

Dynamisch toetsen van onderzoeksvaardigheden op het terrein van Natuur en Techniek bij leerlingen van groep 8 van het basisonderwijs

E. Roelofs, D. Wammes, W. Emons en M. Raijmakers¹

Samenvatting

In een heranalyse van data van een peilingsonderzoek in het najaar van 2015 is nagegaan wat de informatieve meerwaarde kan zijn van dynamisch toetsen ten opzichte van statisch toetsen, voor het vaststellen van onderzoeksvaardigheden uit het domein Natuur & Techniek bij leerlingen uit groep 8 van het basisonderwijs. Hiertoe zijn de data (n=203) gebruikt van één van de praktische opdrachten uit het peilingsonderzoek, de knikkerbaan. Bij de afname van de opdrachten is afhankelijk van de voortgang bij de deeltaken, lees-, proceshulp en inhoudelijke hulp verstrekt, volgens een vast protocol van 'graduated prompts'. Analyses gericht op het vaststellen van de gemaakte progressie in experiment 2 na ondersteunde deelname aan experiment 1 brachten groepen met een verschillend leerpotentieel aan het licht. De grootste geconstateerde progressie maakten leerlingen die in experiment 1 ook na alle hulp niet tot een oplossing van een deelopdracht kwamen, maar daar wel zelfstandig in slaagden in experiment 2. Het relatieve aandeel van niet-oplossers in experiment 1 met deze progressie bedroeg 36 % bij variabelen instellen 42% bij herhaald meten en 56% bij conclusies trekken.

Latente klasse-analyses resulteerden in drie groepen die elk een eigen hulpgeschiedenis doormaakten. Het grootste verschil betrof de gevoeligheid voor hulp bij een cruciale deelvaardigheid voor experimenteren: control-of-variables. Verder bleek bij een groep het geven van lees-/proceshulp een belangrijk aandeel te spelen bij het vinden van oplossingen. De groep leerlingen die het minst profiteerden van hulp bleken minder kennis te hebben van N&T en een negatiever beeld te hebben over de eigen onderzoeksvaardigheden.

Kernwoorden: dynamisch toetsen, natuur en techniek, peilingsonderzoek, formatief evalueren, onderzoekend leren, toegankelijkheid, regulatieve hulp, responsiviteit voor ondersteuning

1 Inleiding

In de afgelopen decennia is in het (basis) onderwijs sprake van een toenemende aandacht voor het aanleren van vaardigheden waarmee leerlingen zélf kennis kunnen verwerven en toepassen. Dit geldt nadrukkelijk voor onderzoeks- en ontwerpvaardigheden in het domein natuur en techniek. Om deze ontwikkeling te bespoedigen is geïnvesteerd in materiaal waarmee leerlingen kunnen onderzoeken en ontwerpen en in de opleiding en nascholing van leerkrachten (De Vries, Van Keulen, Peters, & Walma van der Molen, 2012). Om de effecten daarvan op de onderzoeks- en ontwerpvaardigheden van leerlingen aan het eind van de basisschool te monitoren, zijn sinds de peiling van 2002 praktische opdrachten opgenomen in de Periodieke Peiling van het Onderwijsniveau (PPON) natuur en techniek. De keuze voor praktische opdrachten hing tevens samen met een belangrijk aspect van validiteit, namelijk de mate waarin leerlingen gelegenheid krijgen om vaardigheden te tonen op een manier die dicht bij de wijze ligt waarop die vaardigheden in het dagelijks leven worden gebruikt (National Assessment Governing Board, 2013; OECD, 2014).

De praktische opdrachten uit de Periodieke onderwijspeilingen van 2002 en 2010 leverden weliswaar informatie op over de prestaties op de onderzoeksopdrachten, maar niet over specifieke sterktes en zwaktes in onderliggende onderzoeks- en ontwerpvaardigheden. Bij een incorrect uitgevoerde opdracht kon uit het resultaat niet afgeleid worden welke van de benodigde deelvaardigheden wel of niet werden beheerst. Bovendien kon door de uniforme afnamewijze niet worden vastgesteld of leerlingen om andere oorzaken dan onvoldoende ontwikkelde doelvaardigheden vastliepen in opdrachten, zoals een beperking in noodzakelijke toegangsvaardigheden om überhaupt te kunnen

starten met de opdracht (Beddow, Kurz, & Frey, 2011).

Een methode om meer gedetailleerd informatie te verkrijgen over de onderzoeks- en ontwerpvaardigheden van leerlingen is het gebruik van dynamische toetsen. Met dynamische toetsen wordt zichtbaar gemaakt waartoe leerlingen in de toekomst in staat zijn, wanneer ze tijdens de toets of tussen twee parallelle afnames van een toets, gelegenheid hebben gehad om vaardigheden onder deskundige begeleiding te verwerven en te oefenen. Daarmee sluit dynamische toetsing conceptueel aan bij het concept 'zone van naaste ontwikkeling' (Lidz & Gindis, 2003; Vygotsky, 1978). Het verschil tussen waartoe een leerling zelfstandig in staat is en wat deze kan met hulp van een volwassene, wordt bij dynamische toetsing aangeduid als het leerpotentieel (Lidz & Elliott, 2000). Leerpotentieel wordt in de literatuur nader geïnterpreteerd als de snelheid, de efficiëntie en het gemak waarmee leerlingen zich vaardigheden eigen maken. Een centrale verklarende variabele hierbij is de responsiviteit van de leerling voor ondersteunende interventies, die naar aard, aantal en mate van specificiteit uiteen kunnen lopen (Grigorenko, 2009).

Bij de toepassing van dynamisch toetsen worden twee benaderingen onderscheiden die tevens gepaard gaan met verschillende kwantitatieve maten voor leerpotentieel. Volgens de eerste benadering, het 'sandwich model' (Sternberg & Grigorenko, 2002) krijgt de leerling eerst een pretest voorgelegd. Op basis van de prestaties op deze pretest wordt gerichte instructie gegeven die is afgestemd op geconstateerde sterktes en zwaktes. Vervolgens wordt de post-test voorgelegd, een parallelversie van de pretest. De prestatieverbetering wordt opgevat als een maat voor leerpotentieel.

Volgens de tweede benadering, het 'gelaagde cake-model' (Sternberg & Grigorenko, 2002), krijgt de leerling een reeks opdrachten voorgeschoteld, waarbij onmiddellijk hulp wordt geboden wanneer de leerling niet verder kan met een opdracht. Zodra een oplossing is gevonden, wordt de volgende opdracht gepresenteerd. Als maat van leerpotentieel wordt een kwantificering van de hoe-

veelheid aanwijzingen of scaffolds gehanteerd die leerlingen tijdens de taakuitvoering nodig hadden om de opdracht op te lossen (Fuchs et al, 2011).

Bij vaardigheidsontwikkeling, die gezien kan worden als een fasegewijs proces van integratie van deelvaardigheden (Fischer & Bidell, 2007; Van Geert, 1998), is het met statische toetsing mogelijk om vast te stellen of leerlingen de vaardigheid of de benodigde deelvaardigheden beheersen. Met dynamische toetsing kan daarbovenop zichtbaar worden bij welke leerlingen zich het integratieproces aan het voltrekken is. Bij die leerlingen kan een aanwijzing dat gegevens gecombineerd kunnen worden, voldoende zijn om de taak correct af te ronden. Het kan ook zijn dat de vaardigheid begrijpend lezen nog een beperkende factor is en dat voorlezen een leerling in staat stelt om de opdracht correct uit te voeren. Ook kan blijken dat een onderliggend begrip of principe bij de uit te voeren taak niet werd begrepen, maar dat dit in een vervolgtask wel het geval was, nadat hierover een aanwijzing was gegeven in de fase van ondersteuning.

Leerpotentieel is zelf een dynamische eigenschap. Het vermogen om te profiteren van ondersteuning wordt vastgesteld door prestaties op twee of meer meetpunten te vergelijken, waarbij startniveaus kunnen verschillen. Vervolgens moet een prestatieverbetering tussen de meetpunten gerelateerd worden aan de aard en intensiteit van de hulp die voorafgaand aan een volgend meetpunt wordt gegeven, welke ook varieert tussen leerlingen. Over de verklaring voor deze variaties tussen leerlingen is vooralsnog weinig bekend in de literatuur over dynamisch toetsen, maar deze zal gezocht moeten worden in een combinatie van leerlingkenmerken zoals kennis, ervaring, en attitude maar ook de context waarin dynamische toetsing plaatsvindt (Van der Steen, Steenbeek, & Van Geert, 2012).

Caffrey, Fuchs en Fuchs (2008) concludeerden in een meta-analyse dat het met dynamisch toetsen aangetoonde leerpotentieel unieke variantie verklaart in het voorspellen van toekomstig presteren, bovenop wat statische cognitieve prestatietoetsen al voorspelden. Dergelijke verschillen in gevoeligheid voor

hulpinterventies worden niet zichtbaar bij statische toetsing. De voorspellende kracht van dynamische toetsing is met name vastgesteld op het gebied van begrijpend lezen (Swanson & Howard, 2005; Swanson, 2011) en wiskunde (Fuchs et al., 2008; Fuchs et al., 2011).

2 Ingezette hulpvormen bij dynamische Toetsing

De hulp die eigen is aan dynamische toetsing kan worden vormgegeven met een interactie-model of een interventiemodel (Sternberg & Grigorenko, 2002). Bij het *interactiemodel* heeft de toetsleider de rol van mediator. De interactie met de leerling in de zogeheten ‘Mediated learning experience’ (MLE) verloopt spontaan. De feedback die leerlingen krijgen is niet voorgeschreven en de inhoud en vorm ervan wordt overgelaten aan wat de toetsleider op dat moment het best acht. Bij het *interventiemodel* is sprake van een vooraf beschreven protocol van gestandaardiseerde interventies met een toenemende specificiteit (Fuchs et al., 2011). Die interventies worden volgens een vooraf bepaalde hiërarchie van hints gegeven. Dit wordt ook wel aangeduid als de ‘graduated prompt approach’ (Ferrara, Brown, & Campione, 1986). Bij het interventiemodel wordt telkens afgewacht wat het effect is van een bepaalde eerdere, nog weinig specifieke vorm van hulp, alvorens een volgende meer specifieke hulpvorm wordt aangeboden. Recente studies hebben aangetoond dat het dynamische toetsen volgens het interventiemodel leidt tot de beste voorspelling van leerpotentie (Caffrey, Fuchs, & Fuchs, 2008).

Hulpinterventies bij dynamisch toetsen kunnen aangrijpen op verschillende fasen of aspecten van probleem oplossen door de leerling, zoals oriëntatie op het taakdoel en vereisten, het begrijpen van taakinformatie, het bedenken en plannen van een oplossing, het uitwerken van de oplossing, het monitoren en bijsturen en het reflecteren op de oplossing (Polya, 1945; Schoenfeld, 1992).

Een eerste vorm van hulp in een sequentie van hulpinterventies is *toegangshulp* (Bedow et al., 2011) die met name aangrijpt op het begrijpen van taakinformatie. Voorbeelden

van toegangshulp zijn: het voorlezen van een opdracht of het in andere woorden weergeven van de opdracht (leeshulp). Affectieve hulp is een andere vorm van hulp die geboden kan worden, het gaat daarbij om het op hun gemak stellen van een leerling of het aanmoedigen van een leerling om door te gaan. Dan is er *proceshulp*, welke aangrijpt op regulatieve processen, zoals steun bij het plannen, het richten van de aandacht of het wijzen op de noodzaak om een berekening te controleren.

Vervolgens is er *inhoudelijke hulp* waarmee een deelstap naar de oplossing of een deel van die oplossing wordt gegeven. Hierdoor krijgt de leerling aanknopingspunten om het probleem alsnog op te lossen, wanneer dat zelfstandig niet lukt. Het *modelleren* van een deeltaak is de meest ingrijpende vorm van hulp die gebruikt kan worden in de context van het dynamisch toetsen. Bij modelleren wordt de oplossing op een deeltaak gegeven door de toetsleider. Dit stelt de leerling in staat om verder te gaan met een volgend onderdeel van een opdracht.

3 Onderzoeksdoel

Het doel van dit artikel is om te verkennen in hoeverre het gebruik van dynamische Natuur & Techniek-toetsen, waarbij tijdens de toets verschillende typen hulp worden aangeboden, bijdraagt aan het verkrijgen van een méér verfijnd inzicht in de onderzoeksvaardigheden van leerlingen. Verwacht wordt dat met dynamische toetsen niet alleen informatie wordt verkregen over wat leerlingen in het verleden hebben geleerd, maar ook over hun gevoeligheid voor hulpinterventies, wat gezien kan worden als een indicatie van waartoe zij in staat zijn na het ontvangen van onderwijs.

Voor dit onderzoek zijn afnamedata gebruikt van één van de praktische onderzoekopdrachten uit het peilingsonderzoek Natuur en Techniek (Cito, 2016), namelijk de knikkerbaan, vaak ook aangeduid als de hellingproef. De hellingproef is ontworpen om vast te stellen of leerlingen de Control-of-Variables Strategy (CVS) kunnen toepassen (Klahr & Nigam, 2004; Van der Graaf, Segers, & Verhoeven, 2015; Van Wagenveld,

Segers, Kleemans, & Verhoeven, 2014). In het peilingsonderzoek is deze taak ook gebruikt om vast te stellen of leerlingen rekening houden met mogelijke variatie in meetresultaten en of ze in staat zijn om conclusies te trekken die in overeenstemming zijn met de door hen verzamelde data.

Net als overigens bij de andere opdrachten in het peilingsonderzoek is bij de knikkerbaan het interventiemodel toegepast. We veronderstellen dat de inzet van dit model het mogelijk maakt om groepen leerlingen te onderscheiden die van elkaar verschillen in leerpotentieel, gemeten in termen van prestatieverbetering en gerelateerd aan de aard van de gebruikte hulpvormen. Daarnaast verkennen we of er een verband bestaat tussen die verschillen in leerpotentieel en een aantal leerlingkenmerken.

4 Methode

4.1 Afname en scoring van de knikkerbaanopdracht

4.1.1 *Materialen*

Voor dit artikel zijn de afnamedata gebruikt van de onderzoeksopdracht ‘knikkerbaan’ die onderdeel vormde van een in 2015 uitgevoerde peilingsonderzoek (Cito, 2016; Roelofs, 2017). Voor de taak is de baan in tweevoud aanwezig. Elke baan bestaat uit een helling waar een knikker vanaf kan rollen en een geribbelde uitloopstrook met oplopende nummers. Bij elke knikkerbaan zijn er vier variabelen die elk op twee verschillende manieren kunnen worden ingesteld: de hellingshoek, het startpunt van de knikker, de stroefheid van het oppervlak van de helling en het gewicht van de knikker. Voor de instelling van de laatste variabele zijn twee paar knikkers met een verschillend gewicht beschikbaar.

4.1.2 *Afname*

In het peilingsonderzoek is de knikkerbaan als individuele opdracht én als groepsopdracht afgenomen. In dit artikel worden alleen de data van de individuele afnames gebruikt. De eerste vier deeltaken waren erop gericht om de leerling kennis te laten maken met de vraagstelling en het materiaal. Bij de

eerste deeltaak waren de banen verschillend ingesteld op één van de onafhankelijke variabelen. De leerling werd gevraagd om te voorspellen op welke baan een knikker het verst zou rollen. Bij de tweede deeltaak kregen de leerlingen twee zware en twee lichte knikkers waarmee ze hun voorspelling konden controleren door de knikkers van de vooraf ingestelde banen te laten rollen. De derde deeltaak bestond uit het lezen van een stripverhaal waarin werd omschreven wat een experiment is. Daarna volgden drie korte beschrijvingen van een onderzoek met de vraag bij welke beschrijving de omschrijving ‘experiment’ het best past. In de vierde deeltaak werd de leerling gevraagd om aan de hand van getekende voorbeelden de banen verschillend in te stellen voor telkens één van de variabelen.

Vervolgens werden drie verschillende onderzoeksvaardigheden praktisch getoetst met drie deeltaken die afgeleid zijn uit de Vaardigheden Rubrics Onderzoeken en Ontwerpen (VROO; Keulen, Slot, & Boonstra, 2013): variabelen instellen, meten en concluderen (zie Tabel 1). Dit gebeurde in twee opeenvolgende experimenten. In experiment 1 werd de leerling gevraagd te onderzoeken of de startpositie van de knikker invloed heeft op de afgelegde afstand en in experiment 2 of de hellingshoek van de baan invloed heeft op de afgelegde afstand.

Beide experimenten hadden dezelfde structuur, maar voor de deeltaak ‘meten’ was bij het eerste experiment een tabel met twee kolommen en vijf genummerde rijen beschikbaar om resultaten in te vullen. Bij het tweede experiment waren alleen twee kolommen aangegeven en geen rijen. Leerlingen moesten hier zelf het belang inzien van meerdere metingen.

4.1.3 *Hulpsoorten*

De hulp was, volgens het interventiemodel van dynamische toetsen, vastgelegd in een protocol. Tijdens de eerste drie minuten van de uitvoering van een deeltaak was alleen leeshulp of proceshulp toegestaan. Dit om de leerlingen de kans te bieden de opdracht zelfstandig uit te voeren. Na drie minuten mocht de toetsleider inhoudelijke hulp verstrekken als de leerling daarom vroeg. Na tweederde

Tabel 1

Overzicht van de deeltaken van de opdracht knikkerbaan

VROO	Deeltaak	Criteria voor correcte uitvoering	Inhoudelijke hulp
12. Antwoorden op de onderzoeksvraag verzamelen	variabelen instellen	De te onderzoeken onafhankelijke variabele is verschillend ingesteld en de overige onafhankelijke variabelen zijn gelijk ingesteld	Bij (foutieve) verschillende instelling: "Weet je nu of het alleen aan de [variabele]b) ligt of een bal verder komt? Of kan het ook aan de andere instellingen liggen? Bij gelijke instelling: Als je wit weten of de [variabele] effect heeft, hoe moet je dan de [variabele] instellen? Heb je aan één meting genoeg? Weet je het dan zeker? Heb je nu genoeg metingen om een conclusie te kunnen trekken? Is het eerlijk als je verschillende knikers gebruikt? Wat moet je achter Baan 1 invullen? (Noteren hoe de te onderzoeken variabele op baan 1 resp. baan 2 is ingesteld.t)
13. Gegevens verwerken	meten	Er zijn vijf metingen met hetzelfde paar knikers uitgevoerd. De afstanden zijn correct afgelezen en zijn in de juiste tabelkolom genoteerd ^a .	
14. Conclusies trekken en bediscussiëren	concluderen	De conclusie is in overeenstemming met de gegevens uit de metingen.	Wijs op het (ontbreken van) een patroon in de data. Kijk een naar wat je hebt opgeschreven. Maakt het wat uit of de helling ^b steil of flauw is?

Noot. ^a Er zijn twee kolommen die elk horen bij een waarde van de variabele die wordt onderzocht;
^b Benoemde variabele is startpositie bij experiment 1 en helling bij experiment 2

van de beschikbare tijd bood de toetsleider hulp aan. Nog een minuut later werd deze hulp altijd (ongevraagd) gegeven als leerlingen niet verder kwamen. Als de leerlingen ook na hulp niet tot een oplossing kwamen bij het verstrijken van de beschikbare tijd, werd de oplossing gemodelleerd. Voor de experimenten 1 en 2 van de Knikkerbaan waren resp. 20 en 25 minuten beschikbaar, welke tijd evenredig werd verdeeld over de drie deeltaken van elk experiment. In Figuur 1 is de procedure weergegeven voor een deeltaak van 10 minuten.

De *leeshulp* beperkte zich tot het verduidelijken van de geschreven tekst zonder daar nieuwe informatie aan toe te voegen. De *proceshulp* omvatte niet-specifieke regulatieve steun met aanwijzingen als: "Lees het nog eens goed", "Kun je verder?", "Begrijp je de opdracht?", "Heb je het goed op kunnen schrijven?", "Heb je gecontroleerd of alles goed staat?". Ook het geven van tijdsindicaties viel onder deze regulatieve steun: "Je hebt nog 5 minuten". Ook de affectieve steun is als proceshulp gescoord. Deze betrof

bemoedigende opmerkingen als "Ga zo door", "Je doet het prima", en "Dat is duidelijk".

Het protocol bevat voor de *inhoudelijke hulp* per opdracht specifieke aanwijzingen (zie tabel 1). Inhoudelijke hulp betrof nooit het geven van een volledig antwoord.

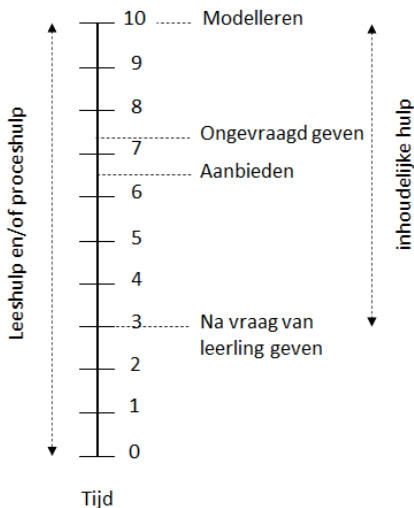
Modelleren is de vorm van hulp waarbij de toetsleider het volledige antwoord geeft of de gevraagde handeling voordoet. Bij variabelen instellen werden de variabelen correct ingesteld door de toetsleider. Bij het meten werden samen met de leerling vijf metingen verricht en bij het concluderen werd een op de waarnemingen gebaseerde conclusie gedicteerd.

4.1.4 Scoring

Of de leerling een specifieke deelvaardigheid al dan niet toonde werd per deeltaak vastgelegd met één of meer prestatie-indicatoren, zoals het correct instellen van de verschillende variabelen. De scoring verliep als volgt. Vlak voordat de begeleiders ertoe besloten om hulp te geven, scoorden ze alle prestatie-

indicatoren voor de betreffende deeltaak onder de conditie ‘vóór hulp’, met hetzij 0 (‘niet voldaan’) of 1 (‘voldaan’). In geval van het verstrekken van inhoudelijke hulp scoorde de begeleider aan het eind van de opdracht-tijd in de kolom ‘na inhoudelijke hulp’ de indicatoren waarop voor hulp 0 was gescoord opnieuw met ofwel 0 (nog steeds niet voldaan) of 1 (nu wel voldaan). Om redenen van praktische uitvoerbaarheid, kon eventuele leeshulp en proceshulp per deeltaak éénmaal worden aangevinkt, ook als die meerdere malen werd gegeven. Leeshulp en proceshulp werden dus niet geregistreerd in combinatie met scores op prestatie-indicatoren. Tot slot werd genoteerd of na alle verstrekte inhoudelijke hulp een niet correct voltooide deeltaak was gemodelleerd.

4.1.5 Afnameprocedure



Figuur 1
Gehanteerde richtlijn voor de inzet van hulpinterventies bij de praktische opdracht bij een opdrachtduur van 10 minuten

In totaal namen 40 derdejaarsstudenten aan een lerarenopleiding basisonderwijs in het oosten van het land de rol van toetsleider op zich. Voor het doel van het peilingsonderzoek, waarvan deze studie deel uitmaakt, is een training verzorgd in het begeleiden en scoren van praktische opdrachten. Voor de afnames van de individuele opdrachten was in een speciaal ingerichte ruimte van de school plaats voor zes individuele praktische

opdrachten voor evenzoveel leerlingen. In de ruimte waren twee toetsleiders aanwezig die elk drie leerlingen begeleidden bij een opdracht. Tussentijds scoorden de toetsleiders de opdrachten. Om de beoordelaarsbetrouwbaarheid te kunnen bepalen is elke praktische opdracht bij ongeveer 9% van de deelnemende leerlingen door een tweede beoordelaar bijgewoond en gescoord, zonder dat deze tweede beoordelaar zelf intervenieerde. De beoordelaarsovereenstemming is vastgesteld met de Gower-coëfficiënt (G; Gower, 1971). Deze was voor de deeltaken minimaal .86 en dus goed te noemen.

4.2 Kennistoets Natuur en Techniek

Bij het peilingsonderzoek is ook een kennistoets Natuur en Techniek afgenomen. Deze schriftelijke toets bestond uit 26 opgaven, verdeeld over zes domeinen van N&T: Leefomgeving, Vorm en functie, Onderzoeken, Weer en klimaat, Werking, Vorm en materiaal (zie voor meer informatie: Cito, 2016).

4.3 Leerlingvragenlijsten

Met behulp van vragenlijsten is informatie ingewonnen over de attitude van leerlingen ten aanzien van Natuur en Techniek en de activiteiten, die zij in de eigen vrije tijd ondernemen ten aanzien van N&T. Voor attitude is gebruik gemaakt van een instrument dat eerder ontwikkeld is door Denessen et al. (2011) en Van Langen en Vierke (2009). Hierbij kreeg de leerling twaalf stellingen voorgelegd over vier aspecten die elk een (betrouwbare) subschaal vormen: a) de mate van plezier die de leerling ervaart met N&T, b) het ingeschatte belang van N&T voor zichzelf en de samenleving, c) plannen om zelf iets te gaan doen met N&T in de eigen toekomst, en d) de ervaren moeilijkheid van N&T.

Daarnaast kregen de leerlingen negen vragen voorgelegd over aan Natuur en Techniek gerelateerde activiteiten die ze in het betreffende schooljaar in hun vrije tijd hadden ondernomen. De antwoordmogelijkheden 1, 2, 3 hebben respectievelijk de betekenis ‘nooit’, ‘één of twee keer’ en ‘vaker dan twee keer’. De schaalscore Vrijtijdsbesteding aan N&T is tot stand gekomen door de gemid-

delde itemscore te berekenen. Deze schaal-score bleek een acceptabele betrouwbaarheid te hebben ($\alpha = .73$).

Na afloop van de praktische opdracht werd de leerlingen gevraagd om aan de hand van een vragenlijst zichzelf te beoordelen op de onderliggende VROO-ontwerp -en onderzoeksvaardigheden. De uitgangstelling was: "Geef nu aan hoe goed je bent in de volgende onderdelen van ontwerpen". Daarop volgden deelactiviteiten van onderzoek of ontwerp. Enkele voorbeelden: "goede redenen geven waarom iets zal gebeuren (voorspelling doen)", "Goed klaarzetten van de twee knikkerbanen bij de experimenten". De leerlingen konden hierbij antwoorden op een vier-puntschaal: 1: 'Daar gaat nog veel mis mee'; 2: 'Dat lukt, maar met enkele fouten'; 3: 'Dat lukt goed, heel soms een fout'; 4: 'Dat lukt perfect'. De schaal bij de door ons gebruikte opdracht Knikkerbaan (8 items) heeft een goede betrouwbaarheid ($\alpha = .80$).

4.4 Data

Aan het onderzoek deden 107 basisscholen mee die representatief waren naar regio, verstedelijking, denominatie, schoolgrootte, en percentage gewichtenleerlingen (zie ITS, Expertisecentrum Nederlands, 2016). Per school zijn acht praktijkopdrachten afgenomen, verdeeld over verschillende natuur en techniek domeinen (Cito, 2016). Drie opdrachten hadden betrekking op onderzoeksvaardigheden, waaronder de Knikkerbaan, en vijf op ontwerpvaardigheden. Per school werkten ongeveer drie leerlingen aan dezelfde praktische opdracht. In totaal hebben 203 leerlingen de knikkerbaan-opdracht uitgevoerd (52% meisjes en 48% jongens). Onder hen bevonden zich 0.4% 9-jarigen, 5.3% 10-jarigen, 68.3% 11-jarigen, 20.7% 12-jarigen en 0.4% 13-jarigen.

4.5 Data-analyses

Het leerpotentieel in de betekenis van het sandwich model is vastgesteld door groepjes leerlingen te onderscheiden die er bij experiment 2 in slaagden de deelopdrachten zelfstandig uit te voeren, maar die in experiment 1 nog enige vorm van hulp nodig hadden, inclusief hen die geen oplossing vonden.

Daarnaast is over de resultaten van beide experimenten een latente klasse analyse (LCA) uitgevoerd om te achterhalen of er groepen leerlingen zijn te onderscheiden met een verwante geschiedenis in de soorten hulp die zij in beide experimenten nodig hadden om tot een oplossing te komen. Deze analyse sluit meer aan bij het gelaagde cake-model van dynamisch toetsen en geeft een beeld van verschillen in responsiviteit voor interventies.

Latente klasse analyse (LCA) is een systematische manier om mensen in te delen in homogene groepen (Magidson & Vermunt, 2004). LCA is het meest verwant met K-means clustering (e.g., Paas, Vermunt, & Bijmolt, 2006), maar met als essentieel verschil dat er gebruik wordt gemaakt van een onderliggend statistisch model. Deze modelmatige aanpak heeft een aantal voordelen. Er hoeven geen ad-hoc beslissingen genomen te worden over het clusteringcriterium, de procedure van samenvoegen, en de keuze voor het aantal clusters. Bovendien worden de individuen probabilistisch aan klassen toegewezen in plaats van een harde opdeling in elkaar uitsluitende clusters. Dit betekent dat in het model rekening wordt gehouden met onzekerheden in klasselidmaatschap. De latente klasse-analyses zijn uitgevoerd met het R-pakket poLCA (Linzer & Lewis, 2011).

Het statistische model achter LCA veronderstelt dat de samenhang tussen indicatoren het gevolg is van het feit dat personen tot verschillende niet-direct waarneembare (latente) subgroepen (klassen) behoren. Binnen een klasse is de veronderstelling dat er geen samenhang is tussen de indicatoren. Dit is de aanname van lokale onafhankelijkheid. Latente klasse-analyse is daarmee nauw verwant aan andere latente-variabelen technieken zoals factor analyse en item-respons theorie (e.g., Heinen, 1996) met als verschil dat in LCA de latent variabele een nominale variabele is, terwijl factoranalyse en item-responsmodellen uitgaan van continue latente variabelen. Dit maakt LCA is bij uitstek geschikt om kwalitatieve en kwantitatieve verschillen te beschrijven (Hillen, 2017).

De vraag is hoeveel klassen er nodig zijn om de samenhang tussen indicatoren in voldoende mate te kunnen beschrijven; dat wil

zeggen, een model dat voldoende past bij de data. Het bepalen van het aantal klassen gebeurt doorgaans stapsgewijs, te beginnen met een model met slechts één klasse (baseline model). Vervolgens wordt er steeds één klasse toegevoegd, en wordt gekeken of de fit van het model verbetert. Hierbij maakt men meestal informatiecriteria zoals AIC, BIC en CAIC (Andrews & Currim, 2003). Als op een gegeven moment de winst in de verklaringkracht niet meer opweegt tegen de complexiteit van het model is het optimale aantal klassen bereikt. Nadat is vastgesteld wat voor de voorliggende data set het optimaal aantal clusters is, worden de respondenten toegewezen aan clusters door middel van klasse-lidmaatschapskansen en worden per klasse responsprofielen afgeleid om de inhoudelijke betekenis van de klassen te duiden.

5 Resultaten

5.1 Prestaties en prestatieverbetering onder verschillende hulpcondities

Experimenteel onderzoek verrichten vereist een combinatie van verschillende deeltaken. De knikkerbaanexperimenten van het peilingsonderzoek bestonden uit drie deeltaken: variabelen instellen, meten en concluderen.

Tabel 2 geeft een overzicht van de aantallen en percentages leerlingen die de deelopdrachten oplostten onder uiteenlopende hulp-

condities, ofwel deze niet oplostten. Met uitsluiting van onvolledige dataseries bleek 8.4% en 25.1% van de leerlingen in staat om resp. experiment 1 en 2 zonder enige hulp correct uit te voeren. Met hulp, maar zonder gemodelleerd te worden was nog eens 12.3% en 37.9% daartoe in staat in de opeenvolgende experimenten. In experiment 1 werd bij 79.3% van de leerlingen minimaal één deeltaak gemodelleerd. Dat percentage ligt bij experiment 2 beduidend lager: 36.9%.

In Tabel 2 is de hulp ook verder uitgesplitst per deeltaak. Bij de eerste deeltaak ‘Variabelen instellen’ werd geen inhoudelijke hulp gegeven.

Aan het percentage leerlingen dat erin slaagt om de opdracht zonder enige hulp correct te voltooien is te zien dat de deeltaken verschillen in moeilijkheidsgraad (zie Tabel 2). Opvallend is dat in tegenstelling tot de deeltaken ‘variabelen instellen’ en ‘concluderen’, bij de deeltaak ‘meten’ minder leerlingen zonder hulp de taak oplossen in experiment 2 (46.3%) dan in experiment 1 (57.1%). Bij deze deeltaak was er bij experiment 1 grafische ondersteuning op het werkblad in de vorm van een invultabel met twee kolommen en vijf genummerde rijen. Bij experiment 2 was er alleen een aanduiding van twee kolommen. Het aandeel leerlingen dat na alle hulpvormen de oplossing voor de deeltaak ‘meten’ gemodelleerd moest krijgen is bij experiment 2 veel hoger dan in experiment 1 (20.2% versus 5.9%). Ook is zichtbaar dat

Tabel 2
Hulpcondities waaronder leerlingen de deelopdrachten correct oplostten*

	Variabelen instellen		Meten		Concluderen		Hele experiment	
	n	%	n	%	n	%	n	%
Experiment 1								
Geen hulp	22	10.8	116	57.1	163	80.3	17	8.4
Uitsluitend lees- of proceshulp	21	10.3	59	29.1	16	7.9	23	11.3
Inhoudelijke hulp maar niet gemodelleerd			16	7.9	8	3.9	2	1.0
Na alle hulpvormen oplossing gemodelleerd	160	78.8	12	5.9	16	7.9	161	79.3
Experiment 2								
Geen hulp	n		n		n		n	
Geen hulp	88	43.3	94	46.3	177	87.2	51	25.1
Uitsluitend lees- of proceshulp	17	8.4	26	12.8	7	3.4	22	10.8
Inhoudelijke hulp maar niet gemodelleerd	54	26.6	42	20.7	6	3.0	55	27.1
Na alle hulpvormen oplossing gemodelleerd	44	21.7	41	20.2	13	6.4	75	36.9

N.B.* Alleen leerlingen van wie op alle deeltaken op beide experimenten data beschikbaar waren (n=203)

Tabel 3

Verschuivingen in de benodigde hulp om tot een correcte oplossing te komen

	Experiment 2		Aanduiding		
	Zelfstandige oplossing	Totaal	n	%	
Experiment 1	n	%	n	%	
<i>Deeltaak variabelen instellen</i>					
Zelfstandige oplossing	18	81.8%	22	10.8%	Niveau vastgehouden
Oplossing na lees- of proceshulp	12	57.1%	21	10.3%	Kleinste progressie
Geen resultaat na hulp, oplossing gemodelleerd	58	36.3%	160	78.8%	Grootste progressie
<i>Deeltaak herhaald meten</i>					
Zelfstandige oplossing	66	56.9%	116	57.1%	Niveau vastgehouden
Oplossing na lees- of proceshulp	19	32.2%	59	29.1%	Kleinste progressie
Oplossing na inhoudelijke hulp	4	25.0%	16	7.9%	Middelgrote progressie
Geen resultaat na hulp, oplossing gemodelleerd	5	41.7%	12	5.9%	Grootste progressie
<i>Deeltaak conclusies trekken</i>					
Zelfstandige oplossing	149	91.4%	163	80.3%	Niveau vastgehouden
Oplossing na lees- of proceshulp	13	81.3%	16	7.9%	Kleinste progressie
Oplossing na inhoudelijke hulp	6	75.0%	8	3.9%	Middelgrote progressie
Geen resultaat na hulp, oplossing gemodelleerd	9	56.3%	16	7.9%	Grootste progressie

een deel van de leerlingen met alleen lees- en proceshulp er in slaagt om een volledige oplossing te vinden voor een deeltaak. Dit geldt met name voor het invullen van de tabel bij de deeltaak 'meten' bij het eerste experiment: 59 leerlingen (29.1%).

In Tabel 3 hanteren we een maat voor leerpotentieel die het dichtst bij het sandwichmodel komt, de grootte van de gemaakte progressie. In de tweede en derde kolom worden resp. de aantallen en percentages leerlingen gerapporteerd die in experiment 2 zonder hulp de deelopdrachten uitvoerden en die in experiment 1 nog een bepaalde vorm van hulp nodig hadden om de opdrachten op te lossen of daarin zelfs niet slaagden. In de vierde en vijfde kolom staan de totale aantallen en percentages van de groepen die in experiment 1 na een bepaalde hulpsoort nodig hadden om tot een oplossing te komen vermeld. Voor de volledigheid is ook steeds de groep leerlingen vermeld die de deelopdrachten al in experiment 1 zelfstandig oploste en of die dat niveau heeft vastgehouden. Bij de te vermelden progressie merken we op dat de aantallen geholpen leerlingen bij opdracht meten en concluderen relatief klein zijn.

Tabel 3 laat zien dat er groepjes leerlingen

kunnen worden onderscheiden met een relatief kleine en grote progressie. Daarbij moet opgemerkt worden dat de aanduiding 'kleinste' relatief is. Het betekent dat de leerling in experiment 1 de taak al grotendeels zelfstandig kon uitvoeren en in experiment 2 geheel zelfstandig. Van de grootste progressie is sprake als leerlingen er ook met hulp niet in slaagden een opdracht op te lossen in experiment 1, maar dat in experiment 2 wel zelfstandig deden. We zien dat 36.3% van de betreffende groep leerlingen deze grootste progressie doormaakt voor de deelopdracht variabelen instellen. Voor meten en conclusies trekken zijn deze percentages grootste progressie resp. 41.7% en 56.3%.

Van een middelgrote progressie is sprake als leerlingen inhoudelijke hulp nodig hadden in experiment 1, maar in experiment 2 de opdracht zelfstandig oplosten. Bij experiment 1 was nog geen sprake van inhoudelijke hulp bij variabelen instellen. Verder boekte 25.0% deze middelgrote progressie bij de deelopdracht herhaald meten en 75% bij de deelopdracht conclusies trekken.

We spreken van de kleinste progressie als leerlingen lees-/proceshulp hulp nodig hadden bij experiment 1, maar in experiment 2 de

Tabel 4

Fitmaten voor LCA met één tot vijf klassen

K	#pars	df	Fitmaten		
			LL	AIC	BIC
1	18	183	-1090.484	2216.967	2276.427
2	39	162	-1028.394	2134.787	2263.616
3	60	141	-990.7418	2101.484	2299.682
4	81	120	-967.1011	2096.202	2363.770
5	102	99	-965.0599	2134.120	2471.057

Noot. #pars is het aantal vrij te schatten parameters in het hele model; df zijn de degrees of freedom; LL is de log-likelihood. Deze grootheden liggen ten grondslag aan de fitmaten AIC en BIC en andere fitmaten en worden derhalve ter informatie gerapporteerd.

opdracht zelfstandig oplossen. 57.1% van deze groep boekte deze kleinste progressie voor de deelopdracht variabelen instellen, 32.2% bij de deelopdracht herhaald meten en 81.3% bij de deelopdracht conclusies trekken.

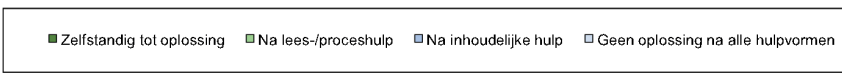
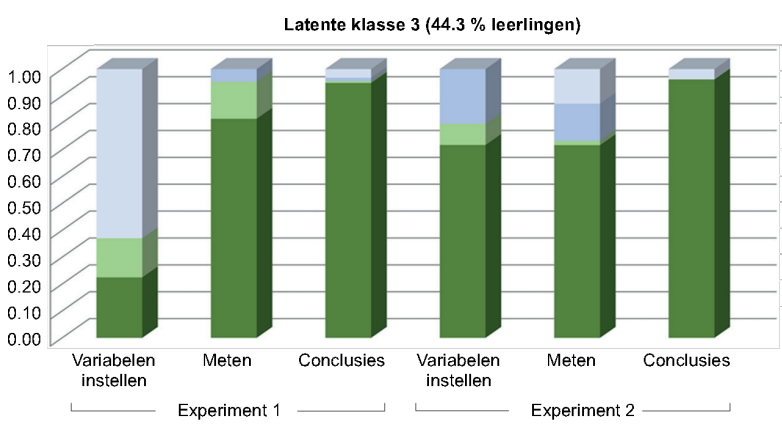
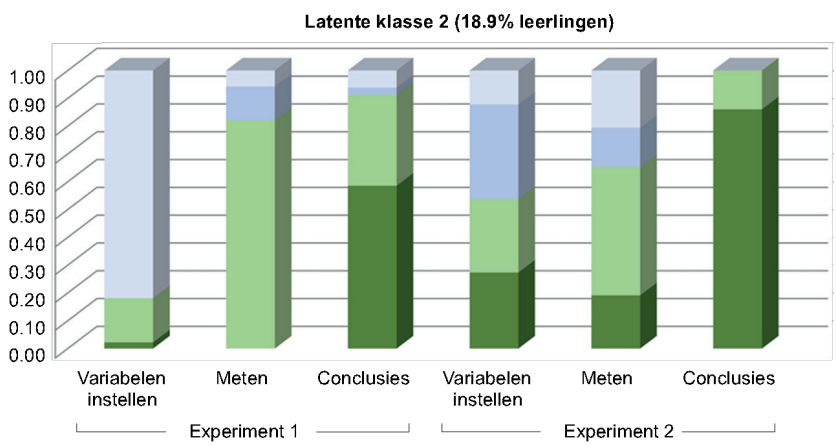
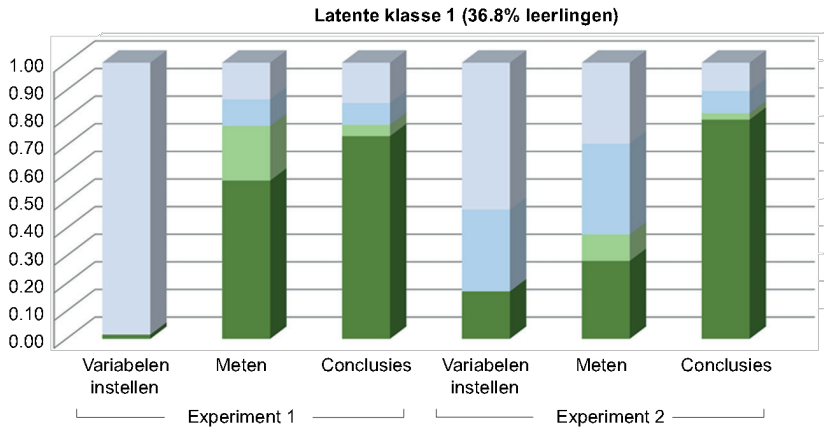
5.2 Latente Klasse Analyses: verschillen in hulpgeschiedenis

Tabel 4 geeft de fitmaten voor de modellen met één tot en met vijf latente klassen, die verwijzen naar groepen leerlingen met een verschillende hulpgeschiedenis in de loop van de twee experimenten met de Knikkerbaan. Het model met de laagste waarde op het criterium is het model met een optimale balans tussen modelfit en complexiteit (zie Tabel 4).

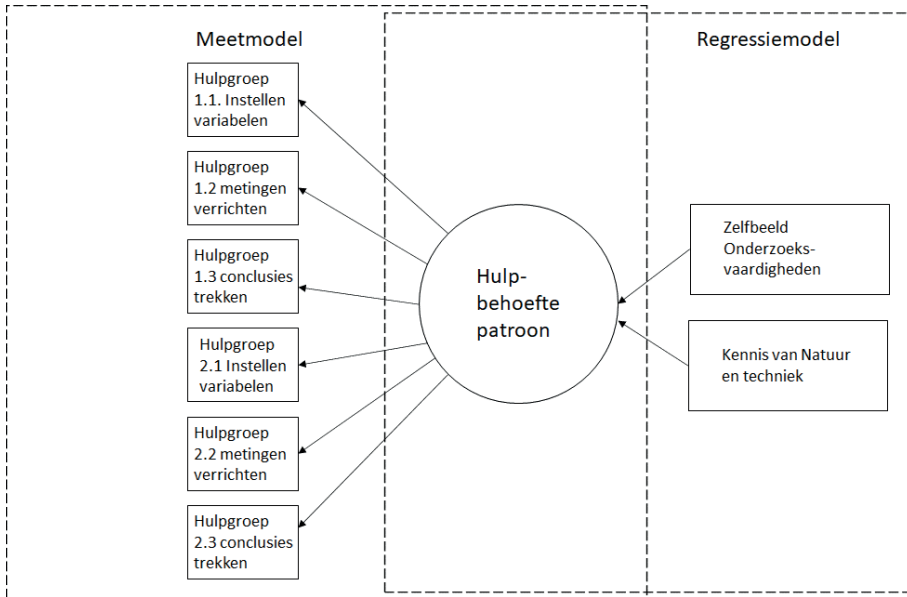
De resultaten voor de AIC en BIC geven geen eenduidig beeld. De AIC suggereert een vier-klasse model, de BIC een twee-klasse model. De discrepantie tussen de AIC en BIC is toe te schrijven aan het verschil in de mate waarin de criteria een penalty leggen op de complexiteit van het model. Omdat de BIC de complexiteit zwaarder meeweegt dan de AIC, zal de BIC doorgaans resulteren in een model met minder klassen dan wanneer de AIC wordt gebruikt voor de modelkeuze (e.g., Van den Bergh, 2017, p. 14). Omdat beide criteria voor- en nadelen hebben, zijn beide criteria gewogen bekeken en is bij de keuze om een drie-klasse-model te gebruiken rekening gehouden met de interpreteerbaarheid ervan.

Figuur 2 laat de latente klasse-profielen zien en de bijbehorende klasse-groottes. De gevonden waarden zijn de conditionele proporties, die per klasse aangeven hoeveel leer-

lingen naar verwachting een bepaalde opdracht met een bepaalde hulpsoort kan voltooien. De klasse-specifieke proporties zullen we aanduiden met $P_{c v_t}$ (hulpsoort), waarbij index c verwijst naar de klasse, v_t naar de vaardigheid (B: baan instellen, M: meten; C: concluderen) op experiment t ($= 1$ of 2). Bijvoorbeeld, binnen klasse 1 zal naar verwachting 2% van de leerlingen zelfstandig de baan in kunnen stellen. We geven dit aan als P_{1B1} (zelfstandig) = .02. Op basis van de conditionele proporties kunnen we de klassen inhoudelijk duiden. Klasse 1 bestaat uit leerlingen (36.8% van de leerlingen) die zowel bij experiment 1 als experiment 2 niet zonder hulp in staat zijn geweest om de knikkerbaan op de juiste manier in te stellen. Een deel van hen slaagt erin in experiment 2 met hulp een oplossing te vinden, P_{1B2} (inhoudelijke) = .30. De moeite met instellen is een uiting van het nog niet onder de knie hebben van de 'control of variables'-vaardigheid (Chen & Klahr, 1999). De leerlingen zijn al wel bij het eerste experiment in staat om de juiste conclusies te trekken, P_{1c1} (zelfstandig) = .73, naar aanleiding van gedane metingen en eventueel gegeven hulp of modellering. Conclusies trekken wordt in experiment 2 nog iets beter gedaan, P_{1c2} (zelfstandig) = .79. Leerlingen in deze klasse slagen er minder goed in een voldoende aantal metingen te verrichten in het tweede experiment, P_{1m2} (zelfstandig) = .28, dan in het eerste experiment, P_{1m1} (zelfstandig) = .57. Waarschijnlijk hangt de verslechterde prestatie samen met het beperken van de grafische ondersteuning op het werkblad in experiment 2.



Figuur 2: Karakterisering van de drie gevonden groepen leerlingen naar hulpstatus op de twee experimenten Knikkerbaan



Figuur 3:
Gebruikt latente klasse-model met covariaten

Leerlingen in klasse 2 (18.9% van de leerlingen) blijken in experiment 1 nog niet in staat om de Knikkerbaan correct in te stellen, P_{2B1} (zelfstandig) = .02. Dat blijkt in experiment 2 in hoofdzaak ook nog niet het geval P_{2B2} (zelfstandig) = .27, wat erop wijst dat ook leerlingen in deze groep de control-of-variables vaardigheid nog niet onder de knie heeft, na het eerdere experiment en de daarbij gegeven hulp.

Opvallend in klasse 2 is de rol die lees- en proceshulp speelt bij het vinden van een correcte oplossing op de deeltaken. Dit zien we vooral bij correct meten. De kans op het verrichten van correcte metingen na lees-/proceshulp bij deze deeltaak bedraagt in experiment 1 P_{2M1} (leeshulp) = .82. Ook bij het trekken van conclusies zien we dit aandeel van lees/proceshulp terug bij beide achtereenvolgende experimenten P_{2c1} (leeshulp) = .33 en P_{2c2} (leeshulp) = .14. We zien wel dat in deze groep het trekken van conclusies vaker zelfstandig verloopt in experiment 2 dan in experiment 1; P_{2c2} (zelfstandig) = .86 versus P_{2c1} (zelfstandig) = .59. We zien dat er weliswaar minder leerlingen lees- en proceshulp nodig hebben bij de deeltaak meten in Experiment 2 dan Experiment 1; P_{2M2} (leeshulp) = .46; P_{2M1} (leeshulp) = .82

We zien echter ook dat meer leerlingen niet tot de juiste oplossing komen in experiment 2; P_{2M2} (geen oplossing) = .21 versus P_{2M1} (geen oplossing) = .06.

Klasse 3 bestaat uit leerlingen (44.3% van de leerlingen) die experiment 1 nog niet goed in staat waren om de knikkerbaan zelfstandig juist in te stellen, P_{3B1} (zelfstandig) = .23, maar vervolgens wel bij experiment 2, P_{3B2} (zelfstandig) = .72, en inclusief lees/proceshulp nog wat vaker, P_{3B2} (zelfstandig of met leeshulp) = .80. Deze leerlingen hebben in tegenstelling tot klasse 1 leerlingen wel geprofiteerd van de hulp bij het inzetten van de control-of-variables vaardigheid. Voor het overige blijken leerlingen in deze groep al bij experiment 1 in staat om een voldoende aantal metingen te verrichten, P_{3M1} (zelfstandig) = .82, zeker na inbegrip van lees-proceshulp, P_{3M1} (zelfstandig of met leeshulp) = .96 en correcte conclusies te verwoorden op basis hiervan, P_{3c1} (zelfstandig) = .95. Wel is ook hier sprake van een lichte terugval in het aandeel dat zelfstandig correct meet in het tweede experiment, P_{3M2} (zelfstandig) = .72, ten opzichte van het eerste experiment P_{3M1} (zelfstandig) = .82. Behalve het grotere pro-

fijt van hulp in het eerste experiment had deze groep leerlingen ook een vaardigere startpositie na afloop van de voorbereidende Knikkerbaanopdrachten, voorafgaand aan de twee volledige experimenten.

Voorspellers voor hulpbehoeften

Bij het boven gerapporteerde latente klasse model zijn ook covariaten inbegrepen die gerelateerd worden aan klasse-lidmaatschap. Er is dan sprake van een structureel-vergelijkingmodel uit twee delen: (i) een meetmodel met een kwalitatieve onderliggende latente variabele en (ii) een multinomiale logistisch regressiemodel. Figuur 3 is een grafische weergave van dit model.

Met het uitgebreide model hebben we onderzocht of we klasse-lidmaatschap – en dus verschillen in hulpbehoeften – konden voorspellen aan de hand van de volgende leerlingkenmerken: kennis van N&T, zelfbeeld van de eigen onderzoeksvaardigheden, de mate van activiteiten die leerlingen ondernemen in hun vrije tijd op het gebied van NT, en de mate waarin Natuur en Techniek een rol spelen in de toekomstplannen. Om te voorkomen dat het aantal prediktoren in het model niet meer in verhouding staat tot de steekproefgrootte, hebben we met behulp van een aantal hiërarchische regressieanalyses de belangrijkste prediktoren geïdentificeerd en die hebben we meegenomen in de finale analyses.

De resultaten staan weergegeven in Tabel 5, waarbij klasse 1 de referentiegroep is.

Tabel 5 geeft de geschatte regressie-coëfficiënten waarmee klasse-lidmaatschap voorspeld kan worden. Als we toetsen op 5% significant-niveau, dan zien we voor zelfbeeld van de eigen onderzoeksvaardigheden een significant effect; naarmate een leerling een positievere zelfinschatting geeft op de eigen

onderzoeksvaardigheden, heeft hij of zij meer kans om tot klasse 2 of 3 te behoren, dan tot klasse 1 (onder constant houding van de overige condities). Er zijn ook significante effecten voor kennis van Natuur en Techniek gevonden. Naarmate leerlingen een hogere score hebben op de toets voor Natuur en Techniek, zullen ze eerder in klasse 3 belanden dan in klasse 1. Voorts geldt ook dat naarmate leerlingen hoger scoren op deze toets ze eerder in klasse 2 dan in klasse 1 belanden. Doordat voor beide kenmerken de regressiecoëfficiënten die het effect ten opzichte van klasse 1 weergeven dicht bij elkaar liggen (bijvoorbeeld 4.51 en 5.30 voor de Kennistoets N&T), waarbij de verschillen kleiner zijn dan de standaardfouten (2.36 en 2.21 voor de Kennistoets), kan worden geconcludeerd dat de leerlingkenmerken niet onderscheidend zijn in het voorspellen of een leerling tot klasse 2 of klasse 3 behoort. De andere prediktoren hadden geen significante bijdrage in de voorspelling van klasse-lidmaatschap.

6 Discussie

6.1 Conclusies

Het doel van deze studie was te verkennen in hoeverre de inzet van dynamisch toetsen informatieve meerwaarde heeft ten opzichte van statisch toetsen bij het vaststellen van onderzoeksvaardigheden bij Natuur en Techniek. Daartoe zijn de data gebruikt van de peilingsstudie van 2015 met betrekking tot de praktische opdracht Knikkerbaan, welke was opgezet volgens de principes van dynamische toetsing. Tijdens een eerste experiment kregen leerlingen na een eerste meting waarbij niet geholpen was waar nodig ondersteuning bij de uitvoering van deelopdrachten. Daarna

Tabel 5
Multinomiale Logistische Regressie

Predictor	Klasse 2 t.o.v. klasse 1			Klasse 3 t.o.v. klasse 1		
	β	s.e.	p	β	s.e.	p
	-10.30	3.16		-12.84	3.25	
Kennistest N&T	4.51	2.36	.058	5.30	2.21	.018
Zelfbeeld onderzoeksvaardigheden	3.15	1.05	.003	4.15	1.06	<.001

kregen ze in een vervolgoexperiment de kans om dezelfde deelopdrachten opnieuw toe te passen, en indien opnieuw nodig ondersteund. Daarbij stonden drie onderzoeksvaardigheden centraal: het manipuleren van variabelen (de 'control of variables strategie'), het verrichten van meerdere metingen bij de verschillende instellingen van de onafhankelijke variabele, en het trekken van conclusies over het effect van een onafhankelijke variabele, op grond van de verrichte metingen.

Analyses volgens het sandwichmodel (Sternberg & Grigorenko, 2002), gericht op het vaststellen van de gemaakte progressie in experiment 2 na ondersteunde deelname aan experiment 1 gaf het beeld van groepen met een verschillend leerpotentieel te zien voor de verschillende deelvaardigheden. We zagen dat van de leerlingen die in experiment 1 ook na alle hulp niet tot een oplossing kwamen een aanzienlijk aandeel daarin wel zelfstandig slaagde in experiment 2 en daarmee een grote progressie, resp. leerpotentieel toonde. De percentages leerlingen met deze progressie verschilden per deelopdracht: 36% bij variabelen instellen, 42% bij herhaald meten en 56% bij conclusies trekken. De percentages van de in experiment 1 minder intensief geholpen leerlingen, die de deelopdrachten in het tweede experiment zelfstandig uitvoerden, en daarmee een minder sterke progressie maakten, volgden voor de deelopdrachten een analogo patroon.

Door middel van latente klasse analyse werd een beeld verkregen van patronen in responsiviteit voor hulpinterventies bij de uitvoering van verschillende deelopdrachten. Deze analyses passen bij het gelaagde-cake-model van dynamisch toetsen. De resultaten wijzen op drie groepen met een verschillende responsiviteit voor hulp. Een eerste klasse van leerlingen heeft weinig profijt van de in Experiment 1 gegeven hulp als het gaat om het instellen van de variabelen in Experiment 2, en tonen daarmee weinig leerpotentieel in het toepassen van de CVS.

Het tegenovergestelde zien we bij de leerlingen in klasse 3. Behalve een kleine groep leerlingen die geheel zelfstandig beide experimenten voltooit, bevat deze klasse een grote groep leerlingen die duidelijk profiteert van

de hulp in experiment 1. Met deze hulp weten deze leerlingen zich de CVS eigen te maken.

Daarnaast onderscheidt zich klasse 2, met relatief veel leerlingen die profiteren van lees- en proceshulp. Een deel van deze leerlingen slaagt er na de verkregen hulp in experiment 1 en leeshulp in experiment 2 de variabelen correct in te stellen. Ook helpt leeshulp hen passende conclusies te trekken naar aanleiding van de metingen.

Het na hulp zelfstandig kunnen toepassen van de CVS is in lijn met de bevindingen van Wagenveld et. al. (2014). Zij constateerden dat na instructie de meeste leerlingen uit groep 8 in staat waren om deze vaardigheid toe te passen. Nu is de bij Experiment 1 gegeven hulp niet direct vergelijkbaar met de door Wagenveld et. al. gegeven goede instructie. Toch had de gegeven hulp bij een deel van de leerlingen een groot effect op de zelfstandig correcte toepassing CVS in experiment 2. De dynamische methodiek maakte dit leerpotentieel zichtbaar. Tegelijkertijd lijkt er ook een groep die niet of nauwelijks profiteert van de hulp, in de literatuur aangeduid als 'non-gainers' (Grigorenko, 2009).

Vanuit een statische invalshoek kan geconcludeerd worden dat slechts weinig leerlingen uit groep 8 de CVS beheersen. Vanuit een dynamische invalshoek kunnen we concluderen dat een aanzienlijk percentage van de leerlingen dit zou kunnen, als ze daartoe passend onderwijs krijgen.

De interpretatie van de bevindingen ten aanzien van (herhaald) meten waren problematisch omdat de opdrachten niet geheel parallel waren. Dit bemoeilijkte vooral de interpretatie vanuit het sandwichmodel van dynamisch toetsen. In experiment 1 werden de leerlingen aangezet tot herhaald meten met een in te vullen tabel waarin de nummers van de metingen 1 t/m 5 expliciet vermeld werden. Bij experiment 2 ontbrak in deze tabel de nummering. Het effect was dat nog maar 57% van de leerlingen die bij experiment 1 de vijf metingen zelfstandig uitvoerden dit ook deed bij experiment 2. Van de leerlingen die op dit punt hulp hadden gehad bij experiment 1, voerde 33% zelfstandig vijf metingen uit bij experiment 2. Hier hebben de veranderde, minder sturende, taakcondi-

ties het effect dat leerlingen niet meer geneigd zijn om herhaald te meten. Mogelijk hangt dit samen met de manier waarop leerlingen afwijkingen in metingen ervaren. In plaats van 'natuurlijke variatie' ervaren zij afwijkingen als een fout die door hen wordt veroorzaakt (Goedhart & Verdonk, 1991). Als leerlingen naar hun idee een meting correct hebben uitgevoerd zijn ze daardoor niet geneigd die te herhalen.

Dat veel leerlingen in groep 8 in staat zijn om de correcte conclusie te trekken uit een experiment met een duidelijk verschillende uitkomst voor de waarde van een specifieke variabele bevestigt de bevindingen van Peikny en Mahler (2013). In hun onderzoek was 89.6% van 13-jarige leerlingen daartoe in staat.

6.2 Perspectieven en aanbevelingen

6.2.1 Validiteit en informatieve meerwaarde

Terugkijkend op de resultaten toont deze studie informatieve meerwaarde van dynamisch toetsen aan boven statisch toetsen. Allereerst geeft de methode van dynamisch toetsen inzicht in waartoe leerlingen in de toekomst zelfstandig in staat zijn, als ze daaraan voorafgaand passende hulp ontvangen bij taakuitvoering. We zagen groepen leerlingen met een verschillend leerpotentieel en ook een verschillende gevoeligheid voor hulpinterventies. Bij statisch toetsen kan slechts gekeken worden naar wat de leerling tot nu toe heeft geleerd.

Verder kunnen door onderscheid te maken tussen toegangsvaardigheden en doelvaardigheden oorzaken van goed en minder goed presteren beter worden onderscheiden. Zo profiteerde een deel van de leerlingen van lees- en proceshulp waardoor ze beter toegang tot de opgave verkregen (Beddow et al, 2011). Bij een statische methodiek zou bij deze leerlingen het niet correct oplossen van de opdrachten onterecht zijn toegeschreven aan beperkt ontwikkelde N&T vaardigheden.

Het onderzoek geeft bemoedigende resultaten als het gaat om de informatieve meerwaarde van dynamisch toetsen. Tegelijkende dit onderzoek ook een aantal beperkingen die de validiteit van uitspraken mogelijk

kunnen bedreigen en nader onderzoek vragen.

Ten eerste was de geschiktheid van de gebruikte deeltaken voor het vaststellen van leerpotentieel nog niet optimaal. Dynamische toetsen dienen voldoende uitdagende en voorts volledig parallelle toetstaken te bevatten om leerpotentieel betrouwbaar en valide vast te kunnen vaststellen. Dit bleek in verschillende opzichten niet zo te zijn: Bij Experiment 1 ontbrak inhoudelijke hulp bij het instellen van de variabelen, bij Experiment 2 ontbrak de nummering in de invultabel voor de metingen, en de deeltaak 'Conclusies trekken' was te gemakkelijk voor de doelgroep. Het verdient dan ook aanbeveling in een volgend peilingsonderzoek zorg te dragen voor taken met een consequente structuur en een moeilijkheidsgraad die ruimte biedt voor prestatieverbetering.

6.2.2 Lees- en proceshulp

Een specifieke bijdrage van deze studie betreft de effecten van lees- en proceshulp. In het peilingsonderzoek is leeshulp toegevoegd op grond van de inzichten uit de toegankelijkheidstheorie van Beddow et al (2011). Zij geven aan dat leerlingen door hiaten in toegangsvaardigheden soms niet toekomen aan de taak waar het eigenlijk om draait in de toets. Deze vaardigheden, zoals het begrijpen van teksten, figuren, en fysieke opstellingen, zijn wel nodig voor de uit te voeren taak, maar vormen op zichzelf geen onderdeel van de te meten doelvaardigheid.

Naast leeshulp werd regulatieve hulp of proceshulp verstrekt, die we ook gerekend hebben tot toegangshulp, zoals de leerling motiveren voor een opdracht, plannen, monitoren en bijsturen. We namen aan dat door deze hulpvormen de opdrachten toegankelijker werden, en dat eventuele onnodige taaklast (of: extrinsieke taaklast; Sweller, 2010) zoveel mogelijk zou worden weggenomen, zonder dat de intrinsieke taaklast kleiner zou worden.

Resultaten van de latente klassenanalyse wezen uit dat één kleinere groep (18.9% van de leerlingen) erin slaagde dankzij lees- en proceshulp oplossingen te vinden bij het verichten van metingen en trekken van conclu-

sies. Interessant hierbij is dat deze hulpvorm een minder grote rol speelde bij de groepen met het laagste en hoogste leerpotentieel. In vervolgonderzoek zou de gevoeligheid voor of behoefte aan verschillende hulpsoorten kunnen worden bekeken in relatie tot leerpotentieel. Hierbij is een opzet wenselijk waarbij naast leeshulp en inhoudelijke hulp, de regulatieve hulp meer verfijnd wordt verstrekt en geregistreerd. Daarbij kan aangesloten worden bij beschikbare kaders voor self-regulated learning (Panadero, 2017). Een beloftevol kader is dat van Winne en Hadwin (1998), waarin analoog aan de 'graduated-prompt' benadering, gradaties van feedback worden onderscheiden. Daarbij kan de regulatie van een leertaak geleidelijk worden overgenomen door een begeleider, als blijkt dat leerlingen er niet in slagen om door middel van eigen 'interne feedback' hun werk adequaat te monitoren en bij te sturen. Afhankelijk van haperingen in de monitoring kan hulp variëren van aanvullen, verfijnen, herstructureren, tot compleet 'overschrijven' van de aanpak van de leerling. Onderzoek naar de gevoeligheid voor deze vormen van regulatieve hulp kan meer licht werpen op verschillen in leerpotentieel tussen leerlingen.

6.2.3 *Leerlingkenmerken en leerpotentieel*

Deze studie heeft kenmerken van leerlingen opgeleverd die samenhangen met leerpotentieel en de daarbij geconstateerde verschillen in gevoeligheid voor hulpinterventies. Leerpotentieel lijkt positief samen te hangen met kennis van Natuur en techniek. Verder bleek dat het zelfbeeld van de eigen onderzoeksvaardigheden positief samenhangt met het leerpotentieel, vooral met betrekking tot de deelvaardigheid van het manipuleren van onafhankelijke variabelen in een experimentopzet teneinde een onderzoeksvraag te kunnen operationaliseren. Gezien het moment van vaststellen is het leggen van een direct oorzakelijk verband tussen zelfbeeld en leerpotentieel niet mogelijk. Vragen over het zelfbeeld werden namelijk gesteld nadat de opdracht was afgesloten. Het is aan te bevelen om dat bij een vervolgonderzoek vooraf te doen.

6.2.4 *Kwaliteit en registratie van hulpinterventies*

Tot slot van dit artikel plaatsten we nog twee kanttekeningen die te maken hebben met de aard en registratie van hulpinterventies. Ten eerste, in de peilingsstudie werden de verschillende hulpinterventies om praktische redenen niet afzonderlijk geregistreerd. Bovendien werd het exacte moment van leeshulp en proceshulp, alsmede de aard ervan niet vastgelegd. Dat bemoeilijkt de interpretatie van de relatie tussen hulpvormen en prestatieverbetering, en daarmee een preciezere duiding van het leerpotentieel. Om een meer fijnmazig beeld te krijgen van gevoeligheid voor uiteenlopende typen hulp, verdient het aanbeveling om in vervolgonderzoek alle hulpinterventies en het moment ervan te registreren, zoals ook gedaan is in studies van Swanson (2011) en Fuchs et al. (2011).

Een tweede kanttekening betreft de rol van de toetsleider als hulpverstrekker. Omdat na de uitvoering van de peiling bleek dat de toetsleiders niet waren geregistreerd op de scoreformulieren konden we in onze analyses geen rekening houden met verschillen tussen toetsleiders. Er zijn evenmin gegevens verzameld over de werkwijzen en mogelijke knelpunten bij het signaleren van hulpbehoeften, het bieden van hulp en het registreren van gegeven hulp. Het verdient dan ook aanbeveling om hulpgedrag door toetsleiders systematischer en meer fijnmazig te onderzoeken, waarbij aandacht is voor de wijze van signaleren van hulpbehoeften, de realisatie van beoogde hulpinterventie, en de mate van sturing richting de oplossing. Voor het valide bepalen van het leerpotentieel is het van belang een goed beeld te hebben van de verleende steun aan de leerling en tot welke leerwinst dit leidde bij de leerling. Daarmee komen we meer te weten over de aard van en omstandigheden waaronder leerpotentieel worden opgeroepen.

Literatuur

- Andrews, R. L., & Currim, I. S. (2003). A comparison of segment retention criteria for finite mixture logit models. *Journal of Marketing Research*, 40, 235–243.
- Beddow, P. A., Kurz, A., & Frey, J. R. (2011). Accessibility theory: guiding the science and practice of test item design with the test-taker in mind. In: S.N. Elliott, R. J. Kettler, P. A. Beddow, & A. Kurz (Eds.), *Handbook of accessible achievement tests for all students: Bridging the gaps between research, practice, and policy* (pp. 163–182). New York: Springer.
- Bergh, M. van den (2017). *Latent class trees*. Dissertation. Tilburg: Tilburg University.
- Caffrey, E., Fuchs, D., & Fuchs, L.S. (2008). The Predictive Validity of Dynamic Assessment *The Journal of Special Education*, 41, 254–270.
- Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, 70, 1098–1120.
- Cito (2016). *Natuur en techniek, technisch rapport over resultaten peil. onderwijs in 2015* Retrieved from <https://www.onderwijsinspectie.nl/onderwerpen/peil-onderwijs/documenten/rapporten/2017/05/31/peil-natuur-en-techniek-technisch-rapport-cito>.
- De Vries, M. J., Van Keulen, H., Peters, S., & Walma van der Molen, J (Eds). (2011). *Professional development for primary teachers in science and technology*. Rotterdam: Sense Publishers
- Denessen, E., Vos, N., Damen, T., Koch, S., Monika, L., & Wigboldus, D. (2011). Explicit and implicit measures of teacher attitudes towards science and technology. In: M.J. de Vries, H. van Keulen, S. Peters, & Walma van der Molen. J. (Eds.), *Professional development for primary teachers in science and technology* (pp. 107–119). Leiden: Brill/Sense.
- Ferrara, R. A., Brown, A. L., & Campione, J. C. (1986). Children's learning and transfer of inductive reasoning rules: Studies of proximal development. *Child Development*, 57, 1087–1099. <https://doi.org/10.2307/1130433>.
- Fischer, K. W., & Bidell, T. R. (2006). Dynamic development of action and thought. In W. Damon & R. M. Lerner & (Eds.), *Handbook of child psychology: Theoretical models of human development* (pp. 313–399). New York: Wiley.
- Fuchs, L. S., Compton, D. L., Fuchs, D., Hollenbeck, K. N., Craddock, C. F., & Hamlett, C. L. (2008). Dynamic assessment of algebraic learning in predicting third graders' development of mathematical problem solving. *Journal of Educational Psychology*, 100, 829–850.
- Fuchs, L. S., Compton, D. L., Fuchs, D., Hollenbeck, K. N., Hamlett, C. L., & Seethaler, P. M. (2011). Two-stage screening for math problem-solving difficulty using dynamic assessment of algebraic learning. *Journal of Learning Disabilities*, 44, 372–380.
- Goedhart, M. J., & Verdonk, A. H. (1991). The development of statistical concepts in a design-oriented laboratory course in scientific measuring. *Journal of Chemical Education*, 68, 1005.
- Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 27, 857–871.
- Grigorenko, E. L. (2009). Dynamic assessment and response to intervention: Two sides of one coin. *Journal of Learning Disabilities*, 42, 111–132.
- Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences*. Thousand Oaks, CA: Sage Publications Inc.
- Hillen, R. (2017). *Modeling psychological attributes: Merits and drawbacks of taxometrics and latent variable mixture models*. Dissertation. Tilburg: Tilburg University. Online beschikbaar via https://pure.uvt.nl/ws/portalfiles/portal/17484431/Hillen_Modeling_23_06_2017.pdf.
- Keulen, H. van, & Slot, E. (2013). *Excellentiebevordering door middel van onderzoekend en ontwerpnd leren: Vaardigheden Rubrics Onderzoeken en Ontwerpen (VROO)*. Den Haag: School aan Zet.
- 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.
- Langen, A. van, & Vierke, H. (2009). *Wat bepaalt de keuze van een natuurprofiel? De invloed van de leerling, de school, de ouders en de peergroup*. Den Haag: Platform Bèta Techniek.
- Lidz, C. S., & Elliott, J. (2000). *Dynamic assessment: Prevailing models and applications* New York, NY: Jai.
- Lidz, C. S., & Gindis, B. (2003). Dynamic assess-

- ment of the evolving cognitive. In A. Kozulin, B. Gindis, V. Ageyev, & S. Miller (Eds.), *Vygotsky's Educational Theory in Cultural Context* (99-116). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511840975.
- Linzer, D. A., & Lewis, J. B. (2011). poLCA: An R-package for polytomous variable latent class analysis. *Journal of Statistical Software*, 42(i10). DOI:10.18637/jss.v042.i10
- Magidson, J. & Vermunt, J. K. (2004). Latent class models. In D. Kaplan (Ed.), *The sage handbook of quantitative methodology for the social sciences* (pp. 175-189). Thousand Oaks, CA: Sage.
- National Assessment Governing Board. (2013). *Technology and engineering framework for the 2014 NAEP*. Retrieved from <http://www.nagb.gov>
- OECD. (2014). *PISA 2012 results: Creative problem solving: Students' skills in tackling real-life problems (volume V)*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264208070-en>
- Paas, L. J., Vermunt, J. K., Bijmolt, T. H. A. (2006). Multi-niveau latente klasse analyse: Met een toepassing bij het simultaan clusteren van landen en consumenten. In L. J. Paas et al. (Red.), *Ontwikkelingen in marktonderzoek* (pp. 161-173), Haarlem: Spaar en Hout.
- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 8, Article 422.
- 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.
- Pólya, G. (1945; 2nd edition, 1957). *How to solve it*. Princeton: Princeton University Press.
- Roelofs, E. C. (2017). *Effecten van hulp bij het toetsen van ontwerp- en onderzoeksvaardigheden. Ontwerp, gebruik en resultaten van dynamische toetsen in de context van een peilingsonderzoek Natuur en Techniek*. Arnhem: Stichting Cito. <https://www.onderwijsinspectie.nl/onderwerpen/peil-onderwijs/documenten/rapporten/2017/05/31/effecten-hulp-toetsen-ontwerp-onderzoeksvaardigheden-cito>.
- Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics. In D. Grouws (Ed.), *Handbook for research on mathematics teaching and learning* (pp. 334-370). New York: MacMillan.
- Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic Testing*. New York: Cambridge University Press.
- Swanson, H. L. (2011). Dynamic testing, working memory, and reading comprehension growth in children with reading disabilities. *Journal of Learning Disabilities*, 44, 358-371.
- Swanson, H. L., & Howard, C. B. (2005). Children with reading disabilities: Does dynamic assessment help in the classification? *Learning Disability Quarterly*, 28, 17-34.
- Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, 22, 123-138.
- Van der Graaf, J., Segers, E., & Verhoeven, L. (2015). Scientific reasoning abilities in kindergarten: Dynamic assessment of the control of variables strategy. *Instructional Science*, 43, 381-400.
- Van der Steen, S., Steenbeek, H., & Paul Van Geert, P. (2012). Using the Dynamics of a Person-Context System to Describe Children's Understanding of Air Pressure. In H. Kloos, B. J. Morris, & J. L. Amaral (Eds.), *Current Topics in Children's Learning and Cognition* (pp. 21-44). New York: InTech. <http://www.intechopen.com/books/current-topics-in-childrens-learning-and-cognition/using-the-dynamics-of-a-person-context-system-to-describe-childrens-understanding-of-airpressure>.
- Van Geert, P. (1998). A dynamic systems model of basic developmental mechanisms: Piaget, Vygotsky, and beyond. *Psychological Review*, 105, 634-677.
- Vygotsky, L. (1978). Interaction between learning and development. *Readings on the Development of Children*, 23, 34-41.
- Wagensveld, B. van, Segers, E., Kleemans, T., & Verhoeven, L. (2014). Child predictors of learning to control variables via instruction or self-discovery. *Instructional Science*, 1-15. <http://dx.doi.org/10.1007/2Fs11251-014-9334-5>.
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated engagement in learning. In D. Hacker, J. Dunlosky, and A. Graesser, *Metacognition in Educational Theory and Practice* (pp. 277-304). Hillsdale, NJ: Erlbaum.

Noot

1. Bij de opzet en uitvoering zijn veel deskundigen betrokken geweest, aan wie we dank verschuldigd zijn. Zonder anderen tekort te doen, willen we noemen: Marleen van der Lubbe, Henk Notté, Tecla Lampe, Bas Hemker, Gunther Maris, opleiders en studenten van de lerarenopleiding van de Hogeschool Arnhem-Nijmegen.

Auteurs

Erik Roelofs is werkzaam als universitair hoofddocent bij de vakgroep Cognition, Data, & Education, bij de Faculteit Behavioural, Management and Social Sciences van de Universiteit Twente. **Dannie Wammes** is werkzaam als docent aan de PABO van de Hogeschool van Arnhem en Nijmegen. **Wilco Emons** is werkzaam als senior onderzoeker bij Cito, afdeling OKI te Arnhem. **Maartje Raijmakers** is werkzaam als hoogleraar bij de Vrije Universiteit Amsterdam, Faculteit der Maatschappij- en Gedragwetenschappen.

Correspondentieadres: W. Emons. Cito. Amsterdamseweg 13, postbus 1034, 6801MG Arnhem; E-mail: wilco.Emons@cito.nl

following a protocol of graduated prompts.

Analyses aimed at determining the progress made in experiment 2 after supported participation in experiment 1 revealed groups with different learning potential. The largest progression observed was made by students who did not reach a solution to a subtask in experiment 1 even after all support, but succeeded independently in doing so in experiment 2. The relative proportion of non-solvers in experiment 1 who made this progression was 36% setting the experimental variables 42% in repeated measurement and 56% drawing conclusions.

Latent class analyses resulted in three groups, each of which went through a different history of support. The biggest difference concerned responsiveness to support with a crucial sub-skill for experimentation: control-of-variables. Furthermore, in one group, giving reading/process support appeared to play an important part in finding solutions. The group of students who benefited the least from help appeared to have less knowledge of N&T and a more negative view of their own research skills.

Keywords: Dynamic assessment, science and technology, accessibility, regulative support, graduated prompts, responsiveness to intervention, inquiry-based learning

Abstract

Dynamic testing in the Dutch National Nature and Technology Assessment among Grade 6 Elementary School Students

In a secondary data analysis based on data from a Dutch national science assessment in the fall of 2015, the added value of dynamic testing for determining research skills in the domain of Nature & Technology of grade 6 students in elementary school was examined compared to static testing. To this end, the research data (n=203) were used from one practical assignment, designed according to the principles of dynamic testing and involved the design and implementation of an experiment and a follow-up experiment with a ramp task, the Marble Run. Depending on progress on the subtasks, reading, process, and substantive help was provided,