en te vergelijken met de ontwikkeling van taal- en rekenvaardigheden. Bij de start van het onderzoek waren deze leerlingen (100 jongens en 82 meisjes) tussen de 6 en 10 jaar oud. In de eerste maanden van 2016 zijn hun vaardigheden voor de eerste keer gemeten. In jaren daarna werden deze metingen herhaald, steeds in de periode tussen januari en maart. De uitval bleef beperkt tot 2 leerlingen in 2017 en 14 leerlingen in 2018.

1.3 Meetinstrumenten

In samenspraak met Het Talent is besloten om drie 21^e-eeuwse vaardigheden te onderzoeken: kritisch denken, digitale geletterdheid en

samenwerken. Bij *kritisch denken* is gekeken naar het vermogen van de leerlingen om te induceren, interpreteren en evalueren. Deze vaardigheden zijn gemeten aan de hand van de 'proefjes' die leerlingen normaal gesproken uitvoeren tijdens de lessen wetenschap en technologie (W&T). Deze proefjes werden begeleid door een onderzoeker die de leerlingen opdrachten gaf, vragen stelde en hun reacties noteerde. Voor *digitale geletterdheid* zijn drie verschillende testen ontwikkeld. De onderdelen 'informatie zoeken' en 'informatie veilig gebruiken' zijn gemeten met een geschreven test waarin de leerlingen hun antwoorden op geprinte schermafbeeldingen konden schrijven of aankruisen. Voor de overige twee onderdelen zijn praktische opdrachten ontwikkeld. Zo moesten de leerlingen voor het onderdeel 'informatie bewerken' de opmaak van een Word bestand veranderen, en bij het onderdeel 'informatie creëren' een PowerPoint presentatie maken. *Samenwerken* gebeurde in teams van drie of vier leerlingen die gezamenlijk aan een ontwerptaak werkten, bijvoorbeeld om een plan voor een nieuwe schooltuin te bedenken. Hun dialogen zijn op video opgenomen en achteraf geanalyseerd om de kwaliteit van de samenwerking te beoordelen.

De vorderingen op het gebied van taal en rekenen zijn gemeten met methode-onafhankelijke voortgangstoetsen. Voor *taalvaardigheid* is gebruik gemaakt van de Cito voortgangstoetsen voor begrijpend lezen. *Rekenvaardigheid* is gemeten met de SchoolVaardigheidsToets Hoofdrekenen (SVT-HR). Voor beide testen zijn normscores beschikbaar, zodat de prestaties en vaardigheidsgroei van de leerlingen van Het Talent konden worden vergeleken met die van een landelijke referentiegroep.

1.4 Resultaten

Bij elke vaardigheid is eerst gekeken of de leerlingen hun prestaties op de testen

jaarlijks wisten te verbeteren. Vervolgens is de vaardigheidsgroei onderling vergeleken om te bepalen of verschillende vaardigheden zich op hetzelfde moment en in gelijke mate ontwikkelen. De belangrijkste resultaten worden hieronder puntsgewijs samengevat.

- De bovenbouwleerlingen van Het Talent zijn bovengemiddeld goed in taal en rekenen en ontwikkelen zich grotendeels in hetzelfde tempo als hun leeftijdsgenootjes van andere scholen.
- De bovenbouwleerlingen van Het Talent ontwikkelen langzaam maar zeker hun 21^e-eeuwse vaardigheden.
 - Deze ontwikkeling is onafhankelijk van de vaardigheidsgroei in taal en rekenen.
 - Deze ontwikkeling gaat met ‘sprongetjes’ die niet voor elke leerling op hetzelfde moment plaatsvinden (maar onafhankelijk zijn van leeftijd).
- De ontwikkeling van vaardigheden in het zoeken van informatie op Internet gaat sneller dan de ontwikkeling van alle andere vaardigheden.
- De prestaties en vaardigheidsgroei zijn onafhankelijk van leerlingkenmerken.

1.5 Conclusies en aanbevelingen

Op basisschool Het Talent ontwikkelen de bovenbouwleerlingen hun 21^e-eeuwse vaardigheden in hetzelfde tempo als hun basisvaardigheden in taal en rekenen. En ze verbeteren hun taal- en rekenvaardigheden jaarlijks even veel als leerlingen van reguliere basisscholen. De verbeteringen van de afzonderlijke vaardigheden gaan echter niet hand-in-hand: leerlingen die bijvoorbeeld veel vooruit gingen in 21^e-eeuwse vaardigheden lieten niet altijd een even grote verbetering in taal- en rekenvaardigheden zien. Uit deze resultaten blijkt dat het PDKS-onderwijsconcept leerlingen in staat stelt

om zich in verschillende vaardigheden te bekwamen zonder dat de ontwikkeling van de ene vaardigheid die van andere vaardigheden in de weg staat.

In de onderzoeksgroep liepen zowel de jaarlijkse prestaties als de vaardigheidsgroei behoorlijk uiteen. Deze individuele verschillen hingen niet samen met algemene leerlingkenmerken zoals leeftijd, geslacht, achtergrond, en sociaaleconomische status.

Op basis van deze resultaten én onze ervaringen tijdens het project worden de volgende aanbevelingen gedaan.

- Leerlingen simpelweg de ruimte bieden om 21^e-eeuwse vaardigheden te gebruiken is al voldoende om deze vaardigheden langzamerhand te verbeteren.
- Voor een grotere verbetering is waarschijnlijk gerichte uitleg en oefening nodig.
- Dit geldt vooral voor ‘evalueren’ en ‘interpreteren’ (kritisch denken) en ‘teamorganisatie’ (samenwerken) omdat de vooruitgang op deze onderdelen relatief traag verloopt. ‘Informatie zoeken’ (digitale geletterdheid) verbetert daarentegen vrij snel zodat hier geen prioriteit aan hoeft te worden gegeven – in elk geval niet klassikaal; aandacht voor individuele leerlingen kan nodig blijven.
- Meet de ontwikkeling van *alle* 21^e-eeuwse vaardigheden; leerlingen die zich goed ontwikkelen in de ene vaardigheid kunnen stagneren in andere vaardigheden.
- Meet 21^e-eeuwse vaardigheden met ‘passende’ meetinstrumenten: praktische opdrachten zijn motiverend en geven leerlingen niet het gevoel dat hun prestaties worden beoordeeld.
- Ga er *niet* van uit dat leerlingen wiens taal- en rekenvaardigheden zich goed ontwikkelen een soortgelijke ontwikkeling in 21^e-eeuwse vaardigheden doormaken.

Klaar
voor
de
toekomst

Eenvoudig
en compleet

1. Inleiding

1.1 Basisschool Het Talent

Basisschool Het Talent heeft een toekomst-gerichte visie op onderwijs die is gebaseerd op één centraal thema: recht doen aan ieder kind. Dit betekent dat de school uitgaat van erkende ongelijkheid. Door vraaggestuurd onderwijs wordt ingespeeld op de verschillen tussen leerlingen: ieder kind start vanuit een eigen beginsituatie en er wordt rekening gehouden met de sociale context waarin het kind functioneert. Deze visie is uitgewerkt in een uniek onderwijsconcept: de pedagogische didactische kernstructuur (PDKS)².

In de PDKS is het onderwijs georganiseerd in zogenaamde spillen: een afgebakend deel van een leerstoflijn dat door de meeste leerlingen in een half jaar kan worden afgerond. Hoe lang een leerling in een spil zit, is afhankelijk van zijn of haar vorderingen en ondersteuningsbehoeften. Anders dan in een jaarklassensysteem kan een leerling voor verschillende leerstofgebieden in een verschillende spil zitten; voor rekenen bijvoorbeeld in spil 17 en voor taal in spil 18. Zo kan iedere leerling zich in zijn of haar

eigen tempo en op zijn of haar eigen niveau ontwikkelen.

Het onderwijs op Het Talent is niet alleen anders georganiseerd dan op de meeste basisscholen, ook wat betreft de leerinhoud worden andere accenten gelegd. Uiteraard wordt voldoende aandacht besteed aan basisvaardigheden in taal en rekenen – elke lesdag begint bijvoorbeeld met een half uur stillezen – maar regelmatig wordt het taal- en rekenonderwijs ingebed in ontwikkelingsgebieden zoals aardrijkskunde, geschiedenis of wetenschap en technologie (W&T). Daarnaast wordt op Het Talent veel aandacht besteed aan 21^e-eeuwse vaardigheden.

1.2 21^e-eeuwse vaardigheden

Sinds het begin van deze eeuw leven we in een informatiemaatschappij waarin parate kennis minder belangrijk wordt gevonden dan zo'n 30 jaar geleden. Het onderwijs moet op deze verandering inspelen om leerlingen goed voor te bereiden op een vervolgopleiding, de arbeidsmarkt en hun rol als burger. Hierbij draait het om zogeheten 21^e-eeuwse vaardig-

ONDERWIJSCONCEPT

Het Talent werkt volgens een uniek onderwijsconcept: de pedagogische didactische kernstructuur (PDKS). In de PDKS wordt de onderwijsinhoud gerangschikt in leerstoflijnen: een hiërarchisch geordende, coherente verzameling kennis en vaardigheden die elk kind zich in zijn of haar eigen tempo eigen kan maken. Met diagnostische toetsen wordt bijgehouden hoe een kind zich ontwikkelt en welke leerstof voor hem of haar het meest geschikt is.

De belangrijkste didactische kenmerken van het onderwijs op Het Talent zijn:

- Tempo- en niveaudifferentiatie, onder andere door het gebruik van individuele weektaken voor elke leerling.
- Een speelse en natuurlijke meting van individuele voortgang op de eigen leerstoflijnen (ter aanvulling op normgebaseerde voortgangstoetsen).
- Kinderen houden een portfolio bij en leren in portfoliogesprekken op hun eigen ontwikkeling te reflecteren.
- Integratie van taal en rekenen in andere ontwikkelingsgebieden.
- Veel aandacht voor 21^e-eeuwse vaardigheden binnen deze ontwikkelingsgebieden.

heden: een verzameling vakoverstijgende vaardigheden met de daarbij behorende kennis, inzicht en houdingen die nodig zijn om nu én in de toekomst te kunnen functioneren in de samenleving. De SLO onderscheidt acht van dit soort vaardigheden: creativiteit, kritisch denken, probleemoplossen, communiceren, samenwerken, digitale geletterdheid, sociale en culturele vaardigheden en zelfregulering.³ Op Het Talent wordt relatief veel aandacht besteed aan samenwerking, digitale geletterdheid en kritisch denken. Vandaar dat de ontwikkeling van deze drie 21^e-eeuwse vaardigheden in het project is onderzocht.

De leerlingen van Het Talent werken veel met elkaar samen omdat de school het belangrijk vindt dat kinderen op een effectieve en respectvolle manier met en van elkaar leren. *Samenwerking* loopt als een rode draad door de leerstoflijnen waardoor leerlingen vrijwel dagelijks oefenen met vaardigheden als goed luisteren, naar elkaars mening vragen, een taakverdeling maken, opbouwende kritiek geven en overeenstemming bereiken.⁴

Digitale geletterdheid verwijst naar het vermogen om digitale informatie en communicatie verstandig te gebruiken en de gevolgen daarvan te beoordelen⁵. Dit vereist technologische vaardigheden, informatievaardigheden en mediawijsheid.⁴ Op Het Talent komen deze laatste twee aspecten veelvuldig aan bod bij wereldoriëntatie, bijvoorbeeld door leerlingen zelf informatie te laten zoeken op Internet. Aan technologische vaardigheden wordt minder aandacht besteed: de meest leerlingen hebben van huis uit al met ICT leren omgaan en bij de start van dit project werd nog geen aandacht besteed aan de meer geavanceerde vaardigheden. Inmiddels gebeurt dit wel en worden facultatieve lessen gegeven

waarin leerlingen leren programmeren en computationeel denken.

Kritisch denken betreft de kunst om juist te oordelen op grond van logische principes, wat een beroep doet op vaardigheden als het interpreteren, analyseren en synthetiseren van informatie⁶. De leerlingen van Het Talent maken uitgebreid kennis met deze vaardigheden bij wereldoriëntatie en W&T.

1.3 Onderzoeksvragen

De leerkrachten en schoolleiding van Het Talent waren benieuwd in hoeverre hun uitwerking van het PDKS-onderwijsconcept leerlingen in staat stelt om 21^e-eeuwse vaardigheden te ontwikkelen. Het ging hierbij om een interne vergelijking ten opzichte van de ontwikkeling van basisvaardigheden in taal en rekenen. Een externe vergelijking met scholen die een ander onderwijsconcept hanteren kon in dit stadium niet worden uitgevoerd.⁷

De interne vergelijking is uitgevoerd aan de hand van de volgende twee onderzoeksvragen

1. In welke mate realiseert het onderwijsconcept van Het Talent de ontwikkeling van 21^e-eeuwse vaardigheden én basisvaardigheden op het gebied van taal en rekenen?
2. Hoe wordt deze ontwikkeling beïnvloed door leerlingkenmerken?

Om deze vragen te beantwoorden, zijn de bovenbouwleerlingen van Het Talent twee jaar lang gevolgd om hun ontwikkeling van 21^e-eeuwse vaardigheden in kaart te brengen. Hun vorderingen zijn vergeleken met de ontwikkeling van basisvaardigheden in taal en rekenen die in dezelfde periode

3 Thijs et al. (2014)

4 Mercer et al., (1999); Saab et al., (2007)

5 KNAW (2012)

6 Facione (1990)

7 Dit zou het beschikbare budget ver te boven gaan. Anders dan bij taal en rekenen, zijn er geen landelijke referentiescores voor de ontwikkeling van 21^e-eeuwse vaardigheden waardoor de arbeidsintensieve longitudinale metingen dus ook op andere scholen zouden moeten worden uitgevoerd.

zijn gemeten met methodeonafhankelijke voortgangstoetsen.

In de komende hoofdstukken wordt eerst een korte beschrijving gegeven van de leerlingen die hebben deelgenomen aan het onderzoek. Hierna wordt de longitudinale meting van de vaardigheden gepresenteerd. Elke vaardigheid wordt eerst in een apart hoofdstuk besproken, waarbij ook aandacht wordt besteed aan de rol van leerlingkenmerken. Hierna volgt een hoofdstuk waarin de relaties tussen de vaardigheden wordt belicht. In het afsluitende hoofdstuk worden de resultaten samengebracht om antwoord te geven op de onderzoeksvragen. De bevindingen worden vertaald in praktische aanbevelingen voor Het Talent – die overigens ook relevant zijn voor scholen die *niet* volgens de PDKS werken. Tot slot worden enkele suggesties voor vervolgonderzoek gegeven.

2. Deelnemers aan het onderzoek

2.1 Keuze voor de onderzoeksgroep

In overleg met Het Talent is besloten om het onderzoek uit te voeren bij leerlingen uit de hoogste spillen (de ‘bovenbouw’). De meeste van deze leerlingen hebben hun hele schooltijd doorgebracht op Het Talent en zijn vertrouwd met het PDKS-onderwijsconcept. Bovendien biedt het onderwijsprogramma voor de hogere spillen de meeste tijd en ruimte om 21^e-eeuwse vaardigheden te ontwikkelen. Om deze ontwikkeling in kaart te brengen zijn op drie momenten gegevens verzameld, steeds met een tussenperiode van één jaar. Hierdoor kon niet de gehele bovenbouw in het onderzoek worden meegenomen: leerlingen waarvan werd verwacht dat ze binnen twee jaar zouden doorstromen naar het voortgezet onderwijs (en dus de tweede en/of derde meting zouden missen), maakten geen deel uit van de onderzoeksgroep. Concreet betekent dit dat het onderzoek is uitgevoerd bij leerlingen die tijdens de eerste meting in 2016 in spil 17 t/m 20 zaten. Vertaald naar het jaarklassensysteem zijn dit grofweg gezien leerlingen uit groep 5 en 6.

2.2 Kenmerken van de leerlingen bij de start van het onderzoek

Bij de start van het project bestond de onderzoeksgroep uit 182 leerlingen, 100 jongens en 82 meisjes, met een gemiddelde leeftijd van 8.5 jaar (spreiding: 82 – 127 maanden). Negentig leerlingen zaten in spil 17 en 18

(groep 5; gemiddelde leeftijd 7.9 jaar), de overige 92 leerlingen in spil 19 en 20 (groep 6; gemiddelde leeftijd 9.0 jaar).

De meeste leerlingen (85%) hadden een Nederlandse achtergrond. Van de 27 leerlingen met een migratieachtergrond hadden 24 een niet-westers herkomstland (o.a. Chili, Ghana, Nepal, Turkije, Zuid-Afrika). Van de resterende drie leerlingen kwam één van de ouders uit Polen, Spanje of Groot-Brittannië.

Het opleidingsniveau van de ouders is gebruikt als sociaaleconomisch kenmerk van de thuissituatie van de leerlingen. Verreweg de meeste ouders (79%) waren hoogopgeleid, wat betekent dat zij een hbo- of universitaire opleiding hebben gevolgd. De ouders met een middelbaar opleidingsniveau (19%) volgden een havo-, vwo- of mbo-opleiding. Alle overige soorten voortgezet onderwijs worden gerekend tot een laag opleidingsniveau; dit betrof 1 procent van de ouders. Ouders die alleen primair onderwijs hebben gevolgd (1%), behoren tot het laagste opleidingsniveau.

Het opleidingsniveau van de ouders wordt door scholen en instanties zoals DUO ook gebruikt voor de berekening van het leerlinggewicht.⁸ Vrijwel alle leerlingen hadden het basisgewicht van nul. Bij drie leerlingen was het leerlinggewicht 0.3 en één leerling had een leerlinggewicht van 1.2. Speciaal voor de analyse van kritisch denken is ook gekeken naar het probleemoplossend vermogen van de leerlingen. Dit leerlingkenmerk is bepaald aan de hand van een



eenvoudige puzzeltaak: de Toren van Hanoi. De gemiddelde score op deze taak was 12 goed opgeloste puzzels. De spreiding in score was redelijk groot: sommige leerlingen konden vrijwel geen enkele puzzel oplossen terwijl andere leerlingen bijna alle 20 puzzels wisten te maken.

2.3 Uitval

Bij een longitudinaal onderzoek is het onvermijdelijk dat leerlingen tussentijds uitvallen omdat zij naar een andere school gaan, bijvoorbeeld als gevolg van een verhuizing. Tabel 2.1 geeft een overzicht van de uitval tijdens het project⁹. Bij de tweede meting bleef de uitval beperkt tot twee leerlingen. Helaas was de uitval voor de derde meting wat groter: 14 leerlingen. Dit kwam deels doordat sommige leerlingen tegen de verwachtingen in waren doorgestroomd naar het voortgezet onderwijs. Andere leerlingen waren overgestapt naar een recent geopende vernieuwingschool dichtbij huis en één leerling kreeg de mogelijkheid om buiten de regio een dansopleiding te gaan volgen. Bij alle drie de metingen waren enkele leerlingen afwezig tijdens de afname van een of meer testen, meestal vanwege ziekte. Omdat de testen op verschillende dagen zijn afgenomen, betrof deze tijdelijke uitval vaak maar een enkele test. Voor deze leerlingen werd in overleg met de leerkrachten minimaal één inhaalmoment gepland. Als zij dan

wederom afwezig waren, hing het van de omstandigheden af of de test alsnog kon worden ingehaald. Zo niet, dan leidde dit tot ontbrekende gegevens.

Ook bij leerlingen die een test wél hadden gemaakt, konden gegevens ontbreken. Bijvoorbeeld doordat een vraag was overgeslagen, een antwoord onleesbaar was, of er iets misging met het opslaan van de video- of computerbestanden.

2.4 Ontbrekende gegevens

Bij alle drie de metingen waren enkele leerlingen afwezig tijdens de afname van een of meer testen, meestal vanwege ziekte. Omdat de testen op verschillende dagen zijn afgenomen, betrof deze tijdelijke uitval vaak maar een enkele test. Voor deze leerlingen werd in overleg met de leerkrachten minimaal één inhaalmoment gepland. Als zij dan wederom afwezig waren, hing het van de omstandigheden af of de test alsnog kon worden ingehaald. Zo niet, dan leidde dit tot ontbrekende gegevens.

Ook bij leerlingen die een test wél hadden gemaakt, konden gegevens ontbreken. Bijvoorbeeld doordat een vraag was overgeslagen, een antwoord onleesbaar was, of er iets misging met het opslaan van de video- of computerbestanden.

Tabel 2.1: Overzicht van de leerlingen die uitvielen tijdens het onderzoek

	Uitval bij de 2 ^e meting (2017)	Uitval bij de 3 ^e meting (2018)
Aantal	2 leerlingen	14 leerlingen
Sekse	0 jongens 2 meisjes	10 jongens 4 meisjes
Spil bij aanvang project	1 spil 17-18 1 spil 19-20	3 spil 17-18 11 spil 19-20
Achtergrond	2 Nederlandse achtergrond 0 migratieachtergrond	11 Nederlandse achtergrond 3 migratieachtergrond

⁹ Leerlingen die tijdens de looptijd van het project zijn ingestroomd op Het Talent, zijn niet toegevoegd aan de onderzoeksgroep.

3. Kritisch denken

3.1 Inleiding

Bij kritisch denken gaat het om het kunnen formuleren van een eigen, onderbouwde visie of mening¹⁰. Kritisch denken is een doelgericht proces dat een beroep doet op vaardigheden als induceren, interpreteren en evalueren¹¹. In dit project is kritisch denken ingebed in W&T-onderwijs en gemeten aan de hand van vaardigheden in wetenschappelijk redeneren. Voorbeelden van dit soort onderzoeksvaardigheden zijn experimenteren, voorspellingen doen, resultaten interpreteren en conclusies trekken¹². In Tabel 3.1 is te zien welke onderzoeksvaardigheden zijn gemeten en welke kritisch denken vaardigheden de leerlingen hierbij moesten gebruiken. Hierbij is sprake van een één op één relatie: de vaardigheden in de rijen van de tabel zijn elkaars synoniem.

Uit eerder onderzoek¹³ blijkt dat kinderen tussen de 8 en 12 jaar in staat zijn deze onderzoeksvaardigheden uit te voeren en

zich hier met de tijd verder in te bekwamen. Bovendien blijkt dat de vaardigheden voorspellen (induceren) en conclusies trekken (infereren) zich geleidelijk maar niet gelijktijdig ontwikkelen. Vanaf een jaar of 6 zijn kinderen prima in staat om conclusies te trekken op basis van resultaten waarin een perfecte samenhang tussen twee variabelen bestaat. Wanneer de gegevens niet of niet volledig causaal zijn, verloopt deze ontwikkeling trager. Het doen van voorspellingen begint zich langzaam te ontwikkelen vanaf het zesde levensjaar en is aan het einde van de basisschool nog voor verbetering vatbaar.

De overige twee vaardigheden, gegevens interpreteren en de kwaliteit ervan evalueren, zijn wat complexer voor een jonge doelgroep. Kinderen tussen de 10 en 14 jaar hebben nauwelijks oog voor de kwaliteit van onderzoeksgegevens. Zij weten nog maar weinig over meetfouten, hoe deze zijn te herkennen en hoe je er mee om moet gaan bij

Tabel 3.1 Overeenkomst tussen onderzoeksvaardigheden en kritisch denken

Induceren	Voorspellingen doen	"Wat denk je dat er gebeurt als je hetzelfde proefje nog een keer doet? Wat zou er dan uitkomen?"
—	Experimenteren ¹	"Bedenk een goede proef om uit te zoeken of het iets uitmaakt of je de bal van het hoge of lage hekje laat rollen"
Interpreteren	Resultaten interpreteren	"Ik heb ook een paar proefjes gedaan om te kijken of de plaats van het hekje wat uitmaakt. Kun jij in je eigen woorden uitleggen wat er uit mijn vijf proefjes komt?"
Evalueren	De kwaliteit van de gegevens beoordelen	"Hier zie je het antwoord van Joris en Anja. Als je kijkt naar hun proefjes in de tabel, wie geloof je dan het meest?"
Infereren	Conclusies trekken	"Ik wilde mijn idee testen dat ballen van het lage hekje verder rollen dan van het hoge hekje. Je hebt de uitkomsten gezien. Moet ik nou mijn idee veranderen?"

¹ Deze vaardigheid was opgenomen in de test om het voor leerlingen actiever en motiverender te maken. Omdat experimenteren geen directe relatie heeft met kritisch denken, zijn de resultaten niet meegenomen in de analyses.

¹⁰ Thijs et al. (2014)

¹¹ American Philosophical Association (1990)

¹² Pedaste et al. (2015)

¹³ Piekny & Maehler (2013); zie ook Piekny & Maehler (2014)

het interpreteren van onderzoeksgegevens¹⁴. Ook het interpreteren van gegevens zelf gaat beter met de jaren. Negenjarigen hebben moeite met het interpreteren van data als het aantal metingen toeneemt; daarnaast besteden zij weinig aandacht aan de variatie binnen en tussen series metingen. Kinderen van 12 zijn hier beter toe in staat, al blijft het voor hen lastig om gegevens te interpreteren als er variatie binnen een dataset bestaat¹⁵.

3.2 Test voor kritisch denken

In dit project is kritisch denken gemeten door leerlingen onder begeleiding een onderzoek te laten bedenken en uitvoeren

waarbij zij hun handelingen, beslissingen en argumenten mondeling konden toelichten. De test bestond uit drie opdrachten per kritisch denken vaardigheid. De opdrachten hadden een oplopende moeilijkheidsgraad omdat de test drie keer is afgenomen, steeds met tussenpozen van één jaar. De geleidelijke opbouw van eenvoudige naar meer uitdagende opdrachten moest ervoor zorgen dat leerlingen bij de eerste testafname in elk geval een paar vragen goed konden beantwoorden en er bij de laatste testafname nog voldoende uitdaging overbleef. Om herkenning van de opdrachten te voorkomen, kregen de leerlingen elk jaar een andere versie van de test die alleen verschilde wat betreft het onderwerp dat zij moesten onderzoeken (zie Tabel 3.2)

Tabel 3.2: Kenmerken van de drie versies van de test voor kritisch denken

	Rollende ballen	Stuiterende ballen	Rijdende auto's
Materiaal			
Onderzoeksvraag	Welke factoren bepalen hoe ver een bal van de helling rolt?	Welke factoren bepalen hoe vaak een bal stuiter?	Welke factoren bepalen hoe ver een auto rijdt?
Variabelen	- Hellingshoek (steil of vlak) - Startpositie van de bal (hoog of laag) - Gewicht van de bal (zwaar of licht) - Oppervlak van de helling (glad of ruw)	- Valhoogte (hoog of laag) - Gewicht van de bal (zwaar of licht) - Samenstelling van de bal (hol of gevuld) - Oppervlak waarop de bal valt (hard of zacht)	- Dikte van de achteras (dik of dun) - Diameter van het elastiekje (klein of groot) - Omtrek van de achterwielen (groot of klein) - Aantal keer dat het elastiekje om de as wordt gewonden ($\frac{1}{4}$ of $\frac{1}{2}$ keer)
Uitkomstmaat	Afstand die de bal van de helling rolt, gemeten als het aantal vakjes	Aantal keer dat een bal stuiter	Afstand die een auto aflegt, gemeten als het aantal vakjes

¹⁴ Kanari & Millar (2004)
¹⁵ Masnick & Morris (2008)

Bij de afname van de test deed de proefleider met individuele leerlingen vier experimenten volgens een vaststaand script. Sommige experimenten werden door de leerlingen bedacht en uitgevoerd, andere experimenten werden door de proefleider gedaan en soms werden de resultaten van een experiment op *hand-outs* gepresenteerd. Het script beschreef de dialoog die de proefleider met de leerling moest voeren en gaf aan wanneer de proefleider iets moest zeggen, doen of noteren. De acties en antwoorden van de leerling werden door de proefleider aangekruist of opgeschreven en zijn achteraf gescoord als goed (1 punt) of fout (0 punten). In totaal konden de leerlingen maximaal 12 punten behalen, drie voor elke vaardigheid.

3.3 Resultaten

Figuur 3.1 laat zien dat de leerlingen elk jaar beter werden in kritisch denken¹⁶. Hun totaalscores stegen jaarlijks met ongeveer 1 punt. Bij leerlingen uit spil 19-20 vlakke deze ontwikkeling op het eind wat af, maar dit effect was niet significant. Waarschijnlijk komt dit door de spreiding in de scores, die op alle toetsmomenten groot was. In beide spellen waren leerlingen te vinden die minder dan twee opdrachten goed hadden, terwijl er ook leerlingen waren met maar twee foute opdrachten. Deze verschillen waren niet te verklaren door sekse, leeftijd, achtergrond, sociaaleconomische status en probleemoplossend vermogen.

Fragment uit het script voor de proefleider



Zet de ballenbaan weer in de beginpositie, dus plat, en verwijder de ballen en hulpstukken.



OK, dan gaan we nog een proefje doen. Kun jij op een goede manier onderzoeken of de hoek van de helling wat uitmaakt voor hoe ver de bal rolt? Ga je gang.



Kruis aan welke variabelen worden gevarieerd:

Hellingshoek: ☐ steil ☐ vlak
 Startpositie: ☐ hoog ☐ laag
 Oppervlak: ☐ glad ☐ ruw
 Bal: ☐ zwaar ☐ licht



Geef de leerling handout 1

¹⁶ Door uitval en ziekte ontbreken de gegevens van 1 leerling in 2016. In de jaren daarna was dit het geval bij respectievelijk 2 en 16 leerlingen.



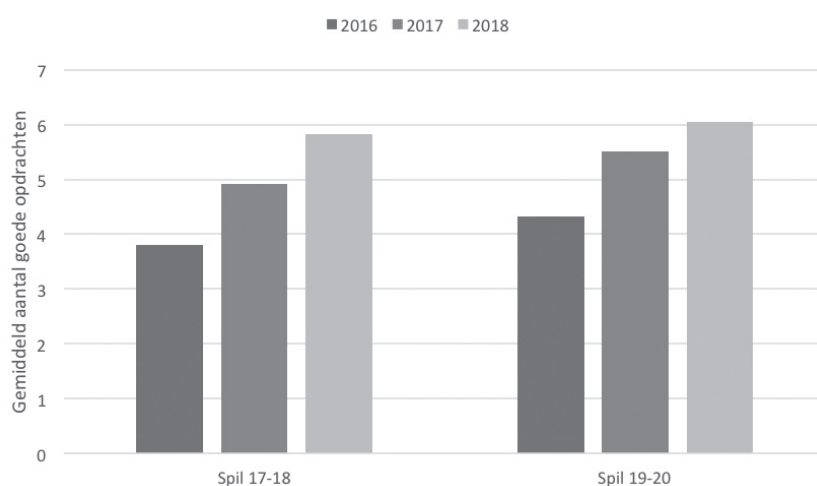
MEER WETEN
OVER DE
ANALYSES?
PAGINA 43

Naast deze individuele verschillen waren er ook verschillen tussen de vaardigheden. In Figuur 3.2 zijn de totaalscores uitgesplitst naar vaardigheid; omdat het scoreverloop in beide spellen gelijk was, zijn ze samengenomen in het staafdiagram. Uit dit overzicht blijkt dat 'induceren' en 'infereren' elk jaar wat beter ging. Voor 'evalueren' was dit alleen het geval tijdens het eerste jaar, en de scores voor 'interpreteren' waren in alle drie de jaren vrijwel gelijk.

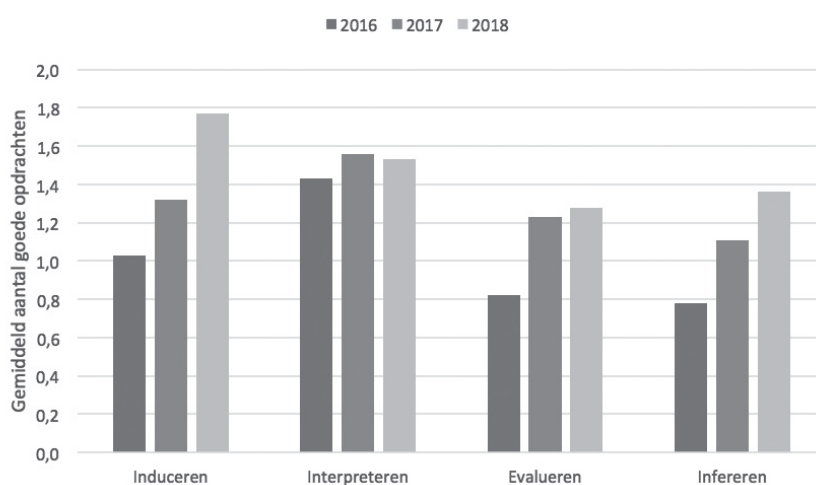
3.4 Conclusies

Kritisch denken vaardigheden ontwikkelen zich langzaam en niet gelijktijdig. Net als in eerder onderzoek blijkt dat leerlingen van Het Talent steeds beter werden in het induceren en infereren. De vaardigheden 'evalueren' en 'interpreteren' zijn daarentegen wat moeilijker en verbeterden niet of nauwelijks. Bovendien bestaan er grote individuele verschillen in (de ontwikkeling van) het vermogen om kritisch te denken. Deze verschillen konden niet worden verklaard door de leerlingkenmerken die in dit project zijn gemeten.

Figuur 3.1: Totaalscores op de kritisch denken test (maximale score = 12 punten).



Figuur 3.2: Scores op de kritisch denken test per vaardigheid.



4. Digitale geletterdheid

4.1 Inleiding

Digitale geletterdheid wordt gedefinieerd als het vermogen om de computer te gebruiken voor het verzamelen, creëren en delen van digitale informatie om effectief te kunnen functioneren in de thuissituatie, op school, op het werk en in de samenleving. Deze definitie is afkomstig uit het ICILS-project waarin de digitale geletterdheid van 14-jarigen uit 21 landen is vergeleken¹⁷. Hierbij is een raamwerk gebruikt dat bestaat uit twee hoofdcategorieën die drie, respectievelijk vier onderdelen bevatten.

Dit raamwerk vertoont veel overeenkomsten met de analyse- en beschrijvingskaders uit soortgelijke projecten. Zo zijn de metingen in het EUKidsOnline project¹⁸ gebaseerd op een typologie waarin onderscheid wordt gemaakt tussen operationele vaardigheden, internetvaardigheden, communicatievaardigheden en productievaardigheden¹⁹. In het kidNET project²⁰ werd eveneens onderscheid gemaakt

tussen operationele vaardigheden en internetvaardigheden; communicatie- en productievaardigheden werden buiten beschouwing gelaten omdat deze buiten het bereik van het onderzoek vielen. In de Verenigde Staten werd in de General Social Survey van 2000 en 2002 speciale aandacht besteed aan het onderwerp 'de informatiemaatschappij'. De vragen die hierover werden gesteld gingen over algemeen computergebruik en het gebruik van email en Internet²¹.

In het huidige project zijn beide hoofdcategorieën uit het ICILS-raamwerk gemeten. Om inhoudelijke en praktische redenen zijn echter niet alle onderliggende onderdelen in de testen opgenomen. In de categorie 'het verzamelen en beheren van digitale informatie' zijn de basisvaardigheden te eenvoudig; de leerlingen van Het Talent hebben thuis geleerd met de computer om te gaan zodat hier op school geen aandacht aan wordt besteed. Het beheren van digitale informatie is daarentegen te moeilijk voor onze doelgroep.


MEER WETEN
OVER DE
INHOUD VAN
DE TESTEN?
PAGINA 45

RAAMWERK UIT HET ICILS-PROJECT

1. **Verzamelen en beheren van digitale informatie**
 - 1.1 Het kunnen omgaan met computers
 - 1.2 Het zoeken, vinden en evalueren van digitale informatie
 - 1.3 Het beheren van digitaal verkregen informatie
2. **Produceren en delen van digitale informatie**
 - 2.1 Het bewerken van digitaal verkregen informatie
 - 2.2 Het creëren van digitale informatie
 - 2.3 Het delen van digitale informatie
 - 2.4 Het veilig en zorgvuldig gebruiken van digitale informat

17 Fraillon et al. (2013); het Nederlandse deel van dit onderzoek wordt beschreven in Meelissen et al. (2014)

18 Ólafsson et al. (2014)

19 Van Deursen et al. (2014)

20 Lazonder et al. (2000), de Vries et al. (2008)

21 <http://www3.norc.ou.edu/GSS/Website/> De vragen over internetgebruik zijn uitgebreid beschreven in Hagattai (2004)

Uit de resultaten van het ICILS-project bleek dat 70 procent van de 14-jarige leerlingen niet weet hoe zij data moeten beheren²². Bovendien komt dit aspect op Het Talent niet aan bod; de leerlingen hebben daar geen eigen computer. Zodoende is slechts één onderdeel uit deze categorie gemeten, namelijk 'het zoeken, vinden en evalueren van digitale informatie'. In de categorie 'het produceren en delen van digitale informatie' is het onderdeel 'het delen van informatie' buiten beschouwing gelaten omdat e-mailen, facebooken, whatsappen, etc. op school verboden is. De overige drie onderdelen uit deze categorie zijn wel in de testen opgenomen.

4.2 Drie testen voor digitale geletterdheid

Voor de onderdelen 'zoeken, vinden en evalueren van digitale informatie' en 'veilig en zorgvuldig gebruiken van digitale informatie' is een geschreven test ontwikkeld die klassikaal werd afgenomen volgens een voorgeschreven procedure; voorafgaand aan de test kregen de leerlingen een korte uitleg die eveneens was

gestandaardiseerd. De test zelf bevatte 16 open vragen — 13 voor het 'zoeken naar informatie' en 3 voor het 'gebruiken van informatie' — waarvoor steeds 1 punt kon worden gehaald. Bij elke vraag werd een scherm afbeelding getoond waarop de leerlingen hun antwoord konden aankruisen of een korte tekst konden schrijven, bijvoorbeeld een URL in de adresregel van een browser of trefwoorden in Google. De leerlingen hebben de vragen in hun eigen tempo beantwoord, er was geen tijdslimiet. Gemiddeld duurde dit 30 tot 45 minuten.

Voor het onderdeel 'bewerken van digitaal verkregen informatie' moesten de leerlingen de opmaak van een Word bestand veranderen. Gezien het beschikbare aantal computers, kon deze opdracht door vijf leerlingen tegelijk worden gemaakt. Na een korte uitleg gingen zij achter een eigen computer zitten waarop een Word bestand was geopend. Op de opdrachtkaart was te zien welke veranderingen zij moesten doorvoeren. Hiervoor kregen ze 10 minuten de tijd. Voor elke goede verandering was 1 punt te verdienen; de maximale score voor deze test was 7 punten.

MEER WETEN
OVER DE KEUZE
VOOR HET
SOORT TEST?
PAGINA 47

WORD OPDRACHT

Oorspronkelijk bestand

Nee hoor, niks mis met mijn gehoor
Door Sandy Stevens

"Hé? Of, zoals het hoort: 'Wat zeg je? Kun je wat harder praten?' Slecht horen is een groot probleem. En niet alleen bij oude mensen. Juist jonge mensen hebben steeds vaker slechte oren. Zelf denken zij vaak dat er niets aan de hand is.

De Nationale Hoorstichting begon in 2003 met de Nationale Hoorstest. Mensen konden aan de telefoon testen hoe het met hun gehoor was. Kort daarna kwamen er ook een website. Zo kon je ook online testen hoe goed je kon horen.

Rap om je gehoor te testen
Ruim één miljoen mensen testten hun gehoor via de Nationale Hoorstest. De stichting is daar erg blij mee. Dat werd getoond op de Stebo-school in Den Haag. Er waren ook politici. Dat zijn mensen uit de politiek die het land besturen. Maar vooral Lange Frans en Baas B. maakten er een feestje van. Zij rapten hun versie van de hoorstest die je kunt doen op de website. Het geluid niet te hard zetten natuurlijk! Want harde muziek is heel slecht voor je oren. Als je je oren beschadigt met lawaai, is dat niet meer te verhelpen. Dus: eens kapot is altijd kapot.



Niks mis?
Mensen die de test deden, dachten vaak van tevoren dat ze heel goed konden horen. Maar soms viel dat erg tegen! Gelukkig gingen de meeste mensen dan meteen naar de dokter. 'Hoorstesten zijn belangrijk', zegt de Hoorstichting. 'Mensen komen er zo achter hoe het met hun oren is. Ook krijgen ze informatie over wat ze kunnen doen als er problemen zijn.'

Opdrachtkaart



Nee hoor, niks mis met mijn gehoor
Door Sandy Stevens

"Hé? Of, zoals het hoort: 'Wat zeg je? Kun je wat harder praten?' Slecht horen is een groot probleem. En niet alleen bij oude mensen. Juist jonge mensen hebben steeds vaker slechte oren. Zelf denken zij vaak dat er niets aan de hand is.

De Nationale Hoorstichting begon in 2003 met de Nationale Hoorstest. Mensen konden aan de telefoon testen hoe het met hun gehoor was. Kort daarna kwamen er ook een website. Zo kon je ook online testen hoe goed je kon horen.

Rap om je gehoor te testen
Ruim één miljoen mensen testten hun gehoor via de Nationale Hoorstest. De stichting is daar erg blij mee. Dat werd getoond op de Stebo-school in Den Haag. Er waren ook politici. Dat zijn mensen uit de politiek die het land besturen. Maar vooral Lange Frans en Baas B. maakten er een feestje van. Zij rapten hun versie van de hoorstest die je kunt doen op de website. Het geluid niet te hard zetten natuurlijk! Want harde muziek is heel slecht voor je oren. Als je je oren beschadigt met lawaai, is dat niet meer te verhelpen. Dus: eens kapot is altijd kapot.

Niks mis?
Mensen die de test deden, dachten vaak van tevoren dat ze heel goed konden horen. Maar soms viel dat erg tegen! Gelukkig gingen de meeste mensen dan meteen naar de dokter. 'Hoorstesten zijn belangrijk', zegt de Hoorstichting. 'Mensen komen er zo achter hoe het met hun oren is. Ook krijgen ze informatie over wat ze kunnen doen als er problemen zijn.'

22 Meelissen et al. (2014)

Bij de opdracht over het 'creëren van digitale informatie' moesten de leerlingen een PowerPoint presentatie maken. Deze opdracht werd aangekondigd in een *flyer* waarin het doel werd uitgelegd en enkele globale aanwijzingen werden gegeven zodat de leerlingen een indruk kregen van het soort presentatie dat ze moesten maken. Deze opdracht werd direct na de Word-opdracht uitgedeeld en toegelicht. De leerlingen hadden 20 minuten de tijd om hun presentatie te maken. De presentaties zijn achteraf beoordeeld op inhoud, technische complexiteit en grafische vormgeving. Voor elk criterium kon maximaal 9 punten worden gehaald; de maximale score was dus 27 punten.

Alle testen zijn drie keer afgenomen, steeds in de eerste maanden van het kalenderjaar. Om de validiteit zo goed mogelijk te waarborgen, zijn drie parallelle versies van elke test gemaakt die uitsluitend verschilden wat betreft het onderwerp waarover informatie werd gezocht (de vragenlijst), het onderwerp van het Word bestand dat moest worden veranderd en het onderwerp waarover een PowerPoint presentatie moest worden gemaakt. Elke leerling kreeg de drie versies in willekeurige volgorde met tussenpozen van een jaar.

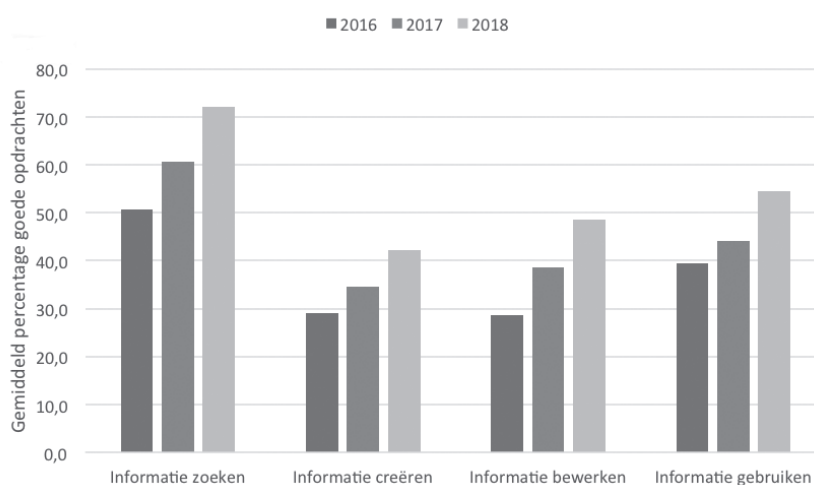
4.3 Resultaten

Omdat de maximale score van de testen verschilde, zijn de gemiddelden omgerekend naar percentages zodat ze onderling te vergelijken zijn. Figuur 4.1 geeft de resultaten van de leerlingen uit spil 17-18; de prestaties in spil 19-20 waren vrijwel gelijk en worden daarom niet apart getoond²³. Er was echter één uitzondering: in 2017 waren hun scores voor 'informatie zoeken' iets meer verbeterd dan bij spil 17-18 (15 i.p.v. 10%). Het jaar daarna waren de scores weer gelijk (72% versus 73 %).

Het scoreverloop in Figuur 4.1 laat zien dat de digitale geletterdheid elk jaar toenam, meestal met iets meer dan 10 procent. Wat verder opvalt is dat de leerlingen hoger scoorden op 'informatie zoeken' dan op de overige drie onderdelen. Of ze hier ook echt beter in waren is moeilijk te zeggen omdat de testen onderling verschilden (geschreven test of opdrachten).

De jaarlijkse toenames zijn gebaseerd op de gemiddelde scores van alle leerlingen uit een spil, wat dus *niet* betekent dat elke leerling elk jaar (even veel) vooruit is gegaan. Bij het onderdeel 'informatie bewerken' bijvoorbeeld,

Figuur 4.1: Percentage goede antwoorden op de testen voor digitale geletterdheid.



23 Door uitval en ziekte ontbreken alle of sommige gegevens van 3 leerlingen in 2016. In de jaren daarna waren er geen of incomplete gegevens voor respectievelijk 5 en 20 leerlingen.



MEER WETEN
OVER DE
ANALYSES?
PAGINA 54

verbeterden 42 leerlingen uit spil 17-18 hun prestatie uit 2016; de meesten (23) met 1 punt, sommigen met 2 punten en een enkeling met 3 of 4 punten. Maar... 34 leerlingen presteerden precies even goed als het jaar daarvoor en bij 10 leerlingen daalde de score met 1 of zelfs 3 punten.

Deze individuele verschillen waren onafhankelijk van leerlingkenmerken. Jongens gingen evenveel vooruit als meisjes, leerlingen met een Nederlandse achtergrond hadden een even grote vaardigheidsgroei als leerlingen met een migratieachtergrond en hun sociaal-economische status had evenmin effect. Bij het 'creëren van informatie' werd echter een negatieve samenhang gevonden met leeftijd: hoe jonger de leerling des te groter de groei tussen 2016 en 2017. De vaardigheidsgroei op de overige onderdelen van digitale geletterdheid was onafhankelijk van leeftijd in beide periodes.

4.4 Conclusies

Bovenbouwleerlingen op Het Talent hebben tijdens het project hun digitale geletterdheid weten te verbeteren. Digitale informatie zoeken, creëren, bewerken en op een veilige manier gebruiken ging elk jaar een beetje beter. Deze vaardigheidsgroei was stabiel en verschilde niet tussen de spillen. Wel waren er tamelijk grote individuele verschillen, die onafhankelijk waren van het geslacht, de achtergrond en sociaaleconomische status van de leerlingen. Hun leeftijd hing samen met de vaardigheidsgroei bij het creëren van digitale informatie tussen 2016 en 2017. Omdat dit een incidenteel verband lijkt te zijn, kan hier verder geen verklaring voor worden gegeven.

5. Samenwerken

5.1 Inleiding

Samenwerken is één van 21^e-eeuwse vaardigheden. De verwachting is dat werknemers in de toekomst goed moeten kunnen samenwerken in (interdisciplinaire) teams. De ontwikkeling van deze vaardigheid begint al op de basisschool, waar leerlingen regelmatig in tweetallen of kleine groepjes aan een taak of opdracht werken. Om deze taak zo goed mogelijk uit te voeren, is het belangrijk dat de groepsgenoten goed met elkaar communiceren en hun kennis en ideeën delen.

Samenwerken is echter meer dan alleen communiceren. Om een gemeenschappelijk

doel te bereiken, moeten de groepsleden hun denken en handelen op elkaar afstemmen en een gemeenschappelijk referentiekader creëren en onderhouden. Hierbij moeten de ideeën van elk groepslid verkend en besproken worden en moeten waar nodig compromissen worden gesloten. Groepsleden moeten een gedeeld begrip van de taak of het probleem opbouwen, waarbij het belangrijk is dat inzichten worden gedeeld en waar nodig uitleg wordt gegeven²⁴. Daarnaast moeten de taken goed worden verdeeld en is het belangrijk dat de sfeer in de groep positief blijft. Samenwerken is dus een activiteit waarbij cognitieve en sociale activiteiten moeten worden gecombineerd.

Tabel 5.1: Samenwerkingsvaardigheden en hun deelvaardigheden

Vaardigheid	Deelvaardigheden
Gedeeld begrip	• Het verkennen van alle perspectieven, kennis en mogelijkheden van de teamleden. • Het werken aan een gemeenschappelijk begrippenkader en gedeelde opvattingen over de taak en eventuele deeltaken. • Met teamleden overleggen over de stappen die je neemt in de planning en uitvoering van de taak. • Monitoren van gedeeld begrip en herstellen indien nodig.
Passende acties	• Het verkennen van de werkzaamheden. • Het identificeren van de activiteiten die nodig zijn voor het succesvol afronden van de taak. • Het uitvoeren van de plannen in overeenstemming met de taak. • Monitoren van de acties, voortgang en de kwaliteit van het werk.
Teamorganisatie	• Het begrijpen en rekening houden met de verschillende rollen en taken in een team. • Het herkennen van de rollen van leden, en ze hierop aanspreken indien nodig. • Het opvolgen en uitvoeren van plannen en afspraken en het uitvoeren van afgesproken werkzaamheden. Hieronder valt ook het oproepen van anderen tot actie. • Monitoren van de gegeven feedback op teamorganisatie. Het aanpassen van de organisatie, afspraken en rollen.

24 Roschelle & Teasley (1995)

In dit project is samenwerken gemeten aan de hand van een ontwerpopdracht die door groepjes van drie of vier leerlingen is uitgevoerd. Deze opdracht is geïnspireerd op het OECD-PISA onderzoek naar samenwerkend probleemoplossen²⁵. De hierboven geschetste kenmerken van samenwerken komen terug in de drie centrale samenwerkingsvaardigheden die door de OECD worden onderscheiden: het tot stand brengen en onderhouden van een gedeeld begrip, het plannen en uitvoeren van acties die bijdragen aan het afronden van de taak, en het tot stand brengen en onderhouden van een goede organisatie van de groep. Elke vaardigheid kan worden opgesplitst in een aantal deelvaardigheden; een overzicht hiervan staat in Tabel 5.1.

5.2 Test voor samenwerken

In dit project zijn de samenwerkingsvaardigheden gemeten door teams van drie of vier leerlingen een ontwerpopdracht te laten uitvoeren. Omdat de teams in drie achtereenvolgende jaren zijn onderzocht, zijn drie versies van de opdracht ontwikkeld waarbij de leerlingen een schooltuin, een schoolshirt of een speeltoestel moesten ontwerpen. Bij elk ontwerp moest het team rekening houden met (soms tegenstrijdige) ontwerpisen die door betrokkenen van binnen en buiten de school waren opgesteld. Het uiteindelijke resultaat moest in een gezamenlijke tekening worden weergegeven. Hoewel de teams elk jaar een andere opdracht kregen, bleven de opzet en structuur van de opdracht gelijk.

De proefleider nam telkens twee teams mee naar een rustige ruimte in de school. Hier introduceerde zij de opdracht volgens een van tevoren vastgesteld script en las de volledige opdracht één keer voor. Vervolgens kreeg iedere leerling een kaartje met daarop één van de ontwerpisen en kon de opdracht beginnen. Tijdens het ontwerpproces moesten de leerlingen ervoor zorgen dat de eis op hun kaartje in het ontwerp werd verwerkt. De teams mochten 20 minuten aan hun ontwerp werken.

De samenwerkingsvaardigheden zijn op groepsniveau geanalyseerd. Van elk team zijn video opnames gemaakt die zijn gescoord met behulp van een codeerschema. Hierbij werd gekeken hoe goed de verschillende ontwerpteams presteerden op het tot stand brengen en onderhouden van gedeeld begrip, het nemen van passende acties, en de teamorganisatie (de drie door OECD onderscheiden categorieën). Voor elke vaardigheid was maximaal 12 punten te behalen.

5.3 Teamsamenstelling

Tijdens het onderzoek werd er gewerkt in teams van drie of vier leerlingen uit dezelfde spil. Bij het samenstellen van de teams is geprobeerd om elk groepje uit zowel jongens als meisjes te laten bestaan. De teamsamenstelling bleef het gehele project gelijk, zodat leerlingen de drie opdrachten steeds met dezelfde medeleerlingen hebben uitgevoerd. Vanwege uitval en ziekte is dit niet altijd gelukt. Werd een team door



MEER WETEN
OVER HET
ONTWERP VAN
DE TEST?
PAGINA 57

ONTWERPTAAK: DE SCHOOLTUIN

Bij het ontwerpen van de schooltuin moeten leerlingen rekening houden met de wens van de **directie**, die het belangrijk vindt dat de tuin er netjes uit ziet. De tuin moet zo worden ingedeeld dat het onkruid gemakkelijk kan worden verwijderd. De **buurtbewoners** vinden het belangrijk dat er aan de buitenzijde van de tuin groen en bloeiende planten te zien zijn, daarnaast geven zij aan dat er geen hoge gebouwen in de tuin mogen staan. De **medeleerlingen** vinden het belangrijk dat er op alle momenten iets eetbaars groeit in de tuin. De **conciërge** is verantwoordelijk voor het tuingereedschap en wil graag dat dit gereedschap op een veilige manier wordt opgeborgen.

25 OECD (2013)

natuurlijk verloop kleiner dan drie leerlingen, dan zijn de prestaties vanaf dat jaar buiten beschouwing gelaten. In 2016 namen in totaal 48 teams deel aan het onderzoek, waarvan 24 uit spil 17-18 en 24 uit spil 19-20. In 2017 viel er één team uit spil 17-18 af. In 2018 was het aantal teams verder afgenomen naar 40, waarvan 21 uit spil 17-18 en 19 uit spil 19-20.

5.4 Resultaten

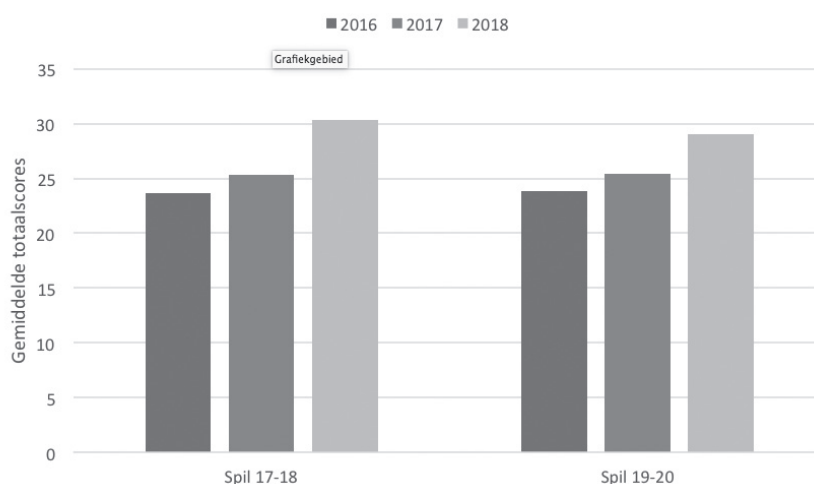
In Figuur 5.1 zijn de totaalscores van de drie samenwerkingsvaardigheden afgebeeld. Aan de hoogte van de staven is te zien dat de teams uit beide spellen elk jaar beter werden in het samenwerken. Wat verder opvalt, is dat de scores in beide spellen vrijwel gelijk zijn. Er is dan ook geen effect van spil gevonden.

In Figuur 5.2 zijn de totaalscores opgesplitst

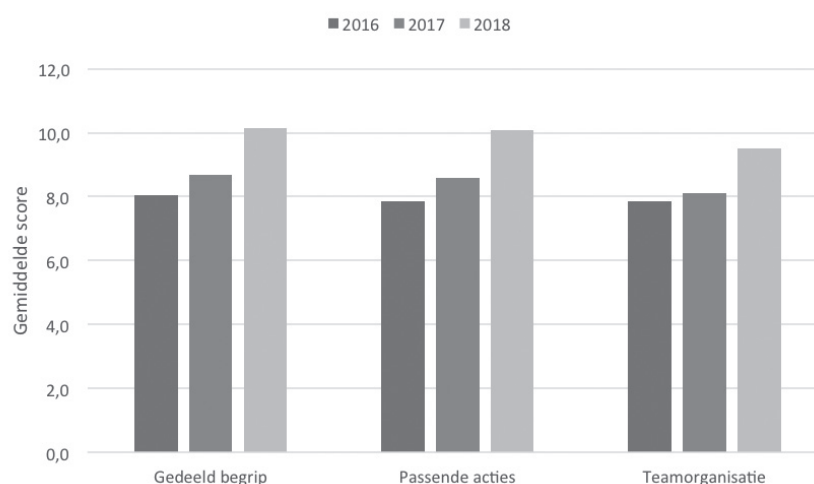
naar vaardigheid. Omdat het scoreverloop in beide spellen vergelijkbaar was, zijn er geen aparte diagrammen gemaakt; de staven tonen het gemiddelde van de gehele onderzoeksgroep. Per deelvaardigheid was de maximale score 12 punten. In de periode 2016-2017 verbeterden de deelvaardigheden 'creëren en onderhouden van gedeeld begrip' en 'uitvoeren van passende acties'; bij 'teamorganisatie' was dit niet het geval. Tussen 2017 en 2018 was er vooruitgang op alle drie de deelvaardigheden. Er is geen effect van spil gevonden, wat betekent dat de teams uit spil 17-18 in beide periodes evenveel vooruit gingen als de teams uit spil 19-20. Wel was er in elke periode een behoorlijke spreiding in de scores. In beide spellen waren teams die boven- en ondergemiddeld vooruit gingen op samenwerking. Gezien het samengestelde karakter van de groepen is het niet mogelijk om deze variatie in scores te verklaren vanuit leerlingkenmerken.

MEER WETEN
OVER DE
ANALYSES?
PAGINA 60

Figuur 5.1: Totaalscores van de samenwerkingsvaardigheden (maximale score = 36 punten)



Figuur 5.2: Scores uit beide spellen op samenwerken per deelvaardigheid.



5.5 Conclusies

Leerlingen op Het Talent werden tijdens het project steeds beter in het samenwerken. In de periode 2016-2017 beperkte deze groei zich tot de deelvaardigheden 'creëren en in stand houden van gedeeld begrip' en 'plannen en uitvoeren van passende acties'. In de periode 2017-2018 was er ook bij de derde deelvaardigheid 'teamorganisatie' vooruitgang te zien, hoewel minder groot dan bij de overige twee deelvaardigheden. Een onderdeel van teamorganisatie is dat leerlingen elkaar houden aan afspraken en aanspreken op gedrag indien nodig. Zonder specifieke voorbereiding of instructie is dit voor veel leerlingen uit deze leeftijdsgroep een lastige vaardigheid²⁶.

26 Baines et al. (2009)

6. Taal en rekenen

6.1 Inleiding

Voor de meting van taal- en rekenvaardigheid is gebruik gemaakt van methode-onafhankelijke voortgangstoetsen, zodat de prestaties van de leerlingen vergeleken konden worden met die van een landelijke referentiegroep. Toch is Het Talent wat sceptisch over dit soort toetsen omdat het een specifieke manier van testen is waar de leerlingen op Het Talent niet aan zijn gewend – hun vorderingen worden meestal op een informele, speelse of authentieke manier beoordeeld. Het voordeel van een vergelijking met normscores woog echter zwaarder. Wel is geprobeerd om kwalitatief hoogstaande toetsen te gebruiken die een positieve Cotaan beoordeling hebben en door de wetgever worden erkend. Voor de rekentoets gold de aanvullende eis dat de opgaven géén beroep mochten doen op bijvoorbeeld lees- en redeneervaardigheden en dus uitsluitend rekenvaardigheid moest meten.

6.2 Test voor taalvaardigheid

Taalvaardigheid is gemeten met de Cito voortgangstoets begrijpend lezen. Deze toets kent verschillende versies met een oplopende moeilijkheidsgraad waardoor de toets als het ware meegroeit met leerlingen. Bij de start van het project hebben de leerlingen uit spil 17-18 de M5 toets gemaakt en de leerlingen uit spil 19-20 de M6 toets. Omdat Het Talent veel ruimte biedt voor individuele ontwikkeling, hebben sommige leerlingen een versie van de toets gemaakt die primair bedoeld is voor een oudere doelgroep. De scores van deze vooruitgetoetste leerlingen zijn omgezet naar de score voor de versie die zij op grond van hun leeftijd hadden moeten maken.

De voortgangstoets begrijpend lezen is jaarlijks afgenomen in de maand januari. De afname gebeurde schriftelijk en in een klassikale toetsopstelling: na een korte uitleg en enkele oefenopgaven lazen de leerlingen informatief-rapporterende en fictionele teksten en beantwoordden zij meerkeuzevragen. De M5 en M6 toets bevatte elk 50 opgaven, de M7 en M8 toets 55 opgaven. De toetsen zijn handmatig nagekeken waarna de ruwe scores zijn vertaald in een vaardigheidsscore die tussen verschillende toetsen en meetmomenten kon worden vergeleken. De vaardigheidsscores zijn bovendien gebruikt om de prestatiegroei te vergelijken met de landelijke referentiegroep.

6.3 Test voor rekenvaardigheid

Rekenvaardigheid is gemeten met de School-VaardigheidsToets HoofdRekenen (SVT-HR). De belangrijkste reden voor deze keuze is dat andere voortgangstoetsen indirect een beroep doen op leesvaardigheid of redeneervaardigheden. Bij de SVT-HR is dit niet het geval doordat leerlingen 'kale' sommen moeten oplossen; de test meet hierdoor de mate waarin rekenkundige bewerkingen als optellen, aftrekken, vermenigvuldigen en delen zijn geautomatiseerd.

De SVT-HR is jaarlijks afgenomen in de maand februari; dit gebeurde schriftelijk en klassikaal, met de bankjes in de toetsopstelling. Na een korte introductie kregen de leerlingen 5 minuten de tijd om zoveel mogelijk van de 200 opgaven te maken. Na afloop werd het aantal goede opgaven geteld en omgezet in een percentielscore omwille van de vergelijking met de landelijke referentiegroep.

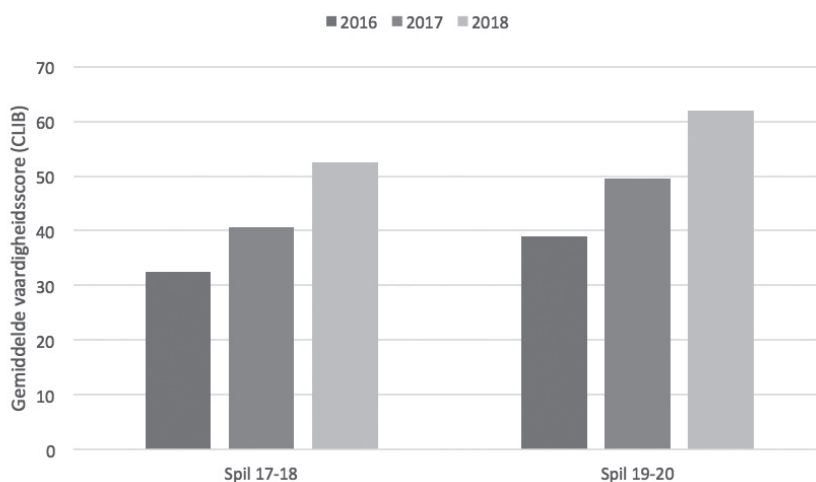
6.4 Resultaten

Figuur 6.1 geeft een overzicht van de vaardigheidsscores voor begrijpend lezen²⁷. In beide spellen werd een gelijkmatige stijging van 8 tot 12 punten gevonden. Wel was de spreiding in de vaardigheidsscores vrij groot, gemiddeld zo'n 14 punten, wat aangeeft dat niet elke leerling zich op hetzelfde moment en in hetzelfde tempo ontwikkelt. Deze verschillen werden niet veroorzaakt door sekse, achtergrond, leeftijd en sociaaleconomische status.

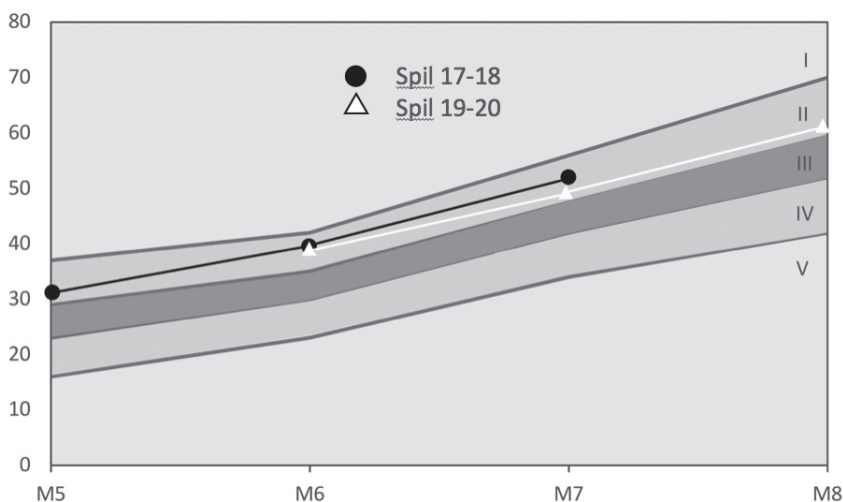
De vooruitgang ten opzichte van de landelijke referentiegroep is afgebeeld in Figuur 6.2.

Hierin is te zien dat de leerlingen van Het Talent gemiddeld genomen in niveaugroep II zitten, dus iets boven het landelijk gemiddelde. De leerlingen uit spil 17-18 gingen aanvankelijk meer vooruit dan hun leeftijdsgenoten: het verschil in vaardigheidsscores tussen de M5 en M6 toets was 8.2 punten tegen 6.5 als landelijk gemiddelde. Van M6 naar M7 was hun vooruitgang iets minder groot dan gemiddeld (11.5 versus 12.4 punten). Bij leerlingen uit spil 19-20 was dit eveneens het geval; hun groei tussen M7 en M8 laat echter zien dat de ontwikkeling zich daarna stabiliseerde. De prestaties op de SVT-HR zijn uitgedrukt

Figuur 6.1: Resultaten van de Cito voortgangstoets begrijpend lezen.



Figuur 6.2: Resultaten van de Cito voortgangstoets begrijpend lezen ten opzichte van de landelijk referentiegroep.



²⁷ Door uitval of tijdelijke afwezigheid hebben niet alle leerlingen de drie testafnames kunnen bijwonen. In 2016 ontbraken de scores van 4 leerlingen; in de daaropvolgende jaren waren dit respectievelijk 3 en 19 leerlingen.

in het aantal goed gemaakte opgaven die vervolgens zijn vertaald in een percentielscore. Figuur 6.3 toont de resultaten²⁸. In 2016 hadden de leerlingen uit spil 17-18 gemiddeld 61 opgaven goed. De laagste score was 17 terwijl de best presterende leerling 122 opgaven goed had. De gemiddelde score kwam overeen met het 70^e percentiel, wat betekent dat 70 procent van de landelijke referentiegroep de test minder goed had gemaakt. Een jaar later lag de gemiddelde score op Het Talent 11 punten hoger, wederom met een grote spreiding (28 – 120 punten). De percentielscore daalde echter van 70 naar 59. In 2018 werd wederom een toename van gemiddeld 11 punten genoteerd, wat ongeveer gelijk was aan de vooruitgang van de landelijke referentiegroep. De scores van de leerlingen op Het Talent liepen dat jaar uiteen van 41 tot 138 punten.

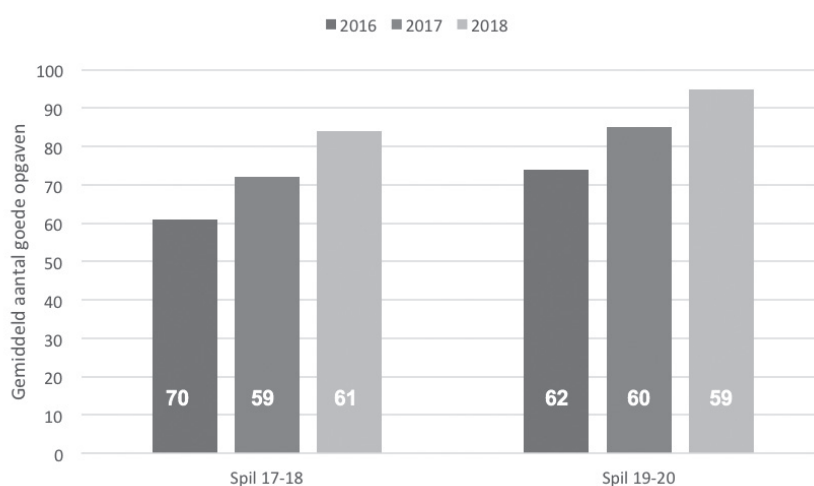
De resultaten voor de leerlingen uit spil 19-20 geven hetzelfde beeld, behalve dat de beginscores in 2016 op het 62^e percentiel lagen. In de jaren daarna stegen de gemiddelde scores met ongeveer 11 punten en werd gelijke tred gehouden met de landelijke referentiegroep. De spreiding in de scores was in alle drie de jaren aanzienlijk en vergelijkbaar met die van de leerlingen uit spil 17-18. Net als bij taalvaardigheid waren

de prestaties in beide spellen onafhankelijk van de leerlingkenmerken sekse, achtergrond, sociaaleconomische status en leeftijd.

6.5 Conclusies

De bovenbouwleerlingen van Het Talent hebben een relatief hoog niveau van taal- en rekenvaardigheid. Op alle meetmomenten vielen de gemiddelde scores in niveaugroep II, wat door het Cito als bovengemiddeld wordt beschouwd. Hieruit blijkt bovendien dat de ontwikkeling gelijke tred houdt met die van de landelijke referentiegroep, hoewel zich enkele kleine schommelingen voordeden. Op de SVT-HR liepen de prestaties van de leerlingen uit spil 17-18 aanvankelijk iets terug (maar dit verschil was niet significant) terwijl zij in diezelfde periode juist relatief veel vooruitgang boekten op de Cito toets begrijpend lezen. De leerlingen uit spil 19-20 behielden op beide toetsen hun niveau ten opzichte van de landelijke referentiegroep. Dit leidt tot de conclusie dat de leerlingen van Het Talent bovengemiddeld goed zijn in taal en rekenen en zich grotendeels in hetzelfde tempo door ontwikkelen als hun leeftijdsgenootjes van andere scholen.

Figuur 6.3: Resultaten van de SVT-HR. De getallen in de staven geven de percentielscore ten opzichte van de landelijk referentiegroep.



²⁸ Door tijdelijke afwezigheid ontbrak de score van 1 leerling in 2016; in de daaropvolgende jaren zorgde uitval en ziekte voor ontbrekende gegevens bij respectievelijk 6 en 19 leerlingen.

7. De vaardigheidsgroei in samenhang

7.1 Inleiding

Als leerlingen beter worden in taal of rekenen, worden ze dan ook beter in, bijvoorbeeld, kritisch denken? En verbeteren leerlingen hun 21^e-eeuwse vaardigheden net zo veel als hun taal- en rekenvaardigheid? Om dit soort vragen te beantwoorden, is de samenhang tussen de vaardigheidsgroei in de periode 2016 tot 2018 onderling vergeleken. Al deze analyses zijn uitgevoerd voor de testen die door de leerlingen individueel zijn gemaakt. Omdat 'samenwerken' op teamniveau is gemeten, kon deze vaardigheid niet betekenisvol worden vergeleken met de overige vaardigheden.

7.2 Samenhang tussen de vaardigheidsgroei

In de eerdere hoofdstukken is de ontwikkeling van elke vaardigheid apart beschreven. In deze paragraaf bekijken we of de mate waarin deze vaardigheden zich ontwikkelen met elkaar samenhangt. Hiervoor is eerst de toename van de scores op een bepaalde test tussen 2016 en 2018 berekend, wat leidde tot een maat voor

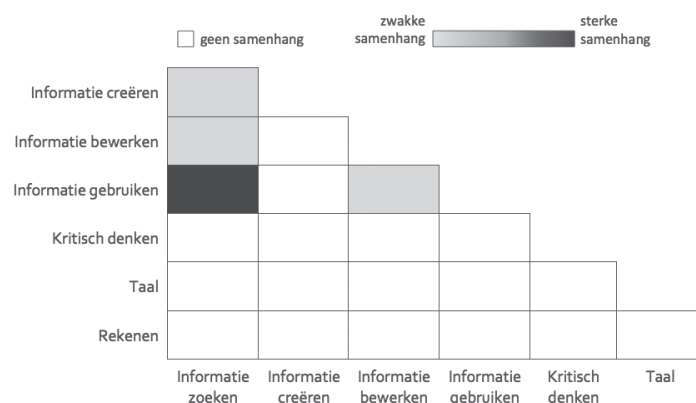
de vaardigheidsgroei op die test. Vervolgens is de groei van de afzonderlijke vaardigheden aan elkaar gerelateerd. De resultaten worden weergegeven in Figuur 7.1.

De matrix in Figuur 7.1 laat zien bij welke vaardigheden de groei met elkaar samenhangt. De sterkte van deze samenhang wordt weergegeven door een kleurtint: hoe donkerder de kleur des te sterker de (positieve) samenhang. Ter illustratie: de cel waarin 'informatie zoeken' en 'informatie gebruiken' samenkomen is donkerrood gekleurd, wat betekent dat de samenhang tussen deze onderdelen vrij sterk was. De samenhang tussen 'informatie zoeken' en 'informatie creëren' was daarentegen tamelijk zwak, wat te zien is aan de lichtrode kleur van deze cel. Als cellen geen kleur hebben, was de samenhang tussen de betreffende vaardigheidsgroei niet statistisch significant.

Tussen taal en rekenen werd geen samenhang in vaardigheidsgroei gevonden. Dit betekent dat niet alle leerlingen die, bijvoorbeeld, veel vooruitgang hadden geboekt op taalvaardigheid, ook een grote verbetering in


MEER WETEN
OVER DE
ANALYSES?
PAGINA 64

Figuur 7.1 Samenhang tussen de vaardigheidsgroei van 2016 tot 2018



rekenvaardigheid lieten zien. De ontwikkeling van taal- en rekenvaardigheid hing evenmin samen met de ontwikkeling van kritisch denken en digitale geletterdheid. De vaardigheidsgroei van deze twee 21^e-eeuwse vaardigheden was eveneens onafhankelijk. Dus: als leerlingen veel beter waren geworden in kritisch denken, dan betekende dat nog niet dat ze ook veel vooruit waren gegaan in digitale geletterdheid en omgekeerd. Bij drie van de vier onderdelen van digitale geletterdheid hing de vaardigheidsgroei wél met elkaar samen, hoewel de sterkte van deze verbanden vaak gering was. Het vierde onderdeel, ‘informatie creëren’, bleek wat betreft vaardigheidsgroei vrijwel helemaal onafhankelijk van de rest.

Vervolgens is ook de vaardigheidsgroei in de opeenvolgende jaren met elkaar gecorreleerd. Deze analyses geven antwoord op de vraag of leerlingen die van 2016 tot 2017 veel, of juist weinig vooruitgang boekten op een bepaalde vaardigheid, het jaar daarna eenzelfde vaardigheidsgroei laten zien. Dit bleek niet het geval te zijn. Sterker nog, voor alle vaardigheden werd een redelijk sterke *negatieve* samenhang gevonden. Hieruit blijkt dat alle vaardigheden zich met ‘sprongetjes’ ontwikkelen: als een leerling tussen 2016 en 2017 relatief weinig vooruitgang in bijvoorbeeld kritisch denken, dan was zijn of haar vooruitgang tussen 2017 en 2018 juist relatief groot. Deze schoksgewijze vaardigheidsgroei bleek

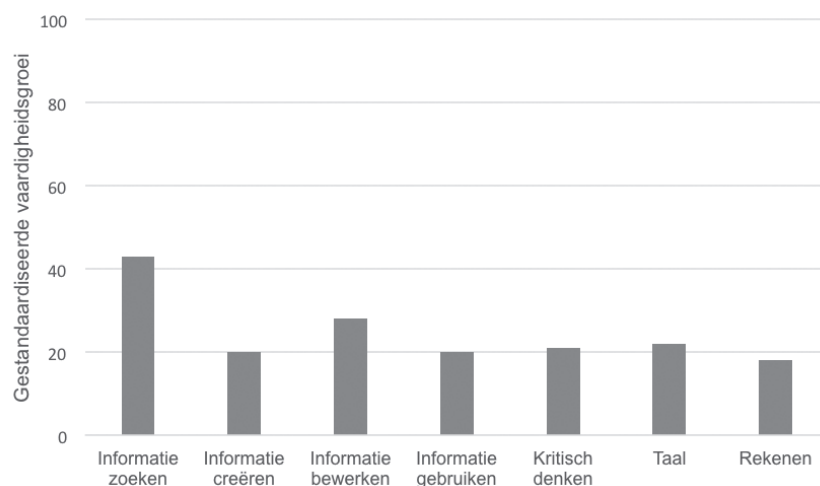
onafhankelijk van leeftijd en spil, wat aangeeft dat de vaardigheden zich niet bij elke leerling op hetzelfde moment even sterk ontwikkelen.

7.3 Verschillen in vaardigheidsgroei

Tot slot is de mate van vaardigheidsgroei onderling vergeleken. Omdat de maximale score van de testen verschilde, is de vaardigheidsgroei eerst omgerekend naar percentages. Hierbij is rekening gehouden met de beginscores uit 2016. Een voorbeeld. Stel dat een leerling 2 punten behaalde op de test voor kritisch denken in 2016. De maximale score op deze test was 12 punten, dus deze leerling kon dan nog 10 punten vooruit gaan. Dit noemen we de groeiruumte, en de resultaten in Figuur 7.2 laten zien welk deel van deze ruimte daadwerkelijk werd benut. Bij een toename van 2 punten, komt dit in ons voorbeeld neer op een gestandaardiseerde vaardigheidsgroei van 20 procent.

De staven in Figuur 7.2 laten zien dat de leerlingen relatief meer vooruit waren gegaan op het onderdeel ‘informatie zoeken’ dan op alle andere vaardigheden. Bij ‘informatie zoeken’ werd gemiddeld 43 procent van de groeiruumte benut, bij de overige vaardigheden lag dit percentage rond de 20 procent.

Figuur 7.2: Gestandaardiseerde vaardigheidsgroei over de periode 2016 – 2018. De hoogte van de staven laat zien hoeveel procent van de mogelijke groeiruumte werd benut.



7.4 Conclusies

De leerlingen hebben elk jaar een aantal testen gemaakt waarmee hun vaardigheden zijn gemeten. Hun vooruitgang op de verschillende testen vertoonde weinig samenhang. Een uitzondering zijn de onderdelen van digitale geletterdheid, waarop de vaardigheidsgroei wél onderling samenhang. Dit zou kunnen betekenen dat leerlingen die 'handig zijn met computers' beter worden in de meeste onderdelen van digitale geletterdheid.

Maar verder ontwikkelden de 21^e-eeuwse vaardigheden zich onafhankelijk van elkaar én van de basisvaardigheden in taal en rekenen. Deze ontwikkeling verliep met 'sprongetjes' en niet bij iedere leerling op hetzelfde moment. De ontwikkeling van de meeste vaardigheden hield gelijke tred, behalve bij 'informatie zoeken': de leerlingen zijn hierin relatief meer vooruit gegaan dan in alle andere vaardigheden.

8. Conclusies en aanbevelingen

8.1 Inleiding

De directie en leerkrachten van Het Talent waren benieuwd in hoeverre hun onderwijs, dat is gestoeld op een innovatief onderwijsconcept, leerlingen in staat stelt om 21^e-eeuwse vaardigheden te ontwikkelen. De school wilde ook graag weten hoe deze ontwikkeling zich verhoudt tot de ontwikkeling van de basisvaardigheden taal en rekenen én hoe de ontwikkeling wordt beïnvloed door algemene leerlingkenmerken zoals leeftijd, achtergrond en geslacht. Om deze vragen te beantwoorden zijn de bovenbouwleerlingen van Het Talent twee jaar lang gevolgd in hun ontwikkeling. In het tekstkader hebben we de belangrijkste resultaten nog even op een rij gezet.

8.2 Beantwoording van de onderzoeksvragen

De eerste drie resultaten uit het tekstkader geven een pasklaar antwoord op de eerste onderzoeksvraag. Op basisschool Het Talent ontwikkelen de bovenbouwleerlingen hun 21^e-eeuwse vaardigheden in hetzelfde tempo als hun basisvaardigheden in taal en rekenen. En hun taal- en rekenvaardigheden verbeteren jaarlijks even veel als bij leerlingen van reguliere basisscholen. Wel is het zo dat de 21^e-eeuwse vaardigheden zich onafhankelijk van de taal- en rekenvaardigheden ontwikkelen. Leerlingen die veel vooruitgaan op bijvoorbeeld taal of rekenen hoeven hun 21^e-eeuwse vaardigheden dus niet even veel te verbeteren. Sommigen doen dit misschien wel, maar anderen niet. Anders dan bij taal en rekenen kan de ontwikkeling van 21^e-eeuwse vaardigheden

niet worden vergeleken met leerlingen van andere scholen²⁹. Hierdoor kan de vaardigheidsgroei bij kritisch denken, samenwerken en digitale geletterdheid niet goed in perspectief worden geplaatst. Wel is in enkele verkennende studies de digitale geletterdheid van leerlingen op andere scholen onderzocht³⁰. De resultaten geven de indruk dat de prestaties van de leerlingen op Het Talent niet duidelijk beter of slechter zijn dan die op andere scholen. Harde conclusies kunnen hier niet aan worden verbonden omdat het om relatief kleine onderzoeken gaat die zijn uitgevoerd door studenten. Bovendien is de vaardigheidsgroei in het studentenonderzoek niet bekeken.

Desondanks wekt het scoreverloop over de jaren nu niet bepaald de indruk dat de 21^e-eeuwse vaardigheden zich in sneltreinvaart ontwikkelen. Op veel testen was de jaarlijkse vaardigheidsgroei tamelijk bescheiden en de prestaties tijdens de laatste meting in 2018 lagen vaak maar rond de helft van het maximaal aantal punten dat kon worden behaald. Misschien komt dit doordat de testen relatief moeilijk waren. Bij de start van het project waren er geen geschikte testen voorhanden om de 21^e-eeuwse vaardigheden te meten, zodat we onze meetinstrumenten zelf hebben moeten ontwikkelen³¹. Hierbij hebben we dankbaar gebruik gemaakt van vragenlijsten, opdrachten en inzichten uit andere projecten, maar die waren meestal bedoeld voor een oudere doelgroep. Dus misschien hebben we hierdoor de lat voor onze leerlingen wat te hoog gelegd. Een andere verklaring zou kunnen zijn dat Het Talent weliswaar veel ruimte biedt voor de ontwikkeling van 21^e-eeuwse vaardigheden, maar hier weinig tot geen expliciete instructie

29 Dit was bij de honorering van dit project al bekend; zie ook voetnoot 7.

30 Dit is het studentenonderzoek dat is beschreven op pagina 61 – 62.

31 Onze overwegingen hierbij zijn beschreven in een blog op <https://didactiefonline.nl/blog/blonz/talentontwikkeling-3>

en begeleide oefening in geeft. Misschien gaat de groei sneller als dit wel gebeurt?

Die ruimte voor 21^e-eeuwse vaardigheden staat de ontwikkeling van taal- en rekenvaardigheid niet in de weg. Dit is een mooi resultaat waarmee de kritiek wordt weerlegd dat kinderen op vernieuwingsscholen alleen maar ‘onzin’ leren en hopeloos achterlopen waar het taal en rekenen betreft³². Op Het Talent is dit zeker niet het geval, wat aangeeft dat je het onderwijs anders kunt organiseren en groepsdoorbrekend kunt werken zonder dat dit de leerprestaties negatief beïnvloedt. Welke elementen uit het PDKS onderwijsconcept hiervoor verantwoordelijk zijn, kan op basis van dit onderzoek niet worden gezegd. Maar we kunnen wel concluderen dat de PDKS als geheel succesvol bijdraagt aan de ontwikkeling van 21^e-eeuwse vaardigheden én taal- en rekenvaardigheden.

De tweede onderzoeksvraag ging over de invloed van leerlingkenmerken op de ontwikkeling van 21^e-eeuwse vaardigheden, taalvaardigheid en rekenvaardigheid. Uit de resultaten bleek dat zowel de prestaties op de testen als de vaardigheidsgroei behoorlijk verschilden tussen de leerlingen. Op zich is dit is niet verwonderlijk omdat het primair onderwijs een zeer diverse leerlingpopulatie heeft. Maar kunnen we een deel van de variatie in prestaties en vaardigheidsgroei verklaren door de kenmerken van de leerlingen? In dit project in elk geval niet: vrijwel alle resultaten waren onafhankelijk van de leeftijd van de leerlingen, de spil waarin ze zaten, hun achtergrond, geslacht, sociaaleconomische status en probleemoplossend vermogen. Voor scholen is dit goed nieuws, maar bij onderzoekers blijft de nieuwsgierigheid knagen (en bij leerkrachten hopelijk ook). Want waardoor worden de aanzienlijke individuele verschillen in prestaties en vaardigheidsgroei dan wél veroorzaakt? Dit is een interessante vraag voor vervolgonderzoek. Op basis van onze resultaten lijkt het zinvol om hierbij te kijken naar specifiekere leerlingkenmerken zoals bijvoorbeeld concentratievermogen, executieve

functies, en motivatie. Maar misschien komt het grillige scoreverloop bij deze jonge doelgroep ook voor een deel door de vorm van de dag...

8.3 Enkele opvallende resultaten

Bij het lezen van dit eindrapport heeft u vast een paar keer gedacht “hé, dat is opmerkelijk!”. Wij hadden dat ook en bespreken in deze paragraaf een aantal resultaten die er voor ons echt uitsprongen. De eerste ‘verrassing’ was de enigszins atypische samenstelling van de onderzoeksgroep. Uit de beschrijving in hoofdstuk 2 blijkt dat de leerlingen van Het Talent gemiddeld gezien aan de bovenkant van de samenleving zitten. De meeste leerlingen komen uit gezinnen waarvan de ouders hoogopgeleid zijn; het aantal leerlingen met een leerlinggewicht groter dan nul was dan ook verwaarloosbaar. Bovendien hadden de meeste leerlingen een Nederlandse achtergrond. Voor een vernieuwingsschool met een innovatief onderwijsconcept is dit misschien niet verwonderlijk, maar het betekent wel dat de leerlingen van Het Talent niet representatief zijn voor de doorsnee-leerling op een willekeurige basisschool. Hierdoor kunnen de resultaten uit dit project niet klakkeloos worden vertaald naar andere scholen, ook omdat de vaak grote individuele verschillen niet konden worden verklaard door de leerlingkenmerken.

Een tweede opvallende uitkomst van ons onderzoek is dat de vaardigheid in het zoeken van informatie op Internet zich meer ontwikkelde dan alle andere vaardigheden. Opmerkelijk, ook omdat dit niet gold voor de andere onderdelen van digitale geletterdheid. Omdat een van deze onderdelen met dezelfde geschreven test is gemeten, lijkt het niet waarschijnlijk dat het soort test hiervoor een verklaring is. Het lijkt aannemelijker dat leerlingen vaker iets opzoeken op Internet – niet alleen op school maar ook thuis – dan dat zij een Word bestand bewerken, een PowerPoint

32 Zie bijvoorbeeld <http://www.beteronderwijsnederland.nl>

presentatie ontwerpen of te maken krijgen met *pop-ups*. Misschien geldt hierbij wel de aloude stelregel ‘oefening baart kunst’.

Tot slot nog een korte bespiegeling over de samenhang tussen de vaardigheidsgroei bij digitale geletterdheid. Tussen de onderdelen ‘informatie zoeken’ en ‘informatie veilig gebruiken’ werd een hoge correlatie gevonden, wat waarschijnlijk komt doordat beide onderdelen met dezelfde vragenlijst en in dezelfde context zijn gemeten. Je zou dan ook verwachten dat de vaardigheidsgroei in het ‘creëren van informatie’ zou samenhangen met de vaardigheidsgroei in het ‘bewerken van informatie’, maar dit was niet het geval. Dit kan komen door de manier waarop de producten van de leerlingen zijn beoordeeld. Bij de Word opdracht is gekeken naar het aantal goede veranderingen, waarvoor uitsluitend ‘technische’ kennis over de commando’s van Word was vereist. Bij de beoordeling van de PowerPoint presentaties werd naast de technische kwaliteiten ook gelet op de inhoudelijke en grafische aspecten.

8.4 Aanbevelingen

Een onderzoeksrapport bevat bijna altijd een paar aanbevelingen voor verder onderzoek en het gebruik van de resultaten in de school of klas. Bij ons project is dit echter wat lastig omdat we een praktijkgericht onderzoek hebben uitgevoerd dat van nature niet is bedoeld om bij te dragen aan theorievorming en waarvan de resultaten niet zonder meer te generaliseren zijn naar scholen die niet aan het onderzoek hebben deelgenomen³³. Desondanks, maar met de nodige voorzichtigheid, willen we op basis van onze resultaten een aantal aanbevelingen doen. De meeste daarvan zijn direct voortgekomen uit de resultaten; een enkele aanbeveling komt uit onze ervaringen met het ontwerpen en afnemen van de testen. Alle aanbevelingen zijn puntsgewijs beschreven in het tekstkader hieronder, waarmee we tevens dit deel van het eindrapport afsluiten

AANBEVELINGEN

- Leerlingen simpelweg de ruimte bieden om 21^e-eeuwse vaardigheden te gebruiken is al voldoende om deze vaardigheden langzamerhand te verbeteren.
- Voor een grotere verbetering is waarschijnlijk gerichte uitleg en oefening nodig.
- Dit geldt vooral voor ‘evalueren’ en ‘interpreteren’ (kritisch denken) en ‘teamorganisatie’ (samenwerken) omdat de vooruitgang op deze onderdelen relatief traag verloopt. ‘Informatie zoeken’ (digitale geletterdheid) verbetert daarentegen vrij snel zodat hier geen prioriteit aan hoeft te worden gegeven – in elk geval niet klassikaal; aandacht voor individuele leerlingen kan nodig blijven.
- Meet de ontwikkeling van alle 21^e-eeuwse vaardigheden; leerlingen die zich goed ontwikkelen in de ene vaardigheid kunnen stagneren in andere vaardigheden.
- Meet 21^e-eeuwse vaardigheden met ‘passende’ meetinstrumenten: praktische opdrachten zijn motiverend en geven leerlingen niet het gevoel dat hun prestaties worden beoordeeld.
- Ga er niet van uit dat leerlingen wiens taal- en rekenvaardigheden zich goed ontwikkelen een soortgelijke ontwikkeling in 21^e-eeuwse vaardigheden doormaken.

33 o.a. Geijssel (2010), NRO (2014).

Klaar
voor
de
toekomst

Verdieping en
verantwoording

1. Probleemoplossend vermogen³⁴

1.1 Toren van Hanoi

De Toren van Hanoi is een taak waarmee het probleemoplossend vermogen kan worden gemeten. De taak bestaat uit een bord met daarop drie paaltjes en een aantal in grootte oplopende schijven. Leerlingen moet de schijven van een beginopstelling tot een eindopstelling brengen, terwijl zij zich aan de volgende regels houden:

- Er mag maar één schijf tegelijk worden verplaatst
- Een schijf mag niet naast het speelbord worden gelegd
- Een grotere schijf mag nooit op een kleinere schijf liggen

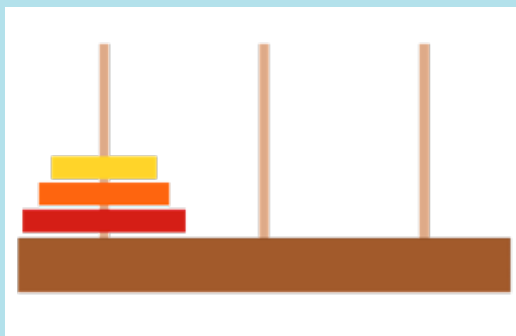
De taak is op de computer afgenomen en bestond uit 20 puzzels die in maximaal 7 minuten moesten worden gemaakt. De

beginopstelling verschilde per puzzel waardoor het aantal stappen dat nodig was om tot de eindopstelling te komen geleidelijk toenam van 3 tot 15 stappen. De eindopstelling was steeds gelijk: alle drie de schijven moesten op het rechterpaaltje staan. Als de eindopstelling was bereikt, ging de leerling automatisch door naar de volgende puzzel. Als de leerling veel stappen zette zonder tot de eindopstelling te komen, werd dezelfde puzzel opnieuw gestart.

Voordat de leerlingen met de taak begonnen, kregen ze een uitleg over de spelregels; hiervoor werd een houten versie van de puzzel gebruikt waarmee de proefleider een aantal puzzels voordeed. Hierna gingen de leerlingen achter hun computer zitten waarop de eerste puzzel al klaarstond. Tijdens de taak werden de acties van de leerlingen door de computer opgeslagen in een log file.

TOREN VAN HANOI

Mogelijke beginopstelling



Vaste eindopstelling



34 Bij het schrijven van dit hoofdstuk is dankbaar gebruik gemaakt van het werk van Erika Schlatter.

1.2 Resultaten

Het probleemoplossend vermogen is uitgedrukt in het aantal goed opgeloste puzzels. Zoals te zien in Tabel 1.1 hadden de leerlingen gemiddeld 12 puzzels opgelost. Het verschil tussen de spillen was minimaal. Dit kon niet worden gezegd van de spreiding: in beide spillen zaten leerlingen die vrijwel geen enkele puzzel konden oplossen maar ook leerlingen die bijna alle puzzels hadden opgelost.

Tabel 1.1: Beschrijvende statistieken van de Toren van Hanoi taak

	<i>N</i>	<i>M</i>	<i>SD</i>	Mediaan	Min.	Max.
Spil 17-18	86	11.67	3.03	12.00	1	18
Spil 19-20	77	12.03	2.78	12.00	4	18

Noot. De maximale score was 20 punten.



TERUG NAAR
DE HOOFDTEKST
PAGINA 14

2. Kritisch denken: keuze voor het soort test

Kritisch denken wordt vaak gemeten met een geschreven test. Een klassiek voorbeeld is de Cornell Critical Thinking Test Series³⁵ waarin studenten 78 vragen moeten beantwoorden die een beroep doen op hun vermogen om deductief te redeneren. Een van de vragen luidt:

Stel dat je weet dat:

Alle studenten die op het platteland wonen
huisdieren hebben
Barbara niet op het platteland woont

Zou de volgende uitspraak dan waar zijn?

Barbara heeft geen huisdieren

Omcirkel je antwoord:

A ja B nee C misschien

Andere testen zoals de Watson-Glaser Critical Thinking Appraisal³⁶ en de California Critical Thinking Skills Test CTST³⁷ bevatten meerkeuzevragen die zijn gesitueerd in een realistische context. Studenten lezen eerst een korte casus en beantwoorden daarna enkele meerkeuzevragen die een beroep doen op de kritisch denken vaardigheden die zijn genoemd in de vorige paragraaf. Een nog authentiekere manier om kritisch denken te meten is een zogeheten *performance assessment*. In plaats van vragen te beantwoorden, schrijven studenten een argumentatieve tekst (bijvoorbeeld een essay of een ingezonden brief naar een krant) waarin zij hun mening over een bepaald onderwerp geven (zoals: zou harde house muziek verboden moeten worden?) en met argumenten onderbouwen. Dit soort opdrachten wordt gebruikt in de Illinois Critical Thinking Essay test³⁸ en de Ennis-Weir Critical Thinking Essay test³⁹.

Onderzoeksvaardigheden worden vaak op eenzelfde manier gemeten. De Integrated Process Skills test (TIPS, versie II)⁴⁰ bevat 36 meerkeuzevragen over het identificeren van variabelen, het operationeel definiëren van die variabelen, hypothesen opstellen, data en grafieken interpreteren en experimenten ontwerpen. Elke vraag wordt voorafgegaan door een korte casus of situatieschets. Voor het experimenteren en interpreteren van resultaten zijn eveneens schriftelijke testen beschikbaar waarin kinderen bijvoorbeeld moeten aangeven of een onderzoeksopzet valide is, en wanneer dit volgens hen niet zo is, moeten aangeven hoe de opzet kan worden verbeterd⁴¹.

Onderzoeksvaardigheden kunnen ook worden gemeten door leerlingen (een deel van) een experiment te laten bedenken en uitvoeren. In de literatuur zijn talloze beschrijvingen van dit soort *performance assessments* te vinden. In het hierboven geciteerde onderzoek van Piekny en Maehler (2013) moesten kinderen bijvoorbeeld beoordelen of rode dan wel groene kauwgum de oorzaak is van tandbederf; tijdens een andere opdracht moesten zij op basis van uiterlijke kenmerken voorspellen welke fictieve dieren tot een bepaalde familie behoren.

Samengevat blijkt dat kritisch denken en onderzoeksvaardigheden op een soortgelijke manier worden gemeten. Grofweg kan hierbij onderscheid worden gemaakt tussen meer en minder authentieke meetmethoden, waarbij *performance assessments* het meest authentiek zijn en testen met meerkeuzevragen het minst authentiek. Voor de doelgroep van dit project zijn deze laatste testen wellicht minder

35 Ennis et al. (1964)

36 Watson & Glaser (1964)

37 Insight Assessment (2013)

38 Finken & Ennis (1993)

39 Ennis & Weir (1985)

40 Okey et al. (1982)

41 Chen & Klahr (1999); Kruit et al. (2018); Mayer et al. (2014)



geschikt: zeker voor de jongste leerlingen is het de vraag of zij in staat zijn de vragen zelfstandig te begrijpen en te beantwoorden. Daarnaast zijn de meeste geschreven testen alleen gevalideerd voor kinderen van 14 jaar en ouder. De voorkeur ging daarom uit naar een *performance assessment*.

Kijkend naar de leeftijd van de leerlingen (bij aanvang 7 tot 10 jaar) lijkt het schrijven van argumentatieve teksten een minder geschikte manier om hun vermogen tot kritisch denken te meten. Enerzijds zijn deze teksten wat minder passend bij W&T-onderwijs, anderzijds zijn dit soort essay testen bestemd voor en gevalideerd bij een oudere doelgroep (meestal

jong volwassenen). Hoewel het mogelijk is om goede scoringsregels te formuleren waarmee de teksten geanalyseerd kunnen worden, kan van de meeste basisschoolleerlingen niet verwacht worden dat zij 'echte' argumentatieve teksten kunnen schrijven. Hun score zou daarom lager uitvallen dan hun daadwerkelijke vermogen tot kritisch denken. Dit leidt tot de conclusie dat kritisch denken in dit project het best kan worden gemeten door leerlingen onder begeleiding een onderzoek te laten bedenken en uitvoeren waarbij zij hun handelingen, beslissingen en argumenten verbaal kunnen uiten.

3. Kritisch denken: ontwerp van de test

3.1 Inhoud van de test

De test voor kritisch denken bevatte vragen over de vaardigheden induceren, interpreteren, evalueren en infereren. In Tabel 3.1 is te zien hoe deze vaardigheden zijn 'vertaald' naar onderzoeksvaardigheden en wat de inhoud van de vragen was. Uit de omschrijvingen wordt bovendien duidelijk hoe de complexiteit van de vragen per vaardigheid toeneemt. In de test zijn de vragen in oplopende moeilijkheidsvolgorde geplaatst; de eerste vragen waren dus relatief eenvoudig en de laatste vragen relatief complex.

In totaal zijn de leerlingen drie keer getest; één keer per schooljaar. Om validiteitsbedreigers als *history* en *content effects* te voorkomen, zijn drie versies van de test ontwikkeld die uitsluitend van elkaar verschilden op grond van het onderwerp dat werd onderzocht (rollende ballen, stuiterende ballen, rijdende auto's). Voorafgaand aan de eerste meting werd door middel van *counterbalancing* bepaald wanneer welke leerling welke versie kreeg. Per jaar kreeg ongeveer een derde van de leerling een bepaalde versie.

Tabel 3.1: Inhoud van de test voor kritisch denken

Vaardigheid	Vragen
Induceren	Gebruik de resultaten van één experiment om de uitkomsten van een soortgelijk experiment te voorspellen
	Gebruik inconsistente resultaten uit een serie identieke experimenten om de uitkomsten van een soortgelijk experiment te voorspellen
	Gebruik inconsistente resultaten uit twee series identieke experimenten om de kans op een bepaalde uitkomst te voorspellen
Interpreteren	Interpreteer de zichtbare uitkomst van één experiment
	Interpreteer de grafische uitkomsten van een serie identieke experimenten
	Interpreteer de numerieke uitkomsten van twee series identieke experimenten
Evalueren	Beoordeel de betrouwbaarheid van een bewering op basis van het aantal waarnemingen
	Beoordeel de betrouwbaarheid van de resultaten op basis van de variatie binnen een serie waarnemingen
	Beoordeel de betrouwbaarheid van de resultaten op basis van de variatie tussen twee series waarnemingen
Infereren	Doe geldige uitspraken op basis van de uitkomsten van één experiment
	Geef de implicaties voor een hypothese op basis van een uitputtende serie experimenten
	Doe geldige uitspraken op basis van de uitkomsten van twee series identieke experimenten

3.2 Proefafnames

Tijdens het ontwerp van de test zijn twee proefafnames gedaan; eerst met vijf leerlingen van Het Talent (11 en 12 jaar oud) die geen deel uitmaakten van de onderzoeksgroep, en een paar maanden daarna met vier 11-jarige leerlingen van een andere basisschool. Na elke proefafname werd de test (i.c. het script voor de proefleider) bijgesteld op basis van de ervaringen. Dit betrof meestal het herformuleren van een vraag of het splitsen van een vraag in twee deelvragen als bleek dat de leerlingen de essentie van de vraag niet begrepen. Ook werden enkele kleine foutjes gecorrigeerd.

3.3 Kwaliteit van de test

De eerste afname van de test, in 2016, is gebruikt om de psychometrische eigenschappen van de test te bepalen. Tabel 3.2 toont de itemkarakteristieken. De gemiddelde moeilijkheidsgraad van de testvragen was .34, met p -waardes die varieerden tussen .01 en .94. Op zich is deze grote spreiding niet verwonderlijk omdat de vragen een oplopende moeilijkheid zouden moeten hebben. Dit bleek inderdaad uit de p -waardes: het gemiddelde van alle eerste vragen was .59, en daalde naar .27 voor de tweede vragen en verder naar .16 voor de derde vragen. Deze geleidelijke toename van complexiteit deed zich voor bij alle vaardigheden behalve 'induceren'.

De discriminatie-index geeft aan in hoeverre een vraag onderscheid kan maken tussen hoog- en laag-presterende leerlingen. Een r_{pb} -waarde van boven de .20 wordt over het algemeen als voldoende beschouwd. Dit was het geval bij 11 van de 12 vragen, waarbij bovendien geldt dat ook vragen met een hoge of lage p -waarde nog voldoende onderscheidend waren. De enige uitzondering was vraag 3 bij het onderdeel 'evalueren'.

Vanwege de oplopende moeilijkheidsgraad van de vragen kon de betrouwbaarheid (of interne consistentie) van de test niet worden bepaald. Een bijkomende reden is dat de gemeten vaardigheden zich niet vanaf hetzelfde moment en in hetzelfde tempo ontwikkelen; op zoek gaan naar interne consistentie is dan weinig zinvol. De betrouwbaarheid van de codering van de antwoorden kon wel worden bepaald. Hiervoor hebben twee onafhankelijke beoordelaars de antwoorden van 22 willekeurig gekozen leerlingen gescoord. Van de in totaal 357 antwoorden kwamen de scores 329 keer overeen (92%), wat neerkomt op een Cohen's κ van .84.

Tot slot zijn de prestaties op de drie versies van de test vergeleken. De gemiddelde scores op de 'rollende ballen' versie ($M = 4.33$, $SD = 1.63$) was iets hoger dan de scores op de 'stuiterende ballen' versie ($M = 4.05$, $SD = 1.68$) en de 'rijdende auto's' versie ($M = 3.86$, $SD = 1.93$), maar de onderlinge verschillen tussen deze drie scores waren niet significant, $F(2, 177) = 1.08$, $p = .341$.

Tabel 3.2: Moeilijkheidsgraad en discriminatie-index van de testvragen

	Induceren		Interpreteren		Evalueren		Infereren	
	p	r_{pb}	p	r_{pb}	p	r_{pb}	p	r_{pb}
Vraag 1	.35	.27	.94	.32	.43	.23	.62	.45
Vraag 2	.29	.38	.34	.33	.39	.40	.07	.21
Vraag 3	.40	.52	.15	.21	.01	.09	.09	.38

4. Statistische analyses van de scores voor kritisch denken

4.1 Vaardigheidsgroei

Tabel 4.1 geeft informatie over de prestaties van de leerlingen uit beide spillen in jaren 2016 tot 2018. Omdat uit onderzoek blijkt dat de kritisch denken vaardigheden zich vanaf een verschillend moment en in een verschillend tempo ontwikkelen⁴², zijn ze per periode afzonderlijk geanalyseerd met mixed-design MANOVA's. In alle analyses werd voldaan aan de eisen van normaliteit en homogeniteit.

Bij de 2016-2017 vergelijking zijn multivariate effecten gevonden van de *within-subject* factor

tijd, $F(4, 174) = 18.96, p < .001, \eta^2 = .30$, van de *between-subject* factor spil, $F(4, 174) = 4.60, p = .001, \eta^2 = .10$, en geen significant interactie effect, $F(4, 174) = 0.07, p = .992, \eta^2 = .00$. Dit betekent dat de leerlingen uit spil 19-20 over het geheel genomen beter presteerden dan leerlingen uit spil 17-18, en dat de prestaties in beide spillen daadwerkelijk en evenredig waren verbeterd. Uit de univariate analyses blijkt dat deze verbetering zich voordeed bij alle vaardigheden, $F(1, 177) > 11.03, p < .001, \eta^2 > .05$, met uitzondering van interpreteren, $F(1, 177) = 3.83, p = .052, \eta^2 = .02$.

Tabel 4.1: Beschrijvende statistieken van de scores op de kritisch denken test

	2016			2017			2018		
	N	M	SD	N	M	SD	N	M	SD
Induceren									
Spil 17-18	90	0.98	0.92	89	1.25	0.86	86	1.79	0.87
Spil 19-20	91	1.09	0.88	91	1.38	0.90	80	1.74	0.91
Interpreteren									
Spil 17-18	90	1.44	0.75	89	1.58	0.64	86	1.59	0.63
Spil 19-20	91	1.40	0.60	91	1.52	0.71	80	1.54	0.66
Evalueren									
Spil 17-18	90	0.67	0.64	89	1.06	0.76	86	1.05	0.62
Spil 19-20	91	0.98	0.70	91	1.41	0.67	80	1.36	0.72
Infereren									
Spil 17-18	90	0.72	0.69	89	1.02	0.69	86	1.33	0.94
Spil 19-20	91	0.84	0.63	91	1.18	0.79	80	1.39	0.80

Noot. De maximale score voor elke vaardigheid was 3 punten.

42 o.a. Koerber et al. (2015); Piekny & Maehler (2013, 2014)

Ook bij de 2017-2018 vergelijking werd een multivariaat effect gevonden van tijd, $F(4, 161) = 10.52, p < .001, \eta^2 = .21$. Het effect van spel was niet significant, $F(4, 161) = 2.39, p = .053, \eta^2 = .10$, en de tijd $\times$ spel interactie evenmin, $F(4, 161) = 0.58, p = .681, \eta^2 = .01$. Dit betekent dat de prestaties in beide spellen daadwerkelijk en evenredig waren verbeterd, terwijl de jongere leerlingen ditmaal niet onderdeden voor de oudere. Uit de univariate analyses blijkt dat de scorewinst zich voordeed bij de vaardigheden 'voorspellen', $F(1, 161) = 34.70, p < .001, \eta^2 = .18$, en 'infereren', $F(1, 161) = 9.38, p = .003, \eta^2 = .05$. De vooruitgang op de overige vaardigheden was niet significant, $F(1, 161) < 1.41, p > .237, \eta^2 < .01$.

4.2 Relatie met leerlingkenmerken

Voor de analyses met leerlingkenmerken zijn *gain* scores berekend door de totaalscores van twee opeenvolgende jaren van elkaar af te trekken. De gemiddelde vooruitgang tussen 2016 en 2017 was 1.13 punten ($SD = 1.98$); het jaar daarop was de vooruitgang wat lager ($M = 0.78, SD = 2.14$). In Tabel 4.2 is te zien dat beide scores onafhankelijk waren van sekse en achtergrond.

De overige leerlingkenmerken zijn uitgedrukt in continue variabelen op ordinaal- of intervalniveau⁴³. Hiervoor zijn correlaties berekend met de beide *gain* scores; de resultaten in Tabel 4.3 laten zien dat geen van deze kenmerken samenhang met de vooruitgang in kritisch denken.

Tabel 4.2: Vooruitgang op de kritisch denken test naar sekse en achtergrond

	Vooruitgang 2016 - 2017					Vooruitgang 2017 - 2018				
	<i>M</i>	<i>SD</i>	<i>F</i>	d.f.	<i>p</i>	<i>M</i>	<i>SD</i>	<i>F</i>	d.f.	<i>p</i>
Sekse			0.72	1,177	.397			1.27	1,164	.262
Jongen	1.25	2.01				0.61	2.11			
Meisje	1.00	1.92				0.99	2.18			
Achtergrond			0.01	1,174	.909			2.66	1,161	.105
Nederlandse	1.10	2.00				0.70	2.10			
Migratie	1.15	1.83				1.46	2.17			

Tabel 4.3: Correlaties tussen de vooruitgang op de kritisch denken test en

	<i>M</i>	<i>SD</i>	Vooruitgang 2016 - 2017	Vooruitgang 2017 - 2018
Leeftijd (maanden)	107.38	9.34	-.05	.11
Achtergrond¹	--	--	.06	.01
Probleemoplossend vermogen	11.84	2.91	.01	.06

¹ Spearman's rho

⁴² Dit geldt ook voor het leerlinggewicht, maar omdat er slechts vier leerlingen waren met een gewicht dat hoger was dan nul, is dit kenmerk niet geanalyseerd.

5. Digitale geletterdheid: inhoud van de testen

Hieronder staat een korte omschrijving van de onderdelen uit het ICILS raamwerk die in het project zijn gemeten.

Het zoeken, vinden en evalueren van digitale informatie

Bij dit onderdeel gaat het om informatievaardigheden. Een leerling is informatievaardig als hij of zij informatie kan zoeken en vinden op Internet en de relevantie, kwaliteit en bruikbaarheid ervan kan beoordelen. Het omgaan met een browser en een zoekmachine is hiervoor een eerste vereiste; daarnaast wordt een beroep gedaan op vaardigheden zoals het filteren van irrelevante zoekresultaten, het lokaliseren van relevante informatie op een webpagina, het beoordelen van de betrouwbaarheid van een Internetbron en het controleren van gevonden gegevens door een cross-check uit te voeren⁴⁴.

Het bewerken van digitaal verkregen informatie

Bij dit onderdeel draait het om het verbeteren van de manier waarop reeds bestaande informatie wordt weergegeven. Voorbeelden zijn het (duidelijker) accentueren van de paragraaftitels in een tekst, het bewerken van afbeeldingen en het vertalen van getallen uit een tabel naar een diagram of grafiek.

Het creëren van digitale informatie

Anders dan bij het vorige onderdeel gaat het hier om het bedenken en maken van

nieuwe digitale producten zoals een verjaardagskalender, een PowerPoint presentatie of een geïllustreerde tekst. De vaardigheden die hiervoor nodig zijn, zijn zowel technisch als cognitief van aard. Om een goede presentatie te kunnen maken, moeten leerlingen niet alleen de meer geavanceerde opties van PowerPoint kunnen gebruiken, maar moeten ze ook in staat zijn om hun presentatie logisch op te bouwen en belangrijke informatie op een korte, heldere manier te verwoorden of verbeelden.

Het veilig en zorgvuldig gebruiken van digitale informatie

Bij dit laatste onderdeel draait het om mediawijsheid: het op een juridisch en ethisch verantwoorde manier omgaan met online informatie. Hieronder valt een breed scala aan onderwerpen zoals het weloverwogen verstrekken van en omgaan met persoonlijke gegevens, het belang van virusscanners en veilige wachtwoorden, het herkennen van 'phishing' en spam-berichten en het respecteren van copyrights. Tabel 5.1 geeft een overzicht van de onderdelen en bijbehorende vaardigheden uit de testen. Bij de meeste onderdelen hebben de vaardigheden een oplopende moeilijkheidsgraad. Zo is het zoeken op één trefwoord relatief eenvoudig, het zoeken op twee trefwoorden iets ingewikkelder, en het geavanceerd zoeken door booleaanse operatoren te gebruiken het meest complex.

⁴⁴ Zie verder Lazonder & Rouet (2008) en Walraven (2008)



TERUG NAAR
DE HOOFDTEKST
PAGINA 19

Tabel 5.1: Inhoud van de testen voor digitale geletterdheid

Categorie	Onderdelen en vaardigheden
Het verzamelen en beheren van digitale informatie	<i>Het zoeken, vinden en evalueren van digitale informatie</i>
	Browser en zoekmachine bedienen
	Zoeken op één trefwoord
	Zoeken op twee trefwoorden
	Geavanceerd zoeken
	Zoekresultaten beoordelen
	Informatie op een website beoordelen
	<i>Het bewerken van digitaal verkregen informatie</i>
	Lettergrootte veranderen
	Tekstopmaak veranderen
Het produceren en delen van digitale informatie	Kleur van de tekst veranderen
	Kantlijn van één paragraaf veranderen
	Eerste regel van een paragraaf laten inspringen
	Regelafstand veranderen
	Tekstterugloop van een afbeelding veranderen
	<i>Het creëren van digitale informatie</i>
	PowerPoint presentatie maken
	<i>Het veilig en zorgvuldig gebruiken van digitale informatie</i>
	Correct reageren op pop-up berichten
	Sociale normen gebruiken bij het posten van een online bericht
	Correct reageren op 'grooming' berichten

6. Digitale geletterdheid: keuze voor het soort test

In grootschalige vergelijkingsstudies wordt digitale geletterdheid doorgaans gemeten met een vragenlijst of survey. Respondenten krijgen een aantal stellingen voorgelegd (bijvoorbeeld: “ik kan een email versturen”) en moeten op een rangordeschaal aangeven hoe vaardig zij zichzelf vinden in het uitvoeren van die taak. Meestal wordt hiervoor gebruik gemaakt van een 5-puntschaal die varieert van “helemaal niet mee eens” tot “helemaal mee eens”⁴⁵. Maar het kan ook anders: in het ICILS project werden bijvoorbeeld de volgende drie antwoordmogelijkheden gegeven: “ik weet hoe dit moet”, “ik weet het nog niet, maar ik kan er wel achter komen” en “ik denk niet dat ik dit kan”⁴⁶. Om de perceptie van de eigen vaardigheden in perspectief te plaatsen, wordt in veel survey’s ook gevraagd naar de frequentie en duur van computer- en internetgebruik; soms wordt ook gevraagd naar het zelfvertrouwen bij het uitvoeren van de genoemde taken.

Uiteraard is het ook mogelijk om digitale geletterdheid op een meer directe manier te meten. Dit gebeurt in de regel door leerlingen vragen te stellen of opdrachten te laten uitvoeren. In het Northstar Digital Literacy Project⁴⁷ zijn een aantal online modules ontwikkeld waarmee volwassenen hun digitale vaardigheden kunnen testen. Elke module bestaat uit een 30-tal korte vragen, vaak geïllustreerd met schermafbeeldingen, die beantwoord moeten worden door iets aan

te klikken (bijvoorbeeld het icoon van een Excel bestand), te typen (bijvoorbeeld: geef het bestand de naam ‘mijnverhaal.doc’) of met de muis te slepen.

Een goed voorbeeld van het gebruik van opdrachten is de Online Reading Comprehension Assessment (ORCA); een instrument waarmee online geletterdheid wordt gemeten⁴⁸. Leerlingen krijgen een serie korte opdrachten, bijvoorbeeld het openen van een email of het kopiëren van relevante informatie van een webpagina, die zij achter de computer uitvoeren. In het ICILS project werden dit soort korte opdrachten ook gebruikt; daarnaast werkten leerlingen aan een omvangrijke open opdracht waarin diverse aspecten van digitale geletterdheid samenkwamen. Een voorbeeld van zo’n opdracht is het maken van een informatiefolder voor een schoolreisje.

Bij de keuze voor een bepaald soort test wordt meestal een afweging gemaakt tussen authenticiteit en haalbaarheid. Wanneer veel personen moeten worden getest of de meting in korte tijd moet worden uitgevoerd, is een survey met perceptievragen de meest haalbare optie. Hoewel een survey weinig authentiek is, geven de uitkomsten in sommige gevallen wel een goed beeld van de daadwerkelijke vaardigheden van volwassen respondenten⁴⁹. Of dit ook geldt voor basisschoolleerlingen is echter niet bekend. Om die reden werd een perceptievragenlijst minder geschikt gevonden

⁴⁵ Zie bijvoorbeeld Van Deursen et al. (2014)

⁴⁶ Meelissen et al. (2014)

⁴⁷ <https://www.digitalliteracyassessment.org/>

⁴⁸ Coiro (2006)

⁴⁹ Van Deursen et al. (2012), Hargittai (2005)



voor de meting van digitale geletterdheid in dit project⁵⁰. Toepassingsvragen en opdrachten leken meer geschikt maar hebben als mogelijk nadeel dat de meting veel tijd kost, zeker wanneer aan een open opdracht wordt gewerkt. Bovendien vergt het beoordelen van dit soort opdrachten veel tijd. Dit leidt tot de conclusie dat digitale geletterdheid in dit project het best kan worden gemeten door leerlingen toepassingsvragen te laten beantwoorden

en aan korte, gestructureerde opdrachten te laten werken. De vragen kunnen zowel op papier als op de computer worden gemaakt; de opdrachten worden vanzelfsprekend achter de computer uitgevoerd. Om de haalbaarheid te vergroten, zullen de testen zo worden ontworpen dat het gedeelte achter de computer in kleine groepjes kan worden afgenomen en de antwoorden snel en eenduidig kunnen worden beoordeeld.

50 Uit het onderzoek van Bedi Tekirdag, student psychologie aan de Universiteit Twente, bleek dat dit waarschijnlijk een goede keuze is geweest. Hij vond dat leerlingen hun vaardigheden op de opdrachten die in het project zijn gebruikt (veel) hoger inschatten dan deze in werkelijkheid zijn. Het onderzoek van Bedi is begeleid door Noortje Janssen.

7. Digitale geletterdheid: ontwerp van de testen

7.1 Geschreven test

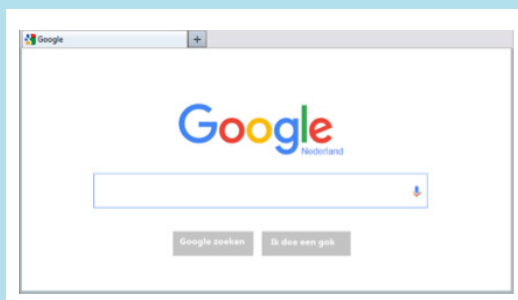
Voor de onderdelen 'zoeken, vinden en evalueren van digitale informatie' en 'veilig en zorgvuldig gebruiken van digitale informatie' is een geschreven test ontwikkeld die uit 17 toepassingsvragen bestond. De eerste 14 vragen gingen over 'informatie zoeken', de laatste 3 vragen over 'informatie gebruiken' (mediawijsheid). Alle vragen konden beantwoord worden door iets op te schrijven of aan te kruisen. Omdat de leerlingen deze test drie keer hebben gemaakt, steeds met tussenpozen van een jaar, zijn drie versies van de test ontwikkeld die alleen verschilden wat betreft het onderwerp waarover informatie werd gezocht. De toepassingsvragen in de drie versies gingen over het maken van een spreekbeurt over respectievelijk onweer, aardbevingen en orkanen. Door de verandering van onderwerp verschilde ook de inhoud van de zoekopdrachten, de zoekresultaten en de getoonde websites. De aard van de vragen bleef echter onveranderd, zoals in Figuur 7.1 wordt geïllustreerd.

Voor de vragen zijn schermafbeeldingen van fictieve browsers en zoekmachines gebruikt die sterk lijken op (maar niet altijd gelijk zijn aan) de gangbare zoekhulpmiddelen. Hiermee werd voorkomen dat de test moest worden aangepast als de interface van de browser en zoekmachine die op Het Talent wordt gebruikt (Internet Explorer, Google Chrome, en Google) gedurende de looptijd van het project zou veranderen. Een bijkomend voordeel was dat de test hierdoor beter in staat is om de daadwerkelijk aanwezige kennis en vaardigheden te meten; leerlingen die thuis of op school gedachteloos internetten door voor hen betekenisloze routines uit te voeren, zouden naar verwachting laag presteren op de test.

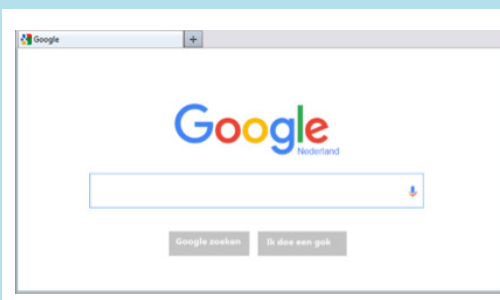
Voor deze test is een antwoordsleutel met scoringsregels opgesteld. Waar mogelijk zijn de zoekopdrachten met Google uitgevoerd om de juistheid van sommige antwoorden te controleren, bijvoorbeeld bij de vraag over het schilderij van Donar of Aeolus uit het tekstkader. De meeste vragen werden als

Figuur 7.1: Voorbeeldvraag uit twee versies van de geschreven test

Bij onweer hoort bliksem. Vroeger dachten mensen dat Donar de god van de donder en bliksem was. Zoek met Google naar een schilderij van Donar. Schrijf in het plaatje hoe je dat doet.



Bij een orkaan hoort wind. Vroeger dachten mensen dat Aeolus de god van de wind was. Zoek met Google naar een schilderij van Aeolus. Schrijf in het plaatje hoe je dat doet.



goed (1 punt) of fout (0 punten) beoordeeld. Voor vragen waar delen van het antwoord goed of fout konden zijn, werd met halve of kwart punten gewerkt zodat een volledig correct antwoord ook hier 1 punt opleverde.

7.2 Opdrachten

Voor het onderdeel 'creëren van informatie' moesten de leerlingen een PowerPoint presentatie maken voor een wedstrijd over de leukste school, het leukste boek, of het lekkerste gerecht van Nederland. De wedstrijd werd aangekondigd in een *flyer* waarin het doel werd uitgelegd en enkele globale aanwijzingen werden gegeven zodat leerlingen een indruk

konden krijgen van het soort presentatie dat ze moesten maken (zie Figuur 7.2).

De PowerPoint presentaties zijn beoordeeld op hun technische, grafische en inhoudelijke kwaliteit. Elke dimensie bestond uit drie aspecten die zijn gescoord aan de hand van een codeerschema. De *technische kwaliteit* had betrekking op het gebruik van de mogelijkheden van PowerPoint om tekst op te maken, afbeeldingen vorm te geven, en animaties of dia-overgangen toe te voegen. Bij de *grafische kwaliteit* werd gelet op de esthetische kwaliteiten van de presentatie als geheel, de vormgeving van de afzonderlijke dia's en de elementen op die dia's. De *inhoudelijke kwaliteit* betrof de informatie over het onderwerp, de manier waarop die informatie

Figuur 7.2: Voorbeeld van een flyer bij de PowerPoint opdracht



WEDSTRIJD DE LEUKSTE SCHOOL VAN NEDERLAND

Is leren saai? Niet op jouw school!
Maar wat maakt jouw school zo leuk? Is het de meester of juf, de schoolreisjes, of wat anders?

Maak een mooie presentatie met PowerPoint om te vertellen waarom jouw school de leukste school van Nederland is.

Wil je weten hoe je een goede presentatie maakt? Lees dan de **tips** in de groene balk naast de foto.

Stichting
LEUKER LEREN

TIPS

Vertel wie jij bent en op welke school je zit

Vertel waarom jouw school zo leuk is

Geef voorbeelden

Gebruik foto's, kleuren en animaties om jouw presentatie zo mooi mogelijk te maken

**STICHTING
LEUKER LEREN**
Grotestraat 23
7459 PG Enschede
www.leukerleren.nl

was gestructureerd en het taalgebruik. Voor elk aspect kon maximaal 3 punten worden gehaald, wat leidt tot een maximum score van 9 punten per dimensie. Hierdoor kon de totaalscore voor deze opdracht variëren tussen de 0 en 27 punten.

Voor de opdracht over het 'bewerken van informatie' zijn drie eenvoudige teksten geselecteerd van respectievelijk 266, 215 en 192 woorden. De structuur van de teksten was gelijk en bestond uit een titel, een auteursregel, een *lead*, drie paragrafen (waarvan de laatste twee een titel hadden) en een afbeelding. De teksten waren ontleend aan de website van het tijdschrift Kidsweek⁵¹ en gingen over de Nationale Hoortest, een Internetdiploma voor kinderen en het instellen van een 'raad van kinderen' door bedrijven. De leerlingen kregen de tekst in een Word bestand en moesten de opmaak veranderen. Hiervoor kregen ze zeven opgaven, die correspondeerden met de vaardigheden uit Tabel 5.1. Met behulp van een codeerschema is elke verandering gescoord als goed of fout. Per goede verandering kon 1 punt worden gehaald waardoor de maximum score voor deze opdracht 7 punten bedroeg.

Beide opdrachten zijn uitgevoerd op de computers van Het Talent waarop aanvankelijk Windows 7 en Office 2010 was geïnstalleerd. In de zomer van 2017 is deze software ge-update naar Windows 10 en Office 365, maar dit had geen gevolgen voor de opdrachten en de scoring van de producten van de leerlingen.

7.3 Proefafnames

De geschreven test is gemaakt door 21 leerlingen van Het Talent (9 tot 11 jaar oud) die niet meededen aan de longitudinale meting. Zij hadden ongeveer 30 minuten nodig om de vragen te beantwoorden en behaalden gemiddeld 10.21 van de 17 punten. De scores liepen uiteen van 5.25 tot 13.75 met een gemiddelde spreiding van 2.10 punten. Uit de reactie van de leerkracht die de test had afgenomen, bleek dat de instructies voor de leerlingen duidelijk waren. Wel gaven de

leerlingen enkele suggesties om de vragen verder te verbeteren. Sommige leerlingen merkten op dat de schermafbeeldingen verouderd waren of er anders uitzagen dan op hun eigen computer. Een korte uitleg hierover is toegevoegd aan de instructies. Andere suggesties hebben geleid tot aanscherpingen van de formulering van de vragen, bijvoorbeeld het vervangen van 'plaatje' door 'tekstvak' als geen schermafbeelding werd getoond. Tijdens het scoren van de antwoorden werd duidelijk dat sommige codeerregels versoepeld moesten worden. Zo is de zin "hoe werkt onweer" toegevoegd als goed antwoord bij de vraag over zoektermen en is het weggelijken van het pop-up venster goed gerekend bij de vraag over het reageren op pop-up berichten.

De opdracht om het Word bestand aan te passen, is uitgevoerd door 3 leerlingen van 10 en 11 jaar van een basisschool uit Hengelo. Zij maakten de opdracht in ongeveer 10 minuten en hadden 3 of 4 goede veranderingen doorgevoerd. Op basis van hun reacties en prestaties is de uitleg op de opdrachtkaart verduidelijkt en zijn enkele termen gewijzigd. Zo bleek de term 'kantlijn' onbekend te zijn; dit is veranderd in 'alle regels 1 cm naar links'. Een tweede verbetering is dat 'dik drukken' is vervangen door 'vet maken'. Vanwege het open karakter van de PowerPoint opdracht, is besloten hiermee geen proefafname te doen.

7.4 Kwaliteit van de geschreven test

De gegevens uit de proefafname zijn gebruikt om de itemkarakteristieken van de geschreven test te bepalen. De gemiddelde moeilijkheidsgraad van de vragen over het onderdeel 'informatie zoeken' was .58, met waarden die varieerden tussen .04 en .95. Voor het onderdeel 'informatie gebruiken' was de gemiddelde moeilijkheidsgraad .57 met een iets minder grote spreiding (zie Tabel 7.1). Deze waarden worden als acceptabel beschouwd. De discriminatie-index uit Tabel 7.2 geeft aan in hoeverre een vraag onderscheid kan maken tussen goed- en slecht-presterende leerlingen.

51 <http://www.kidsweek.nl/>

Waardes van boven de .20 worden over het algemeen als voldoende onderscheidend beschouwd. Dit was het geval bij 12 van de 17 vragen.

Op basis van deze resultaten bestond twijfel over de kwaliteit van vraag 1, 4, 5, 8, 10, 12, 13 en 15. Omdat de deelnemers aan de proefafname ouder waren dan de leerlingen uit de onderzoeksgroep (en minder in aantal), is een definitief besluit pas na de eerste meting genomen. Uit die resultaten bleek dat de moeilijkheidsgraad van de meeste vragen ongeveer .20 tot 0.30 lager lag dan tijdens de proefafname. Het onderscheidend vermogen van de vragen was voldoende ($.20 < r_{pb} < .73$), behalve bij vraag 12 ($r_{pb} = .00$). Deze vraag is daarom niet meegenomen in de longitudinale analyses.

De data van de proefafname zijn ook gebruikt om de interbeoordelaarsovereenstemming te berekenen. Twee onafhankelijke onderzoekers hebben de antwoorden van de leerlingen beoordeeld; hun scores kwamen in 97 procent van de gevallen overeen (Cohen's $\kappa = .94$). De interne consistentie van de vragenlijst is niet berekend omdat de vragen over

drie relatief onafhankelijke onderwerpen gingen (het zoeken, vinden en evalueren van digitale informatie) die bovendien van elkaar verschilden in complexiteit, wat wordt bevestigd door de scores in Tabel 7.1.

De validiteit van de test is bepaald door de scores op de drie versies met elkaar te vergelijken. De meeste punten werden gehaald op de versie over aardbevingen ($M = 10.68$, $SD = 1.27$). De scores op de versie over orkanen ($M = 10.29$, $SD = 2.16$) en onweer ($M = 9.64$, $SD = 2.91$) waren iets lager maar deze verschillen waren niet significant, $F(2, 17) = 0.42$, $p = .665$. Bij een tweede validiteitsmeting zijn de scores op de testvragen vergeleken met de prestaties op identieke praktische taken⁵². Hiervoor zijn de vragen over 'informatie zoeken' vertaald naar praktische opdrachten die zijn uitgevoerd door 43 leerlingen van een basisschool in Enschede (tussen de 10 en 13 jaar oud, $M = 11.05$). Hun score op de geschreven test ($M = 7.03$, $SD = 1.92$) was wat lager die van leerlingen van Het Talent. Op de praktische opdrachten ($M = 8.60$, $SD = 1.17$) presteerden zij wat beter dan op de geschreven test, maar belangrijker is dat deze scores voldoende samenhang vertoonden, $r(43) = .47$, $p < .001$.

Tabel 7.1: Moeilijkheidsgraad van de testvragen

	Vraag													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Informatie zoeken	.95	.64	.48	.95	.81	.52	.67	.86	.48	.81	.24	.04	.05	.57
Informatie gebruiken	.86	.29	.57	—	—	—	—	—	—	—	—	—	—	—

Tabel 7.2: Discriminatie-index van de testvragen

	Vraag													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Informatie zoeken	.11	.51	.57	-.19	.29	.68	.58	-.01	.48	.35	.04	.26	.19	.59
Informatie gebruiken	.82	.47	.82	—	—	—	—	—	—	—	—	—	—	—

52 Dit onderzoek is onder begeleiding van Ard Lazonder uitgevoerd door Petra ter Denge, Nadjia Erdinger, Jeroen Pierik en Hugo Uiterwijk Winkel, studenten psychologie aan de Universiteit Twente, in het kader van een tweedejaars vak.

7.5 Kwaliteit van de opdrachten

Omdat de PowerPoint en Word opdracht niet, of met een klein aantal leerlingen zijn geprobeerd, is de kwaliteit ervan bepaald op basis van de gegevens uit de eerste meting in 2016. De item-karakteristieken in Tabel 7.3 laten zien dat de eerste drie veranderingen in het Word bestand relatief eenvoudig waren; vanaf opgave 4 kon vrijwel geen enkele leerling de wijzigingen correct doorvoeren. Hierdoor bood de Word opdracht voldoende gelegenheid om vaardigheidsgroei te meten, temeer omdat het onderscheidend vermogen van alle opgaven groot was.

Door de toenemende complexiteit van de veranderingen was het niet betekenisvol om de betrouwbaarheid (interne consistentie) van de Word opdracht te bepalen. De betrouwbaarheid van de scoring van de veranderingen is wel berekend, op basis van de bestanden van 15 willekeurig geselecteerde leerlingen. Twee onafhankelijke beoordelaars kwamen tot dezelfde scores bij 104 van de 105 veranderingen (Cohen's $\kappa = .98$).

Er was geen verschil tussen de drie versies van de opdracht. Leerlingen die de tekst over de Nationale Hoortest hadden bewerkt, haalden weliswaar een iets lagere score ($M = 1.90$, $SD = 1.10$) dan de leerlingen die de overige twee teksten hadden gekregen ($M = 2.15$), maar dit marginale verschil was niet significant,

$F(2, 177) = 1.02$, $p = .362$. De validiteit van de opdracht werd verder ondersteund door cross-sectionele verschillen⁵³. In een apart onderzoek onder 81 leerlingen van twee Twentse basisscholen, is gekeken of leerlingen uit groep 8 ($n = 37$) meer goede veranderingen konden doorvoeren dan leerlingen uit groep 6 ($n = 44$). Dit bleek inderdaad het geval te zijn, $t(79) = 3.66$, $p < .001$. Het verschil tussen beide groepen was ruim anderhalf punt ($M_{\text{groep 6}} = 2.39$, $SD = 1.26$; $M_{\text{groep 8}} = 3.78$, $SD = 2.12$).

Bij de PowerPoint opdracht is de interbeoordelaarsovereenstemming berekend op basis van de 2016 data. De presentaties van 29 willekeurig geselecteerde leerlingen zijn beoordeeld door twee onafhankelijke onderzoekers; de Cohen's κ was voldoende (.72). Net als bij de Word opdracht waren de scores op de drie versies vergelijkbaar. De presentaties over het leukste gerecht ($M = 8.21$, $SD = 2.82$) werden gemiddeld ongeveer 1 punt hoger beoordeeld dan de presentaties over de leukste school ($M = 7.17$, $SD = 2.69$) en het leukste boek ($M = 7.30$, $SD = 2.89$). Dit verschil was niet statistisch significant, $F(2, 177) = 2.37$, $p = .097$. In het flankerend onderzoek op de Twentse basisscholen werd geen verschil gevonden tussen de presentaties van leerlingen uit groep 6 ($M = 9.14$, $SD = 3.21$) en groep 8 ($M = 9.73$, $SD = 3.84$), $t(79) = 0.76$, $p = 0.45$. Dit resultaat moet voorzichtig geïnterpreteerd worden omdat de interbeoordelaarsovereenstemming in dit onderzoek matig was, Cohen's $\kappa = .62$.

Tabel 7.3: Moeilijkheidsgraad en discriminatie-index van de Word opdrachten

	<i>p</i>	<i>r_{pb}</i>
1. Lettergrootte veranderen	.79	.56
2. Tekstopmaak veranderen	.41	.58
3. Kleur van de tekst veranderen	.67	.66
4. Kantlijn van één paragraaf veranderen	.06	.34
5. Eerste regel van een paragraaf laten inspringen	.01	.23
6. Regelaafstand veranderen	.07	.33
7. Tekstterugloop van een afbeelding veranderen	.07	.37

53 Dit onderzoek is onder begeleiding van Noortje Janssen uitgevoerd door Jannick Pohlmann, Francis Seeger en Bedi Tekirdag, studenten psychologie aan de Universiteit Twente, in het kader van een tweedejaars vak.

8. Statistische analyses van de scores voor digitale geletterdheid

8.1 Vaardigheidsgroei

Tabel 8.1 geeft een overzicht van de prestaties op de vier onderdelen van digitale geletterdheid. Omdat er niet vanuit kan worden gegaan dat deze vaardigheden zich elk jaar in hetzelfde tempo ontwikkelen, zijn de periodes 2016-2017 en 2017-2018 apart geanalyseerd met mixed-design MANVOA's. In deze analyses was 'tijd' de *within-subject* variabele en 'spil' de *between-subject* variabele. De afhankelijke variabelen 'testscore' voldeden allen aan de eisen van normaliteit en homogeniteit van variantie. De vergelijking van de scores uit 2016 en 2017

leverde een multivariaat effect van tijd, $F(4, 165) = 57.28, p < .001, \eta^2 = .58$, en geen effect van spil, $F(4, 165) = 1.14, p = .339, \eta^2 = .03$. Dit betekent dat de testscores in beide spellen niet van elkaar verschilden en zijn toegenomen in de tijd. Uit de significante tijd $\times$ spil interactie, $F(4, 165) = 2.51, p = .044, \eta^2 = .06$, blijkt verder dat deze toename niet in beide spellen even groot was. Uit de univariate analyses blijkt dat dit komt doordat de leerlingen uit spil 19-20 meer vooruitgang boekten op het onderdeel 'informatie zoeken' dan de leerlingen uit spil 17-18, $F(1, 168) = 4.32, p = .039, \eta^2 = .03$. Bij de overige onderdelen waren de interactie-

Tabel 8.1: Beschrijvende statistieken van de scores op de digitale geletterdheid testen

	2016			2017			2018		
	N	M	SD	N	M	SD	N	M	SD
Informatie zoeken									
Spil 17-18	90	6.58	2.00	89	7.89	1.94	86	9.39	2.12
Spil 19-20	89	6.27	2.17	91	8.21	2.20	78	9.52	2.06
Informatie creëren									
Spil 17-18	89	7.83	2.84	88	9.34	2.62	84	11.42	2.85
Spil 19-20	91	7.24	2.79	89	9.75	2.27	77	11.61	2.36
Informatie bewerken									
Spil 17-18	88	2.01	1.11	88	2.69	1.15	85	3.40	1.13
Spil 19-20	92	2.12	1.11	90	3.14	1.16	77	3.64	1.14
Informatie gebruiken									
Spil 17-18	90	1.19	0.72	89	1.33	0.70	86	1.63	0.69
Spil 19-20	89	1.04	0.81	91	1.47	0.78	78	1.68	0.67

Noot. Maximale scores: informatie zoeken (13), informatie creëren (27), informatie bewerken (7) en informatie gebruiken (3).

effecten niet significant, $F(1, 168) < 3.55$, $p > .061$, $\eta^2 < .02$. Wel werden voor deze vaardigheden significante effecten van tijd gevonden, $F(1, 168) > 17.62$, $p < .001$, $\eta^2 > .10$.

Bij de 2017-2018 vergelijking is een multivariaat effect gevonden van tijd, $F(4, 151) = 43.78$, $p < .001$, $\eta^2 = .54$, geen effect van spil, $F(4, 151) = 0.66$, $p = .620$, $\eta^2 = .02$, en geen interactie-effect, $F(4, 151) = 0.20$, $p = .936$, $\eta^2 = .01$. De univariate effecten van tijd waren significant voor alle vier de onderdelen, $F(1, 154) > 14.27$, $p < .001$, $\eta^2 > .09$. Dit betekent simpelweg dat alle testcores waren verbeterd en dat er geen verschil tussen de spullen was, noch qua hoogte van de scores en noch qua vaardigheidsgroei.

8.2 Relatie met leerlingkenmerken

Voor de vier onderdelen van digitale geletterdheid is de vooruitgang in vaardigheid berekend door de scores van twee opeenvolgende jaren van elkaar af te trekken. Tussen 2016 en 2017 maakten de jongens evenveel voortgang als de meisjes op alle vier de vaardigheden, $F(4, 165) = 0.13$, $p = .971$, $\eta^2 = .00$. De achtergrond van de leerlingen was evenmin van invloed op de vaardigheidsgroei, $F(4, 165) = 1.82$, $p = .128$, $\eta^2 = .04$, ondanks dat de vooruitgang van de leerlingen met een migratieachtergrond wat hoger was dan die van hun Nederlandse medeleerlingen (zie Tabel 8.2)

Tabel 8.2: Vooruitgang op de digitale geletterdheid testen naar sekse en achtergrond (2016-2017)

	Informatie zoeken		Informatie creëren		Informatie bewerken		Informatie gebruiken	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Sekse								
Jongen	1.60	2.00	1.19	2.82	0.85	1.48	0.28	0.92
Meisje	1.75	2.19	2.07	2.83	0.76	1.33	0.30	0.88
Achtergrond								
Nederlandse	1.51	2.11	1.19	2.84	0.64	1.38	0.23	0.91
Migratie	2.28	1.82	2.57	3.03	1.22	1.13	0.64	0.73

Tabel 8.3 toont de gemiddelde vaardigheidsgroei over de periode 2017-2018. De effecten van sekse, $F(4, 151) = 1.75, p = .141, \eta^2 = .04$, en achtergrond, $F(4, 151) = 1.13, p = .346, \eta^2 = .03$, waren wederom niet significant, wat betekent dat de vaardigheidsgroei op alle vier de onderdelen niet verschilden tussen de seksen en evenmin tussen leerlingen met een verschillende achtergrond.

De vooruitgang in beide periodes was eveneens onafhankelijk van sociaaleconomische status: de Spearman's rho correlatiecoëfficiënt van deze ordinale variabele varieerde tussen

de $-.11$ en $-.02, p > .150$. Voor leeftijd werd alleen een significant verband gevonden met de vaardigheidsgroei in het creëren van informatie over de periode 2016-2017, $r = -.356, p < .001$. Uit deze negatieve correlatie blijkt dat leerlingen die ouder zijn tijdens het eerste jaar minder vooruitgingen dan leerlingen die jonger zijn. Het jaar daarop was dit effect verdwenen, $r = -.072, p = .365$. De vaardigheidsgroei op de overige onderdelen van digitale geletterdheid was onafhankelijk van leeftijd in beide periodes, $-.019 < r < .028, p > .728$.

Tabel 8.3: Vooruitgang op de digitale geletterdheid testen naar sekse en achtergrond (2017-2018)

	Informatie zoeken		Informatie creëren		Informatie bewerken		Informatie gebruiken	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Sekse								
Jongen	1.64	2.24	1.74	2.81	0.84	1.25	0.33	0.80
Meisje	1.50	2.04	2.54	3.15	0.54	1.22	0.19	0.93
Achtergrond								
Nederlandse	1.63	2.21	2.14	3.02	0.67	1.25	0.32	0.86
Migratie	1.25	1.71	2.00	2.93	0.83	1.20	-0.06	0.91

9. Samenwerken: ontwerp van de test

9.1 De ontwerpopdracht

In het onderzoek is gebruik gemaakt van een ontwerpstaak. Leerlingen werkten in groepjes van drie of vier aan een ontwerpopdracht. Om herkenning te voorkomen, kregen de teams elk jaar een andere opdracht. De aard van het ontwerp verschilde, de opzet en structuur van de opdracht werd constant gehouden. Bij elke opdracht werden de teams gevraagd een ontwerp tekening te maken. Het ontwerp was telkens gerelateerd aan de school. De onderwerpen betroffen het maken van een ontwerp voor de schooltuin, een speeltoestel voor op het schoolplein of een sportshirt voor schooltoernooien. Hierbij moesten de teams rekening houden met de belangen van interne betrokkenen (bijvoorbeeld de schooldirectie en medeleerlingen) en externe betrokkenen (bijvoorbeeld de buurtbewoners). Iedere teamlid kreeg een opdrachtkaartje met daarop een bepaalde wens of eis van een van de betrokken partijen, waarvoor hij of zij de verantwoordelijkheid moest dragen. De leerling moest zorgen dat dit aspect in het gezamenlijke eindproduct goed naar voren kwam. Door individuele leerlingen verantwoordelijk te maken voor een deel van de ontwerpisen, wordt afhankelijkheid gecreëerd en zullen leerlingen moeten overleggen om tot een goed eindontwerp te komen. Indien één leerling zijn of haar taak niet voldoende serieus neemt of niet wil overleggen, kan dit nadelige gevolgen hebben voor het groepsproces en het eindproduct.

De belangen van de betrokken partijen lijken in eerste instantie moeilijk met elkaar te verenigen. Maar onmogelijk was dit niet: door goed met elkaar te overleggen konden de leerlingen tot een goed compromis komen. Bij het ontwerpen van de schooltuin wilde de

conciërge bijvoorbeeld een voorziening om het gereedschap veilig op te bergen, maar deze voorziening mocht van de buurtbewoners niet te hoog zijn. Leerlingen konden er dus voor kiezen om een lage opbergbak of minischuurtje in de tuin te plaatsen. Of een (gedeeltelijk) ondergrondse berging te maken.

9.2 Proefafnames

De drie ontwerpstaakjes zijn afgenomen bij een viertal leerlingen (10 en 11 jaar oud) die niet op Het Talent zaten. Na afloop van elke proefafname werden het script en de opdracht aangepast. In de meeste gevallen betrof dit de formulering van de opdracht of het script. Een voorbeeld hiervan is het gebruik van het woord logo. Niet alle leerlingen in de proefafname wisten wat het woord logo betekende. Op basis hiervan is in de definitieve versie van het script gevraagd of leerlingen weten wat een logo is en hier een voorbeeld van kunnen geven. Indien dit lastig is, geeft de proefleider een korte uitleg.

9.3 Codering van de samenwerking

Het codeerschema dat is gebruikt voor het beoordelen van de samenwerking, is in sterke mate gebaseerd op de codering die door de OECD is voorgesteld voor het meten van zogeheten *collaborative problem-solving skills*⁵⁴. In navolging van dit rapport heeft de codering zich gericht op de volgende drie centrale aspecten van samenwerken:

- Het tot stand brengen en onderhouden van een gedeeld begrip van het probleem/de opdracht.
- Samen met anderen passende stappen

nemen gericht op het oplossen van het probleem.

- Het tot stand brengen en onderhouden van een goed georganiseerd team.

Om te komen tot meetbare samenwerkingsvaardigheden zijn deze drie hoofdaspecten gecombineerd met vier probleemsoplossingsprocessen (zie Tabel 9.1) die door de OECD worden onderscheiden⁵⁵. *Exploratie en begrip* heeft betrekking op het verkennen van de opdracht en de kennis en vaardigheden die aanwezig zijn in het team. *Representeren en formuleren* heeft betrekking op het formuleren van doelen, het uitleggen en expliciteren van informatie, acties en rollen en waar nodig het vastleggen hiervan. *Plannen en uitvoeren* heeft betrekking op het bespreken van plannen, taken en deeltaken en de uitvoering van deze taken. Het uitvoeren van plannen en deelplannen is tevens onderdeel van dit proces. *Monitoren en reflecteren* heeft betrekking op het formatief en summatief evalueren van het gedeelde begrip, de passende acties en de teamorganisatie.

De vaardigheden worden concreet in het samenwerkingsproces van de verschillende leerlingen. Hierbij is gekeken naar de communicatie én naar de acties. Leerlingen kunnen informatie delen door deze te bespreken in het team, maar ze kunnen ook iets laten zien op de tekening of het kaartje met de taakeisen aan een ander teamlid laten lezen.

In tegenstelling tot de collaborative problem-solving taken die door de OECD zijn ontwikkeld voor de PISA-2016 vergelijkingsstudie, is de codering in dit project gericht op de prestatie van de groep als geheel (i.t.t. de prestaties van de afzonderlijke groepsleden). Bij de ontwerp opdrachten werkten de leerlingen van Het Talent met elkaar samen en niet zoals in de OECD-PISA taken met een virtuele medeleerling op de computer. Acties die door één bepaalde leerling worden uitgevoerd, kunnen het verloop en de scoring van de samenwerking beïnvloeden. Als bijvoorbeeld één leerling het initiatief neemt om de taken te verdelen, kan de rest van de teamleden dit niet meer doen zodat hun individuele score voor dit onderdeel nul punten zou opleveren. Om dit soort oneerlijke situaties te voorkomen, zijn de samenwerkingsvaardigheden voor het team als geheel bepaald.

9.4 Werkwijze

Per opdracht is van elk team een video-opname gemaakt. Voorafgaand aan de codering zijn deze opnames verdeeld in 4 tijdsblokken van 5 minuten (de totale tijd van de sessie bedroeg 20 minuten). Voor elk tijdsblok is bij de codering een lege versie van de matrix uit Tabel 9.1 gebruikt. De beoordelaars hadden tevens de beschikking over een tabel met beschrijvingen en voorbeelden van situaties en gedragingen die horen bij de cellen uit de matrix. Voor elke cel werd een score van 0 (niet aanwezig),

1 (laag), 2 (gemiddeld), of 3 (hoog) punten gegeven. De laagste score werd alleen gegeven als er niet werd samengewerkt, of de leerlingen in het geheel niet taakgericht bezig waren.

Ter illustratie volgt de uitwerking voor Cel C1. De score 1 (laag) werd gegeven als er sprake was van communicatie die niet taakrelevant is maar waarop leerlingen wel reageren. Of wanneer er sprake was van acties of uitingen die verwarring creëerden zonder verder uitleg. De score 2 (gemiddeld), werd gegeven als de acties en uitingen contextueel relevant waren (inspelen op de taak en op vragen en informatie van teamgenoten) en waar indien nodig uitleg en verklaringen werden gegeven. De score 3 (hoog) werd gegeven als leerlingen actief relevante informatie deelden, ook wanneer hier niet om werd gevraagd. Bijvoorbeeld als

leerlingen uitleg gaven op het moment dat zij inschatten dat dit belangrijk was. Een ander voorbeeld is het actief vragen naar de manier waarop een teamlid de taak begrijpt, of omgaat met informatie of taken.

Voor elke cel werden scores vastgesteld en vervolgens zijn per tijdsblok de gemiddelde scores voor de drie centrale vaardigheden berekend. Het totaal van de gemiddelde scores over de vier tijdsblokken was de score voor de betreffende vaardigheid. De maximale score bedroeg 12 punten.

Een tweede codeur heeft 20 procent van de data uit 2016 gecodeerd. Er is een Cohen's κ berekend om de mate van overeenstemming tussen de eerste en tweede codeur te berekenen. Deze bedroeg .80.



TERUG NAAR
DE HOOFDTEKST
PAGINA 24

Tabel 9.1: Matrix met uitwerking van de samenwerkingsvaardigheden

	1. Gedeeld begrip	2. Passende acties	3. Teamorganisatie
A. Exploratie en begrip	(A1) Het verkennen van de perspectieven en mogelijkheden van alle teamleden.	(A2) Het verkennen van het type samenwerking of communicatie dat nodig is om het probleem op te lossen.	(A3) Het begrijpen van de verschillende rollen en taken in het team.
B. Representeren en formuleren	(B1) Het werken aan een gemeenschappelijke opvatting/ begrip over de taak en de sub-taken.	(B2) Het identificeren van de verschillende taken en stappen die nodig zijn om het probleem op te lossen.	(B3) Het beschrijven van de organisatie van de rollen en de regels, het protocol.
C. Plannen en uitvoeren	(C1) Met anderen spreken over de uitvoering van de taken.	(C2) Het uitvoeren van de plannen.	(C3) Het opvolgen/ uitvoeren van de afspraken en uitvoeren van de rollen. Hieronder valt bijvoorbeeld het oproepen van andere leerlingen tot actie.
D. Monitoren en reflecteren	(D1) Het monitoren van het gedeelde begrip en eventueel herstellen indien dit nodig is.	(D2) Het monitoren van de acties en het evalueren van de kwaliteit van de oplossing.	(D3) Het monitoren van feedback, het aanpassen van de organisatie, afspraken en rollen.

10. Statistische analyses van de scores voor samenwerken

10.1 Vaardigheidsgroei

Tabel 10.1 geeft informatie over de scores van de ontwerpteams uit beide spellen. Deze scores zijn geanalyseerd met behulp van mixed-design MANOVA's, waarbij is voldaan aan de eisen van normaliteit en homogeniteit van de variantie. In de uitgevoerde analyses was 'tijd' de *within-subject* factor en 'spil' de *between-subject* factor.

Om inzicht te krijgen in de vaardigheidsgroei zijn de scores uit de drie projectjaren stapsgewijs vergeleken. Bij de vergelijking van de scores uit 2016 en 2017 werd een significant effect gevonden van de *within-subject* factor tijd, $F(3, 43) = 6.75$, $p = .001$, $\eta^2 = .32$, en geen

significant effect van de *between-subject* factor spil, $F(3, 43) = 1.04$, $p = .384$, $\eta^2 = .07$. Er was ook geen significant interactie effect, $F(3, 43) = 1.81$, $p = .159$, $\eta^2 = .11$. Dit betekent dat de teams in 2017 beter konden samenwerken dan het jaar daarvoor en dat er geen verschillen waren tussen de spellen.

Uit univariate analyses blijkt dat de teams beter waren geworden in het creëren en onderhouden van gedeeld begrip, $F(1, 45) = 6.32$, $p = .016$, $\eta^2 = .12$, en het uitvoeren van passende acties, $F(1, 45) = 8.93$, $p = .005$, $\eta^2 = .17$. Op de deelvaardigheid 'teamorganisatie' werd in deze periode geen significante vooruitgang geboekt, $F(1, 45) = 1.09$, $p = .445$, $\eta^2 = .01$.

Tabel 10.1: Beschrijvende statistieken van de scores op de samenwerkingsopdracht

	2016			2017			2018		
	N	M	SD	N	M	SD	N	M	SD
Gedeeld begrip									
Spil 17-18	24	8.14	1.67	23	8.73	1.69	21	10.37	1.44
Spil 19-20	24	7.94	2.07	24	8.63	1.76	19	9.92	1.96
Passende acties									
Spil 17-18	24	7.86	1.77	23	8.67	1.77	21	10.20	1.50
Spil 19-20	24	7.93	1.76	24	8.51	1.86	19	9.96	1.89
Teamorganisatie									
Spil 17-18	24	7.75	2.11	23	7.95	1.50	21	9.73	1.74
Spil 19-20	24	7.96	2.03	24	8.28	1.16	19	9.15	2.10

Noot. De maximale score voor elke vaardigheid was 12 punten.



TERUG NAAR
DE HOOFDTEKST
PAGINA 26

Bij de vergelijking van de scores uit 2017 en 2018 zijn multivariate effecten van de *within-subject* factor tijd gevonden, $F(3, 36) = 6.13, p = .002, \eta^2 = .34$, maar niet van de *between-subject* factor spil, $F(3, 36) = 0.31, p = .819, \eta^2 = .03$. Er werd geen interactie effect gevonden, $F(3, 36) = 2.12, p = .103, \eta^2 = .16$. Uit de univariate analyses blijkt dat de verbetering van de samenwerking zich voordeed bij het 'creëren en onderhouden van gedeeld begrip', $F(1, 38) = 53.23, p < .001, \eta^2 = .32$, het 'uitvoeren van passende acties', $F(1, 38) = .53.41, p < .001, \eta^2 = .29$, en 'teamorganisatie', $F(1, 38) = 42.72, p = .002, \eta^2 = .22$.

11. Statistische analyses van de scores voor taal en rekenen

11.1 Vaardigheidsgroei

Tabel 11.1 geeft informatie over de vaardigheidsscores van beide spellen in de periode 2016 tot 2018. Om te bepalen of de getoonde toename in scores significant was, zijn mixed-design ANOVA's uitgevoerd met de vaardigheidsscores als *within-subject* factor en spel als *between-subject* factor. Omdat er niet vanuit kan worden gegaan dat de vaardigheidsgroei elk jaar even groot is, zijn de jaren afzonderlijk geanalyseerd. In alle analyses werd voldaan aan de eisen van normaliteit en homogeniteit.

Bij begrijpend lezen was de vaardigheidsgroei tussen 2016 en 2017 significant, $F(1, 173) = 139.74$, $p < .001$, $\eta^2 = .45$, en in beide spellen even groot, $F(1, 173) = 2.16$, $p = .143$, $\eta^2 = .01$. De groei over de periode 2017-2018 was eveneens significant, $F(1, 160) = 157.91$, $p < .001$, $\eta^2 = .50$, en wederom niet afhankelijk van de spellen, $F(1, 160) = 1.00$, $p = .317$, $\eta^2 = .01$.

Bij rekenen verschilden de 2016 en 2017 scores significant, $F(1, 174) = 137.15$, $p < .001$, $\eta^2 = .44$, en was de groei x spel interactie niet significant, $F(1, 174) = 0.18$, $p = .871$, $\eta^2 = .00$. Ook in de periode 2017-2018 werd een significante toename gevonden, $F(1, 157) = 128.72$, $p < .001$, $\eta^2 = .45$, en een niet-significante interactie, $F(1, 157) = 0.02$, $p < .880$, $\eta^2 = .00$.

11.2 Vooruitgang t.o.v de landelijke referentiegroep

Bij begrijpend lezen bestaan referentiescores voor de groei tussen de meetmomenten. De groei tussen M5 en M6 bedraagt landelijk gezien 6.5 punten; de leerlingen uit spel 17-18 behaalden een gemiddelde groei van 8.2 punten. Uit een *one-sample* t-test blijkt dat dit verschil significant is, $t(84) = 4.19$, $p = .001$. De groei tussen M6 en M7 is over beide spellen samen berekend en bedroeg 10.4 punten, wat

Tabel 11.1: Beschrijvende statistieken van de vaardigheidsscores voor taal en rekenen

	2016			2017			2018		
	N	M	SD	N	M	SD	N	M	SD
Cito begrijpend lezen									
Spil 17-18	86	32.50	13.19	89	40.72	12.38	85	52.54	12.57
Spil 19-20	92	38.95	16.45	90	49.53	13.43	78	61.90	18.81
SVT-HR									
Spil 17-18	89	61.12	18.20	87	72.34	19.24	85	84.41	19.72
Spil 19-20	92	74.13	21.16	89	85.09	23.57	78	94.92	28.08

significant minder was dan de landelijke toename van 12.4 punten, $t(182) = -2.30$, $p = .022$. Van M7 naar M8 was de groei op Het Talent weer bovengemiddeld (14.2 versus 11.0 punten) maar dit verschil bleek niet significant te zijn, $t(71) = 1.76$, $p = .083$.

Bij rekenen is gekeken naar de verandering in percentielscores. Bij leerlingen uit spil 17-18 daalde de percentielscore aanvankelijk met 3.4 punten, wat betekent dat ze iets minder snel vooruit gingen dan hun leeftijdsgenoten. Dit verschil was echter niet significant, $t(86) = -1.89$, $p = .062$. Van 2017 tot 2018 steeg hun percentielscore met 1.70 punten, wat wederom niet significant was, $t(82) = 0.99$, $p = .326$. In spil 19-20 waren de percentielscores tamelijk stabiel, met een variatie van -0.9 in de eerste periode en 1.7 in de tweede periode. Beide verschuivingen waren niet significant, $t(89) = -0.40$, $p = .689$, respectievelijk, $t(77) = 0.79$, $p = .431$.

11.3 Relatie met leerlingkenmerken

Voor de analyses met leerlingkenmerken zijn *gain* scores berekend door de vaardigheidscores van twee opeenvolgende jaren van elkaar af te trekken. In Tabel 12.2 is te zien dat de vaardigheidsgroei voor taal en rekenen in beide periodes tussen de 9 en 13 punten lag. Deze toename was onafhankelijk van sekse, $F < 3.33$, $p > .070$, en achtergrond, $F < 0.37$, $p > .545$.

De vaardigheidsgroei bleek evenmin afhankelijk van de leeftijd van de leerlingen, $-.13 < r < -.01$, $p > .110$, en de sociaaleconomische status, $-.13 < \text{Spearman's } \rho < -.06$, $p > .120$. Voor leerlinggewicht zijn geen correlaties berekend omdat bijna alle leerlingen, vier uitgezonderd, een leerlinggewicht van 0 (nul) hadden.

Tabel 11.2: Vooruitgang op de vaardigheidsscores naar sekse en achtergrond

	Vooruitgang 2016 - 2017				Vooruitgang 2017 - 2018			
	Taal		Rekenen		Taal		Rekenen	
	M	SD	M	SD	M	SD	M	SD
Sekse								
Jongen	9.54	11.60	9.64	11.98	12.19	13.19	13.56	15.06
Meisje	9.49	9.41	12.55	12.66	13.32	12.62	9.76	9.99
Achtergrond								
Nederlandse	9.34	10.67	10.61	12.20	12.80	13.05	12.01	13.30
Migratie	10.08	11.21	12.19	13.42	12.95	12.35	11.33	13.26

12. De onderlinge samenhang tussen de resultaten

In de voorgaande hoofdstukken is steeds gekeken naar de ontwikkeling van één bepaalde vaardigheid. In dit hoofdstuk worden de vaardigheden aan elkaar gerelateerd. Hierbij wordt eerst ingegaan op de samenhang tussen de scores op de diverse testen in elk van de drie projectjaren. Vervolgens wordt de samenhang in de ontwikkeling van de gemeten vaardigheden beschreven. Alle analyses zijn uitgevoerd met vier vaardigheden: kritisch denken, digitale geletterdheid, taal en rekenen. De vijfde vaardigheid, samenwerken, is buiten beschouwing gelaten omdat hierbij is gekeken naar teamprestaties (i.t.t. de prestaties per individuele leerling).

12.1 Samenhang tussen de testcores

Tabel 12.1 toont de correlaties tussen de testcores uit 2016. In dat jaar hingen de prestaties op de vier onderdelen van digitale geletterdheid

tamelijk consistent met elkaar samen; alleen 'informatie veilig gebruiken' correleerde niet met 'informatie creëren' en 'informatie bewerken'. De relatief hoge correlatie tussen 'informatie zoeken' en 'informatie veilig gebruiken' zou mogelijk veroorzaakt kunnen zijn doordat beide onderdelen met eenzelfde soort test zijn gemeten.

In 2016 waren de scores op de test voor kritisch denken onafhankelijk van de scores op de testen voor digitale geletterdheid. Wel werd een positief verband gevonden met taalvaardigheid. Rekenvaardigheid tenslotte vertoonde een matige samenhang met de meeste onderdelen van digitale geletterdheid, maar niet met kritisch denken. Tussen rekenen en taalvaardigheid bestond een bescheiden correlatie.

Een jaar later waren de meeste van deze correlaties nog steeds significant (zie Tabel 12.2). Uitzonderingen waren het verband

Tabel 12.1: Correlaties tussen de testcores uit 2016

	1.	2.	3.	4.	5.	6.	7.
1. Informatie zoeken	—						
2. Informatie creëren	.36**	—					
3. Informatie bewerken	.25**	.34**	—				
4. Informatie gebruiken	.61**	.14	.15	—			
5. Kritisch denken	-.08	-.02	-.06	.01	—		
6. Taal	.15	.04	-.01	.16*	.23**	—	
7. Rekenen	.25**	.28**	.18*	.11	.00	.24**	—

* $p < .05$ (tweezijdig) ** $p < .001$ (tweezijdig)

tussen 'informatie zoeken' en 'informatie creëren' en tussen taalvaardigheid en rekenvaardigheid. In Tabel 12.3 is te zien dat dit verband in 2018 weer was hersteld. Kritisch denken en taalvaardigheid correleerde in 2018 nog steeds alleen maar met elkaar, en bij rekenvaardigheid ging de significante samenhang met 'informatie zoeken' verloren.

12.2 Samenhang tussen de vaardigheidsgroei

Omdat de (totaal)scores voor alle vaardigheden lineair toenamen, is eerst de vaardigheidsgroei berekend over de gehele looptijd van het project.

Tabel 12.4 toont de correlaties tussen de groei van de afzonderlijke vaardigheden tussen 2016 en 2018. Hierin is te zien dat taal, rekenen en kritisch denken zich onafhankelijk van elkaar én van de onderdelen van digitale geletterdheid ontwikkelden. Tussen deze laatste onderdelen werden wel significante correlaties gevonden. Een uitzondering hierop is de vaardigheidsgroei voor 'informatie creëren'. Dit kan mogelijk verklaard worden door de beoordeling van de PowerPoint presentaties die de leerlingen bij dit onderdeel hebben gemaakt; hierbij werd namelijk niet alleen op de technische kwaliteiten gelet maar ook op inhoudelijke en grafische aspecten. Om een genuanceerder beeld te krijgen, is ook

Tabel 12.2: Correlaties tussen de testscores uit 2017

	1.	2.	3.	4.	5.	6.	7.
1. Informatie zoeken	—						
2. Informatie creëren	.36**	—					
3. Informatie bewerken	.25**	.34**	—				
4. Informatie gebruiken	.61**	.14	.15	—			
5. Kritisch denken	-.08	-.02	-.06	.01	—		
6. Taal	.15	.04	-.01	.16*	.23**	—	
7. Rekenen	.25**	.28**	.18*	.11	.00	.24**	—

* $p < .05$ (tweezijdig) ** $p < .001$ (tweezijdig)

Tabel 12.3: Correlaties tussen de testscores uit 2018

	1.	2.	3.	4.	5.	6.	7.
1. Informatie zoeken	—						
2. Informatie creëren	.26**	—					
3. Informatie bewerken	.30**	.23**	—				
4. Informatie gebruiken	.63**	.14	.25**	—			
5. Kritisch denken	.14	.09	-.05	.09	—		
6. Taal	.01	.01	-.08	.16	.34**	—	
7. Rekenen	.12	.16*	.18*	-.15	-.02	-.04	—

* $p < .05$ (tweezijdig) ** $p < .001$ (tweezijdig)

de vaardigheidsgroei per opeenvolgend jaar geanalyseerd. Uit de resultaten in Tabel 12.5 blijkt dat veel van de significante correlaties tussen de onderdelen van digitale geletterdheid zijn verdwenen. Kennelijk is de samenhang tussen de vaardigheidsgroei niet zo sterk dat dit ook van jaar tot jaar merkbaar is. Wat verder opvalt, is dat de toename van de scores op een bepaalde test *negatief* met elkaar correleerde (de groene cellen in de tabel)⁵⁶. Dit betekent dat als leerlingen het ene jaar veel vooruitgaan op een bepaalde vaardigheid, die groei het jaar daarop wat stagneert.

Dit doet vermoeden dat er een soort 'glazen leeftijdsplafond' zou kunnen bestaan: de ontwikkeling vlakt immers op een bepaald moment af. Om te achterhalen of dit inderdaad het geval is, zijn de correlaties ook per spel berekend. Hieruit bleek geen verschil met de resultaten uit Tabel 12.5, noch wat betreft de hoogte van de correlaties noch wat betreft hun statistische significantie. Uit partiële correlaties waarin werd gecontroleerd voor de leeftijd van de leerlingen (in maanden) kwam evenmin een ander patroon naar voren.

Tabel 12.4: Correlaties tussen de vaardigheidsgroei tussen 2016 en 2018

	1.	2.	3.	4.	5.	6.	7.
1. Informatie zoeken	—						
2. Informatie creëren	.24**	—					
3. Informatie bewerken	.20*	.09	—				
4. Informatie gebruiken	.50**	.01	.18*	—			
5. Kritisch denken	-.13	.07	.08	.02	—		
6. Taal	-.04	.10	.02	.06	-.11	—	
7. Rekenen	.02	.10	-.03	.12	-.04	.14	—

* $p < .05$ (tweezijdig) ** $p < .001$ (tweezijdig)

Tabel 12.5: Correlaties tussen de vaardigheidsgroei per opeenvolgend jaar

	1.	2.	3.	4.	5.	6.	7.
1. Informatie zoeken	-.49**	.07	.12	.56**	.04	-.09	.14
2. Informatie creëren	.05	-.40**	.10	.05	-.01	-.04	.07
3. Informatie bewerken	.02	.06	-.45**	.00	-.03	-.14	.14
4. Informatie gebruiken	.47**	.15	-.05	-.48**	-.01	.00	.11
5. Kritisch denken	-.13	.03	-.06	-.11	-.42**	-.09	-.03
6. Taal	.07	.03	.02	.02	-.11	-.42**	.08
7. Rekenen	-.04	.13	-.08	.04	-.04	-.11	-.33**

Noot. De cellen linksonder bevatten de correlaties voor de periode 2016-2017, de cellen rechtsboven voor de periode 2017-2018. De cellen op de diagonaal tonen de samenhang tussen de groei in beide periodes voor de betreffende vaardigheid.

* $p < .05$ (tweezijdig) ** $p < .001$ (tweezijdig)

55 Voor de vaardigheid 'samenwerken' is eveneens een negatieve correlatie gevonden tussen de vaardigheidsgroei van de teams tussen 2016-2017 en 2017-2018. Deze was echter niet statistisch significant, $r = -.20$, $p = .229$.

12.3 Verschillen in de mate de vaardigheidsgroei

Tot slot is gekeken of de vaardigheden zich in gelijke mate ontwikkelen. Hiervoor is wederom de vaardigheidsgroei tussen 2016 en 2018 gebruikt. Omdat de vaardigheden met verschillende testen zijn gemeten, is hun groei eerst gestandaardiseerd. Hiervoor is bij iedere leerling de geobserveerde toename in score op elke test gedeeld door de maximaal mogelijke toename. In formule:

vaardigheidsgroei' = vaardigheidsgroei ÷ (maximale testscore – testscore in 2016)

Het resultaat van deze berekening is de proportie vooruitgang ten opzichte van de maximaal mogelijke vooruitgang; in Tabel 12.6 zijn deze waarden omgerekend naar percentages (vermenigvuldigd met 100). Voor

het onderdeel 'informatie zoeken' benutten de leerlingen gemiddeld 43 procent van hun mogelijkheid tot groei, voor de overige vaardigheden lagen deze waarden rond de 20 procent.

De resultaten zijn geanalyseerd met een univariate ANOVA met het soort vaardigheid als onafhankelijke variabele en de gestandaardiseerde vaardigheidsgroei als afhankelijke variabele. Uit de resultaten bleek dat de vaardigheden zich niet in gelijke mate hadden ontwikkeld, $F(6, 1135) = 10.34, p < .001, \eta^2 = .05$. Om te bepalen waardoor dit verschil werd veroorzaakt, zijn post hoc analyses uitgevoerd met een Hochberg *step-up* correctie voor meervoudige vergelijkingen. Hieruit bleek dat 'informatie zoeken' significant meer was verbeterd dan alle andere vaardigheden ($.001 < p < .005$). De overige verschillen waren niet statistisch significant ($p > .095$).



TERUG NAAR
DE HOOFDTEKST
PAGINA 30

Tabel 12.6: Gestandaardiseerde vaardigheidsgroei (%) van 2016 tot 2018

	1.	2.	3.	4.	5.	6.	7.
1. Informatie zoeken	-.49**	.07	.12	.56**	.04	-.09	.14
2. Informatie creëren	.05	-.40**	.10	.05	-.01	-.04	.07
3. Informatie bewerken	.02	.06	-.45**	.00	-.03	-.14	.14
4. Informatie gebruiken	.47**	.15	-.05	-.48**	-.01	.00	.11
5. Kritisch denken	-.13	.03	-.06	-.11	-.42**	-.09	-.03
6. Taal	.07	.03	.02	.02	-.11	-.42**	.08
7. Rekenen	-.04	.13	-.08	.04	-.04	-.11	-.33**

* $p < .05$ (tweezijdig) ** $p < .001$ (tweezijdig)

13. Geraadpleegde literatuur

American Philosophical Association (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. "The Delphi Report," Committee on Pre-College Philosophy. ERIC Document Reproduction No. ED 315 423.

Baines, E., Rubie-Davies, C., & Blatchford, P. (2009). Improving pupil group work interaction and dialogue in primary classrooms: Results from a year-long intervention study. *Cambridge Journal of Education*, 39, 95-117.

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, 70, 1098-1120.

Coiro, J. (2006). *Measuring Internet comprehension: Accessing, evaluating, and communicating information*. Poster gepresenteerd bij de jaarlijkse bijeenkomst van de Educational Research Association, Chicago, IL.

van Deursen, A., van Dijk, J., & Peters, O. (2012). Proposing a survey instrument for measuring operational, formal, information and strategic internet skills. *International Journal of Human-Computer Interaction*, 28, 827-837.

van Deursen, A. J. A. M., Helsper, E. J. & Eynon, R. (2014). *Measuring digital skills: From digital skills to tangible outcomes project report*. Verkregen van www.oii.ox.ac.uk/research/projects/?id=112

Ennis, R. H., Gardiner, W. L., Morrow, R., Paulus, D., & Ringel, L. (1964). *Cornell critical thinking test series*. Urbana Champaign: University of Illinois.

Ennis, R. H., & Weir, E. (1985). *The Ennis-Weir critical thinking essay test*. Pacific Grove: Midwest publications.

Eshet-Alkalai, E. (2004). Digital literacy: A conceptual framework for survival skills in the digital era. *Journal of Educational Multimedia and Hypermedia*, 13, 93-106.

Facione, P.A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. ERIC Document Reproduction No. ED 315-432.

Finken, M., & Ennis, R. H. (1993). *Illinois critical thinking essay test*. Urbana Champaign: University of Illinois.

Fraillon, J., Schulz, W., & Ainley, J. (2013). *International computer and information literacy study: Assessment framework*. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).

Geijssel, F. (2010). Praktijkgericht onderzoek: Goed voor de onderwijskundige, de school, de docent en de leerling? *Pedagogische Studiën*, 87, 288-296.

Hargittai, E. (2004). Classifying and coding online actions. *Social Science Computer Review*, 22, 210-227.

Hargittai, E. (2005). Survey measures of web-oriented digital literacy. *Social Science Computer Review*, 23, 371-379

Insight Assessment (2013). *California Critical Thinking Skills Test: CCTST test manual*. San Jose, CA: Insight Assessment/California Academic Press.

Kanari, Z., & Millar, R. (2004). Reasoning from data: How students collect and interpret data in science investigations. *Journal of Research in Science Teaching*, 41, 748-769.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, 15, 661-667.

Kruit, P., Oostdam, R., van den Berg, E., & Schuitema, J. (2018). Performance assessment as a diagnostic tool for science teachers. *Research in Science Education*, doi: 10.1007/s11165-018-9724-9

KNAW (2012). *Digitale geletterdheid in het voortgezet onderwijs: Vaardigheden en attitudes voor de 21ste eeuw*. Amsterdam: KNAW.

Lazonder, A. W., & Rouet, J.-F. (2008). Information-problem solving instruction: Some cognitive and metacognitive issues. *Computers in Human Behavior*, 24, 753-765.

Lazonder, A. W., van der Meij, H., & de Vries, B. (2000). *Wizzkids of digibeten? Een onderzoek naar Internetvaardigheden van leerlingen uit groep 7 en 8 van het basisonderwijs*. Ongepubliceerd rapport, Universiteit Twente, Vakgroep Instructietechnologie.

Masnick, A. M., & Morris, B. J. (2008). Investigating the development of data evaluation: The role of data characteristics. *Child Development*, 79, 1032-1048.

Mercer, N., Wegerif, R., & Dawes, L. (1999). Children's talk and development of reasoning in the classroom. *British Educational Research Journal*, 25, 95-111.

Ministerie van Onderwijs, Cultuur en Wetenschappen (2006). *Nieuwe gewichten-regeling basisonderwijs*. Zoetermeer: CFI.

Mooij, T. (2007). Contextual learning theory: Concrete form and a software prototype to improve early education. *Computers & Education*, 48, 100-118.

OECD (2011). *The OECD guide to measuring the information society*. Verkregen van <http://browse.oecdbookshop.org/oecd/pdfs/free/9311021e.pdf>

OECD (2013). *PISA 2015 draft collaborative problem solving framework*. Verkregen van <http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf>

Okey, J. R., Wise, K. C., & Burns, J. C. (1982). *Integrated process skills test II*. University of Georgia, Department of science education.

Mayer, D., Sodian, B., Koerber, S., & Schwiippert, K. (2014). Scientific reasoning in elementary school children: Assessment and relations with cognitive abilities. *Learning and Instruction*, 29, 43-55.

Meelissen, M. R. M., Punter, R. A., & Drent, M. (2014). *Digitale geletterdheid van leerlingen in het tweede leerjaar van het voortgezet onderwijs. Nederlandse resultaten van ICILS-2013*. Enschede: Universiteit Twente, Vakgroep OMD.

NRO (2014). *Call for proposals langlopend onderwijsonderzoek: 2015 eerste ronde*. Den Haag: Nederlandse Organisatie voor Wetenschappelijk Onderzoek.

Ólafsson, K., Livingstone, S., & Haddon, L. (2014). *Children's use of online technologies in Europe: A review of the European evidence base* (revised edition). London: LSE/EU Kids Online.

Pedaste, M., Mäeots, M., Siiman, L. A., de Jong, T., van Riesen, S. A. N., Kamp, E. T., ... Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, 14, 47-61.

Piekny, J., Gruber, D., & Maehler, C. (2014). The development of experimentation and evidence evaluation skills at preschool age. *International Journal of Science Education*, 36, 334-354.

Piekny, J., & Maehler, C. (2013). Scientific reasoning in early and middle childhood: The development of domain-general evidence evaluation, experimentation, and hypothesis generation skills. *British Journal of Developmental Psychology*, 31, 153-179.

Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem-solving. In C. E. O'Malley (Ed.), *Computer-supported collaborative learning* (pp. 69-97). Berlijn: Springer-Verlag.

Saab, N., van Joolingen, W.R., & van Hout-Wolters, B.H.A.M. (2007). Supporting communication in a collaborative discovery learning environment: The effect of instruction. *Instructional Science*, 35, 73-98.

Sijtsma, K. (2009). Over misverstanden rond Cronbachs alfa en de wenselijkheid van alternatieven. *De Psycholoog*, 44(10), 561-567.

Thijs, A., Fisser, P., & van der Hoeven, M. (2014). *21e eeuwse vaardigheden in het curriculum van het funderend onderwijs*. Enschede: SLO.

Van de Keere, K., & Vervaet, S. (2013). *Leren is onderzoeken*. Leuven: LannooCampus.

Voogt, J., & Pareja Roblin, N. (2010). *21st century skills: Discussienota*. Enschede: Universiteit Twente.

de Vries, B., van der Meij, H., & Lazonder, A. W. (2008). Supporting reflective web searching in elementary schools. *Computers in Human Behavior*, 24, 649-665.

Walraven, A. (2008). *Becoming a critical websearcher. Effects of instruction to foster transfer*. Proefschrift Open Universiteit Nederland, Heerlen.

Watson, G., & Glaser E. M. (1964). *Watson-Glaser critical thinking appraisal*. New York: Harcourt, Brace and World.