

De ontwikkeling van de diagnostische tussentijdse toets (2012-2017)

Sanneke Schouwstra (redactie)

cito.nl



Auteurs: Arjan Aarnink; Paul Drijvers; Joke Hofstee; Laura van Hofwegen; Herbert Hoijtink; Kirsten van Ingen; Karen Keune; Patrick de Klein; Jesse Koops; Roelien Linthorst; Peter van Os; Daniel van der Palm; Sanneke Schouwstra; Uriël Schuurs; Irene van Stiphout; Wilma Vrijs; Ingrid Williams

Eindredactie: Sanneke Schouwstra

Screening: Sandrijn Vernooij

© Stichting Cito Instituut voor Toetsontwikkeling Arnhem (2017)

Niets uit dit werk mag zonder voorafgaande schriftelijke toestemming van Stichting Cito Instituut voor Toetsontwikkeling worden openbaar gemaakt en/of verveelvoudigd door middel van druk, fotokopie/reprografie, scanning, computersoftware of andere elektronische verveelvoudiging of openbaarmaking, microfilm, geluidskopie, film- of videokopie of op welke wijze dan ook.

Met medewerking van:

Toetsdeskundigen

Engels

Margreet van Aken; Jonna Bos; Rick Godschalk; Evelyn Reichard; Natalie Schols; Ingrid Williams, Wilma Vrijs

Nederlands

Ron van den Beemt; Tom Duindam; Hiske Feenstra; Karin Heesters; Laura van Hofwegen; Kirsten van Ingen; Roelien Linthorst; Uriël Schuurs

Wiskunde

Saskia van Boven; Nynke Brandsma; Madelon Groenheiden; Kees Lagerwaard; Ger Limpens; Melanie Steentjes; Irene van Stiphout; Paul Drijvers

Assistentie

Frederik de Boer; Esmeralda Craamer; Jacqueline Drijber; Monique Esvelt; Jorien van Haaften; Delphine Kerkmeester; Linda Ravelli; Merle Spruijt; Evelien Verhelst; Sandrijn Vernooij; Wietske Vriezen; Rieneke van der Wal

Wetenschappelijk onderzoekers

Servaas Frissen; Herbert Hoijtink; Karen Keune; Jesse Koops; Daniël van der Palm; Ivailo Partchev; Erik Roelofs; Sanneke Schouwstra

IT

Arjan Aarnink; Floris Briolas; Wilco Budding; Stefan Garcia; Martijn van Haren; Martijn Hendriksen; Marcel Hoekstra; Nard van Kessel; Patrick de Klein; Marcel Lagerwerf; Mark Martinot; Peter van Os

Management en projectleiding ontwikkeling DTT

Anne-Marie Anthonissen; Godelieve van den Boogaard; Joke Hofstee; Sanneke Schouwstra; Lody Smeets; Wilma Vrijs; Anke Weekers

Stuurgroep

Anton Beguin; Aukje Bergsma; Alice Groen; Bert Ijsveld; Ed Kremers; Paul van der Molen; Mark Molenaar; Cor Sluijter

Met dank aan:

Constructiegroepleden:

Engels

Arthur Gelink; Marieke Nijhof; Lotte Poll; Merel Thijink; Aline van Engelenburg; Nicolette Hardon; Hans Hillen; Jacqueline Pantophlet; Hafize Topcu; Maleene Wilson

Nederlands

Rinske Groenier; Irma Nelck; Miranda Paulen; Albert Pouwels; Renée van de Schans; Manon Tiehuis

Wiskunde

Ans van der Ark; Jeroen Duineveld; Sietske Greeuw; Mark Henkens; Marieke Hoving; Mascha Klerx; Jan-Otto Kranenborg; Eric Lemson; Frits Lemstra; Lianne Maerman; Martin Mollema; Gerrit Ros; Wendy Smit; Lieke Verschueren

Experts voor de standaardbepaling:

Engels

Martina Acharrat - Winter; Andrea Barthel; Iris Berends; Wilfred Berkhof; Jos Bohnen; Ioana Bona; Adrienne Buschmann; Bert Christiaans; Rukiye Dutar; Yaira Eekhof; Jiska Evers; Nagad Farah; Gienke Former; Aukje de Gier; Martien de Graaf; Monique Harleman; Helen Hol; Claudia Huiskes; Loes Janssen; Evelien van Kats; Sven Klippel; Jan Kortman; Joyce Marien; Hetty Mulder; Salina Nowak; Kitty van Ool; Jacqueline Pantophlet; Nancy Pinas; Iris Praster; Anke van den Reek; Mirvat Salman; Edwin van Schie; Tim Simons; Marjolijn Spijker; R. Tieben; Nyree Vermey; Kitty Weterings; Deidre van de Wijdeven; Marriet Wijma

Nederlands

L. Aarab; Mieke Balemans; Peter van Ballegooijen; Martin Bouma; Frederique van Breugel; Robert Chamalaun; Claudia Delfgou; Annelies van Diepen; Kelly Diepenbroek; Marije Doedens; Gjalt van Duinen; Ank Gakpo-van Bommel; Esther van Gorkum; Rien de Grauw; Marc Groen; Miranda Hansen; Waldo Happee; Miriam Heijs; Sonja Heynsdijk-Koch; Sarah Hock; Frans Hoyinck; Marian Kleverkamp; Maud Koenen; Mara Kuijpers; Miranda van Loon; Anne-Marie Maartens; Jaap Molenhuis; Sandra Molhoek; Sonja Mombarg; Arva Muradin; Nienke Nagelmaeker; Rachel Ooft; Bas van Pelt; Mariska Prevoo; Sophie Raaijmakers-Rietbroek; Larissa Rutten; Angela Schellekens; W.R. Scheppink; Janneke Sleenhof; Kim van Soest-Meijer; Jos van Son; Liesbeth Stolk; William Verkade; R. Verkerke; K.R. Verschuren; Lydia van Vliet; Inge de Vries; Elly Wilting

Wiskunde

Ajrir; Azougagh; Carla Beckschebe; Mohammed Beladel; Berghuis; A. Brinker; Karel Broekhuis; K. van Camfort; Chi Ki Cheung; Coen van Diepen; Martha van Eijk; Nienke Flier; Claudia van Geffen; Madelon Groenheiden; Annie van der Heijden; J.C. Heilig; E. Heitmeijer; Floortje Holten; Jongerius; Soraya Kamps-Soebhag; Edward van Kervel; Fayola Klaris; Gert de Kleuver; Michelle Kuijt; Ton Lecluse; H. van Loon; Marko Meijer; Mariska Mesken; Majelle Meulman; Saro Mottes; Elena Maria Fonseca Neves; Marion van den Nieuwenhuijzen; Marleen Peddemors-Greftenhuis; Wineke Ritzema; H.M.J. Rorink-Heerink; C. J. Smit; Wybe Stavenga; Roel Veerbeek; Jade Vlaming; Jurgen van der Wal; Annemarie Westerbeek; Mark Wiendels; Michael van Woerkom; Vincent Ykema; A. Zomerman; Jan-Willem Zuiderduijn

Inhoudsopgave

1	Achtergrond en proces van toetsontwikkeling van de DTT Sanneke Schouwstra	5
1.1	Achtergrond van de DTT	5
1.2	Uitgangspunten en kenmerken van de DTT	5
1.3	Het proces van de toetsontwikkeling	7
1.4	Realisatie van het toetsontwerp	10
1.5	Literatuur	11
2	Opgaveconstructie Wilma Vrijs en Sanneke Schouwstra	13
2.1	Het proces van opgaveconstructie	13
2.2	Opgavetypen binnen de DTT	17
2.3	Tot slot	22
2.4	Literatuur	22
3	De responsverwerking: coderen in plaats van scores Joke Hofstee	23
3.1	Begripsdefinitie: items en responsen	23
3.2	Groeperen van interacties in de auteursomgeving	24
3.3	Het leerlingmodel in de auteursomgeving	26
3.4	Koppeling van interacties aan het leerlingmodel in de auteursomgeving	26
4	DTT Nederlands Uriël Schuurs, Kirsten van Ingen, Roelien Linthorst en Laura van Hofwegen	31
4.1	Schrijfvaardigheid Nederlands	31
4.2	Leesvaardigheid Nederlands	42
4.3	Reflecterende paragraaf	46
4.4	Literatuur	47
5	DTT Engels Ingrid Williams en Wilma Vrijs	49
5.1	Schrijfvaardigheid Engels	49
5.2	Leesvaardigheid	53
5.3	Literatuur	57
6	Beoordelen van open schrijftaken Roelien Linthorst en Karen Keune	59
6.1	Inleiding	59
6.2	Geautomatiseerd beoordelen van schrijfvaardigheid	61
6.3	Onderzoek naar beoordelingsmodellen	61
6.4	Ontwikkeling van open schrijftaken en eerste verkenning samenhang	63
6.5	Conclusies en aanbevelingen	66
6.6	Literatuur	67
7	DTT wiskunde Irene van Stiphout	69
7.1	Leerlingmodel	69
7.2	Operationalisatie	71
7.3	Literatuur	74
8	Automatische beoordeling wiskunde Paul Drijvers, Joke Hofstee	75
8.1	Probleemstelling	75

8.2	Werkwijze	75
8.3	Resultaten	76
8.4	Toekomstperspectieven	80
8.5	Literatuur	80
9	Uitwisseling van informatie tussen de auteursomgeving en Facet Arjan Aarnink.....	81
9.1	Inleiding	81
9.2	Dutch Exam Profile	81
9.3	Toepassing van standaarden in de keten	82
9.4	Noodzakelijke afwijkingen en extensies	82
9.5	Toekomstige ontwikkelingen	84
9.6	Literatuur	85
10	Dataverwerking en itemanalyses Jesse Koops	87
10.1	Inleiding	87
10.2	Dataverwerking	88
10.3	Itemanalyses en sleutelcorrectie	90
10.4	Literatuur	92
11	Standaardbepaling Sanneke Schouwstra, Jesse Koops en Karen Keune.....	93
11.1	Doel van standaardbepaling	93
11.2	De 3DC methode	93
11.3	De experts	93
11.4	De opgaven	94
11.5	De taak van de experts	95
11.6	Vaststelling van de standaard	97
11.7	Conclusie	98
11.8	Literatuur	99
12	Psychometrische benadering binnen de DTT Daniel van der Palm, Herbert Hoijtink.....	101
12.1	Model	101
12.2	Ad-hoc-kalibratie, kalibratie en herkalibratie	101
12.3	Optimalisatie van de blokindeling door middel van simulatie	104
12.4	Optimalisatie van de priors.....	107
12.5	Literatuur	108
13	Adaptiviteit Sanneke Schouwstra, Peter van Os, Joke Hofstee.....	109
13.1	Gradaties van adaptiviteit.....	109
13.2	Adaptieve procedure met blokken.....	110
13.3	Adaptief verloop	112
13.4	Adaptieve regels	113
13.5	Zaaien	114
13.6	Adaptieve architectuur	115
13.7	Conclusie	117
13.8	Literatuur	117

14	De creatie van een nieuwe rapportagevorm voor de DTT Sanneke Schouwstra en Patrick de Klein	119
14.1	Stadia in de ontwikkeling van de rapportagevorm	119
14.2	Basis van de DTT-rapportage	120
14.3	De mock-up	121
14.4	Het definitieve product	124
14.5	Literatuur	126
15	Resultaten van de DTT-afname 2017 Sanneke Schouwstra, Jesse Koops, Daniel van der Palm.....	127
15.1	Doorlopen paden in de adaptieve afname	127
15.2	Toetsduur en antwoortijd.....	129
15.3	Toetsuitkomsten	130
15.4	Itemanalyses en (her-)kalibratie	133
15.5	Blokindeling voor oplevering 2018	133
15.6	Literatuur	135
16	Nawoord Sanneke Schouwstra	137
16.1	Literatuur	139
17	Bijlagen	141
17.1	Bijlage Voorbeelden van opgavetypen.....	141
17.2	Bijlage accuratesse, grenswaardes en priors bij de opgeleverde blokindelingen	161
17.3	Bijlage Visualisaties van de doorlopen paden per vak en schoolsoort/leerweg (per package).....	176
17.4	Bijlage Resultaten van de enquête over het rapportagemodel	197

1 Achtergrond en proces van toetsontwikkeling van de DTT

Sanneke Schouwstra

1.1 Achtergrond van de DTT

Het idee voor de ontwikkeling van een diagnostische tussentijdse toets (DTT) komt voort uit een advies van de Onderwijsraad begin 2011. De Onderwijsraad adviseerde dat scholen aan het einde van het tweede leerjaar voor alle leerlingen bepalen waar zij staan om tussentijds de vinger goed aan de pols te houden. “De tussentijdse niveaubepaling moet waar nodig leiden tot een bijstelling van het onderwijsprogramma voor (individuele) leerlingen” (Onderwijsraad, februari 2011, p. 14). Dit advies werd opgevolgd in het Actieplan Beter Presteren (mei 2011). In dit plan staat dat de toenmalige minister van Onderwijs, Cultuur en Wetenschap (OCW), Marja van Bijsterveldt-Vliegenthart, een diagnostisch tussentijdse toets voor de doorstroomrelevante vakken wil invoeren als stimulans tot de verdere realisatie van een cultuur van opbrengstgericht werken, die is gericht op de optimalisatie van de leerresultaten van alle leerlingen.

Op 19 juni 2013 diende de toenmalige staatssecretaris van OCW, Sander Dekker, een wetsvoorstel in voor de invoering van een leerlingvolgsysteem en een DTT in het voortgezet onderwijs. Volgens het wetsvoorstel (2013, p. 9) zou de DTT verplicht ingevoerd worden voor alle leerlingen in alle onderwijsvormen in het tweede leerjaar van het vmbo en in het derde leerjaar van het havo en vwo. Deze DTT had primair een formatieve functie (2013, p. 14). De DTT was erop gericht om ten behoeve van de docent zoveel mogelijk informatie te vergaren over de prestaties van de leerling, opdat de docent gericht actie kan ondernemen voor verbetering en maatwerk (p. 23).

In opdracht van het Ministerie van OCW is Stichting Cito (hierna: Cito) in 2012 begonnen met de ontwikkeling van de DTT, onder leiding van het College voor Toetsen en Examens (CvTE). De DTT moest gebaseerd zijn op de tussendoelen, een beschrijving van vakinhouden en vakvaardigheden in een doorlopende leerlijn (SLO, 2012), en afgenomen worden in Facet, het digitale systeem voor het afnemen van centrale toetsen en examens (beheerd door DUO).

Na een levendig politiek debat over het nut en de noodzaak van toetsing in het funderend onderwijs heeft het Ministerie van OCW in 2014 een herziene opdracht verstrekt aan het CvTE voor de ontwikkeling en implementatie van DTT. Volgens deze opdracht moest de DTT in een driejarige pilot op vrijwillige basis doorontwikkeld worden samen met scholen. Aan het einde van de pilotperiode zou een prototype van de DTT opgeleverd worden.

In juni 2016 is in een algemeen overleg van de vaste commissie voor Onderwijs, Cultuur en Wetenschap met de staatssecretaris besloten om het prototype DTT aan het einde van de pilot aan de markt over te dragen (verslag algemeen overleg, 17 augustus 2016). In de daaropvolgende maanden is het ministerie in gesprek gegaan met verschillende aanbieders van toetsen en leermiddelen. Voor de geïnteresseerden zijn bijvoorbeeld informatiesessies georganiseerd waarin presentaties en demonstraties zijn gegeven over DTT. Begin juni 2017 werd bekend gemaakt dat twee organisaties vanaf schooljaar 2017-2018 een diagnostische tussentijdse toets gaan aanbieden. Een half jaar voor het einde van de pilotperiode, in juni 2017 is begonnen met de oplevering van de DTT zodat het ministerie van OCW de toets kon overdragen aan marktpartijen. In dit rapport wordt de ontwikkeling van het prototype DTT beschreven dat in 2017 opgeleverd is.

1.2 Uitgangspunten en kenmerken van de DTT

De uitgangspunten voor de DTT worden beschreven in het wetsvoorstel Leerlingvolgsysteem en diagnostische tussentijdse toets voortgezet onderwijs (Kamerstuk 33661, 2013). Er was een landelijke verplichte DTT voorzien voor de doorstroomrelevante vakken Nederlandse taal, Engelse taal en wiskunde, alsmede op het terrein van rekenen. De toets was voor alle leerlingen aan het einde van de onderbouw in vmbo (leerjaar 2) en havo/vwo (leerjaar 3). De methode-onafhankelijke, landelijk genormeerde toets (wetsvoorstel, p. 19) moest op de niveaus vmbo-bb, vmbo-kb, vmbo-gt, havo en vwo worden aangeboden.

De DTT had een formatieve functie en was diagnostisch van aard. De nog vast te stellen tussendoelen moesten richtinggevend zijn voor de inhoud van DTT. De tussendoelen moesten geformuleerd worden op de vijf beheersingsniveaus en moesten geënt worden op de kerndoelen van de onderbouw, de referentieniveaus, de eindtermen (exameneisen) en voor Engels op het Europees Referentiekader (p. 24).

Verder moesten de mogelijkheden om de toets adaptief en digitaal af te nemen onderzocht worden (p. 23, p. 42). In een voorstudie moest het idee van een diagnostische tussentijdse toets verder uitgewerkt worden en beproefd worden op zijn haalbaarheid. Een dergelijke voorstudie werd van groot belang geacht voor het welslagen, omdat het een complex en uitdagend instrument was dat een sterk beroep zou doen op het innovatieve vermogen van de ontwikkelaars.

1.2.1 Kenmerken van de DTT

Formatieve functie

De instrumenten die men inzet bij formatief evalueren om informatie te krijgen, zijn niet per definitie anders dan zogenaamde summatieve toetsen. Wanneer toetsresultaten gebruikt worden om leerlingen te plaatsen, selecteren, certificeren of classificeren is sprake van een summatieve functie. Een toets vervult een formatieve functie als de resultaten gebruikt worden om feedback te geven en het leerproces verder te sturen en richting te geven (Heritage, 2007; Roelofs & Schouwstra, 2012; Van der Kleij, Vermeulen, Eggen, & Veldkamp, 2013).

Er zijn verschillende strategieën om aan informatie te komen voor formatief evalueren zoals spontane evaluaties tijdens een les en geplande beoordelingsmomenten tijdens lessen (Heritage, 2007). Kenmerkend is dat dergelijke strategieën om aan informatie te komen verbonden zijn aan de onderwijsmethode die de docent hanteert. De DTT is echter een andere vorm van formatieve evaluatie: methode-onafhankelijke formatieve evaluatie (Roelofs & Schouwstra, 2012, p. 17).

Diagnostisch van aard

Om die formatieve functie te kunnen vervullen wordt de informatie verzameld met behulp van een diagnostische toets die de leerbehoeftes van leerlingen in kaart brengt. Oorspronkelijk waren diagnostische toetsen vooral bedoeld om leerlingen met speciale onderwijsbehoeftes te identificeren (Van der Kleij, Vermeulen, Eggen, & Veldkamp, 2013), zoals leerlingen met leerproblemen of juist excellente leerlingen die meer uitdaging nodig hebben. Kenmerkend aan de DTT is dat de toets juist voor álle leerlingen bedoeld is. Elke leerling heeft zijn eigen onderwijsbehoeftes en door daarop in te spelen kan elke leerling geholpen worden het beste uit zichzelf te halen.

Er kunnen verschillende soorten diagnoses onderscheiden worden op grond waarvan een advies voor interventie gegeven kan worden. Allereerst is er een verhelderende diagnose: Beheerst de leerling de gestelde onderwijsdoelen globaal of niet? Dit is het soort diagnose dat doorgaans bij een summatieve toets wordt gegeven. Bij een onderkende diagnose, de tweede soort, wordt een sterkte en zwakte profiel gegeven: Beheerst de leerling de kennis, deelvaardigheden en strategieën die noodzakelijk zijn voor een prestatie op niveau? Het derde soort diagnose is de verklarende diagnose, die aangeeft hoe het komt dat de leerling sterk of zwak is binnen een specifiek gebied. Een dergelijke verklarende diagnose is een vergaande ambitie die grondig wetenschappelijk (leer-)psychologisch onderzoek en kennis vereist. Binnen de DTT hebben we ons derhalve voornamelijk gericht op de tweede soort van diagnoses: een onderkende diagnose.

Er zijn verschillende soorten diagnostische toetsen (Rupp, Templin en Henson, 2010). Veel diagnostische toetsen zijn gebonden aan specifieke lesmethodes (Oomens et al., 2017), maar er zijn ook zogenaamde cognitieve diagnostische toetsen. Dit zijn toetsen die ontworpen zijn om meer fijnmazige informatie te geven over cognitieve sterkte en zwaktepunten van leerlingen (Leighton & Gierl, 2007). Dergelijke toetsen zijn gebaseerd op zogenaamde cognitieve modellen of leerlingmodellen. De DTT is zo een toets gebaseerd op leerlingmodellen. Een leerlingmodel beschrijft de aard van de vaardigheid die centraal staat. In het model wordt een samenhangende structuur beschreven van kennisaspecten, deelvaardigheden en/of strategieën die noodzakelijk zijn voor een prestatie op niveau of die onderdeel vormen van de vaardigheid.

Adaptieve toets

Om die fijnmazige informatie en vele diagnoses te kunnen geven moest de DTT adaptief zijn. In een gewone (lineaire) toets zou een enorme hoeveelheid aan opgaven en antwoorden van een leerling nodig zijn om alle diagnoses te kunnen stellen. Een dergelijke hoeveelheid opgaven afnemen bij leerlingen is praktisch niet haalbaar, maar in een adaptieve toets kan effectiever getoetst worden. Een adaptieve toets is een toets die zich automatisch aanpast aan het antwoordgedrag van de individuele leerling. Na elke beantwoorde opgave wordt nagegaan hoeveel we al weten over de sterkte en verbeterpunten van de leerlingen en welke opgaven nog voorgelegd moeten worden om accurate diagnoses te kunnen stellen.

Een tweede reden voor de adaptiviteit was, dat oorspronkelijk ook het idee was dat een leerling moest kunnen laten zien dat hij/zij een vak op een hoger niveau beheerste (wetsvoorstel, p. 23). Na de herziene opdracht in 2014 heeft het Ministerie van OCW echter besloten dat in het prototype een dergelijke opstroom-indicatie (nog) niet gegeven hoefde te worden omdat een dergelijke opstroom-indicatie extra afnames (in havo2) zou vereisen.

Digitale afname

Een volledig adaptieve afname, waarin na elk opgave geëvalueerd wordt wat er al bekend is over de leerling en welke opgaven vervolgens voorgelegd moeten worden, vereist een digitale afname en het automatisch evalueren van de leerlingantwoorden. De automatische evaluatie van de leerlingantwoorden maakt het ook mogelijk om vrijwel direct na de afname de toetsuitslag gereed te hebben. De digitale afname, de automatische evaluatie en de snelle rapportage van uitkomsten zouden ook de last voor docenten moeten beperken (wetsvoorstel, p. 22).

Een digitale afname heeft ook voor de leerlingen voordelen. Leerlingen groeien op in het digitale tijdperk. Veel scholen werken tegenwoordig met digitale leermiddelen, tablets en computers. Een digitale toets sluit daarom beter aan bij de leefwereld van de leerlingen. Een digitale toets maakt het ook mogelijk om nieuwe vormen van opgaven te ontwikkelen, die meer gebruikmaken van de digitale mogelijkheden waaraan leerlingen tegenwoordig gewend zijn.

In opdracht van het ministerie van OCW is in 2011 begonnen met de bouw van Facet, de software voor de afname van digitale examens en toetsen (CvTE, 2017). Het was de bedoeling dat de DTT afgenomen zou worden in dit nieuwe digitale afnamesysteem.

Kader 1-1. Overzicht van de uitgangspunten en kenmerken van de DTT

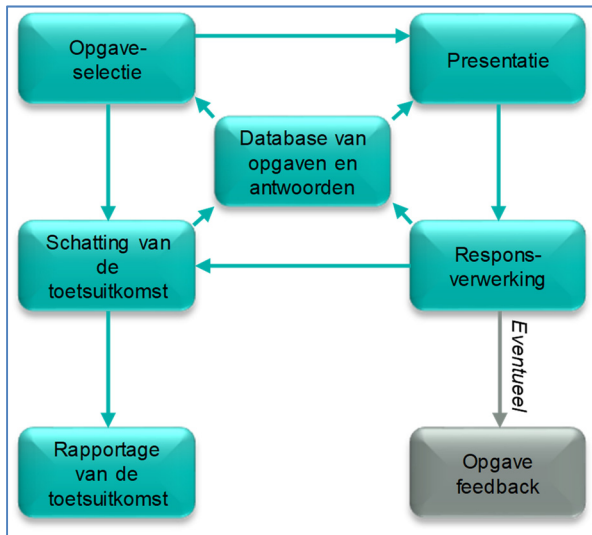
- Voor de doorstroomrelevante vakken Nederlandse taal, Engelse taal en wiskunde/rekenen
- Op de niveaus vmbo-bb, vmbo-kb, vmbo-gt, havo en vwo
- Voor alle leerlingen aan het einde van de onderbouw in vmbo (leerjaar 2) en havo/vwo (leerjaar 3)
- Methode-onafhankelijk, landelijk genormeerd
- Formatieve functie
- Diagnostisch van aard: gebaseerd op leerlingmodellen: voor alle leerlingen een onderkende diagnose van de kennisaspecten, deelvaardigheden en strategieën die noodzakelijk zijn voor een prestatie op niveau
- Adaptief
- Digitaal (in Facet)

1.3 Het proces van de toetsontwikkeling

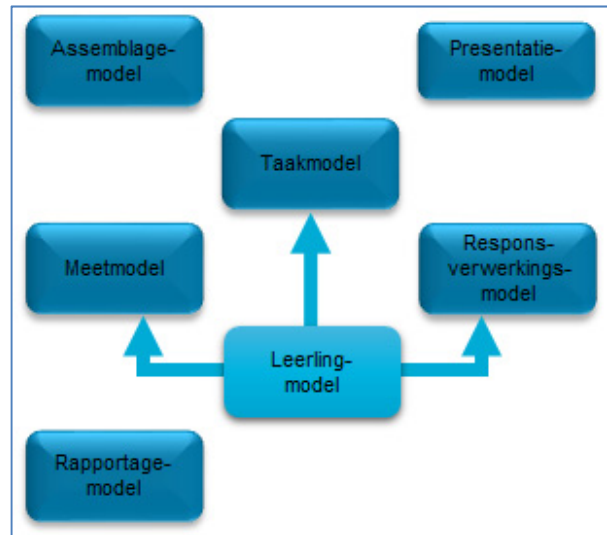
1.3.1 Toetsontwerp

De toetsontwikkeling is begonnen met een voorstudie waarin het concept van een diagnostische tussentijdse toets is uitgewerkt en beproefd op zijn haalbaarheid (Roelofs & Schouwstra, 2012). Voor de ontwikkeling van de DTT is een aanpak van toetsontwikkeling opgezet, die aansluit bij het evidence-centered design (ECD; Mislevy, Steinberg & Almond, 1999; Mislevy, Almond & Lukas, 2003). Het ECD-raamwerk dwingt ontwikkelaars om een heldere argumentatie op te zetten bij alle activiteiten en processen die deel uitmaken van het toetsen van (complexe) vaardigheden (Roelofs & Schouwstra, 2012).

Volgens Mislevy, Almond en Lukas (2003) kan de toetsafname weergegeven worden als een cyclus van een aantal processen (zie Figuur 1-1). Voor elk van deze processen moet een blauwdruk ontwikkeld worden in het toetsontwerp (zie Figuur 1-2). Tijdens de voorstudie is een exemplarisch toetsontwerp gemaakt door toetsdeskundigen Engels, Nederlands en wiskunde, psychometrici, onderwijskundigen en ICT'ers. Er zijn meerdere deelmodellen onderscheiden in het toetsontwerp waaraan simultaan gewerkt is: het leerlingmodel, het responsverwerkingsmodel, het meetmodel, het taakmodel, het rapportagemodel, het assemblagemodel, het presentatiemodel en het afnamemodel (zie Figuur 1-2).



Figuur 1-1. Processen in de toetsafname (bewerking van Mislevy, Almond en Lukas, 2003)



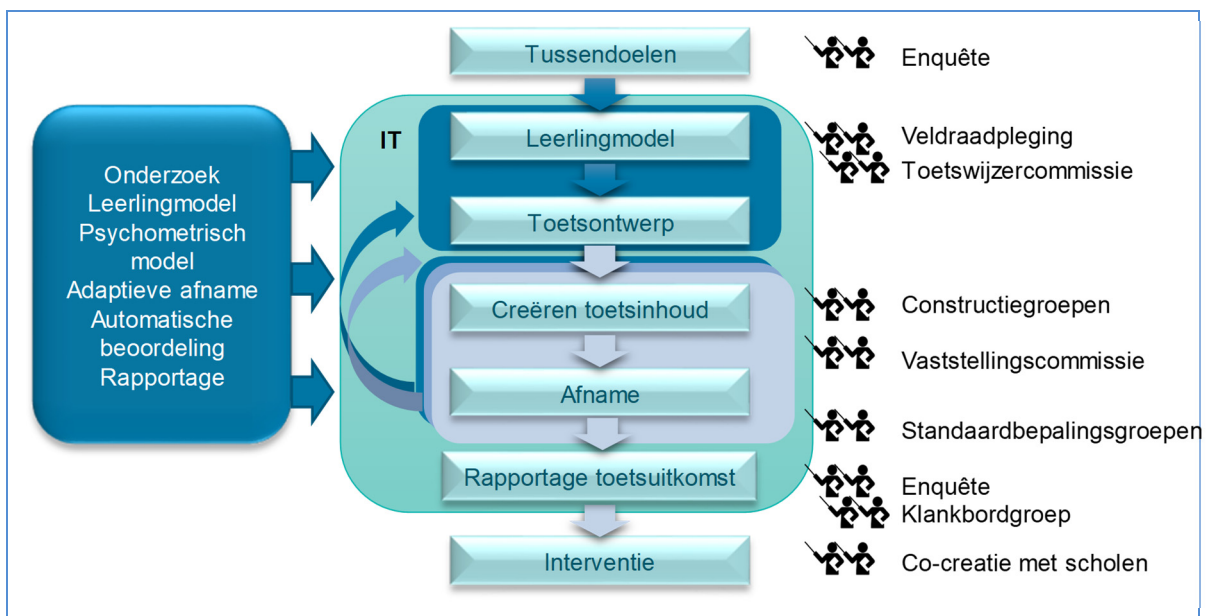
Figuur 1-2. Toetsontwerp

- Het hart van het raamwerk wordt gevormd door het *leerlingmodel*. Zoals eerder aangegeven beschrijft het leerlingmodel kennisaspecten, deelvaardigheden en/of strategieën die noodzakelijk zijn voor een prestatie op niveau.
- Het *taakmodel* is de basis voor de opgave-ontwikkeling. Het beschrijft met wat voor taken en situaties de vaardigheden van de leerling geobserveerd kunnen worden en welke responsen van leerlingen verwacht kunnen worden.
- Het *presentatiemodel* beschrijft hoe deze opgaven worden voorgelegd aan de leerlingen.
- Het *responsverwerkingsmodel* beschrijft hoe de responsen worden geëvalueerd en hoe bijvoorbeeld scores of codes worden toegekend aan de antwoorden.
- In het *meetmodel* worden de waarnemingen verbonden aan de vaardigheid waarin men geïnteresseerd is en wordt de standaardbepaling of normering beschreven. In het ECD-model van Mislevy, Steinberg en Almond (1999) worden het responsverwerkingsmodel (scoring) en het meetmodel samengenomen in één model: het bewijsmodel.
- In het *assemblagemodel* wordt beschreven hoe opgaven geselecteerd worden en de toets wordt samengesteld. Voor de DTT wordt hier dus de adaptieve afname beschreven.
- In het *rapportagemodel* wordt beschreven hoe de rapportage wordt vormgegeven zodat de rapportage voldoet aan de gebruikerswensen en functie van de toets. In het oorspronkelijke ECD-raamwerk is geen sprake van een rapportagemodel, maar de rapportage is cruciaal voor het vervullen van de formatieve functie en is daarom aan het ontwerp van de DTT toegevoegd.
- Het *afnamemodel* beschrijft hoe de verzameling van modellen tezamen werken bij de toetsafname. Het beschrijft ook zaken die dwars door alle andere modellen spelen, zoals het platform waarop wordt afgenomen.

1.3.2 Fases in de toetsontwikkeling

Het hele proces van ontwikkeling van de DTT heeft gefaseerd plaatsgevonden (zie Figuur 1-3). Gedurende het gehele proces zijn vakexperts en met name docenten en scholen, betrokken bij de ontwikkeling van de DTT geweest.

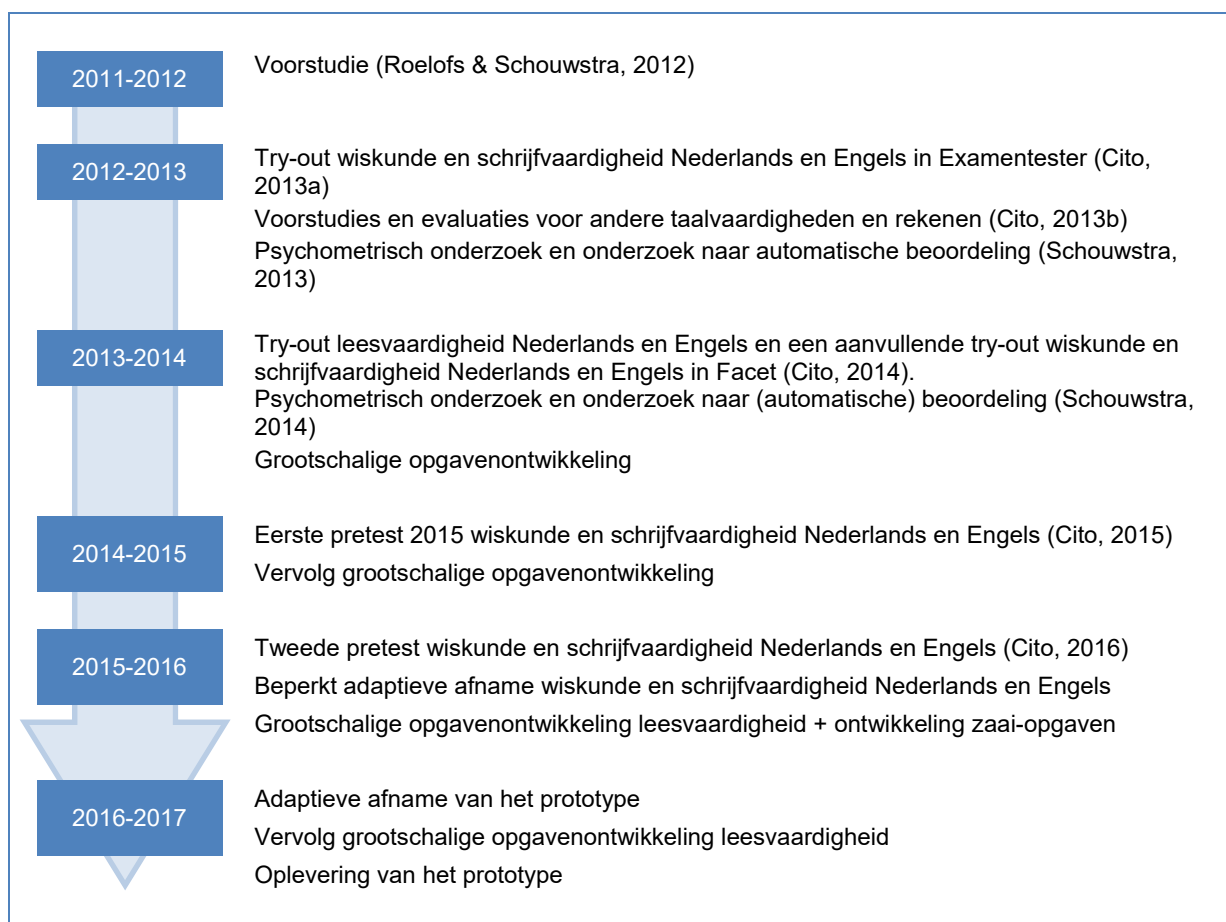
1. De eerste fase is het formuleren van de tussendoelen door SLO (SLO, 2012). Voor het formuleren van die tussendoelen heeft SLO vakexperts geraadpleegd onder andere via een digitale enquête.
2. Vervolgens heeft de voorstudie plaatsgevonden (Roelofs & Schouwstra, 2012), waarin het concept van een diagnostische toets en de aanpak van de toetsontwikkeling zijn uitgewerkt.
 - a. Eerst zijn leerlingmodellen ontwikkeld. De tussendoelen zijn geanalyseerd om relevante kennis- en vaardigheidsaspecten te identificeren voor het leerlingmodel. Daarnaast hebben per vak veldraadplegingen plaatsgevonden over de eerste aanzet van de leerlingmodellen. Uiteindelijk hebben de toetswijzercommissies van het CvTE, bestaande uit vakexperts, de definitieve leerlingmodellen uitgewerkt.
 - b. Aansluitend is een exemplarisch toetsontwerp ontwikkeld.
3. Op grond van het eerste toetsontwerp zijn tijdens een try-out exemplarische opgaven ontwikkeld en afgenomen in kleinschalige afnames waaraan scholen vrijwillig deelnamen. Op grond van de bevindingen kon het toetsontwerp en de specificaties aangescherpt worden. Tijdens de try-out vond ook verder onderzoek plaats voor kwaliteitsborging, bijvoorbeeld simulatie-onderzoek om het aantal benodigde opgaven te bepalen.
4. Na de try-out begon de grootschalige ontwikkeling van opgaven met constructiegroepen bestaande uit vakdocenten en is de opgavebank gevuld. De vaststellingscommissie van het CvTE bepaalde welke opgaven uitgetest konden worden in de pretest.
5. De vastgestelde opgaven zijn in twee pretesten afgenomen. Hierna zijn de afgenomen opgaven gebruikt in een toetsgeoriënteerde standaardbepaling met docenten. Op grond van die standaarden en de pretestgegevens konden de opgaven gekalibreerd worden.
6. Daarna heeft de eerste afname plaatsgevonden, in de beoogde opzet en infrastructuur.
7. Na de eerste afname kan de operationele fase volgen, waarin de toets jaarlijks afgenomen wordt en onderhouden wordt.



Figuur 1-3. Ontwikkeling van de DTT

Sommige fases verliepen deels gelijktijdig omdat de DTT zeer snel gerealiseerd moest worden. De eerste afname was in het wetsvoorstel voorzien in 2015, maar na de herziene opdracht van het Ministerie van OCW in 2014 is de ontwikkeling getemporeerd en heeft de eerste beperkt adaptieve afname plaatsgevonden in 2016.

De ontwikkeling van de DTT was uitgewerkt in een groeimodel. Volgens dit groeimodel besloeg een fase één jaar en zou elk jaar de ontwikkeling van een andere taalvaardigheid starten. Na de herziene opdracht van het Ministerie van OCW in 2014 is besloten om de ontwikkelde opgaven in twee kleinere afnames met vrijwillige deelname te pretesten in plaats van in één hele omvangrijke afname. Ook de doorontwikkeling van leesvaardigheid is getemporeerd en is niet verder gegaan met de ontwikkeling van de toets voor luistervaardigheid. Na het besluit om de DTT over te dragen in 2016 heeft het Ministerie van OCW besloten om het pretesten van de opgaven leesvaardigheid te beëindigen, omdat dit niet in 2017 tot een afgerond onderdeel van het prototype zou leiden. Hieronder een overzicht van de doorlopen fases per schooljaar (Figuur 1-4).



Figuur 1-4. Tijdlijn van de DTT-ontwikkeling

1.4 Realisatie van het toetsontwerp

In de volgende hoofdstukken wordt beschreven hoe het toetsontwerp is uitgewerkt tot het prototype dat is opgeleverd in 2017. In het eerste deel van dit rapport wordt vakoverstijgend uiteengezet hoe de opgavenontwikkeling (Hoofdstuk 2) en responsverwerking (Hoofdstuk 3) is aangepakt. Vervolgens worden per vak de leerlingmodellen en de operationalisatie (realisatie van de taakmodellen) beschreven. Speciale aandacht wordt besteed aan de open opgaven bij talen en bij wiskunde.

In de Hoofdstukken 4 tot en met 8 wordt de vakspecifieke ontwikkeling besproken. Voor de leesbaarheid van de separate hoofdstukken was enige herhaling van de ontwikkelingsstappen noodzakelijk.

In Hoofdstuk 4 wordt de ontwikkeling van de DTT Nederlands besproken en aansluitend in Hoofdstuk 5 de ontwikkeling van de DTT Engels. In Hoofdstuk 6 wordt het onderzoek naar de beoordeling van open

schrijftaken bij Nederlands en Engels besproken. Na de talen wordt in Hoofdstuk 7 de ontwikkeling van de DTT wiskunde besproken en in Hoofdstuk 8 de ontwikkeling van automatische beoordeling bij wiskunde.

In het derde deel van dit rapport wordt ingegaan op de techniek en de psychometrie. Allereerst wordt in Hoofdstuk 9 beschreven hoe de informatie tussen de auteursomgeving waarin de opgaven staan wordt uitgewisseld met de digitale afname-omgeving Facet. Hierna wordt beschreven hoe de gegevens die terugkwamen zijn verwerkt en geanalyseerd (Hoofdstuk 10), de standaardbepaling (Hoofdstuk 11) beschreven en het psychometrische model dat ten grondslag ligt aan de toetsuitkomsten (Hoofdstuk 12). In Hoofdstuk 13 wordt de adaptieve toetssamenstelling uiteen gezet.

In het laatste deel van het rapport wordt ingegaan op de resultaten. In Hoofdstuk 14 wordt beschreven hoe een nieuwe rapportagevorm voor de toetsuitkomsten is ontwikkeld. Tenslotte worden de resultaten van de volledige adaptieve afname in 2017 gepresenteerd (Hoofdstuk 15) en het rapport wordt afgesloten met een nawoord (Hoofdstuk 16).

1.5 Literatuur

- Cito (2013a). *Diagnostische tussentijdse toets: Verslag try-out 2013*. Arnhem: Cito.
- Cito (2013b). *Diagnostische tussentijdse toets: Voorstudies en evaluaties 2013*. Arnhem: Cito.
- Cito (2014). *Diagnostische tussentijdse toets: Verslag try-out 2014*. Arnhem: Cito.
- Cito (2015). *Diagnostische tussentijdse toets: Verslag pretest 2015*. Arnhem: Cito.
- Cito (2016). *Diagnostische tussentijdse toets: Verslag pretest 2016*. Arnhem: Cito.
- College voor Toetsen en Examens (2017). *Facet geslaagde examensoftware*. Utrecht: College voor Toetsen en Examens.
- Heritage, M. (2007). Formative Assessment: What Do Teachers Need to Know and Do. *Phi Delta Kappan*, 89, 140-145.
- Leighton, J., & Gierl, M. (Red.). (2007). *Cognitive Diagnostic Assessment for Education: Theory and Applications*. Cambridge University Press.
- Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A Brief Introduction to Evidence-centered Design*. Princeton, NJ: Educational Testing Service.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (1999). *Evidence-Centered Assessment Design*. Princeton, NJ: Educational Testing Service.
- Onderwijsraad (2011). *Advies: Naar hogere leerprestaties in het voortgezet onderwijs*. Den Haag: de Onderwijsraad.
- Oomens, M., Exalto, R., de Jong, A., Scholten, F., Veldkamp, B., Janse, R., & Scheerens, J. (2017). *Marktonderzoek formatief evalueren: Een onderzoek naar vraag en aanbod*. Oberon.
- Roelofs, E., & Schouwstra, S. (Red.). (2012). *Diagnostische tussentijdse toets: Verslag van de voorstudie*. Arnhem: Cito.
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic Measurement: Theory, Methods, and Applications*. New York: The Guilford Press.
- Schouwstra, S. (Red.). (2013). *Diagnostische tussentijdse toets: Onderzoek 2013*. Arnhem: Cito.
- Schouwstra, S. (Red.). (2014). *Diagnostische tussentijdse toets: Onderzoek 2014*. Arnhem: Cito.

SLO: Nationaal expertisecentrum leerplanontwikkeling (april 2012). *Concept-tussendoelen kernvakken onderbouw vo*. Geraadpleegd op <http://www.slo.nl/nieuws/00291/>.

Van Bijsterveldt-Vliegenthart, M. (2011, 5 mei). *Actieplan beter presteren* [bijlage Kamerbrief]. Geraadpleegd van <https://www.rijksoverheid.nl/documenten/kamerstukken/2011/05/23/actieplan-vo-beter-presteren>.

Van der Kleij, F., Vermeulen, J., Eggen, T., & Veldkamp, B. (2013). *Leren van Toetsen: Een cyclisch proces*. Arnhem: ToetsWijzer|Stichting Cito Instituut voor Toetsontwikkeling.

Verslag van een algemeen overleg over de diagnostische tussentijdse toets (2016, 17 augustus).

Geraadpleegd van https://www.tweedekamer.nl/debat_en_vergadering/commissievergaderingen/details?id=2016a02431

Voorstel Wijziging van de Wet op het voortgezet onderwijs, de Wet voortgezet onderwijs BES, de Wet college voor toetsen en examens alsmede de Wet op de expertisecentra in verband met onder meer de invoering in het voortgezet onderwijs van een leerlingvolgsysteem, een diagnostische tussentijdse toets en verplichte deelname aan internationaal vergelijkend onderzoek (leerlingvolgsysteem en diagnostische tussentijdse toets voortgezet onderwijs), (2013, 19 juni), Kamerstuk 33661, Geraadpleegd op <https://zoek.officielebekendmakingen.nl/kst-33661-4.html>.

2 Opgaveconstructie

Wilma Vrijs en Sanneke Schouwstra

2.1 Het proces van opgaveconstructie

Het maken van opgaven voor een formatieve, diagnostische toets via het evidence-centered design (zie Hoofdstuk 1) vereist een andere manier van werken dan het maken van een summatieve toets zoals een examen. Voor de toetsdeskundigen die de DTT hebben ontwikkeld kwam dat verschil vooral tot uiting in de volgende onderdelen:

- **Een leerlingmodel** vormde de basis voor de toets: een model waarin alle (deel)aspecten van de te diagnosticeren vaardigheid beschreven zijn. De leerlingmodellen voor de drie vakken van de DTT staan beschreven in de Toetswijzer (CvTE, 2014);
- **Een taakmodel** diende als input voor de context in opgaven, meer authentieke interacties (opgavetypen) zijn toegepast en waar nodig zijn innovatieve opgavetypen ontwikkeld om een meer directe en authentieke bevraging mogelijk te maken. In de Toetswijzer is beschreven welke opgavetypen voorkomen in de DTT (CvTE, 2014);
- **Het coderen** van de antwoordruimte (alle mogelijke antwoorden), gebaseerd op het leerlingmodel. Het gaat niet om een goed/foutscore per opgave. Bij de codering wordt bepaald op welke (deel)vaardigheden een opgave betrekking heeft en welke leerlingresponsen welke diagnostische waarde hebben.

Er kunnen vier stadia onderscheiden worden in de opgaveconstructie. De opgaveconstructie is begonnen met het ontwikkelen van prototypische opgaven tijdens de voorstudie (Roelofs & Schouwstra, 2012). Met name in dit eerste stadium van opgaveconstructie zijn de toetsdeskundigen begeleid door onderwijskundig specialisten. Na de voorstudie is in het tweede stadium een selectie aan opgavetypen uitgetoetst in kleinschalige try-outs (Cito, 2013a; Cito, 2014). Zo hadden de toetsdeskundigen een beeld van welke opgaven goed functioneerden in relatie tot de leerlingmodellen. Hierna is een tweejarig traject van grootschalige opgaveconstructie ingezet dat in dit hoofdstuk beschreven wordt. Na de grootschalige opgaveconstructie is in het vierde stadium jaarlijks (tweemaal) geconstrueerd voor verversing van de opgavebank.

2.1.1 Constructieopdrachten

De basis voor de grootschalige opgaveconstructie zijn de constructieopdrachten van het CvTE. In de constructieopdrachten van het CvTE is beschreven *hoeveel* opgaven (responsen) er gemaakt moesten worden per vak, leerweg/schoolsoort en (deel)aspect om een accurate diagnose te stellen of de leerling dit onderdeel onder, op of boven het gewenste niveau beheerst. De *inhoud* van de opgavebanken is gebaseerd op de leerlingmodellen per vak zoals die geadviseerd zijn door de toetswijzercommissies. Met een opgave wordt hier een toetsvraag op één beeldscherm bedoeld. Een opgave kan bestaan uit meerdere responsen.

In Tabel 2-1 is het aantal te construeren responsen per vak/vaardigheid schematisch weergegeven (inclusief een marge voor opgaven die na de pretest afvallen en een beperkt aantal opgaven voor een voorbeeldtoets per vak). Het streven was om binnen elke leerweg/schoolsoort voor elk deelaspect opgaven te ontwikkelen die 24 responsen zouden opleveren. Om op termijn eventueel af- en opstroomindicaties te kunnen geven (en de standaardbepaling), gaat de constructieopdracht uit van 50% overlap tussen de opgavebanken voor aangrenzende leerwegen/schoolsoorten. De gerealiseerde overlap tijdens de afname kan groter of kleiner zijn afhankelijk van het aantal beschikbare opgaven. Bij de vmbo bb-leerlingen worden bijvoorbeeld de vmbo bb-opgaven afgenomen en een deel van de vmbo kb-opgaven.

De aantallen responsen in de constructieopdrachten van het CvTE (uit 2013, 2014 en 2015) zijn steeds berekend door het aantal deelaspecten van het leerlingmodel te vermenigvuldigen met het aantal responsen voor de leerwegen/schoolsoorten inclusief 50% overlap. Bijvoorbeeld, de berekening voor Schrijfvaardigheid Engels vmbo:

9 deelaspecten x 48 responsen (12 geschikt voor vmbo-bb, 12 geschikt voor vmbo-bb en vmbo-kb, 12 geschikt voor vmbo-kb en vmbo-gt, 12 geschikt voor vmbo-gt) = 432 responsen

Tabel 2-1. Overzicht van aantallen te construeren opgaven conform constructieopdrachten

Vak	Schoolsoort	Responsen volgens constructieopdracht grootschalige ontwikkeling	Op te leveren responsen per jaar voor verversing
Schrijfvaardigheid Nederlands	vmbo	576	115
	havo/vwo	468	94
Leesvaardigheid Nederlands	vmbo	528	106
	havo/vwo	396	79
Schrijfvaardigheid Engels	vmbo	432	86
	havo/vwo	324	65
Leesvaardigheid Engels	vmbo	672	134
	havo/vwo	504	101
Wiskunde	vmbo	576	115
	havo/vwo	540	108

Na het afronden van de opgavebanken werden er jaarlijks opgaven geconstrueerd ter verversing conform de constructieopdracht van het CvTE uit 2015. Een opgavebank dient altijd verversd te worden omdat na verloop van tijd de opgaven, en met name de context van de opgaven, verouderd zijn en omdat er opgaven nodig zijn die aansluiten bij de veranderende belevingswereld van de leerlingen. Door jaarlijks een deel te verversen kan de constructie en de analyses van de opgaven gespreid plaatsvinden. Verversing en uitbreiding van de opgavebank is ook noodzakelijk om de kwaliteit, efficiëntie en bruikbaarheid van de toets en diagnoses te handhaven en verhogen (zie ook de sectie Kwaliteitsrichtlijnen in de volgende paragraaf).

In het laatste jaar van de opgaveconstructie is er per vak een set van voorbeeldopgaven ontwikkeld. Doel hiervan was om een illustratieve opgave te hebben bij de verschillende diagnoses die de DTT biedt, bijvoorbeeld: de leerling presteert naar verwachting boven niveau op dit deelaspect. De opgave is dan illustratief voor opgaven die goed onderscheiden tussen leerlingen die op en boven niveau presteren op dat betreffende deelaspect. Zo is er voor elk vak per leerweg/schoolsoort en deelaspect, een opgave ontwikkeld die illustratief is voor een diagnose onder/op niveau en voor een diagnose op/boven niveau.

2.1.2 Fases in de grootschalige opgaveconstructie

Voor de grootschalige opgaveconstructie was er voor elk vak een team van circa vijf toetsdeskundigen beschikbaar; vakspecialisten die over het algemeen zelf voor de klas hebben gestaan en met een specialisatie op vmbo of havo/vwo. Zij zorgden voor de interne begeleiding van het traject: het realiseren van de aantallen opgaven binnen de vastgestelde planningen, maar ook het meedenken over toetssamenstellingen en het meeschrijven aan wetenschappelijke rapporten. Daarnaast adviseerden de toetsdeskundigen de vaststellingscommissies van het CvTE over de opgaven tijdens de vaststellingen. De vakteams werden aangestuurd door een projectleider opgaveconstructie en daarnaast begeleid door een inhoudelijk projectleider en een projectleider ICT.

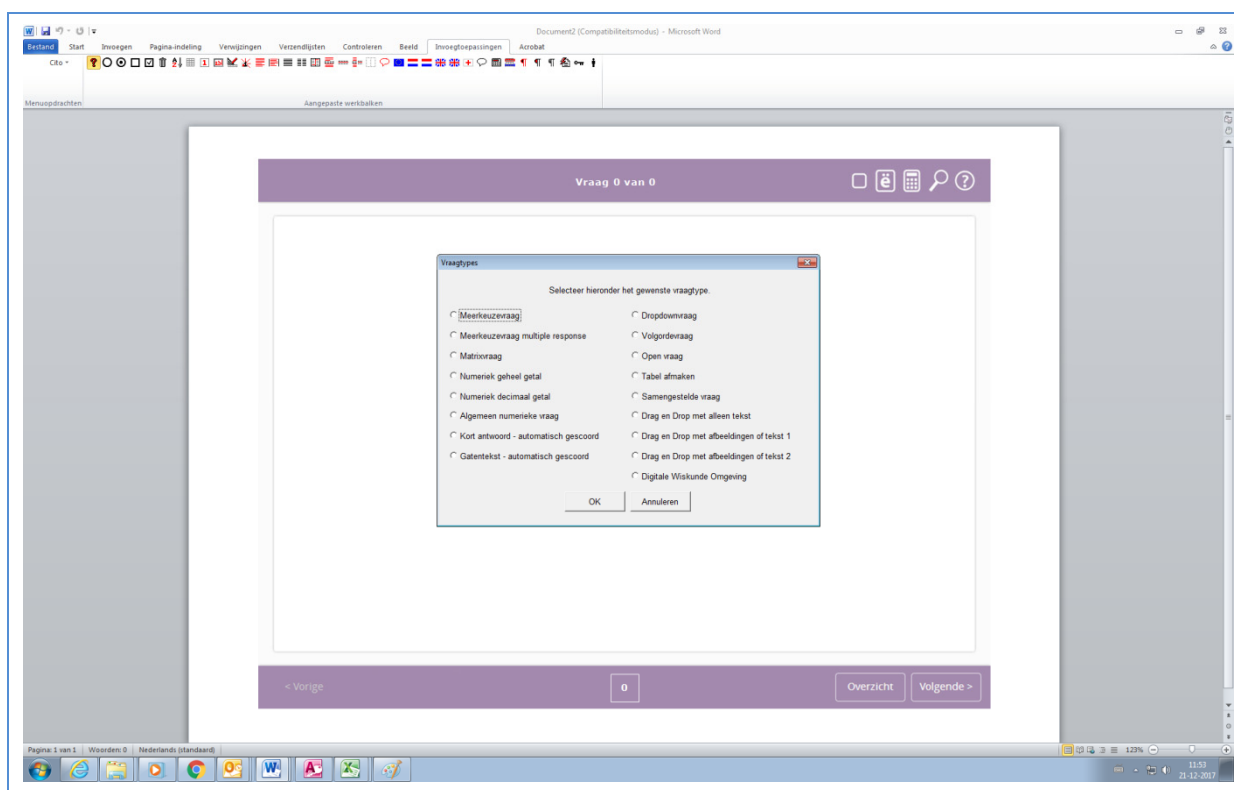
De constructiegroepen maakten onder begeleiding van de vakteams eerst basisopgaven. Deze opgaven werden vervolgens meerdere keren gescreend en verbeterd op grond van de bevindingen uit de screening. Nadat de opgaven goed waren bevonden door de vakteams zijn de opgaven voorgelegd aan de vaststellingscommissie van het CvTE. De vastgestelde opgaven werden vervolgens gepretest zodat de kwaliteit van de opgaven nagegaan kon worden met behulp van empirische gegevens en hierna konden de opgaven gekalibreerd worden. Na de pretest volgde de allerlaatste vaststellingsprocedure voor adaptieve afname van de DTT.

Basisopgaven gemaakt door constructiegroepen

Voor de ontwikkeling van de basisopgaven is gewerkt met constructiegroepen. Deze bestaan uit docenten die lesgeven aan de leerwegen/schoolsoorten en in de leerjaren waarvoor de DTT beoogd is. Zo is geborgd dat de opgaven aansluiten bij het niveau en de beleavingswereld van de leerling. Alle vakken hebben gekozen voor twee aparte constructiegroepen; één voor vmbo en één voor havo/vwo. De docenten uit de constructiegroepen zijn begeleid door de toetsdeskundigen. De toetsdeskundigen en docenten kwamen regelmatig bij elkaar om specifieke opgaven en de werkwijze in het algemeen te bespreken.

De docenten in de constructiegroep zijn ingewerkt via een generieke opleiding toetstechniek van een dag, en een maatwerkprogramma van een halve dag specifiek gericht op DTT en de operationalisatie van het leerlingmodel van het betreffende vak. Vervolgens kregen de docenten een ontwikkelopdracht op basis van het leerlingmodel, bijvoorbeeld: maak vijf opgaven over *Afstemming op publiek en doel*. De docent leverde een eerste versie van een opgave op aan de toetsdeskundige. De toetsdeskundige gaf dan feedback en de docent paste de opgave aan. Een tweede versie werd voorgelegd en de toetsdeskundige bekeek de opgave nog een keer. Gemiddeld genomen maakte een docent circa 5 opgaven per week en had hij daarvoor ongeveer een dagdeel per week ter beschikking.

Voor het maken van de basisopgaven maakten de docenten gebruik van een speciaal voor dit doel ontwikkeld Word-sjabloon. Dit sjabloon hielp hen bij het selecteren van een opgavetype (zie Figuur 2-1) en bij het invullen van alle benodigde metadata bij de opgave. De sjablonen maakten het ook mogelijk om nieuwe vormen van opgaven te maken, ruim voordat dit technisch gerealiseerd was in de auteursomgeving en afspeelomgeving. De Word-sjablonen vergemakkelijkten tevens het werk van de toetsondersteunende medewerkers die de opgaven invoerden in de opgavebank, doordat alle opgaven in een gestandaardiseerde vorm aangeleverd werden. De docenten voerden niet zelf de opgaven in, omdat dit een technische training en ervaring met de auteursomgeving vereist.



Figuur 2-1. De keuze voor een opgavetype in het Word-sjabloon

Screenen

Als de toetsdeskundige tevreden was over de opgaven, werden deze gescreend door andere toetsdeskundigen. Het proces van screening bestond in ieder geval uit:

- Een vakinhoudelijke screening door de toetsdeskundigen van het desbetreffende vakteam.
- Een toetstechnische screening, ook door toetsdeskundigen van het vakteam.
Bij de vakinhoudelijke en toetstechnische screening werd er gekozen voor 'kruislingse' screening, waarbij de toetsdeskundigen met als specialisatie havo/vwo de vmbo-opgaven screenden, en andersom. Zo werd ook aansluiting tussen de leerwegen/schoolsoorten geborgd.
- Een screening op taalgebruik/Nederlands door een afdeling die verantwoordelijk is voor de logistieke productie van de toetsen in het kader van kwaliteitsbeheer.
- Een screening (door diezelfde afdeling) op naleving van overeengekomen richtlijnen ten aanzien van lay-out en taalgebruik, bijvoorbeeld: de bron staat altijd links, de vraag altijd rechts. De vraag is altijd dikgedrukt.
- Technische screenings door de toetsdeskundigen en medewerkers van het Bedrijfsbureau, bedoeld om te testen of de losse opgaven en de toetspakketten als geheel naar behoren functioneren op aspecten als bediening (werkt de sleepopdracht?) scoring/codering (wordt het juiste antwoord goed gerekend) enzovoort.

Tijdens het proces werd de opgave ingevoerd in de softwareapplicatie voor het maken van toetsen en beheren van opgavebanken (de auteursomgeving Questify Builder). Eén van de screenings vond plaats in de auteursomgeving en een laatste technische screening gebeurde in het digitale afnamesysteem (Facet). Het doel hiervan is om de opgave nog een laatste keer te bekijken precies zoals de leerling hem ziet.

Kwaliteitsrichtlijnen

Tijdens de ontwikkeling van opgaven en bij het screenen is gelet op onderstaande kwaliteitsaspecten, die geformuleerd zijn in de constructieopdracht van CvTE uit 2013.

Kader 2-1. Kwaliteitsaspecten in de constructieopdracht van CvTE uit 2013

Validiteit en authenticiteit

In welke mate wordt daadwerkelijk het beoogde aspect van het leerlingmodel gemeten? Dat wil ook zeggen dat bekeken wordt in hoeverre geen andere onbedoelde deelvaardigheden of variabelen een belangrijke rol spelen bij het goed kunnen beantwoorden van de opgave.

Betrouwbaarheid, moeilijkheidsgraad en onderscheidend/discriminerend vermogen

Is de opgave niet te makkelijk of te moeilijk voor de beoogde leerweg? Maakt de opgave goed onderscheid tussen leerlingen die onder, op of boven niveau zitten? In de regel zijn opgaven die door bijna alle leerlingen van de beoogde leerweg goed of fout worden gemaakt weinig informatief en onderscheidend.

Diagnostische informatiewaarde

In welke mate zijn de (sub)opgaven en responscategorieën informatief over de beheersing van een of meer aspecten van het leerlingmodel? De vraag is of alle mogelijke antwoorden/antwoordcategorieën van de leerlingen goed te koppelen zijn aan de mate van beheersing van de beoogde aspecten van het leerlingmodel.

Efficiëntie

Is de responstijd van de opgave niet te groot, gelet op de maximaal beschikbare tijd voor de toets als geheel?

Functionaliteit van bronmaterialen en de taaksituaties

- Is het bronmateriaal (teksten, afbeeldingen, video en audio) nodig en functioneel, gelet op het toetsdoel?
- Is er voldoende variatie in bronmaterialen en taaksituaties? Een taaksituatie is de context waarin de opgave plaatsvindt, denk bij een schrijfoopdracht bijvoorbeeld aan het invullen van een formulier of het schrijven van een brief of verslag.
- Is het bronmateriaal niet te complex, niet te omvangrijk en voldoende herkenbaar voor de leerling?
- Is het bronmateriaal toelaatbaar?

De vaststelling

Een opgave die was ingevoerd in de auteursomgeving en die alle screeningen doorlopen had, werd voorgelegd aan een vaststellingscommissie. Het CvTE heeft deze vaststellingscommissie ingesteld voor elk vak: Nederlands, Engels en wiskunde. Elke vaststellingscommissie bestond uit vier docenten die lesgeven aan de leerwegen/schoolsoorten en in de leerjaren waarvoor de DTT beoogd is, en een voorzitter. Samen hadden zij de taak te bepalen of opgaven dan wel complete toetsen goedgekeurd zijn.

Gespreid door het jaar werden vaststellingsdagen gepland. De vakteams leverden voor elke vaststellingsdag een set opgaven aan de vaststellingscommissie die gekeurd moest worden. De opgaven werden digitaal voorgelegd via de afnameomgeving die de scholen ook gebruiken, zodat de leden van de vaststellingscommissie alle opgaven digitaal konden bekijken op de manier zoals de leerling ze ook te zien zou krijgen.

De vaststellingscommissie beoordeelde elke opgave en gaf indien nodig suggesties voor verbetering. Ter vergadering werden de opgaven besproken en goed- of afgekeurd. Bij afgekeurde opgaven werd soms een aanpassing voorgesteld waardoor de opgave alsnog goedgekeurd kon worden.

De vaststellingscommissie lette bij haar beoordeling op aspecten zoals:

- De moeilijkheid van de opgave: past de moeilijkheid bij de beoogde leerweg/schoolsoort?
- De operationalisatie: is dit een goede manier om het betreffende deelaspect van het leerlingmodel te bevragen?
- De afwisseling binnen de set van opgaven en binnen (deel)aspecten: is er voldoende afwisseling in typen opgaven en inhoud?

In de auteursomgeving hielden de toetsdeskundigen bij welke opgaven waren vastgesteld. Er werd regelmatig gekeken of er voldoende materiaal was goedgekeurd om de aantallen opgaven, noodzakelijk voor accurate diagnoses, in de constructieopdracht te kunnen halen. Indien nodig konden dan meer opgaven geconstrueerd worden.

Pretesten, analyseren en adaptieve afname binnen het proces van toetsconstructie

Na de vaststelling is er een design gemaakt van alle opgaven (Cito, 2015, Hoofdstuk 3; Cito, 2016, Hoofdstuk 3), dat door de toetsdeskundigen inhoudelijk bekeken is. Vervolgens werden de opgaven in twee opeenvolgende jaren gepretest.

Na de pretest zijn toets- en itemanalyses uitgevoerd om de kwaliteit van de opgaven te kunnen beoordelen (Cito, 2015; Cito, 2016). De toetsdeskundigen beoordeelden de resultaten aan de hand van een aantal criteria (zie Hoofdstuk 10 over Dataverwerking en itemanalyses). Alle opgaven waarbij meer dan de helft van de responsen adequaat waren qua moeilijkheid, onderscheidend vermogen en samenhang met de overige responsen van hetzelfde deelaspect werden automatisch goedgekeurd. Alle opgaven die mogelijk problematisch waren werden door de toetsdeskundigen inhoudelijk geïnspecteerd en besproken met de vaststellingscommissie. Die besloot of een opgave werd goedgekeurd voor opname in de adaptieve toets of niet.

2.2 Opgavetypen binnen de DTT

Bij de DTT is gestreefd naar de ontwikkeling van meer authentieke opgaven om directer te kunnen toetsen. Dit is van belang bij diagnostische opgaven, om aan de hand van de aspecten van het leerlingmodel 'bewijs' te kunnen verzamelen over de vaardigheid van de leerling (Evidence Centered Design). Er is een grote diversiteit aan opgavetypen gecreëerd, die het mogelijk maakte om de vaardigheid van de leerlingen op een directe manier aan te spreken en om leerlingen een afwisselende toets aan te bieden om hun vaardigheid te tonen (zie Figuur 2-2). Bij authentiek toetsen wordt geprobeerd om de vorm en inhoud van de toets zo veel mogelijk overeen te laten komen met hoe de leerlingen de gemeten vaardigheid in echte situaties zouden toepassen. Volgens Frey, Petersen, Edwards, Pedrotti en Peyton leggen onderzoekers steeds meer nadruk op het belang van authentiek toetsen (2005). Daarnaast is het met digitaal toetsen mogelijk om meer complexe, dynamische en construct-relevante opgaven aan te bieden (Jodoin, 2003). Cheng en Basu beweren daarnaast in hun artikel uit 2006 over het gebruik van multimedia in opgaven dat

innovatieve vraagvormen motivatie en betrokkenheid van leerlingen kunnen verhogen (geciteerd in Goosen & Vernooij, 2017).

De meeste opgavetypen zijn bruikbaar voor alle vakken, enkele zijn vakspecifiek voor talen of wiskunde ontwikkeld. In de meeste opgavetypen kunnen tekst, afbeeldingen, audio, video en wiskundige expressies worden toegevoegd door de constructeurs. Bij enkele vormen is juist een afbeelding de basis voor de te construeren interactie. De diverse interacties zijn grofweg onder te verdelen in klikken, slepen en invullen (typen).

Op basis van het leerlingmodel en het taakmodel bepaalt de toetsdeskundige welke opgavetype het meest geschikt is om het betreffende aspect te bevragen. In sommige gevallen zijn er meerdere opties mogelijk. Er is speciaal voor de DTT een aantal opgavetypen ontwikkeld die een directere bevraging mogelijk maken. Bij de talen zijn dit de alinea-opgave (zie Figuur 2-3) en het opgavetype waarbij je tekstelementen kunt aanklikken (zoek het woord met dezelfde betekenis, zoek de spelfout, klik het signaalwoord aan waaruit dit blijkt, enzovoort). Er is tevens een vorm ontwikkeld waarbij de leerling niet alleen markeert in de tekst, maar ook kan wijzigen c.q. verbeteren, de correctie-opgave (zie Figuur 2-4).

Bij wiskunde zijn er verschillende opgavetypen ontwikkeld waarbij de leerling diverse wiskundige interacties kan uitvoeren, bijvoorbeeld een grafiek tekenen of het correcte meetkundige figuur tekenen of aanpassen (zie Hoofdstuk 8 over Automatische beoordeling wiskunde). Deze bijzondere opgavetypen zijn tot stand gekomen met input van de toetsdeskundigen.

Selecteer een itemlay-outtemplate

Selecteer een itemlay-outtemplate in onderstaand grid en klik op 'OK'

Code	Titel	Itemtype	Aantal referenties	Beschrijving
DTT (29)				
Cito.CTE.Choice.Inline.DC	Dropdownkeuzevakken twee kolommen	Kort antwoord	308	Dropdownkeuzevakken twee kolommen
Cito.CTE.Choice.Inline.SC	Dropdownkeuzevakken een kolom	Kort antwoord	10	Dropdownkeuzevakken een kolom
Cito.CTE.Combined.Inline.DC	Diverse inline controls twee kolommen	Kort antwoord	116	Diverse inline controls twee kolommen
Cito.CTE.Combined.Inline.SC	Diverse inline controls een kolom	Kort antwoord	6	Diverse inline controls een kolom
Cito.CTE.CustomInteraction.DC	Hm15 customInteraction - twee kolommen	Kort antwoord	32	Hm15 customInteraction - twee kolommen
Cito.CTE.CustomInteraction.SC	Hm15 customInteraction - een kolom	Kort antwoord	89	Hm15 customInteraction - een kolom
Cito.CTE.GapMatch.Inline.DC	Tekst drag and drop twee kolommen	Kort antwoord	233	Tekst drag and drop twee kolommen
Cito.CTE.GapMatch.Inline.SC	Tekst drag and drop een kolom	Kort antwoord	33	Tekst drag and drop een kolom
Cito.CTE.Gaps.Inline.DC	Instelbare invulvakken twee kolommen	Kort antwoord	83	Instelbare invulvakken twee kolommen
Cito.CTE.Gaps.Inline.SC	Instelbare invulvakken een kolom	Kort antwoord	4	Instelbare invulvakken een kolom
Cito.CTE.Geogebra.DC	Geogebra item - twee kolommen	Kort antwoord	1	Geogebra item - twee kolommen
Cito.CTE.Geogebra.SC	Geogebra item - een kolom	Kort antwoord	21	Geogebra item - een kolom
Cito.CTE.MC.DC	Meerkeuze item twee kolommen	Meerkeuze	130	Multiple-choice keuze item met een opmaak in twee kolommen
Cito.CTE.MC.SC	Meerkeuze item een kolom	Meerkeuze	1	Multiple-choice keuzetiem met een opmaak in een kolom
Cito.CTE.MR.DC	Multiple response item twee kolommen	Meerkeuze	5	Multiple response keuzetiem met een opmaak in twee kolommen
Cito.CTE.MR.SC	Multiple response item een kolom	Meerkeuze	1	Multiple response keuzetiem met een opmaak in een kolom
Cito.Generic.DivideInAlineas.DC	Cito.Generic.DivideInAlineas.DC	Meerkeuze	100	
Cito.Generic.GraphicGapMatch.Categorize.DC	Hotspot item twee kolommen	Hotspot	23	Hotspot item met een opmaak in twee kolommen
Cito.Generic.GraphicGapMatch.Categorize.SC	Hotspot item een kolom	Hotspot	33	Hotspot item met een opmaak in één kolom
Cito.Generic.GraphicGapMatch.DC	Hotspot item twee kolommen	Hotspot	133	Hotspot item met een opmaak in twee kolommen
Cito.Generic.GraphicGapMatch.SC	Hotspot item een kolom	Hotspot	35	Hotspot item met een opmaak in één kolom
Cito.Generic.Hotspot.DC	Hotspot item twee kolommen	Hotspot	46	Hotspot item twee kolommen
Cito.Generic.Hotspot.SC	Hotspot item een kolom	Hotspot	51	Hotspot item een kolom
Cito.Generic.Hottext.Corrections.DC	Hottext item twee kolommen	Kort antwoord	278	Hottext item met correctievakken twee kolommen
Cito.Generic.Hottext.DC	Hottext item twee kolommen	Meerkeuze	941	Hottext item twee kolommen
Cito.Generic.Matrix.DC	Matrix twee kolommen	Meerkeuze	729	Matrix item met een opmaak in twee kolommen
Cito.Generic.Matrix.SC	Matrix een kolom	Meerkeuze	22	Matrix item met een opmaak in 7n kolom
Cito.Generic.Order.DC	Volgorde item, twee kolommen	Volgorde	550	Volgorde item, twee kolommen
Cito.Generic.Order.SC	Volgorde item, een kolom	Volgorde	39	Volgorde item, een kolom

Figuur 2-2. De verschillende opgavetypen beschikbaar als templates in de opgevebank van de auteursomgeving

DTT Nederlands schrijfvaardigheid havo-vwo voorbeeldopgaven **Vraag 7 van 15**

I van 3 alinea's geplaatst.

Niveau: havo

Het einde van de gloeilamp

Ruim honderd jaar lang werden alle huizen en kantoren in Nederland verlicht met behulp van de gloeilamp. De lampen waren niet duur en waren gemakkelijk in gebruik. Ging er een lamp kapot, dan draaide je er gewoon een nieuw peertje in. Aan het tijdperk van de gloeilamp is echter een einde gekomen.

Sinds 1 september 2011 mogen winkels geen heldere gloeilampen van 60 watt of meer inkopen. De handel in matte gloeilampen is vanaf 1 september 2009 al verboden. Hierdoor zijn er steeds minder gloeilampen in de winkel te koop. En sinds 1 september 2012 mogen winkels hun bestaande voorraden nog wel verkopen, maar kunnen ze geen nieuwe gloeilampen meer inkopen. Vanaf dat moment is in de hele Europese Unie de handel in gloeilampen verboden. De gloeilamp is natuurlijk een grote slokop waar het om energie gaat. Maar 8% van de stroom die een gloeilamp krijgt, wordt omgezet in licht; de rest wordt omgezet in warmte. Bij spaarlampen en ledlampen krijg je veel meer licht voor de stroom die je erin stopt. Toch zijn veel mensen ook kritisch over de spaarlamp: daar zit kwik in en de lampen gaan veel minder lang mee dan wordt beloofd. Met al die weggegooid spaarlampen komt er dus veel van dat gevaarlijke, giftige kwik in ons milieu. Dat nadeel had die goede oude gloeilamp dan weer niet...

Je schrijft een nieuwsbericht over het einde van de gloeilamp.

Verdeel de tekst in vier alinea's. Klik de zinnen aan waarmee de nieuwe alinea's moeten beginnen.

1 2 3 4 5 6 **7** 8 9 10 11 12 ...

Inleveren Volgende >

Figuur 2-3. Voorbeeld van een alinea-opgave

DTT Nederlands schrijfvaardigheid havo-vwo voorbeeldopgaven **Vraag 13 van 15**

Niveau: havo

De selectie

Ik vind *De selectie* een interessant en spannend boek dat veel onverwachte gebeurtenissen bevat. Het boek is interessant, omdat er veel dingen in gebeuren die mij aanspreken. Het boek is bovendien anders dan andere boeken, omdat America, de hoofdpersoon, zich af en toe een beetje brutaal gedraagt en soms zelfs leugens **1. verteld**.

Het boek is spannend, omdat de schrijfster allerlei details geeft als er iets uitgelegd moet worden of er iets belangrijks gebeurt. Dat heeft voor veel spanning gezorgd toen ik het boek las.

Het boek zat vol onverwachte dingen. Als de hoofdpersoon met iets belangrijks bezig waren of iets leuks bedachten, dan kwam er altijd wel iets tussen, waardoor ze hun plannen moesten wijzigen.

Kortom, het boek sprak mij erg aan. Zo erg zelfs, dat ik de twee vervolgverhalen onlangs ook aan mijn boekenkast toegevoegd heb. Zodra er weer een deel verschijnt, wil ik dat ook gaan lezen. Ik heb echt mijn hart verpand aan deze serie!

Je hebt een boekverslag geschreven over het boek *De selectie* van Kiera Cass. Jouw docent geeft een deel van het verslag terug met het verzoek de spelling van de werkwoorden nog eens goed te controleren.

Klik de fout gespelde werkwoorden aan en verbeter ze.

1 verteld

... 4 5 6 **7** 8 9 **10** 11 12 **13** 14 15

Inleveren Volgende >

Figuur 2-4. Voorbeeld van een correctie-opgave

2.2.1 Overzicht van digitale diagnostische opgavetypen

Hierna volgt een tabel met alle gebruikte opgavetypen binnen de DTT met een beknopte beschrijving. In Bijlage 17.1 staan voorbeelden van elk opgavetype.

Tabel 2-2. Gebruikte opgavetypen binnen de DTT

Opgavetype	Beschrijving	Interactie
Alinea-opgave	Een opgavetype waarbij de leerling een tekst in alinea's onderverdeelt. De alinea'scheidingen worden direct zichtbaar in de tekst. Geschikt om de leerling zelf tekststructuur te laten aanbrengen.	Klikken (meerkeuze)
Categoriseer-opgave	De leerling sleept gegeven elementen (tekst, afbeeldingen) in vakken per categorie. Geschikt om elementen te sorteren, denk bijvoorbeeld aan formeel/informeel bij tekstsoorten.	Slepen
Combinatie-opgave	Een opgavetype waarbij meerdere soorten interacties (bijvoorbeeld een meerkeuzevraag en een invulvraag) op één scherm worden gecombineerd.	Diversen
Correctie-opgave	De leerling selecteert en verbetert onderdelen van een tekst. Geschikt om grammatica en spelling te toetsen.	Klikken en invullen
Dropdown-opgave	Een opgavetype waarbij de leerling een antwoord kiest uit een keuzelijst. De leerling ziet daarna direct het effect van zijn keuze in de tekst. Dit opgavetype wordt veel gebruikt als woordenschat-vraag waarbij de leerling het meest geschikte woord invoeft in de tekst.	Klikken (meerkeuze)
DWO-opgave	De leerling kan zelf grafiepunten of grafieken tekenen als antwoord.	Diversen
Geogebra-opgave	De leerling kan diverse wiskundige interacties uitvoeren als antwoord. Bijvoorbeeld het correcte meetkundige figuur tekenen of aanpassen.	Diversen
Hotspot-opgave	De leerling klikt het juiste antwoord aan op een afbeelding.	Klikken
Kort open-opgave	De leerling typt zelf een antwoord in. Dit opgavetype wordt bij talen gebruikt om spelling te toetsen. Bij wiskunde-opgaven kunnen de leerlingen input geven in de vorm van getallen en formules, via een speciale formule-editor.	Invullen
Markeer-opgave	De leerling selecteert actieve onderdelen van een tekst. Geschikt om antwoorden te zoeken in de tekst, bijvoorbeeld een citaat of fout gespelde woorden.	Klikken
Matrix-opgave	De leerling kiest uit (meestal) twee mogelijkheden, zoals 'wel' of 'niet'. Geschikt voor het bevragen van details.	Klikken (meerkeuze)
Meerkeuze-opgave	De leerling kiest het juiste antwoord uit een lijst van alternatieven (teksten, afbeeldingen, filmpjes, geluidsbestanden) of uit actieve gebieden in een afbeelding.	Klikken (meerkeuze)
Multiplerespons-opgave	De leerling kiest één of meer antwoorden uit een lijst van alternatieven (teksten, afbeeldingen, filmpjes, geluidsbestanden). Deze vorm komt veel voor met opgaven over internetformulieren en is in feite ook een vorm van een sorteervraag.	Klikken (meerkeuze)
Sleep-opgave	De leerling sleept een antwoord in de vorm van tekst. Geschikt om te matchen of te sorteren/categoriseren.	Slepen
Sleep-opgave afbeelding	De leerling sleept een antwoord. Geschikt om plaatjes aan woorden te koppelen.	Slepen
Volgorde-opgave	De leerling sleept tekstdelen of getallen in een nieuwe volgorde.	Slepen

2.2.2 De keuze voor een opgavetype

Zoals gezegd vormden het leerlingmodel en het taakmodel de basis voor een opgave. Vervolgens kwam de afweging welk opgavetype het meest geschikt was. Enkele afwegingen bij het kiezen van het meest geschikte opgavetype:

- Kies de interactievorm die de leerling in staat stelt om zo authentiek, direct en efficiënt mogelijk de te toetsen (deel)vaardigheid te tonen.

- Maak een afweging ten aanzien van efficiëntie bij complexere interacties zoals de volgorde-opgave: neemt deze niet te veel afnametijd in beslag in relatie tot de informatie die de opgave oplevert?
- Er zijn diverse opgavetypen mogelijk waarbij de leerlingen het effect van de interactie direct zien in de tekst. Bijvoorbeeld door een antwoord uit een keuzelijst in te vullen, er een antwoord in te slepen of zelf een antwoord in te typen. Het is authentieker wanneer het effect te zien is, maar dit is vaak ook makkelijker.
- Het opgavetype kort open-opgave heeft als specifiek kenmerk dat de leerling zelf het antwoord invult; dat is over het algemeen moeilijker dan een woord uit een lijst kiezen of slepen.
- Zorg voor een goede balans tussen afwisseling voor de leerling versus eenheid van stijl en interactie.
- De DTT is uniek in de wijze waarop de opgaven automatisch 'beoordeeld' worden. Tegelijkertijd nadenken over de opgave én de 'beoordeling' bevordert de consistentie van de opgaven.

De toetsdeskundigen hebben zoveel mogelijk per opgave bekeken welke vorm het meest authentiek en geschikt was. Uit de eerste try-out en pretest zijn een paar ervaringsgegevens naar voren gekomen (Cito, 2013a; Cito, 2014; Cito, 2015; Cito, 2016):

- Als de leerling (tekst)elementen op volgorde moet zetten zoals bij chronologische gebeurtenissen (tekstbegrip) of getallen van laag naar hoog, dan is een volgorde-opgave een goede keus. Hoewel een dergelijke opgave wat meer tijd kost, is dit een directe manier van bevragen.
- Als de leerling (tekst)elementen moet categoriseren (formeel/informeel taalgebruik of wel/geen priemgetallen) dan is een categoriseer-opgave een goede keus.
- Een aspect als spelling kan via vrije tekstvakken bevraagd worden of via het opgavetype waarbij de leerling een tekstfout kan verbeteren in de tekst zelf (Correctie-opgave).
- Een rekenopgave waarbij de leerling zelf het antwoord moet intypen, is moeilijker dan een meerkeuzevraag.
- Als een leerling inzicht in tekststructuur moet laten zien, kies dan bijvoorbeeld de alinea-opgave waarbij de leerling een tekst zelf in alinea's verdeelt.
- Ook het ontwerp van de opgave kan bijdragen aan de authenticiteit: als de leerling een formulier moet invullen, zorg dan dat de opgave eruitziet als een formulier.

Opgavekenmerken en codering

Bij het construeren van een opgave was het niet alleen van belang om na te denken over het te bevragen construct en de beoordeling daarvan. Middels coderen moest een koppeling gelegd worden tussen het juiste (of bijna juiste) antwoord en het betreffende aspect van het leerlingmodel. Bijvoorbeeld, bij een meerkeuze-opgave moest aangegeven kunnen worden dat het juiste antwoord D informatie geeft over de beheersing van deelaspect 2.4 uit het leerlingmodel Nederlands Schrijven. De mogelijkheid tot het coderen van opgaven is speciaal ontwikkeld voor DTT. In het volgende hoofdstuk (3) wordt de codering uitgebreid besproken.

Bij elke opgave moest tevens worden aangegeven:

- Het aspect van het leerlingmodel dat bevraagd wordt;
- De leerweg/schoolsoort waarvoor deze bedoeld is;
- Welke tussendoelen met de opgave worden afgedekt;
- Bij wiskunde werd het soort scoring aangegeven.

Daarnaast was een aantal kenmerken van belang vanuit het oogpunt van opgavebankbeheer, zoals in welke pretest de opgave opgenomen is geweest en wanneer deze voor het eerst in de adaptieve toets kwam.

2.3 Tot slot

Het ontwikkelen van grote hoeveelheden diagnostische innovatieve opgavetypen vergde veel creativiteit en flexibiliteit van de toetsdeskundigen, zeker ook gezien de krappe ontwikkelplanning. Het is desalniettemin gelukt om afwisselende toetsen te ontwikkelen die resulteren in een authentieke bevraging én een betrouwbaar resultaat blijken op te leveren. In de volgende hoofdstukken wordt de ontwikkeling per vak besproken.

2.4 Literatuur

Cheng, I., & Basu, A. (2006). Improving multimedia innovative item types for computer based testing. In *Eighth IEEE International Symposium on Multimedia*, 557-566. doi:10.1109/ism.2006.92

Cito (2013a). *Diagnostische tussentijdse toets: Verslag try-out 2013*. Arnhem: Cito.

Cito (2014). *Diagnostische tussentijdse toets: Verslag try-out 2014*. Arnhem: Cito.

Cito (2015). *Diagnostische tussentijdse toets: Verslag pretest 2015*. Arnhem: Cito.

Cito (2016). *Diagnostische tussentijdse toets: Verslag pretest 2016*. Arnhem: Cito.

College voor Toetsen en Examens (2014). *Publieksversie toetswijzer diagnostische tussentijdse toets voor Nederlands, Engels en wiskunde*. Utrecht: College voor Toetsen en Examens.
<https://www.pilotdt.nl/documenten/publicaties/2014/12/15/toetswijzer-dtt>

Frey, B. B., Petersen, S., Edwards, L., Pedrotti, J. T., Peyton, V. (2005). Item-writing rules: Collective wisdom. *Teaching and Teacher Education*, 21(4), 357-364. doi:10.1016/j.tate.2005.01.008

Goosen, R. & Vernooij, S. (2017). *Innovatieve vraagvormen in de DTT schrijfvaardigheid Engels: Een onderzoek naar de geschiktheid van meer authentieke vraagvormen*. Ongepubliceerd stagerapport. Arnhem: Cito.

Jodoin, M. G. (2003). Measurement efficiency of innovative item formats in computer-based testing. *Journal of Educational Measurement*, 40(1), 1-15. doi:10.1111/j.1745-3984.2003.tb01093.x.

Roelofs, E., & Schouwstra, S. (Red.). (2012). *Diagnostische tussentijdse toets: Verslag van de voorstudie*. Arnhem: Cito.

3 De responsverwerking: coderen in plaats van scoren

Joke Hofstee

Om veel diagnoses te kunnen stellen in de gegeven toetstijd is de DTT adaptief (zie Hoofdstuk 1). Om tot betrouwbare diagnoses te kunnen komen binnen de beschikbare toetstijd is het belangrijk dat opgaven in een toets efficiënt zijn, dat wil zeggen dat een opgave maximale informatie geeft en minimale tijd vraagt om te beantwoorden. Daarnaast is het voor de validiteit belangrijk dat opgaven authentiek zijn, dat wil zeggen dat de taak die de leerling moet uitvoeren om de opgave te beantwoorden zo dicht mogelijk ligt bij de vaardigheid die de toets beoogt te meten en zo weinig mogelijk niet relevante vaardigheden meet. Een klein voorbeeld: het invoegen van witregels in een tekst om alinea's te onderscheiden is aanmerkelijk authentiek dan het intypen van regelnummers waartussen een alineascheiding geplaatst zou moeten worden. Om een opgavebank te creëren met zo efficiënt en authentiek mogelijke opgaven was het noodzakelijk om nieuwe opgavetypen te ontwikkelen, zoals beschreven in het vorige hoofdstuk (zie Hoofdstuk 2).

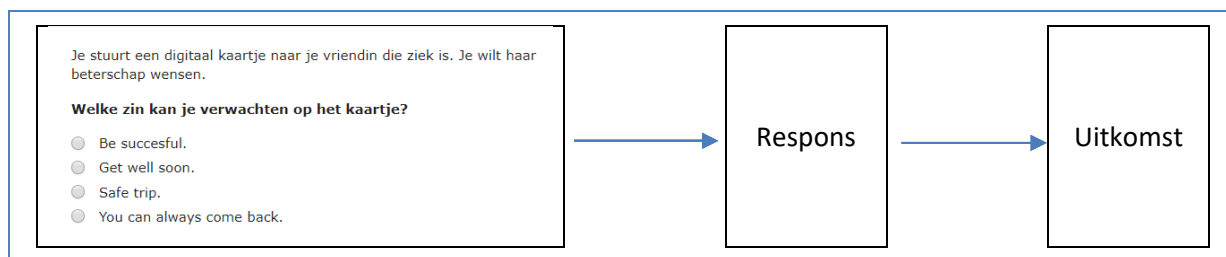
Naast de inhoud van de opgaven was er ook een noodzaak om een nieuwe manier van responsverwerking van de opgaven te implementeren. Een opgave kan bestaan uit meerdere onderdelen (interacties), die allemaal hun eigen koppeling aan de aspecten uit het leerlingmodel kunnen hebben. Per onderdeel van een opgave moet aangegeven worden welke antwoorden correct zijn en hoe die antwoorden gekoppeld zijn aan het leerlingmodel. Deze koppeling van antwoorden aan de verschillende aspecten uit het leerlingmodel worden bij de DTT coderingen genoemd. Een verschil met scoren is dat de coderingen niet simpelweg opgeteld worden om tot een eindscore of somscore per deelaspect of hoofdaspect te komen, maar gebruikt worden om de kans te berekenen dat een leerling per aspect onder, op of boven niveau zit (zie verder in Hoofdstuk 12). In deze berekening worden de antwoorden op de opgavenonderdelen niet per definitie samengenomen, maar ze kunnen gebruikt worden als separate observaties. Hiermee won de DTT aan efficiëntie.

Hieronder wordt verder toegelicht hoe deze wijze van coderen is opgebouwd en is vormgegeven in de auteursomgeving van de opgaven en welke invloed deze coderingen hebben op de werkwijze bij het construeren van nieuwe opgaven.

3.1 Begripsdefinitie: items en responsen

Zowel in de literatuur over toetsen als in de psychometrische literatuur wordt het begrip 'item' gebruikt, alhoewel niet altijd hetzelfde verstaan wordt onder dit begrip. Ook in de QTI-standaard ('Question & Test Interoperability', zie Hoofdstuk 9) die gehanteerd wordt in de auteursomgeving en de afnamespeler, Facet, wordt het begrip 'item' gehanteerd.

Een eenvoudige meerkeuze-opgave wordt door toetsdeskundigen, psychometrici en in de auteursomgeving een item genoemd. Zo is bij het item in Figuur 3-1 één interactie mogelijk, namelijk de keuze van een alternatief. Na het beantwoorden van de opgave is er één respons, die wordt beoordeeld als correct (1) of incorrect (0): één uitkomst. Bij meer complexe opgaven met meerdere onderdelen die tot meerdere responsen en tot één of meerdere uitkomsten kunnen leiden is het begrip 'item' niet meer zo eenduidig.



Figuur 3-1. Een item: een meerkeuze-opgave

Bijvoorbeeld, in Figuur 3-2 staat een opgave met drie onderdelen. Dit is in de QTI-standaard een item met drie interacties. QTI definieert een item als een groep interacties¹ die samen wordt weergegeven. In het voorbeeld in Figuur 3-2 kan elk respons onafhankelijk van de andere responsen gegeven worden en onafhankelijk van de andere responsen goed of fout zijn en beoordeeld worden. Er is voor psychometrici dan sprake van drie itemresponsen.

In het vervolg van dit hoofdstuk worden de term 'opgave' gebruikt voor de taak die op een scherm wordt weergegeven en een inhoudelijke eenheid is.

In deze tekst ontbreken enkele leestekens.

Kies op de open plekken de formulering met de juiste leestekens.

Driekoningen

In delen van Nederland vieren de kinderen op 11 november Sint-Maarten, maar in Duitsland gebeurt dat de kinderen lopen daar niet met een lampion langs maar vieren er op 6 januari wel het feest Driekoningen. De kinderen lopen dan verkleed als drie koningen van huis tot huis en spreken de volgende tekst "Christus Mansionem Benedicat". Ze doen dit om het huis te zegenen.

Figuur 3-2. Illustratie itemdefinities in QTI: een item (blauw omlijnd) met drie interacties (rood omlijnd)

3.2 Groeperen van interacties in de auteursomgeving

De auteursomgeving in Questify is voor de DTT aangepast zodat responsen zowel onafhankelijk als per groep beoordeeld en gekoppeld kunnen worden aan één of meerdere hoofd- en deelaspecten.

De constructeur moet bij elke opgave bepalen of responsen onafhankelijk moeten worden beoordeeld of dat ze als geheel beoordeeld moeten worden.

Responsen met een onderlinge afhankelijkheid moeten over het algemeen als geheel worden beoordeeld, denk bijvoorbeeld aan een opgave waarin zowel een getal als een schaal moet worden gekozen; het getal (bijvoorbeeld 100) kan dan niet onafhankelijk worden beoordeeld van de gekozen schaal (bijvoorbeeld 'cm'). De score-editor van de auteursomgeving, zie Figuur 3-3, biedt daarom de mogelijkheid om de responsen te groeperen.

¹ De volledige namen in QTI zijn 'assessmentItem' en 'interaction'

[illegible]

Figuur 3-3. Voor de DTT aangepaste score-editor in de auteursomgeving waarin groeperen van responsen mogelijk is

Een voorbeeld waarin de afweging of responsen samen dan wel los van elkaar moeten worden beoordeeld lastiger is, is weergegeven in Figuur 3-4. Overwegingen in dit geval zijn:

- Is er een heel sterke inhoudelijke samenhang in de interacties? (niet, het zijn in feite drie aparte vragen)
- Hoe vaak kunnen de antwoordalternatieven gekozen worden: één keer of vaker? (één keer, na het geven van elke respons is er een alternatief minder)
- Hoeveel antwoordalternatieven worden aangeboden: meer dan te maken keuzes? (hier meer, elke interactie blijft een meerkeuze-opgave)

In dit geval gaat het dus om drie onafhankelijke responsen die ieder apart kunnen bijdragen aan de diagnose. Daarom moeten de responsen los van elkaar beoordeeld worden. Elk van de drie responsen moet in Questify vervolgens verbonden worden aan het leerlingmodel.

You are writing three emails for different occasions.

Match the appropriate endings with the occasions.

You describe to your friend the wonderful place you are staying in.	
You write to your teacher to apologise for being late handing in your project.	
You write to your friend who is going backpacking in South America.	

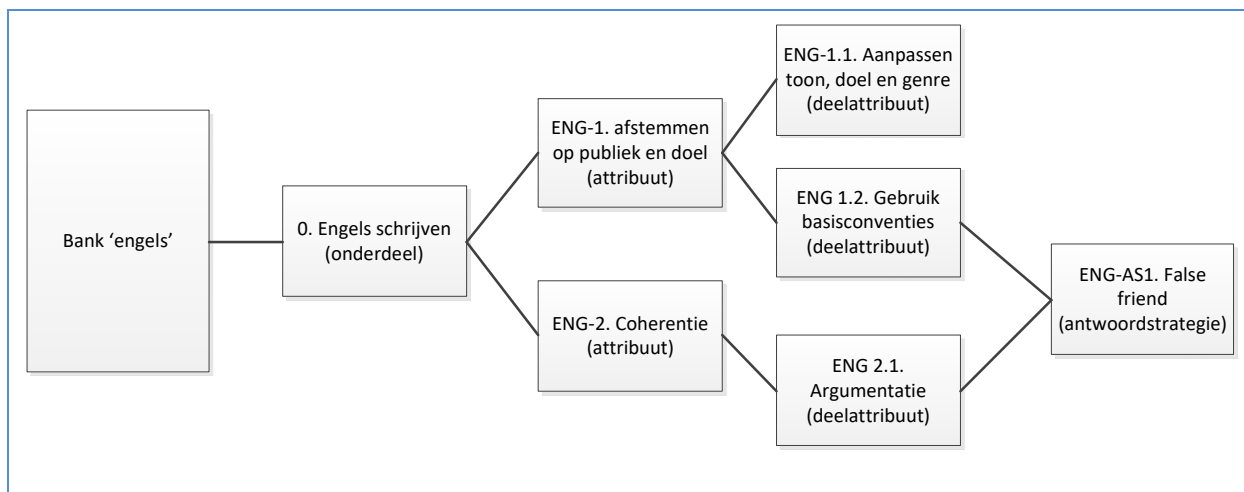
Enjoy yourself Let bygones be bygones

Thank you for understanding Wish you were here With sympathy

Figuur 3-4. Voorbeeld van een opgave met meerdere interacties

3.3 Het leerlingmodel in de auteursomgeving

De leerlingmodellen in de DTT beschrijven meerdere aspecten per vaardigheid. Deze kunnen onderverdeeld zijn in hoofdaspecten en deelaspecten met een boomstructuur, zoals hieronder te zien is in een *hypothetisch* voorbeeld voor schrijfvaardigheid Engels.



Figuur 3-5. Voorbeeld van een *hypothetische* boomstructuur voor schrijfvaardigheid Engels

Naast de hoofdaspecten en deelaspecten is het ook mogelijk om antwoordstrategieën te definiëren. In het schema in Figuur 3-5 wordt bijvoorbeeld 'False friend' aangegeven. 'False friend' is een foutieve antwoordstrategie waarbij de betekenis van een Engels woord wordt verward met een woord uit een andere taal (zoals Nederlands) dat er qua spelling of vorm veel op lijkt. Denk bijvoorbeeld aan het woord 'actual', dat verward kan worden met het Nederlandse woord 'actueel' maar waarvan de betekenis 'werkelijk' is. Antwoordstrategieën zijn in de DTT tot nu toe niet beschreven in de leerlingmodellen, maar in Questify wel mogelijk gemaakt. In de auteurs en afname-omgeving is het definiëren van een compleet leerlingmodel², zoals hierboven, de samenhang met opgaven en interacties en de verwerking in een adaptieve module mogelijk gemaakt (zie ook Hoofdstuk 9).

3.4 Koppeling van interacties aan het leerlingmodel in de auteursomgeving

In de auteursomgeving moet de constructeur een inhoudelijke koppeling maken tussen het leerlingmodel en de verschillende responsen. In de ontwikkelde coderingseditor kan aangegeven worden over welk aspect en deelaspect het antwoord informatie geeft. Bijvoorbeeld, bij een meerkeuze-opgave kan men aangeven dat het correcte antwoord B is. In de coderingseditor geeft men vervolgens aan dat het correcte antwoord informatie geeft over hoofdaspect 1 (*Afstemming op publiek en doel*) en deelaspect 1.1. (*Toonzetting en register afstemmen*), zie Figuur 3-6.

² Een leerlingmodel wordt in de auteursomgeving conceptstructuur genoemd en de codering, conceptcodering.

Score-editor

Scoringmethode : Dichotoom

Naam	Operator	Waarde
multiChoiceScoring	=	B

Coderingseditor

multiChoiceScoring

Name	B	A	C	D
☒ ENG-1	1	0	0	0
☒ ENG-1.1	1	0	0	0
☐ ENG-1.2				
☐ ENG-2				
☐ ENG-3				
☐ ENG-4				

Figuur 3-6. Codering van een meerkeuze-opgave

Bij opgaven met meerdere interacties is het coderen wat ingewikkelder. De constructeur moet bepalen welke antwoorden mogelijk zijn en wat deze indiceren. Elke interactie kan tot een separate respons leiden die correct of incorrect is, of de responsen kunnen gegroepeerd worden en tezamen als correct of incorrect beoordeeld worden. Per uitkomst moet de constructeur bepalen voor welke hoofd- en deelaspecten het antwoord van de leerling bij kan dragen aan de diagnose. Bijvoorbeeld in de volgorde-opgave in Figuur 3-7 zijn alle responsen gegroepeerd en vormen zij tezamen een uitkomst. Er is één volgorde die correct gerekend wordt (B, E, D, A, C) en alle andere volgordes worden als incorrect beoordeeld. De correcte volgorde geeft informatie over hoofdaspect 2 en deelaspect 2.1.

Voor interacties waar zeer veel antwoorden mogelijk zijn, zoals bij correctie-opgaven, geldt dat deze voor verwerking eerst gecategoriseerd moeten worden op basis van een regel. Dit kan betekenen dat een aantal responsen gegroepeerd worden en één uitkomst als correct wordt gecodeerd en alle andere uitkomsten als fout. Maar er kan ook gebruik worden gemaakt van bepaalde regels of voorbewerkingen van het antwoord, bijvoorbeeld het verwijderen van witruimte, transformeren naar kleine letters ('lowercase'), of het bepalen van wiskundige equivalentie met een *computer algebra system*. De zogenaamde coderingseditor is hiertoe specifiek voor de DTT ontwikkeld.

Het is ook mogelijk in de coderingseditor aan te geven dat meerdere uitkomsten (volgordes) correct zijn. Bijvoorbeeld in Figuur 3-8 is aangegeven dat twee verschillende volgordes correct zijn en alle andere volgordes incorrect. Het is ook mogelijk om de uitkomstenruimte niet in twee categorieën in te delen (correct en incorrect), maar in meerdere categorieën, bijvoorbeeld correct, een beetje correct en incorrect. Bijvoorbeeld in Figuur 3-9 staat dat de eerste volgorde helemaal correct is (categorie 2), de tweede volgorde een beetje correct (categorie 1), en alle andere volgordes incorrect zijn.

De taak van een constructeur is dus aanzienlijk uitgebreid. Naast het ontwikkelen van de opgaven, moet de constructeur bedenken welke antwoorden mogelijk zijn, of antwoorden voor de beoordeling bewerkt moeten worden, welke antwoorden samen beoordeeld moeten worden, welke antwoorden correct (of beetje correct) zijn, en over welke aspecten uit het leerlingmodel die correcte antwoorden iets zeggen. Om al deze taken mogelijk te maken is de standaard score-editor in de auteursomgeving aangepast.

Next week you have to present a poster with classmates.

Put the instructions for making the poster in the correct order.

Divide the subtopics and select relevant information for the presentation.

Next week you are going to give a poster presentation in groups of four.
The following steps will help you to prepare.

Put the pieces together and come up with a design for the poster.

Search for information on the Internet and make subtopics.

Start by deciding on a topic for the presentation.

Score-editor

Scoringmethode: Dichotoom

Naam	Operator	Waarde
Positie 1	=	B
Positie 2	=	E
Positie 3	=	D
Positie 4	=	A
Positie 5	=	C

Coderingseditor

orderScoring.A & orderScoring.B & orderScoring.C & orderScoring.D & orderScoring.E

Name	B&E&D&A&C	{+}&{+}&{+}&{+}&{+}	⊖ ⊕
☐ ENG-1			
☒ ENG-2	1	0	
- ☒ ENG-2.1	1	0	
- ☐ ENG-2.2			
☐ ENG-3			
☐ ENG-4			

Figuur 3-7. Een volgorde-opgave waarbij één volgorde correct is

Score-editor

Scoringmethode: Dichotoom

Naam	Operator	Waarde		Naam	Operator	Waarde
Positie 1	=	B	of	Positie 1	=	B
Positie 2	=	E		Positie 2	=	E
Positie 3	=	D		Positie 3	=	D
Positie 4	=	A		Positie 4	=	C
Positie 5	=	C		Positie 5	=	A

Coderingseditor

orderScoring.A & orderScoring.B & orderScoring.C & orderScoring.D & orderScoring.E

Name	B&E&D&A&C	B&E&D&C&A	{+}&{+}&{+}&{+}&{+}&{+}
☐ ENG-1			
☒ ENG-2	1	1	0
☒ ENG-2.1	1	1	0
☐ ENG-2.2			
☐ ENG-3			
☐ ENG-4			

Figuur 3-8. Codering van een volgorde-opgave waarbij meerdere volgordes correct zijn

Score-editor

Scoringmethode: Dichotoom

Naam	Operator	Waarde		Naam	Operator	Waarde
Positie 1	=	B	of	Positie 1	=	B
Positie 2	=	E		Positie 2	=	E
Positie 3	=	D		Positie 3	=	D
Positie 4	=	A		Positie 4	=	C
Positie 5	=	C		Positie 5	=	A

Coderingseditor

orderScoring.A & orderScoring.B & orderScoring.C & orderScoring.D & orderScoring.E

Name	B&E&D&A&C	B&E&D&C&A	{+}&{+}&{+}&{+}&{+}&{+}
☐ ENG-1			
☒ ENG-2	2	1	0
☒ ENG-2.1	2	1	0
☐ ENG-2.2			
☐ ENG-3			
☐ ENG-4			

Figuur 3-9. Codering van een volgorde-opgave waarbij één volgorde correct is, een andere volgorde een beetje correct en de rest incorrect

4 DTT Nederlands

Uriël Schuurs, Kirsten van Ingen, Roelien Linthorst en Laura van Hofwegen

Bij de start van het project is verkend welke vaardigheden en deelaspecten voor een diagnostische toets voor Nederlands in de onderbouw vmbo en havo/vwo het eerst in aanmerking zouden komen. Besloten is te starten met het verkennen van de mogelijkheden van een diagnostische schrijfvaardigheidstoets. De belangrijkste reden dat niet voor leesvaardigheid is gekozen, was dat leesvaardigheid in de praktijk de meest getoetste vaardigheid is en dat uitwerking van deze vaardigheid in het kader van de voorstudie naar verwachting weinig vernieuwende inzichten zou opleveren. Schrijfvaardigheid staat echter te boek als een onderbelichte vaardigheid in het voortgezet onderwijs en er zijn mogelijkheden voor geautomatiseerde afname en beoordeling, al liggen die in eerste instantie misschien niet voor de hand. Voor lezen en luisteren lijkt dit weinig problematisch; derhalve was er weinig reden te veronderstellen dat er bij deze vaardigheden automatiseringsproblemen overwonnen zouden moeten worden in het vervolg van het project. De vaardigheden spreken en gesprekken voeren daarentegen zijn uiterst moeilijk geautomatiseerd te toetsen binnen de kaders van het project. In het vervolg van het project zijn respectievelijk de mogelijkheden van diagnostische leesvaardigheids-, luistervaardigheids- en spreekvaardigheidstoetsing wel verkend. Alleen voor schrijven en voor lezen zijn vervolgens opgaven ontwikkeld en gepretest. In dit hoofdstuk beperken we ons daarom tot de schrijf- en de leesvaardigheid.

4.1 Schrijfvaardigheid Nederlands

4.1.1 Leerlingmodel schrijfvaardigheid Nederlands

Een belangrijk kenmerk van het voor de DTT gekozen leerlingmodel is dat hierin niet alleen aandacht wordt gegeven aan kenmerken van het schrijfproduct, maar dat ook nadruk wordt gelegd op aspecten van het schrijfproces. Uit onderzoeken van de laatste twee decennia is gebleken dat schrijfvaardigheid een complexe vaardigheid is, die - net als taalverwerving in het algemeen - geen lineaire opbouw van vaardigheden laat zien en dat een schrijver gedurende het schrijfproces voortdurend bezig is met verschillende, interacterende deelprocessen, zoals het plannen van de verdere tekst, het onder woorden brengen van zijn gedachten en het reviseren van eerder geschreven tekstfragmenten (Rijlaarsdam et al., 2011). Een bedreven schrijver doet dit bewust als onderdeel van zijn schrijfstrategieën. In het leerlingmodel voor schrijven wordt aangesloten bij het model van een onderzoeksgroep van Educational Testing Service (Princeton) dat de huidige wetenschappelijke inzichten in lees- en schrijfprocessen goed samenvat (Deane et al., 2008; Deane, 2011). Deze groep ontwikkelt sinds 2007 formatieve en summatieve toetsen voor schrijven die informatie opleveren over de kwaliteit van zowel het schrijfproduct als het schrijfproces van de leerling. Zie voor een uitgebreide beschrijving van het model het rapport van de voorstudie DTT (Roelofs & Schouwstra, 2012).

Een bijkomende consequentie van de keuze voor het model van Deane is dat de Lees- en de Schrijfvaardigheid op vergelijkbare wijze zijn verkaveld in hoofd- en deelaspecten en diagnostische uitspraken mogelijk maken op het niveau van de verschillende 'taalgebruiklagen' die Deane onderscheidt. Ook luisteren en waarschijnlijk ook spreken kunnen in ditzelfde systeem worden ondergebracht (zie de toetswijzer, CvTE, 2014; Cito, 2013b). Bij het opstellen van het leerlingmodel schrijfvaardigheid is ook gebruikgemaakt van de door SLO opgestelde tussendoelen (2012). De SLO-tussendoelen corresponderen goed met het model van Deane. SLO hanteert echter geen onderliggend leerlingmodel waarin over cognitieve activiteiten van leerlingen wordt gerept. Wel zijn er ideeën over wat een vmbo-, havo-, of vwo-leerling zou moeten kunnen. Er is – onder andere door de toetswijzercommissie – op gecontroleerd dat de hoofdaspecten tezamen de tussendoelen dekken. Omgekeerd is het niet eenvoudig om alle individuele opgaven toe te wijzen aan een individueel tussendoel.

Op basis van het model van Deane en na een raadpleging bij docenten via internet is in 2014 door de Toetswijzercommissie een leerlingmodel schrijfvaardigheid Nederlands vastgesteld dat uit vier hoofdaspecten bestaat, met daarin drie of vier deelaspecten (zie Tabel 4-1). Voor de uiteindelijke versie van het leerlingmodel dat werd vastgesteld, is op bescheiden schaal met diverse varianten ervan geëxperimenteerd. Onder andere zijn deelaspecten en mogelijke operationalisering daarvan voorgelegd

aan docenten Nederlands met de vraag te reflecteren op de herkenbaarheid, de zinnigheid en de bruikbaarheid ervan in de onderwijspraktijk. Ook de mate van uitgebreidheid was een punt van discussie: zo bestond het leerlingmodel aanvankelijk uit vijf hoofdaspecten en in totaal 23 deelaspecten (zie Roelofs & Schouwstra, 2012, p. 71). Op basis van discussies binnen de Toetswijzercommissie Nederlands en de consultatie van docenten Nederlands via internet, is de definitieve versie van dit leerlingmodel vastgesteld. Hieronder volgt een korte beschrijving van ieder hoofdaspect en de bijbehorende deelaspecten.

Tabel 4-1. Leerlingmodel schrijfvaardigheid Nederlands (CvTE, 2014)

Hoofdaspect	Deelaspect
1 Retorische vaardigheden (doel en publiek)	1.1 Voorkennis en informatiebehoefte bij de lezers inschatten 1.2 Toonzetting afstemmen op de lezer (o.a. formeel-informeel) 1.3 Schrijfdoel bepalen (bv. informeren, overtuigen, uitleggen, uitnodigen)
2 Tekststructurele vaardigheden (structuur)	2.1 Tekstelementen kiezen, rekening houdend met het genre 2.2 Passende volgorde van tekstelementen bepalen en correcte indeling en lay-out aanbrengen 2.3 Samenhang tussen tekstelementen aanbrengen (coherentie) 2.4 Een standpunt weergeven en van passende argumenten voorzien (alleen havo/vwo)
3 Linguïstische vaardigheden (woord- en zinsniveau)	3.1 Een correcte zinsbouw hanteren 3.2 Een passende en cohesieve schrijfstijl hanteren 3.3 Passend en gevarieerd woordgebruik laten zien
4 Orthografische vaardigheden (spelling en interpunctie)	4.1 Correct spellen van werkwoorden 4.2 Correct toepassen van overige regelgeleide spelling 4.3 Leestekens en hoofdletters correct hanteren

Retorische vaardigheden (doel en publiek)

Hier gaat het om de vraag in hoeverre leerlingen in staat zijn doelgerichte teksten te schrijven die zijn afgestemd op het gevraagde publiek. De opgaven zijn gericht op de keuzes die schrijvers moeten maken. Onder het eerste deelaspect vallen opgaven waarbij ingeschat moet worden welke informatie wel en niet in de tekst moet worden opgenomen. Het tweede deelaspect gaat over het kiezen en (consequent) hanteren van de juiste toonzetting, waaronder het verschil tussen formeel en informeel taalgebruik. Opgaven bij het derde deelaspect bevragen de keuze voor en het gebruiken van een passend schrijfdoel (o.a. informeren, overtuigen, uitleggen, uitnodigen).

Tekststructurele vaardigheden (structuur)

Hierbij gaat het om de vraag in hoeverre leerlingen hun teksten kunnen opbouwen tot een logisch, coherent geheel. De opgaven zijn gericht op keuzes die schrijvers moeten maken bij het structureren van hun teksten. Het eerste deelaspect betreft het kiezen van juiste tekstelementen binnen het genre van de tekst. Onder het tweede deelaspect vallen opgaven waarbij een logische volgorde moet worden aangebracht in de tekstelementen en het toepassen van de bijbehorende lay-out-technische aspecten. Voor havo en vwo gaat het daarbij ook om het plaatsen van tussenkopjes. Het derde deelaspect bevat opgaven waarin het aanbrengen van inhoudelijke samenhang tussen de tekstelementen centraal staat, bijvoorbeeld via inhoudswoorden en woordvelden. Het vierde deelaspect bevat opgaven waarbij een standpunt met passende argumenten gegeven moet worden (alleen voor havo en vwo), waaronder opgaven waarbij argumenten voor en tegen een standpunt gegeven moeten worden.

Linguïstische vaardigheden (woord- en zinsniveau)

Hier gaat het om het formuleren van woorden en zinnen. Onder het eerste deelaspect vallen opgaven waarin foute zinsbouw herkend en eventueel verbeterd moet worden. Het tweede deelaspect gaat over het gebruiken van de juiste woorden om de samenhang in een gegeven tekst te bevorderen, zoals voeg- en verwijswaarden en synoniemen en hyponiemen die de lexicale samenhang bevorderen. Bij opgaven bij het derde deelaspect wordt getoetst of een schrijver in een gegeven context het juiste woord weet te kiezen, waarbij bij havo/vwo ook sprake kan zijn van figuurlijk taalgebruik.

Orthografische vaardigheden (spelling en interpunctie)

Het betreft hier de vraag in hoeverre leerlingen correct spellen en interpuncteren. Het eerste deelaspect bevat opgaven over de spelling van werkwoorden, zoals de persoonsvorm, de infinitief, het voltooid deelwoord van regelmatige en onregelmatige werkwoorden en bijvoeglijk gebruikte werkwoorden. Onder het tweede deelaspect vallen opgaven over de spelling van regelgeleide woorden, waarbij het gaat om de spellingkwesties die in het referentiekader taal genoemd worden, zoals meervoudsvorming, verkleinwoorden, achtervoegsels en de tussen-n. Opgaven horende bij het derde deelaspect toetsen het correct gebruik van leestekens en hoofdletters.

Zoals opgemerkt is er in de aanloophase ook gewerkt aan een leerlingmodel dat 23 deelaspecten omvatte (Roelofs & Schouwstra, 2012). Nadere bestudering leerde dat in die variant een aantal deelaspecten zoveel inhoudelijke overlap vertoonde, dat er indikking kon plaatsvinden. Uiteindelijk is het leerlingmodel schrijfvaardigheid beperkt tot 12 deelaspecten en voor havo/vwo 13 deelaspecten (CvTE, 2014). Deze keuze is ook gemaakt met het oog op de praktische consequentie dat voor elk deelaspect een groot aantal opgaven moest worden ontwikkeld teneinde elke deelvaardigheid apart goed te kunnen meten.

4.1.2 Operationalisatie schrijfvaardigheid Nederlands

Om de DTT schrijfvaardigheid te operationaliseren is per deelaspect een aantal opgaven geconstrueerd. In deze paragraaf gaan we eerst in op de gang van zaken bij het constructieproces; daarna bespreken we de opgaven van de verschillende deelaspecten voor schrijfvaardigheid en gaan we in op enkele bijzonderheden bij de gebruikte opgavetypen.

Het constructieproces

Voor de ontwikkeling van de DTT schrijfvaardigheid zijn er 545 opgaven (met 1900 responsen) gemaakt die in twee pretests of als zaai-opgave zijn uitgetoetst. Zoals in Hoofdstuk 2 beschreven zijn de opgaven ontwikkeld door twee constructiegroepen van docenten uit het voortgezet onderwijs (een constructiegroep voor vmbo en een constructiegroep voor havo/vwo) aan de hand van de deelaspecten. De ontwikkelde opgaven zijn met de docenten besproken en waar nodig bijgesteld, waarna een screeningronde plaatsvond waarbij elke opgave door minstens twee toetsdeskundigen werd beoordeeld en waar nodig bijgesteld. Nadat de opgaven in de opgavebank (Questify) waren geplaatst is iedere opgave voorgelegd aan de vaststellingscommissie Nederlands DTT. Elk lid van deze commissie gaf voorafgaande aan de vergadering per opgave diens indrukken door aan de Cito-medewerkers. Ter vergadering werden de unaniem goedgekeurde opgaven gefiatteerd zonder verdere bespreking. De opgaven die door de meerderheid afgekeurd waren werden ook niet uitgebreid besproken; wel werd toegelicht om welke redenen deze opgaven niet geschikt werden bevonden. De opgaven die in principe geschikt werden bevonden, maar in huidige vorm niet goed genoeg waren, werden uitgebreider besproken zodat die konden worden verbeterd en in een daaropvolgende vergadering nog eens aan de orde konden worden gesteld.

In 2013 en 2014 heeft een try-out plaatsgevonden waarin verschillende opgavetypen zijn uitgetoetst. Daarnaast zijn open schrijfopdrachten ontwikkeld, zodat de schrijfvaardigheid van leerlingen ook op een 'natuurlijkere' manier in beeld kon worden gebracht. Het uiteindelijke doel was om docentoordelen over de prestaties op deze open schrijftaken te vergelijken met die op de meer gesloten opgaven en waar mogelijk als derde variabele ook de geautomatiseerde beoordelingen van leerlingteksten in de vergelijking te betrekken. De open schrijftaken worden, inclusief de beoordelaarsproblematiek die deze met zich brengen, in Hoofdstuk 6 uitgebreider besproken.

Op basis van de ervaringen uit deze try-outs zijn bepaalde opgaven doorontwikkeld en zijn nieuwe opgaven geconstrueerd. In Tabel 4-2 is het totale aantal goedgekeurde opgaven en responsen te zien per aspect per leerweg/schoolsoort. Na de adaptieve afname van 2017 (na twee pretesten en de eerste keer zaaien) zijn er in totaal 476 opgaven (1596 responsen) goedgekeurd voor afname. Gemiddeld zijn er per deelaspect in een leerweg/schoolsoort 29 opgaven die samen 52 responsen opleveren. Een aantal opgaven is in meerdere leerwegen/schoolsoorten afgenomen. Van alle opgaven is 50% (42% van de responsen) gericht op één leerweg/schoolsoort, de andere helft van de opgaven zijn geschikt voor twee of drie leerwegen/schoolsoorten. In de laatste kolom van Tabel 4-2 staan alleen de aantallen unieke opgaven (zonder overlap) vermeld.

Tabel 4-2. Aantal goedgekeurde opgaven en responsen na de adaptieve afname 2017 per deelaspect van het leerlingmodel Schrijfvaardigheid Nederlands

	vmbo-bb	vmbo-kb	vmbo-gt	havo	vwo	Totaal uniek
	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen
NED-1.1	9 (54)	8 (43)	9 (44)	6 (37)	3 (17)	22 (123)
NED-1.2	11 (33)	11 (37)	14 (47)	14 (31)	15 (29)	41 (111)
NED-1.3	13 (50)	17 (72)	11 (51)	14 (25)	18 (31)	43 (118)
NED-2.1	15 (40)	13 (33)	15 (32)	16 (24)	16 (32)	51 (110)
NED-2.2	12 (12)	15 (15)	19 (19)	17 (32)	21 (37)	65 (84)
NED-2.3	12 (50)	13 (50)	14 (53)	9 (28)	7 (30)	30 (110)
NED-2.4				16 (43)	15 (24)	24 (54)
NED-3.1	10 (29)	15 (52)	13 (44)	15 (67)	16 (71)	42 (149)
NED-3.2	15 (56)	15 (62)	10 (48)	9 (30)	11 (32)	42 (161)
NED-3.3	12 (46)	12 (43)	13 (40)	12 (61)	13 (41)	39 (150)
NED-4.1	10 (51)	12 (64)	12 (62)	8 (49)	8 (49)	23 (124)
NED-4.2	13 (72)	14 (79)	14 (73)	12 (70)	11 (64)	31 (172)
NED-4.3	11 (57)	8 (47)	8 (55)	8 (42)	6 (30)	23 (130)
Totaal	143 (550)	153 (597)	152 (568)	156 (539)	160 (487)	476 (1596)

Noot. Opgaven die op meerdere deelaspecten laden worden slechts bij een deelaspect vermeld in de tabel.

Opgavetypen

De 13 deelaspecten binnen schrijfvaardigheid worden bevraagd door middel van een groot aantal verschillende opgavetypen zoals dropdown-opgaven, sleep-opgaven, kort open-opgaven en correctie-opgaven. Bij elke opgave is er nagedacht over de aard van de vaardigheid die opgeroepen moet worden en welk type interactie hier het dichtst bij in de buurt komt. Zo past een kort open-opgave bijvoorbeeld heel goed bij opgaven die vragen naar spelling en komt een sleep-opgave dicht in de buurt van opgaven over structuur en woordvolgorde.

Bepaalde opgavetypen leenden zich meer voor gebruik bij bepaalde deelaspecten dan andere, omdat ze de mentale activiteiten tijdens het schrijfproces beter lijken te representeren. Bijvoorbeeld hebben we bij deelaspect 3.3 *Passend en gevarieerd woordgebruik laten zien* vaak gebruik gemaakt van een opgavetype waarin leerlingen via een rolmenu (dropdown-opgave) het best passende woord kunnen kiezen, in de veronderstelling dat dit opgavetype goed representeert op welke manier een schrijver tijdens het formuleren van een tekst kan kiezen uit verschillende woorden of uitdrukkingen die in principe allemaal even geschikt zijn, maar waarvan er één binnen de specifieke context de voorkeur heeft. Bij deelaspect 1.1 *Voorkennis en informatiebehoefte bij de lezers inschatten* hebben we vaker gebruikgemaakt van multiplerespons-opgaven, omdat daarmee de situatie wordt nagebootst waarbij een schrijver in de planningsfase van het schrijfproces reflecteert op de vraag welke brokken informatie relevant zijn om te vermelden in de tekst, gegeven het schrijfdoel en het lezerspubliek. Op deze manier hebben we steeds zorgvuldig de opgavetypen gekozen die het best leken te passen bij een nauwgezette nabootsing van het natuurlijke schrijfproces zoals dat beschreven is in de theorieën over schrijfvaardigheid (vgl. Flower & Hayes, 1981; Bereiter & Scardamalia, 1987; Deane 2011).

Om bepaalde cognitieve deelactiviteiten zo goed mogelijk te benaderen, zijn er op ons verzoek ook specifieke nieuwe opgavetypen ontwikkeld (zie ook Hoofdstuk 2), zoals de alineagave waarbij de leerling met de muis in een gegeven tekst kan klikken om aan te geven op welke manier de tekst in alinea's moet worden verdeeld. Bij iedere muisklik in de tekst verschijnt op de aangegeven plek een nieuwe alineagrens, op dezelfde manier als in een normale tekstverwerker. Om vergissingen te kunnen herstellen kan de leerling op dezelfde plek klikken, waarna de nieuwe alineagrens weer wordt opgeheven. Een ander opgavetype dat specifiek voor dit project is ontwikkeld, betreft de correctie-opgave. Daarbij kan de leerling met een muisklik aangeven waar een fout in een formulering zit; daarna opent zich een klein invulveld waarin de leerling de verbetering kan ingeven. Dit opgavetype is met name ingezet bij deelaspecten 3.3 *Passend en gevarieerd woordgebruik laten zien*, 4.1 *Correct spellen van werkwoorden* en 4.2 *Correct toepassen van overige regelgeleide spelling*.

Overigens blijven er nog wensen op het gebied van opgavetypen. Bij de constructiegroepen en toetsdeskundigen is met name behoefte aan een opgavetype waarbij een leerling kan laten zien in staat te zijn bepaalde zinsbouwproblemen te detecteren en deze op meer dan één manier op te lossen. Dat is bijvoorbeeld aan de orde bij de zogenoemde incongruentiefouten, waarbij onderwerp en persoonsvorm niet hetzelfde getal hebben: dit soort fouten kan worden verbeterd door ofwel het onderwerp van enkelvoud naar meervoud om te zetten (of omgekeerd) of door de persoonsvorm om te zetten van meervoud naar enkelvoud (of omgekeerd); beide mogelijkheden leiden tot een juist antwoord. Iets soortgelijks geldt voor fouten als gevolg van onjuist gebruik van verwijzwoorden. Fouten met verwijzwoorden en met getalscongruentie blijken veelvuldig voor te komen in leerlingteksten (vgl. Dirksen, Schellens, & Schuurs, 1987). In het kader van het nu ontwikkelde project bleek het te ambitieus om een opgavetype te ontwikkelen waarmee deze deelcompetentie in beeld gebracht kan worden.

In Tabel 4-3 zijn de totale aantallen goedgekeurde opgaven uitgesplitst naar opgavetype. Hierin is te zien dat de markeer-opgave, meerkeuze-opgave, dropdown-opgave en sleep-opgave de meest gebruikte opgavetypes zijn.

Tabel 4-3. Aantal goedgekeurde opgaven en responsen na de adaptieve afname 2017 per opgavetype

	vmbo-bb	vmbo-kb	vmbo-gt	havo	vwo	Totaal uniek
	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen
Alinea-opgave	2 (2)	3 (3)	4 (4)	3 (6)	3 (6)	11 (14)
Correctie-opgave	6 (30)	7 (34)	9 (46)	19 (121)	21 (121)	34 (191)
Dropdown-opgave	27 (94)	25 (95)	25 (97)	15 (47)	13 (40)	70 (243)
Hotspot-opgave	0 (0)	0 (0)	0 (0)	2 (10)	2 (10)	2 (10)
Kort open-opgave	18 (95)	23 (126)	19 (106)	10 (56)	8 (47)	39 (212)
Markeer-opgave	23 (74)	28 (86)	28 (86)	37 (111)	35 (102)	91 (261)
Matrix-opgave	17 (107)	18 (108)	13 (78)	9 (65)	6 (34)	38 (241)
Meerkeuze-opgave	15 (15)	17 (17)	18 (18)	37 (37)	43 (43)	88 (88)
Multiple respons-opgave	7 (40)	7 (41)	8 (47)	7 (47)	5 (32)	23 (139)
Sleep-opgave	20 (85)	18 (80)	16 (74)	11 (33)	18 (46)	48 (165)
Volgorde-opgave	8 (8)	7 (7)	12 (12)	6 (6)	6 (6)	32 (32)
Totaal	143 (550)	153 (597)	152 (568)	156 (539)	160 (487)	476 (1596)

Operationalisatie van de aspecten van schrijfvaardigheid Nederlands

Hoofdaspect 1 Retorische vaardigheden (doel en publiek)

Bij de eerste pretest hadden de vmbo-opgaven voor hoofdaspect 1 een goede, gemiddelde moeilijkheidsgraad (gemiddelde p-waarde tussen .57 en .70). De opgaven bij dit hoofdaspect waren in de tweede pretest wat eenvoudiger, maar het onderscheidend vermogen van de opgaven was goed. Wel was er in deze tweede pretest van 2016 weinig verschil in moeilijkheidsgraad tussen de afzonderlijke opgaven binnen dit hoofdaspect. Een factor die daartoe wellicht heeft bijgedragen, is de na eerste pretest gemaakte afspraak om bij vmbo het aantal op te sporen antwoorden in de vraag te vermelden (zoals in de opgave in Figuur 4-1). Bij opgaven die hoofdaspect 1 toetsen lijkt dit echter niet altijd wenselijk, omdat dit de leerlingen mogelijk te veel stuurt.

DTT Nederlands schrijfvaardigheid havo-vwo voorbeeldopgaven

Vraag 1 van 15

Niveau: havo

Je schrijft een tekst voor de website van jouw zangkoor. Met de tekst wil je nieuwe leden werven voor dit koor. De tekst mag niet te lang worden.

Welke elementen moet je in ieder geval noemen in je tekst? Klik vier elementen aan.

- ☐ aangeven hoe het lesprogramma van het koor opgebouwd is
- ☐ aangeven waar en wanneer het koor repeteert
- ☐ aangeven wie het koor opgericht heeft en wanneer dat is gebeurd
- ☐ benadrukken dat het koor bestaat uit een enthousiaste groep jongeren van 12 tot 18 jaar
- ☐ informatie geven over het type muziek dat het koor zingt
- ☐ meedelen hoe je je kunt afmelden voor een repetitie
- ☐ vertellen dat nieuwe leden geen ervaring hoeven te hebben met zingen

1 2 3 4 5 6 7 8 9 10 11 12 ...

Inleveren Volgende >

Figuur 4-1. Voorbeeld van een havo-opgave die deelaspect 1.1 toetst waarbij het aantal op te sporen antwoorden is aangegeven

Voor de havo- en vwo-leerlingen bleken bij de eerste pretest in 2015 de opgaven bij hoofdaspect 1 relatief eenvoudig (gemiddelde $p > .80$). Bij de constructie van de opgaven voor de tweede pretest in 2016 is getracht wat moeilijkere opgaven voor dit aspect te construeren. Bij deelaspect 1.2 *Toonzetting afstemmen op de lezer* en 1.3 *Schrijfdoel bepalen* is dit redelijk goed gelukt. Bij de opgaven voor deelaspect 1.1 *Voorkennis en informatiebehoefte* bij de lezers inschatten bleken de opgaven, hoewel minder dan tijdens de eerste pretest, nog steeds aan de makkelijke kant. Een mogelijke verklaring kan zijn dat havo- en vwo-leerlingen in leerjaar 3 al goed in staat zijn om te schatten welke voorkennis lezers hebben en wat hun informatiebehoefte is. Een andere verklaring zou kunnen zijn dat het lastig is bij dit deelaspect opgaven te construeren die én inhoudelijk uitdagend genoeg zijn voor havo- en vwo-leerlingen én volledige consensus opleveren bij de beoordeling ervan. Omdat de opgaven gesloten zijn vanwege de automatische scoring, moet steeds gekozen worden voor eenduidige opgaven waarbij de antwoorden onbetwistbaar goed en de afleiders onbetwistbaar fout zijn. Dit maakt dat de opgaven voor havisten en vwo'ers voor dit deelaspect wellicht aan de eenvoudige kant zijn (zie Figuur 4-2 voor een voorbeeld van een te eenvoudige opgave voor deelaspect 1.3 op havo/vwo).

Import User
dti-ne-hv-0093-SC - Picknick
Vraag 1 van 1

Samen met enkele klasgenoten organiseer jij een picknick voor het goede doel. Jullie bedenken samen wat er in de flyer moet staan om iedereen zo goed mogelijk te informeren.

Welke zinnen moeten in ieder geval in jullie flyer komen? Klik deze zinnen aan.

- ☐ Dat vinden we leuk om te doen.
- ☐ De burgemeester vindt het goed.
- ☐ De opbrengst van de drankjes gaat naar het Rode Kruis.
- ☐ De picknick vindt plaats op 15 mei vanaf 14.00 uur.
- ☐ Dit organiseren we met met een heel stel jongens en meisjes.
- ☐ Locatie: het sportveld achter de school.
- ☐ Neem zelf broodjes mee, wij verkopen drankjes.
- ☐ We hebben tijd om zo'n picknick te organiseren.
- ☐ We weten nog niet of het dan droog blijft.

Figuur 4-2. Voorbeeld van een havo-opgave voor deelaspect 1.1 die te eenvoudig bleek en afgekeurd is

Hoofdaspect 2 Tekststructurele vaardigheden (structuur)

De moeilijkheidsgraad en het onderscheidend vermogen van de opgaven die deelaspecten 2.1 *Tekstelementen kiezen, rekening houdend met het genre* en 2.3 *Samenhang tussen tekstelementen aanbrengen* toetsen zijn goed. De opgaven voor deelaspect 2.2 *Passende volgorde van tekstelementen bepalen en correcte indeling en lay-out aanbrengen* daarentegen bleken echter relatief moeilijk. Het is duidelijk dat het in een bepaalde volgorde slepen van een aantal alinea's een complexe taak is, met name voor vmbo-leerlingen. Er zijn aanwijzingen dat het 'vastzetten' van een aantal alinea's – waardoor de leerling meer sturing krijgt – en het verminderen van het aantal te slepen alinea's effect heeft op de moeilijkheidsgraad: opgaven waarbij hiervan sprake is, zijn makkelijker gebleken dan opgaven waarbij dit niet het geval is (zie Figuur 4-3 voor een voorbeeld van een opgave die deelaspect 2.2 toetst waarbij twee alinea's zijn 'vastgezet'). Ook het genre waartoe de tekst behoort, lijkt de moeilijkheidsgraad te beïnvloeden. Naar dit deelonderwerp is flankerend onderzoek geïnitieerd waarover in maart 2018 gerapporteerd zal worden (Schreurs & Schuurs, te verschijnen).

🔊 🔍 🗨️ ❓



Je hebt in de les techniek een instructie geschreven voor het bouwen van een vogelhuisje. Als je de tekst nog eens doorleest, zie je dat de verschillende stappen van de aanwijzingen door elkaar staan.

Sleep de stappen naar de juiste plek.

Een aantal stappen is al gegeven.

Maak na het zagen van de onderdelen in het midden van de voorwand het vlieggat. Dit is een rond gat met een doorsnee van 3 cm.

Lijm dan de voorkant tussen de zijwanden en spijker deze vast. Aan de bovenkant van de kast moeten nu alle wanden even hoog zijn, zodat het deksel goed aansluit.

Smeer als laatste alle gaatjes in de kast dicht met houtlijm.

Draai aan de onderkant in het midden van elk van de vier wanden een koperen schroefhaak. Plaats dan de bodem op deze haken.

Lijm dan als eerste de achterwand tussen de zijwanden en spijker de zijwanden vervolgens met een klein spijkertje vast aan de achterwand.

Zaag de volgende onderdelen op maat: een voor- en een achterwand, twee zijwanden, de bodem en het deksel.

Plaats het huisje op een geschikte plaats. Hang het nooit in de zon, want dan krijgen de vogels het te warm.

Figuur 4-3. Voorbeeld van een vmbo-kb opgave die deelaspect 2.2 toetst waarbij drie alinea's zijn 'vastgezet'

Bij havo/vwo bleken de opgaven voor deelaspect 2.2 *Passende volgorde van tekstelementen bepalen en correcte indeling en lay-out aanbrengen* en deelaspect 2.3 *Samenhang tussen tekstelementen aanbrengen* in 2016 flink moeilijker dan die in 2015. Hier lijken echter geen conclusies aan verbonden te kunnen worden, aangezien er voor deelaspect 2.2 in de eerste pretest maar een beperkt aantal opgaven afgenomen is en voor deelaspect 2.3 de aantallen in 2016 juist vrij laag waren.

In de resultaten van de pretest in 2015 op havo/vwo viel op dat er - tegen de verwachting in - relatief veel opgaven correct beantwoord werden bij deelaspect 2.4 *Een standpunt weergeven en van passende argumenten voorzien* (alleen getoetst bij havo/vwo). Dit bleek met name veroorzaakt te worden door eenvoudige afleiders in de meerkeuze- en multiplerespons-opgaven. In de eerste pretest werden conform het advies van de Toetswijzercommissie alleen opgaven gebruikt waarbij de leerling het standpunt in een argumentatie moest kunnen aanwijzen. Omdat deze taakstelling door de vaststellingscommissie als te eenvoudig werd gezien voor havo/vwo is het arsenaal aan opgaven uitgebreid. In de pretest van 2016 zijn - in afwijking van wat is opgenomen in de SLO-tussendoelen, maar op uitdrukkelijk verzoek van docenten en in samenspraak met de DTT vaststellingscommissie - dan ook andere aspecten van argumentatie bevestigd, zoals

- Het selecteren van de juiste standpunten en argumenten in een kort tekstje, met behulp van onder andere dropdown-opgaven;
- Het invullen van argumentatieschema's, met behulp van sleep-opgaven waarbij het centrale standpunt en de bijbehorende argumenten naar de juiste plek in een argumentatieschema moeten worden gesleept;
- Het aangeven welke argumenten al dan niet passen bij een gegeven standpunt, met behulp van matrix-opgaven.

Hoewel het aantal afgenomen opgaven in de pretest van 2016 beperkt was, lijkt geconcludeerd te kunnen worden dat met name de dropdown-opgaven en de sleep-opgaven met argumentatieschema's goede alternatieven zijn voor afname bij havo en vwo (zie onderstaand voorbeeld in Figuur 4-4 van een opgave waarmee argumentatieve vaardigheden worden getoetst).

Volgens sommige mensen is het een goed idee om contant geld af te schaffen nu er zoveel mensen zijn die pinnen, contactloos betalen of digitaal geld overmaken. Jij bent het hier helemaal niet mee eens en schrijft een tekst om anderen van je standpunt te overtuigen. Voordat je het betoog schrijft, maak je een schema waarin je je argumentatie uitwerkt.

Sleep de zinnen naar de juiste plek in het argumentatieschema.

Let op! Er staan meer zinnen dan je moet gebruiken.

Let op! De volgorde van argument 1, 2 en 3 maakt niet uit.

Standpunt:	_____
Argument 1:	_____
Argument 2:	_____
Argument 3:	_____
Tegenargument:	_____

Als er een storing is, is het fijn om contact geld bij je te hebben.

Contant geld heb je voorlopig nog wel nodig voor zaken als parkeerautomaten en collectes aan de deur.

Contant geld is natuurlijk van alle tijden en zal nooit verdwijnen.

Contant geld moet blijven bestaan.

Het is veiliger voor winkeliers om niet zo veel contant geld in de kassa te hebben.

Mensen hebben al bijna nooit meer contant geld in hun portemonnee.

Met contant geld heb je beter in de gaten hoeveel geld je daadwerkelijk uitgeeft.

Figuur 4-4. Voorbeeld van een argumentatieschema waarmee deelaspect 2.4 getoetst wordt in havo

Hoofdaspect 3 Linguïstische vaardigheden (woord- en zinsniveau)

Bij hoofdaspect 3 is de gemiddelde moeilijkheidsgraad van de opgaven goed bij alle leerwegen/schoolsoorten. Ook het gemiddeld onderscheidend vermogen is in orde. In 2015 waren opgaven bij deelaspect 3.1 *Een correcte zinsbouw hanteren* aan de moeilijke kant. Dat had bij vmbo onder meer te maken met het inzetten van een aantal opgaven voor alle leerwegen, die voor vmbo bb toch echt te moeilijk bleken. In de pretest in 2016 zijn de opgaven voor deelaspect 3.1 nog steeds wat moeilijker dan die voor de andere deelaspecten, maar het verschil is niet meer opvallend groot. Overigens blijkt het aangeven van fouten als getalsincongruentie tussen onderwerp en persoonsvorm of onjuiste/onduidelijke verwijzingen voor vmbo-leerlingen relatief moeilijk. De afspraak om het aantal op te sporen elementen in de vraag te noemen, lijkt - in tegenstelling tot bij hoofdaspect 1 - hier wel een juiste keuze: het geeft de leerlingen wat meer steun, maar leidt niet tot te gemakkelijke opgaven (zie Figuur 4-5). De gekozen vraagvormen en opgavetypen lijken een goede manier om dit deelaspect te toetsen.

DTT Nederlands schrijfvaardigheid vmbo voorbeeldopgaven

Vraag 7 van 12

Niveau: GT

De sportdag

Vorige week hadden we sportdag op school. Die was erg geslaagd. Je kon tijdens de sportdag allerlei sporten beoefenen zoals handbal, honkbal, voetbal, netbal en minivoetbal. Je zelfs kon een geweldige wandeling met een quiz maken!

Voor de brugklassers was het de eerste sportdag. **Ging het voor deze leerlingen dus vooral om samenwerking met de klas en de docenten.**

Hierdoor leerden de docenten de leerlingen ook beter kennen.

Aan het einde van de dag was er ook nog een leuke afsluiter voor alle leerlingen. Het was voor jou vast ook een sportieve en fijne dag!

Je schrijft een stukje over de sportdag voor de website van de school. Drie zinnen zijn niet goed geformuleerd.

Klik de drie zinnen aan die niet goed zijn geformuleerd.

1 2 3 4 5 6 7 8 9 10 11 12

Inleveren Volgende >

Figuur 4-5. Voorbeeld van een vmbo-gt opgave die 3.1 toetst waarbij het aantal op te sporen elementen in de vraag wordt genoemd

Opvallend was dat ook de havo- en vwo-leerlingen veel moeite hadden met opgaven voor deelaspect 3.1 *Een correcte zinsbouw hanteren* tijdens de pretest in 2015. In die schoolsoorten werd dit deelaspect met name getoetst met behulp van opgaven waarbij leerlingen zinsbouwfouten moesten opsporen én verbeteren, en met behulp van opgaven waarbij verschillende deelaspecten tegelijkertijd getoetst werden. Beide taken bleken voor leerlingen lastig. De meest voor de hand liggende verklaring hiervoor is dat de meeste soorten zinsbouwfouten die in opgaven geoperationaliseerd waren pas in de bovenbouw van het voortgezet onderwijs behandeld worden, zoals ook bleek uit analyse van de drie meest gebruikte lesmethodes Nederlands tijdens de voorstudie (Roelofs & Schouwstra, 2012). In de pretest van 2016 zijn vooral opgaven afgenomen waarbij leerlingen alleen gevraagd werd om fouten op te sporen. Met dit type opgave lijken leerlingen beter uit de voeten te kunnen. Fouterkenning is echter pas een eerste stap: juist bij schrijfvaardigheid gaat het ook om fouten te kunnen verbeteren en (uiteindelijk) om fouten te kunnen vermijden (vgl. Schuurs, 1990). Leerlingen lijken zich dus op het gebied van de zinsbouw nog te ontwikkelen in de bovenbouw van het voortgezet onderwijs.

De opgaven bij deelaspect 3.2 *Een passende en cohesieve schrijfstijl hanteren* leken in de pretest in 2016 ook wat eenvoudig te zijn, maar in die pretest is er slechts een beperkt aantal opgaven afgenomen, waardoor hier niet te veel conclusies aan te verbinden zijn.

Hoofdaspect 4 Orthografische vaardigheden (spelling en interpunctie)

Ook bij dit hoofdaspect zijn de gemiddelde moeilijkheidsgraad en het onderscheidend vermogen van de opgaven goed. Bij havo-vwo is het onderscheidend vermogen van deze opgaven zelfs het best van alle aspecten.

In de pretest van 2015 viel het deelaspect 4.3 *Leestekens en hoofdletters correct hanteren* op door een hoge moeilijkheidsgraad. Om deze vaardigheid op de meest directe manier te kunnen toetsen, was er voor de pretest van 2015 voor gekozen om kort open-opgaven af te nemen waarbij leerlingen zelf leestekens in een gegeven tekstje moesten plaatsen. Dit leverde echter een grotere variëteit aan antwoorden op dan waarin op voorhand was voorzien: leerlingen vulden niet alleen het beoogde leesteken in, maar herhaalden

bijvoorbeeld ook woorden of plaatsten onvoorziene spaties. Het bleek praktisch onmogelijk om al deze varianten, die in essentie wel goed waren, automatisch te evalueren. Om die reden zijn er tijdens de pretest in 2016 andere typen opgaven gebruikt om dit deelaspect te toetsen, waaronder opgaven waarbij leerlingen vanaf een zogenaamd 'leestekenplankje' leestekens naar de juiste plekken in een tekst moesten slepen of waarbij leerlingen via een dropdown-menu het juiste leesteken moesten selecteren. Deze vorm lijkt voor vmbo-gt veel geschikter (gemiddelde p-waarden tussen .66 en .79). De dropdown-opgaven blijken voor havisten en vwo'ers aan de eenvoudige kant en om die reden is één van de twee opgaven voor vwo ook afgekeurd na de pretest. Het slepen van leestekens blijkt een iets moeilijkere vraagvariant te zijn en biedt wellicht mogelijkheden voor vervolgopgaven. In Figuur 4-6 is een voorbeeld van een opgave voor leestekens opgenomen.

DTT Nederlands schrijfvaardigheid vmbo voorbeeldopgaven

Vraag 12 van 12

Niveau: KB

Lisanne heeft haar mentor Klaas geïnterviewd voor een artikel op de schoolwebsite. Er ontbreekt nog een aantal leestekens in de tekst.

Sleep steeds het juiste leesteken naar de open plek.

» Let op: je mag leestekens vaker gebruiken en je hoeft niet alle leestekens te gebruiken.

Interview met Klaas van Berkel

Klaas werkt sinds 2011 op het Viersprong College. We vragen Klaas hoe hij vroeger zelf als leerling was. Klaas antwoordt: "Ik was altijd redelijk rustig maar ik hield wel van een geintje af en toe." verder vertelt Klaas dat hij vroeger niet altijd goed oplette in de klas. Dat hadden wij nou niet van hem verwacht.

□ □ □ ?

1 2 3 4 5 6 7 8 9 10 11 12

Inleveren Volgende >

Figuur 4-6. Voorbeeld van een opgave waarbij leerlingen leestekens naar de juiste plek in de tekst moeten slepen

Na de twee pretesten en de eerste keer zaaien is in totaal 13% van de uitgeteste opgaven (16% van de responsen) afgekeurd (zie Tabel 4-4 en Tabel 4-5). Opvallend is dat bij deelaspect 2.3 *Samenhang tussen tekstelementen aanbrengen* geen enkele opgave is afgekeurd en bij 2.4 *Een standpunt weergeven en van passende argumenten voorzien* en 4.1 *Correct spellen van werkwoorden* slechts één opgave is afgekeurd (zie Tabel 4-4). Bij deelaspect 1.1 *Voorkennis en informatiebehoefte bij de lezers inschatten* zijn juist vrij veel opgaven afgekeurd, met name omdat ze te makkelijk waren (zoals eerder beschreven). We zien ook dat relatief veel markeer-opgaven zijn afgekeurd. Dit was een nieuw opgave-type, waarmee de constructiegroepen en toetsdeskundigen nog ervaring moesten krijgen.

Tabel 4-4. Aantal afgekeurde opgaven en responsen na de adaptieve afname 2017 per deelaspect van het leerlingmodel

Deelaspect	vmbo-bb Aantal opgaven en responsen	vmbo-kb Aantal opgaven en responsen	vmbo-gt Aantal opgaven en responsen	havo Aantal opgaven en responsen	vwo Aantal opgaven en responsen	Totaal uniek Aantal opgaven en responsen
NED-1.1	3 (13)	5 (30)	5 (23)	7 (48)	7 (48)	15 (91)
NED-1.2	6 (24)	3 (10)	2 (5)	3 (7)	2 (6)	9 (31)
NED-1.3	1 (3)	1 (3)	1 (3)	1 (7)	1 (7)	2 (10)
NED-2.1	4 (6)	3 (3)	3 (3)	2 (4)	1 (1)	6 (10)
NED-2.2	5 (7)	3 (3)	3 (3)	0 (0)	0 (0)	9 (11)
NED-2.3	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
NED-2.4				1 (6)	0 (0)	1 (6)
NED-3.1	2 (6)	2 (6)	2 (7)	2 (4)	2 (4)	5 (13)
NED-3.2	1 (1)	2 (6)	3 (10)	1 (1)	0 (0)	5 (16)
NED-3.3	0 (0)	0 (0)	1 (6)	2 (16)	1 (8)	3 (22)
NED-4.1	1 (4)	1 (4)	0 (0)	0 (0)	0 (0)	1 (4)
NED-4.2	2 (4)	2 (4)	2 (4)	2 (8)	2 (8)	4 (12)
NED-4.3	4 (36)	5 (47)	6 (59)	2 (19)	2 (19)	9 (81)
Totaal	29 (104)	27 (116)	28 (123)	23 (120)	18 (101)	69 (307)

Opmerking. Opgaven die op meerdere deelaspecten laden worden slechts bij een deelaspect vermeld in de tabel.

Tabel 4-5. Aantal afgekeurde opgaven en responsen na de adaptieve afname 2017 per opgavetype voor de beoogde leerwegen/schoolsoorten

	vmbo-bb Aantal opgaven en responsen	vmbo-kb Aantal opgaven en responsen	vmbo-gt Aantal opgaven en responsen	havo Aantal opgaven en responsen	vwo Aantal opgaven en responsen	Totaal uniek Aantal opgaven en responsen
Alinea-opgave	3 (5)	2 (2)	2 (2)	0 (0)	0 (0)	4 (6)
Correctie-opgave	4 (36)	4 (36)	4 (43)	3 (26)	3 (26)	8 (72)
Dropdown-opgave	3 (10)	1 (2)	1 (6)	3 (9)	2 (6)	7 (25)
Kort open-opgave	1 (4)	1 (4)	0 (0)	0 (0)	0 (0)	1 (4)
Markeer-opgave	9 (16)	10 (29)	11 (32)	7 (16)	6 (15)	20 (56)
Matrix-opgave	3 (19)	4 (25)	2 (10)	3 (27)	3 (27)	9 (64)
Meerkeuze-opgave	0 (0)	0 (0)	1 (1)	2 (2)	1 (1)	3 (3)
Multiple respons-opgave	1 (4)	1 (9)	2 (17)	5 (40)	3 (26)	8 (61)
Sleep-opgave	2 (7)	2 (7)	3 (10)	0 (0)	0 (0)	3 (10)
Volgorde-opgave	3 (3)	2 (2)	2 (2)	0 (0)	0 (0)	6 (6)
Totaal	29 (104)	27 (116)	28 (123)	23 (120)	18 (101)	69 (307)

4.2 Leesvaardigheid Nederlands

4.2.1 Leerlingmodel leesvaardigheid Nederlands

Naar analogie van het model voor schrijfvaardigheid is er een leerlingmodel opgesteld voor leesvaardigheid. Het eerste voorstel telde, net als dat voor het model voor schrijfvaardigheid, vijf hoofdaspecten. Met het oog op een noodzakelijke beperking van het aantal te toetsen deelaspecten – per onderscheiden deelaspect moet immers een voldoende aantal opgaven kunnen worden geconstrueerd en

gepretest – is het oorspronkelijke model ingedikt tot een leerlingmodel met vier hoofdaspecten en in totaal 11 deelaspecten (zie Tabel 4-6). Dit leerlingmodel is na een veldraadpleging op internet door de Toetswijzercommissie in januari 2014 vastgesteld.

Tabel 4-6. Leerlingmodel leesvaardigheid Nederlands (CvTE, 2014)

Hoofdaspect	Deelaspect
1 Retorische vaardigheden (doel en publiek)	1.1 De aard van de informatiebron, het algemene schrijfdoel en de bedoelingen van specifieke tekstdelen herkennen en interpreteren 1.2 Informatie uit teksten interpreteren en conclusies trekken 1.3 Informatie(bronnen) beoordelen op betrouwbaarheid en bruikbaarheid voor het leesdoel en op inhoudelijke kwaliteit
2 Tekststructurele vaardigheden (structuur)	2.1 Hoofd- en bijzaken onderscheiden 2.2 Opbouw en samenhang van tekstelementen herkennen 2.3 Essentiële tekstelementen herkennen en begrijpen
3 Linguïstische vaardigheden (woord- en zinsniveau)	3.1 De samenhang tussen zinnen herkennen en begrijpen 3.2 De betekenis van complexe tekstgedeeltes, formuleringen en voor de tekst relevante woorden begrijpen ³
4 Strategische vaardigheden (leesstrategieën)	4.1 Systematisch informatie opzoeken die bruikbaar is gezien het leesdoel 4.2 Uiterlijke kenmerken van een tekst gebruiken om de structuur en de gedachtegang te begrijpen 4.3 De betekenis van onbekende woorden afleiden uit delen van het woord, de zin of de tekst

Bij dit leerlingmodel zijn enkele verduidelijkingen en specificaties aangebracht door de Toetswijzercommissie. Bij deelaspect 1.1 kan het gaan om de schrijfdoelen informeren, instrueren of betogen. Deelaspect 2.2 heeft betrekking op de coherentie ofwel de *inhoudelijke* samenhang in een tekst, terwijl het bij 3.1 gaat om cohesie, dus de morfosyntactische samenhang in de tekst. Cohesie kan worden versterkt door gebruikmaking van verwijswaarden, lexicale samenhang door middel van synoniemen en hyponiemen en door voegwoorden. Cohesie wordt door professionele schrijvers doorgaans bewust aangebracht teneinde de inhoudelijke coherentie te markeren. Met 'essentiële tekstelementen' bij deelaspect 2.3 wordt bedoeld op feiten en meningen, standpunt en argumenten en op schema's, tabellen en grafieken. Bij 4.1 moet gedacht worden aan informatie opzoeken in de mediatheek, op internet en in naslagwerken en bij 4.2 gaat het om zaken als de titel, tussenkopjes, het gebruik van verschillende lettertypes en witregels.

De gedeeltelijke parallelle met het leerlingmodel voor Schrijfvaardigheid is opvallend: de eerste drie hoofdaspecten betreffen – als bij Schrijven – de retorische, de tekststructurele en de linguïstische vaardigheden. In de plaats van orthografische vaardigheden – die bij Lezen geen toetsbare vaardigheid lijken te zijn op het niveau van voortgezet onderwijs – is er in dit leerlingmodel een hoofdaspect opgenomen dat betrekking heeft op leesstrategieën.

4.2.2 Operationalisatie leesvaardigheid Nederlands

De hoofd- en deelaspecten van het leerlingmodel zijn als volgt in de DTT uitgewerkt. In de Publieksversie Toetswijzer diagnostische tussentijdse toets (CvTE, 2014) zijn voor alle deelaspecten enkele voorbeelden van opgaven opgenomen. In deze paragraaf beperken we ons daarom tot een nadere toelichting op de deeltaalvaardigheden die we beogen te meten en tot opmerkingen die betrekking hebben op manier waarop de deelaspecten zijn geoperationaliseerd.

³ Inclusief het begrijpen van figuurlijk taalgebruik (alleen voor havo/vwo)

Operationalisatie van de aspecten van leesvaardigheid Nederlands

Hoofdaspect 1 Retorische vaardigheden (doel en publiek)

Bij *Retorische vaardigheden (doel en publiek)* gaat het om de vraag in hoeverre leerlingen in staat zijn te herkennen met welk doel en voor welk publiek de tekst is geschreven, welke informatie wordt verstrekt en om de vraag in hoeverre zij de in de tekst gebruikte informatie(bron) kunnen beoordelen.

In de categorie *Retorische vaardigheden* zijn de volgende soorten opgaven ondergebracht:

- Opgaven waarin het herkennen en interpreteren van de aard van de informatiebron, het algemene schrijfdoel en de bedoelingen van specifieke tekstdelen centraal staat: herkent de lezer het soort tekst en het algemene schrijfdoel (zoals informeren, instrueren of betogen) en kan de lezer aangeven welke bedoelingen bepaalde tekstgedeeltes hebben?
- Opgaven waarin informatie uit de tekst geïnterpreteerd moet worden en conclusies moeten worden getrokken: kan de lezer specifieke informatie en meningen uit de tekst interpreteren en herkent de lezer conclusies die in de tekst staan (havo) dan wel kan de lezer zelf conclusies trekken op basis van gegevens uit de tekst (vwo)?
- Opgaven waarin het beoordelen van informatie(bronnen) op betrouwbaarheid, bruikbaarheid en inhoudelijke kwaliteit centraal staat: kan de lezer aangeven of de informatie(bron) betrouwbaar dan wel bruikbaar is met het oog op het leesdoel, en kan de lezer aangeven of de kwaliteit van de informatie(bronnen) voldoende of onvoldoende is?

Hoofdaspect 2 Tekststructurele vaardigheden (structuur)

Bij *Tekststructurele vaardigheden (structuur)* gaat het om de vraag in hoeverre leerlingen de structuur van een tekst kunnen doorgronden en inzetten bij het begrijpen en interpreteren ervan.

In de rubriek *Tekststructurele vaardigheden* zijn de volgende soorten opgaven onderbracht:

- Opgaven waarbij het onderscheiden van hoofd- en bijzaken centraal staat: kan de lezer de belangrijkste informatie in een tekst aanwijzen?
- Opgaven waarbij de opbouw en de samenhang tussen tekstelementen centraal staat: kan de lezer tekstverbanden en inhoudelijke verbanden herkennen?
- Opgaven waarbij het herkennen en begrijpen van essentiële tekstelementen centraal staat: kan de lezer feiten, meningen, standpunt en argumenten, schema's, tabellen en grafieken herkennen en begrijpen?

Hoofdaspect 3 Linguïstische vaardigheden (woord- en zinsniveau)

Bij *Linguïstische vaardigheden (woord- en zinsniveau)* gaat het om de vraag in hoeverre leerlingen zinnen en woorden begrijpen.

In de rubriek *Linguïstische vaardigheden* zijn de volgende soorten opgaven ondergebracht:

- Opgaven waarin het herkennen en begrijpen van de samenhang tussen zinnen centraal staat: kan de lezer met behulp van verwijzwoorden, voegwoorden, synoniemen en hyponiemen betekenis geven aan zinnen en hun onderlinge verbanden?
- Opgaven waarin de betekenis van complexe tekstgedeeltes, formuleringen en voor de tekst relevante woorden centraal staat: begrijpt de lezer de betekenis van complexe woorden en zinnen? Begrijpt de lezer figuurlijk taalgebruik (alleen voor havo en vwo)?

Hoofdaspect 4 Strategische vaardigheden (leesstrategieën)

Bij *Strategische vaardigheden (leesstrategieën)* gaat het om de vraag in hoeverre leerlingen gebruik kunnen maken van strategieën bij het proces van betekenisgeving.

In de categorie *Strategische vaardigheden (leesstrategieën)* zijn de volgende soorten opgaven ondergebracht:

- Opgaven waarin het systematisch opzoeken van bruikbare informatie centraal staat: kan de lezer met het oog op het leesdoel de juiste informatie selecteren, bijvoorbeeld in de mediatheek, op internet en in naslagwerken?
- Opgaven waarin het gebruikmaken van de uiterlijke kenmerken van de tekst om de structuur en de gedachtegang ervan te begrijpen centraal staat: kan de lezer onder andere de titel, tussenkopjes, het gebruik van verschillende lettertypes, witregels, illustraties, etc. gebruiken om de structuur van de tekst te doorgronden en/of de gedachtegang van de tekst te begrijpen?
- Opgaven waarin het afleiden van de betekenis van onbekende woorden centraal staat: kan de lezer de betekenis van onbekende woorden afleiden uit delen van het woord, de zin of de tekst?

Tijdens besprekingen van de Toetswijzercommissie en tijdens de opgaveconstructie bleek het onderscheid tussen de deelaspecten 1.2, 2.3, 3.2 en 4.3 lastig te operationaliseren. Om die reden worden de verschillen tussen deze deelaspecten hier nader toegelicht.

- Deelaspect 1.2 betreft interpretatie van de tekst als geheel. Opgaven bij dit deelaspect komen in essentie neer op de vraag “wat zegt de tekst nu eigenlijk?”.
- Deelaspect 2.3 focust op specifiek begrip van een aantal specifiek benoemde tekstelementen (het onderscheid tussen feit en mening, argumenten en conclusie en begrip van schema's, tabellen en grafieken).
- Deelaspect 3.2 betreft begrip van een echt complex tekstgedeelte, bijvoorbeeld van een echt moeilijke zin of van moeilijke woorden die voor de tekst echt relevant zijn.
- Deelaspect 4.3 daarentegen heeft betrekking op onbekende woorden in algemene zin: van die woorden moet de leerling de betekenis kunnen afleiden uit delen van het woord, de zin of de tekst, maar die woorden hoeven niet per se een belangrijke rol te spelen in de tekst. Onderscheidend voor deelaspect 4.3 is dus dat de leerling de woordbetekenis uit de context moet kunnen destilleren.

In de Toetswijzercommissie is vóór definitieve vaststelling van het leerlingmodel Lezen al aan de orde geweest dat enkele prototypische voorbeeldopgaven voor schrijfvaardigheid niet zouden misstaan als voorbeeldopgaven voor leesvaardigheid. Naar aanleiding daarvan is door de Toetswijzercommissie een voorstel uitgewerkt waarin Lezen en Schrijven werden gecombineerd tot één gezamenlijk domein Schriftelijke Taalvaardigheden, waarin een groot aantal deelaspecten zoals nu geformuleerd voor Lezen en voor Schrijven afzonderlijk, zouden worden gecombineerd. Daarvoor was vanuit zowel theoretisch perspectief (vgl. bv. Shanahan, 2016) als de onderwijspraktijk het nodige voor te zeggen. Omdat deze globaliserende benadering haaks staat op de doelstelling om zo specifiek mogelijk diagnostisch te toetsen is deze gedachtelijn weer verlaten. De consequentie daarvan is geweest dat gedurende de hele projectgang de operationalisaties voor leesvaardigheid en voor schrijfvaardigheid elkaar soms erg dicht naderden.

In totaal zijn 506 opgaven voor leesvaardigheid ontwikkeld (zie Tabel 4-7). Meer dan de helft van de opgaven zijn Meerkeuze-opgaven en ruim een kwart Markeer-opgaven (zie Tabel 4-8). Voor een beschrijving van de bij leesvaardigheid gebruikte opgavetypen verwijzen we naar de voorbeelden in de vorige paragraaf en in Hoofdstuk 2.

Tabel 4-7. Aantal vastgestelde opgaven en responsen per deelaspect van het leerlingmodel leesvaardigheid Nederlands

	vmbo-bb	vmbo-kb	vmbo-gt	havo	vwo	Totaal uniek
Deelaspect	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen
NED 1.1	12 (13)	12 (21)	14 (17)	16 (24)	15 (23)	54 (78)
NED 1.2	16 (36)	19 (39)	15 (28)	29 (51)	24 (31)	81 (140)
NED 1.3	5 (5)	11 (13)	3 (3)	11 (31)	10 (15)	31 (53)
NED 2.1	6 (10)	12 (19)	9 (11)	14 (28)	12 (19)	40 (66)
NED 2.2	2 (2)	5 (7)	2 (2)	16 (16)	19 (20)	32 (35)
NED 2.3	4 (5)	17 (32)	16 (36)	13 (16)	10 (16)	41 (70)
NED 3.1	10 (13)	10 (14)	3 (4)	9 (9)	11 (13)	31 (37)
NED 3.2	8 (8)	14 (14)	14 (17)	23 (24)	23 (24)	57 (61)
NED 4.1	23 (29)	14 (16)	3 (4)	20 (23)	15 (17)	59 (70)
NED 4.2	5 (5)	5 (5)	6 (6)	16 (24)	6 (12)	29 (41)
NED 4.3	7 (8)	15 (16)	19 (19)	17 (17)	11 (11)	51 (52)
Totaal	98 (134)	134 (196)	104 (147)	184 (263)	156 (201)	506 (703)

Tabel 4-8. Aantal vastgestelde opgaven en responsen leesvaardigheid Nederlands per opgavetype

	vmbo-bb	vmbo-kb	vmbo-gt	havo	vwo	Totaal uniek
	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen
Categoriseer-opgave	1 (5)	1 (5)	0 (0)	0 (0)	0 (0)	1 (5)
Dropdown-opgave	0 (0)	1 (1)	0 (0)	1 (2)	0 (0)	2 (3)
Hotspot-opgave	11 (13)	6 (8)	1 (2)	18 (30)	5 (7)	32 (47)
Kort open-opgave	0 (0)	0 (0)	0 (0)	1 (1)	2 (5)	2 (5)
Markeer-opgave	26 (28)	44 (54)	31 (37)	37 (44)	42 (52)	138 (163)
Matrix-opgave	6 (22)	4 (18)	5 (21)	10 (43)	3 (11)	22 (89)
Meerkeuze-opgave	46 (46)	66 (66)	62 (62)	105 (105)	90 (90)	273 (273)
Multiple respons-opgave	3 (10)	8 (34)	5 (25)	7 (27)	6 (16)	22 (85)
Sleep-opgave	5 (10)	4 (10)	0 (0)	4 (8)	7 (17)	13 (30)
Sleep-opgave afbeelding	0 (0)	0 (0)	0 (0)	1 (3)	1 (3)	1 (3)
Totaal	98 (134)	134 (196)	104 (147)	184 (263)	156 (201)	506 (703)

4.3 Reflecterende paragraaf

Terugkijkend mogen we stellen dat de ontwikkelingsgang van de DTT Nederlands een aantal belangrijke ontwikkelingen in gang gezet heeft. In de eerste plaats is het vakinhoudelijk van belang dat de Schrijfvaardigheid in 13 deelaspecten is opgedeeld op een manier die wetenschappelijk verantwoord is én die door veel docenten Nederlands wordt herkend als waardevol in hun vakgebied. Hetzelfde geldt voor de opdeling van de Leesvaardigheid in 12 deelaspecten, waar doorgaans maar drie niveaus worden gehanteerd door docenten (macro-, meso- en microniveau, die grofweg corresponderen met tekstueel

niveau, alineaniveau en zinsniveau). De gemaakte indelingen geven docenten de mogelijkheid om - ongeacht de didactische filosofie die wordt aangehangen en los van de gebruikte methode - op elk van de onderscheiden deelaspecten individuele feedback te geven aan de leerlingen. Formatieve evaluatie houdt niet op bij de aangereikte toetsinstrumenten, het begint juist in de klas (vgl. Wiliam, 2012) en docenten kunnen veel profijt hebben van wat in de DTT is ontwikkeld voor schrijf- en leesvaardigheid.

Daarnaast zijn er de honderden opgaven die zijn geconstrueerd en die een operationalisatie vormen van de onderscheiden deelaspecten. Deze opgaven vormen het bewijs dat de aanvankelijk op theoretische inzichten gemaakte indelingen bestaansrecht hebben: ze blijken vertaalbaar in opgaven die door docenten als zinnig zijn ervaren en blijken in pretests op het juiste niveau afgestemd. De opgaven hebben ook een goede kwaliteit. Met relatief weinig opgaven kan in de adaptieve afname een goed beeld gegeven worden van de vaardigheden van leerlingen. Dit komt omdat het onderscheidend vermogen van de opgaven zonder meer goed te noemen is, zoals blijkt uit simulatiestudies (zie Hoofdstuk 15).

Er zijn ook kwesties die zijn blijven openstaan en waar vervolgonderzoek wenselijk is. We noemen twee zaken:

In het project is een veelheid aan opgavetypen gebruikt en op verzoek van de ontwikkelaars zijn er nieuwe opgavetypen ontwikkeld. Uit interviews met scholen (CvTE, te verschijnen) kwam naar voren dat de creatieve vraagstelling werd gewaardeerd. De docenten waren te spreken over de nieuwe vormen van opgaven en volgens de docenten hebben de leerlingen de opgaven als motiverend ervaren. Tegelijkertijd kan men zich afvragen of de verscheidenheid aan opgavetypen niet een onbedoelde - negatieve - invloed uitoefent op de prestatieniveaus van de leerlingen. Het is denkbaar dat de beoogde vaardigheden veel nauwkeuriger gemeten worden wanneer de aandacht van de leerling niet voortdurend verstrooid raakt tussen inhoudelijke overwegingen en de technische vraag naar hoe in weer een nieuw opgavetype het antwoord gegeven moet worden. Onderzoek op dit terrein ontbreekt en het lijkt de moeite waard om aan de hand van hardop-denken-protocollair onderzoek na te gaan of een leerling gericht te werk kan gaan in een toets met bijvoorbeeld maar drie soorten opgaven.

De relatie tussen de diagnoses op deelaspecten en het vermogen om goed te schrijven is niet helder. Het is verleidelijk om uit te gaan van de veronderstelling dat de diagnoses boven niveau op de deelaspecten een hoge mate van schrijf- respectievelijk leesvaardigheid indiceren, maar zeker weten doen we dat niet. Er is onderzoek nodig naar de relatie tussen de diagnoses op deelaspecten in de DTT en globale oordelen van docenten over de lees- en schrijffprestaties van de leerlingen.

4.4 Literatuur

Bereiter, C., & Scardamalia, M. (1987). *The Psychology of Written Composition*. Hillsdale, NJ: Erlbaum.

Cito (2013b). *Diagnostische tussentijdse toets: Voorstudies en evaluaties 2013*. Arnhem: Cito.

College voor Toetsen en Examens (2014). *Publieksversie toetswijzer diagnostische tussentijdse toets voor Nederlands, Engels en wiskunde*. Utrecht: College voor Toetsen en Examens.
<https://www.pilotdt.nl/documenten/publicaties/2014/12/15/toetswijzer-dtt>

College voor Toetsen en Examens (te verschijnen). *Eindrapport project DTT*. Utrecht: College voor Toetsen en Examens.

Deane, P. (2011). *Writing Assessment and Cognition*. Princeton, New Jersey: Educational Testing Service.

Deane, P., Odendahl, N., Quinlan, Th., Fowles, M., Welsh, C. & Bivens-Tatum, J. (2008). *Cognitive Models of Writing: Writing Proficiency as a Complex Integrated Skill*. Research Report ETS RR-08-55. Princeton, New Jersey: Educational Testing Service.

Dirksen, A., P. J., Schellens, & U. Schuurs (1987). Vragen fouten in de zinsbouw om grammatica-onderwijs? *Spektator*, 16, p. 380-393.

- Flower, L. & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication*, 32, 365-387.
- Rijlaarsdam, G.C.W., van den Bergh, H., Couzijn, M., Janssen, T., Braaksma, M., Tillema, M., Van Steendam, E. & Raedts, M. (2011). Writing. In S. Graham, A. Bus, S. Major & L. Swanson (Eds.), *Application of Educational Psychology to Learning and Teaching*. APA Handbook Volume 3
- Roelofs, E., & Schouwstra, S. (Red.). (2012). *Diagnostische tussentijdse toets: Verslag van de voorstudie*. Arnhem: Cito.
- Schreurs, Z., & Schuurs, U. (te verschijnen). *Bevraging van tekststructuur: het effect van genre en het aantal te plaatsen elementen op de moeilijkheidsgraad*. Arnhem: Cito.
- Schuurs, U. (1990). *Leren schrijven voor lezers* (Dissertatie). Universiteit Twente.
- Shanahan, T. (2016). Relationships between Reading and Writing Development. In Ch. MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of Writing Research*. New York, The Guilford Press, 2016.
- SLO: Nationaal expertisecentrum leerplanontwikkeling (April 2012). *Concept-tussendoelen kernvakken onderbouw vo*. Geraadpleegd op <http://www.slo.nl/nieuws/00291/>.
- Sluismans, D., Joosten-ten Brinke, D., & Van der Vleuten, C. (2013). *Toetsen met leerwaarde: een reviewstudie naar de effectieve kenmerken van formatief toetsen*. Maastricht: Universiteit Maastricht.
- William, D. (2012). *Embedded Formative Assessment*. Bloomington: Solution Tree Press.

5 DTT Engels

Ingrid Williams en Wilma Vrijs

5.1 Schrijfvaardigheid Engels

5.1.1 Leerlingmodel schrijfvaardigheid Engels

Net zoals bij de DTT Nederlands vormt het model van Deane de basis voor een algemeen leerlingmodel van taalvaardigheid (Deane et al., 2008; Deane, 2011). Dit model is echter niet zonder meer toepasbaar op Engels, omdat het ontwikkeld is voor eerste taalverwerving. Daarnaast komt een aantal aspecten van schrijven op onderbouwniveau niet aan de orde. Tijdens de voorstudie (Roelofs & Schouwstra, 2012) kreeg het leerlingmodel voor schrijfvaardigheid vijf hoofdaspecten toebedeeld: *Afstemming op publiek en doel*; *Coherentie*; *Woordenschat en woordgebruik*; *Grammatica, spelling en interpunctie* en *Ondersteunende vaardigheden*. Tijdens het vaststellen van het leerlingmodel heeft de toetswijzercommissie ervoor gekozen het vijfde aspect (*Ondersteunende vaardigheden*) te laten vervallen, omdat deze vaardigheid impliciet terugkomt in andere deelaspecten⁴. Het vastgestelde leerlingmodel Schrijfvaardigheid Engels onderscheidt daarom de volgende vier hoofdaspecten: *Afstemming op publiek en doel*; *Coherentie*; *Woordenschat en woordgebruik*; *Grammatica, spelling en interpunctie*.

Ten behoeve van het identificeren van deelaspecten zijn de door SLO opgestelde concept-tussendoelen, gebaseerd op het ERK, geanalyseerd op toetsbaarheid. Bij de analyse is tevens gekeken naar het onderzoek over de koppeling van het curriculum Moderne Vreemde Talen aan het Europees Referentiekader (Van Hest, Beltman & Kleintjes, 2001). Om een zo volledig mogelijk beeld te krijgen van gebruikelijke aspecten bij schrijfvaardigheid Engels is tevens gekeken naar de algemene ‘can do statements’ van het ERK en voor de aansluiting bij het leerlingmodel van Nederlands ook naar de aspecten van commissie Meijerink (2008). Dit heeft uiteindelijk geleid tot het onderscheiden van acht deelaspecten in het leerlingmodel Schrijfvaardigheid Engels.

Gecombineerd vormen de hoofdaspecten en deelaspecten het leerlingmodel Schrijfvaardigheid Engels, dat is samengevat in Tabel 5-1. Omdat taalverwerving schoksgewijs verloopt, moet deze tabel niet als een sequentie of hiërarchie worden opgevat; van een lineaire opbouw in vaardigheden is geen sprake.

Tabel 5-1. Leerlingmodel schrijfvaardigheid Engels (CvTE, 2014)

Hoofdaspect	Deelaspect
1. Afstemming op publiek en doel	1.1 Kan toonzetting en register afstemmen op publiek en schrijfdoel. 1.2 Kan conventies behorend bij een tekstsoort gebruiken.
2. Coherentie	2.1 Kan tekststructuur en verbanden aanbrengen. 2.2 Kan bij de taak passende structuurwoorden gebruiken: gebruik van voegwoorden en verwijswwoorden.
3. Woordenschat en woordgebruik	3.1 Kan bij de taak passende woorden en woordcombinaties gebruiken. 3.2 Kan het woordgebruik functioneel variëren (inclusief als compenserende strategie).
4. Grammatica, spelling en interpunctie	4.1 Kan woordvolgorde en zinsconstructie functioneel hanteren. 4.2 Kan passende spelling en interpunctie hanteren.

Bij het hoofdaspect *Afstemming op publiek en doel* gaat het om de vraag in hoeverre leerlingen in staat zijn hun tekst zo af te stemmen op het schrijfdoel en het publiek dat de tekst communicatief effectief is: komt de beoogde boodschap op de beoogde wijze over? Het eerste deelaspect doet een beroep op het vermogen om de toonzetting en het register af te stemmen op het beoogde publiek en schrijfdoel. Het tweede

⁴ In de deelaspecten 3.2 “*Kan het woordgebruik functioneel variëren*” en 4.1 “*Kan woordvolgorde en zinsconstructie functioneel hanteren*”.

deelaspect is gericht op de conventies die bij een specifieke tekstsoort horen: kan de schrijver een brief op gebruikelijke wijze beginnen of afsluiten?

Het tweede hoofdaspect, *Coherentie*, gaat om het aanbrengen van logische en coherente structuur in teksten. Het eerste deelaspect doet een beroep op het vermogen om tekststructuur en verbanden aan te brengen. Het tweede deelaspect doet een beroep op het vermogen om structuurwoorden correct en passend te gebruiken. Denk hier bijvoorbeeld aan het gebruik van opsommingen, verwijzingen, voeg- en verwijzwoorden.

Bij het derde hoofdaspect, *Woordenschat en woordgebruik*, gaat het om de vraag in hoeverre leerlingen hun woordenschat en woordgebruik kunnen inzetten om adequaat een taak uit te voeren: is de schrijver in staat om in de context passende woorden te gebruiken? Onder het eerste deelaspect vallen opgaven die vragen om woorden, woordcombinaties en idiomatische uitdrukkingen te gebruiken in verschillende contexten. Het tweede deelaspect doet een beroep op het vermogen om woordgebruik functioneel te variëren. Dit betreft ook het gebruik van compenserende strategieën: kan de schrijver synoniemen en omschrijvingen gebruiken?

Bij het hoofdaspect *Grammatica, spelling en interpunctie* gaat het om het hanteren van zinsconstructies, woordvolgorde en spellings- en interpunctieregels. Het eerste deelaspect doet een beroep op het vermogen om woordvolgorde en zinsconstructie adequaat toe te passen. Het tweede deelaspect is gericht op correcte spelling en passende interpunctie.

5.1.2 Operationalisatie schrijfvaardigheid Engels

De opgaven zijn ontwikkeld door twee constructiegroepen bestaande uit docenten uit het voortgezet onderwijs: een vmbo-constructiegroep en een havo/vwo-constructiegroep. Aan de hand van de deelaspecten zijn opgaven gemaakt die ernaar streven alleen die bepaalde deelvaardigheid te bevragen. Alle opgaven zijn voorgelegd aan een vaststellingscommissie bestaande uit experts uit het voortgezet onderwijs (zie Hoofdstuk 2). In de pretesten zijn alleen de opgaven getoetst die uiteindelijk door de vaststellingscommissie zijn goedgekeurd. Na de pretesten zijn de opgaven opnieuw voorgelegd aan de vaststellingscommissie en heeft de commissie de opgaven goedgekeurd of afgekeurd met behulp van de statistische analyses van de pretest.

In Tabel 5-2 is het totale aantal goedgekeurde opgaven en responsen weergegeven per deelaspect per leerweg/schoolsoort. Na de adaptieve afname van 2017 (na twee pretesten en de eerste keer zaaien) zijn er in totaal 731 opgaven (1294 responsen) goedgekeurd voor afname. Gemiddeld zijn er per deelaspect in een leerweg/schoolsoort 29 opgaven die samen 52 responsen opleveren. Een aantal opgaven is in meerdere leerwegen/schoolsoorten afgenomen. Van alle opgaven is 51% (49% van de responsen) gericht op één leerweg/schoolsoort, de andere 49% van de opgaven zijn geschikt voor twee of drie leerwegen/schoolsoorten. Deze zijn in de uitsplitsing op niveau bij elke leerweg/schoolsoort te zien in het totale aantal. In de laatste kolom van het totale aantal unieke opgaven en responsen zijn deze echter slechts één keer meegeteld.

Tabel 5-2. Aantal goedgekeurde opgaven en responsen na de adaptieve afname 2017 per deelaspect van het leerlingmodel schrijfvaardigheid Engels

	vmbo-bb	vmbo-kb	vmbo-gt	havo	vwo	Totaal uniek
	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen
ENG-1.1	27 (51)	25 (62)	24 (76)	26 (68)	19 (43)	73 (171)
ENG-1.2	22 (48)	26 (56)	23 (47)	29 (76)	19 (44)	75 (178)
ENG-2.1	28 (37)	32 (42)	30 (38)	19 (30)	15 (24)	62 (88)
ENG-2.2	31 (37)	32 (49)	36 (52)	31 (54)	32 (56)	106 (159)
ENG-3.1	31 (71)	30 (80)	27 (68)	46 (86)	40 (74)	113 (235)
ENG-3.2	26 (59)	33 (67)	31 (64)	28 (40)	28 (42)	95 (164)
ENG-4.1	27 (31)	30 (51)	32 (47)	32 (51)	19 (37)	94 (142)
ENG-4.2	30 (34)	38 (54)	33 (47)	35 (53)	29 (42)	113 (157)
Totaal	222 (368)	246 (461)	236 (439)	246 (458)	201 (362)	731 (1294)

Opmerking. Opgaven die meerdere deelaspecten meten worden slechts bij één deelaspect vermeld in de tabel.

De deelaspecten worden bevraagd door middel van een groot aantal verschillende opgavetypes zoals: dropdown-opgaven, sleep-opgaven, korte open-opgaven en opgaven waarin leerlingen stukken tekst moeten selecteren en herschrijven. Omdat de DTT adaptief is, zijn alle ontwikkelde opgavetypes automatisch te beoordelen. De opgavetypes verschillen binnen de deelaspecten. Wel blijken bepaalde opgavetypes meer geschikt dan andere voor het bevragen van bepaalde deelvaardigheden. De aard van de taak bepaalt vaak of een actie authentiek aanvoelt of juist niet zo goed past bij een bepaalde gevraagde deelvaardigheid. Bij elke opgave is er nagedacht over de aard van de vaardigheid die opgeroepen moet worden en welk type interactie tot de meest directe manier van bevraging leidt. Zo past een korte openopgave bijvoorbeeld heel goed bij opgaven die vragen naar spelling en blijkt een sleepactie zeer geschikt voor opgaven over structuur en woordvolgorde. Ook zijn er typen interacties die juist het bevragen van een vaardigheid bemoeilijken. Neem bijvoorbeeld het deelaspect *Woordenschat en woordgebruik*. Dit aspect is erg moeilijk te meten in een korte open-opgave zonder hierbij de spelling van een woord te bevragen. Het is onmogelijk alle varianten in schrijfwijze op te nemen in het antwoordmodel. In zulke gevallen is er gekozen voor een opgavetype dat het bevragen van de specifieke deelvaardigheid niet tegenwerkt.

In Tabel 5-3 zijn de totaal aantal goedgekeurde opgaven uitgesplitst naar opgavetype. Hierin is te zien dat de dropdown-opgave, meerkeuze-opgave, sleep-opgave, korte open-opgave en de volgorde-opgave de meest gebruikte opgavetypes zijn.

Tabel 5-3. Aantal goedgekeurde opgaven en responsen na de adaptieve afname 2017 per opgavetype

	vmbo-bb	vmbo-kb	vmbo-gt	havo	vwo	Totaal uniek
	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen
Alinea-opgave	1 (3)	1 (3)	1 (3)	0 (0)	0 (0)	2 (6)
Categoriseer-opgave	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Combinatie-opgave	1 (2)	2 (4)	1 (2)	3 (9)	3 (8)	6 (16)
Correctie-opgave	1 (2)	0 (0)	0 (0)	0 (0)	0 (0)	1 (2)
Dropdown-opgave	55 (83)	63 (133)	71 (133)	89 (149)	73 (127)	224 (383)
Hotspot-opgave	0 (0)	0 (0)	0 (0)	2 (8)	2 (8)	2 (8)
Korte open-opgave	28 (29)	37 (48)	24 (32)	30 (37)	25 (28)	97 (119)
Markeer-opgave	6 (6)	6 (10)	2 (2)	3 (8)	4 (12)	12 (24)
Matrix-opgave	3 (18)	6 (31)	9 (45)	13 (77)	9 (57)	25 (147)
Meerkeuze-opgave	62 (62)	61 (61)	55 (56)	57 (57)	50 (50)	195 (196)
Multiplerespons-opgave	6 (29)	7 (28)	7 (27)	6 (32)	1 (7)	15 (70)
Sleep-opgave	35 (95)	36 (95)	39 (96)	29 (67)	22 (53)	98 (241)
Sleep-opgave met afbeelding	4 (19)	6 (27)	5 (21)	0 (0)	0 (0)	8 (36)
Volgorde-opgave	20 (20)	21 (21)	22 (22)	14 (14)	12 (12)	46 (46)
Totaal	222 (368)	246 (461)	236 (439)	246 (458)	201 (362)	731 (1294)

Na de twee pretesten en de eerste keer zaaien is in totaal 6% van de uitgeteste opgaven (7% van de responsen) afgekeurd (zie Tabel 5-4 en Tabel 5-5). Dit waren voornamelijk opgaven waarbij te veel leerlingen het goede antwoord gaven waardoor de opgave geen onderscheid kon maken tussen vaardige en minder vaardige leerlingen. Er waren ook enkele opgaven die juist te moeilijk waren en daardoor geen onderscheid konden maken. Opvallend is dat bij deelaspect 3.2 *Woordgebruik functioneel variëren* slechts één opgave is afgekeurd (zie Tabel 5-4), alle overige 95 opgaven zijn goedgekeurd.

Tabel 5-4. Aantal afgekeurde opgaven en responsen na de adaptieve afname 2017 per deelaspect van het leerlingmodel

	vmbo-bb	vmbo-kb	vmbo-gt	havo	vwo	Totaal uniek
	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen
ENG-1.1	1 (1)	0 (0)	0 (0)	0 (0)	1 (5)	3 (7)
ENG-1.2	2 (12)	5 (10)	4 (9)	1 (1)	1 (1)	9 (24)
ENG-2.1	0 (0)	0 (0)	2 (2)	1 (3)	0 (0)	4 (6)
ENG-2.2	2 (2)	3 (4)	2 (3)	1 (1)	1 (1)	9 (10)
ENG-3.1	2 (2)	2 (7)	3 (14)	1 (3)	2 (2)	11 (31)
ENG-3.2	0 (0)	0 (0)	1 (5)	0 (0)	0 (0)	1 (5)
ENG-4.1	2 (2)	1 (1)	1 (1)	2 (3)	4 (9)	9 (17)
ENG-4.2	0 (0)	1 (1)	1 (1)	0 (0)	1 (4)	6 (10)
Totaal	9 (19)	12 (23)	14 (35)	6 (11)	10 (22)	52 (110)

Tabel 5-5. Aantal afgekeurde opgaven en responsen na de adaptieve afname 2017 per opgavetype voor de beoogde leerwegen/schoolsoorten

	vmbo-bb	vmbo-kb	vmbo-gt	havo	vwo	Totaal uniek
	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen
Alinea-opgave	0 (0)	1 (3)	0 (0)	0 (0)	0 (0)	1 (3)
Correctie-opgave	0 (0)	0 (0)	0 (0)	1 (5)	1 (5)	1 (5)
Dropdown-opgave	2 (2)	3 (4)	3 (4)	10 (19)	4 (11)	16 (27)
Hotspot-opgave	0 (0)	1 (3)	0 (0)	0 (0)	0 (0)	1 (3)
Korte open-opgave	3 (3)	2 (2)	1 (1)	3 (4)	1 (1)	9 (10)
Markeer-opgave	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Matrix-opgave	0 (0)	0 (0)	0 (0)	1 (4)	0 (0)	1 (4)
Meerkeuze-opgave	2 (2)	2 (2)	2 (2)	4 (4)	3 (3)	10 (10)
Multiple respons-opgave	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Sleep-opgave	0 (0)	2 (2)	5 (19)	2 (2)	0 (0)	7 (21)
Sleep-opgave afbeelding	2 (12)	1 (6)	1 (7)	0 (0)	1 (6)	4 (25)
Volgorde-opgave	0 (0)	0 (0)	2 (2)	0 (0)	0 (0)	2 (2)
Totaal	9 (19)	12 (22)	14 (35)	21 (38)	10 (26)	52 (110)

5.2 Leesvaardigheid

5.2.1 Leerlingmodel leesvaardigheid Engels

Het cognitief model van Deane - dat de taakprocessen die een rol spelen bij taalverwerving beschrijft en verklaart - vormt eveneens de basis voor het leerlingmodel leesvaardigheid. Daarnaast is er aansluiting gezocht met het ERK-kader en de tussendoelen, waarin niveaus van taalcompetentie worden beschreven in toenemende complexe taaltaaksituaties. Waar mogelijk, is ervoor gekozen aan te sluiten bij de aspecten en onderliggende theoretische onderbouwing van het leerlingmodel schrijfvaardigheid. Dit heeft uiteindelijk geleid tot de volgende vier hoofdaspecten: *Doelgericht lezen*; *Tekstfuncties herkennen*; *Tekststructuur begrijpen*; *Woordenschat en Woordgebruik*. In Tabel 5-6 is een overzicht van het leerlingmodel Lezen weergegeven.

Bij het eerste hoofdaspect, *Doelgericht lezen*, gaat het om de vraag in hoeverre leerlingen in staat zijn om op basis van leesstrategieën informatie te vinden die nodig is om een bepaald leesdoel te bereiken. De deelvaardigheden die worden getoetst bij andere hoofdaspecten overlappen met deze leesstrategieën. De leerling past een bepaalde leesstrategie toe om de opgaven binnen een bepaald deelaspect te beantwoorden. Het eerste hoofdaspect bestaat uit drie deelaspecten: *globaal lezen*, *zoekend lezen* en *intensief lezen*. Bij *globaal lezen* gaat het bijvoorbeeld om opgaven die een beroep doen op het vermogen om de hoofdgedachte van een tekst af te leiden. Het tweede deelaspect, *zoekend lezen*, daarentegen, is een strategie die toegepast kan worden als er details uit een tekst gehaald moeten worden: de lezer zoekt selectief naar relevante details uit de tekst. Tenslotte gaat het bij *intensief lezen* om het vermogen om fijnere betekenismatigheden uit de tekst te kunnen halen, zoals bijvoorbeeld het onderscheiden van hoofd- en bijzaken.

Het tweede hoofdaspect, *Tekstfuncties herkennen: tekstdoel en tekstsoort herkennen*, gaat om de vraag in hoeverre leerlingen in staat zijn te herkennen met welk doel en voor welk publiek een tekst geschreven is. Dit is vervolgens onderverdeeld in de aspecten *het doel van de schrijver herkennen* en *tekstinformatie gebruiken en conclusies trekken*. Bij *het doel van de schrijver herkennen* gaat het erom of de lezer kan

bepalen of de schrijver wil informeren, overtuigen, instrueren of amuseren. Anderzijds doen de opgaven die bij het tweede deelaspect horen een beroep op het vermogen om vergelijkingen, classificaties en illustraties te gebruiken om informatie te achterhalen en conclusies te trekken over een tekst.

Bij het derde hoofdaspect, *Tekststructuur begrijpen: informatie uit teksten halen*, gaat het om de vaardigheid waarmee op basis van structuurelementen informatie uit een tekst gehaald kan worden. Dit zijn bij het eerste deelaspect opgaven die vragen de belangrijkste gedachte uit een tekst te herkennen en/of te verwoorden: *hoofdgedachte van de tekst afleiden*. Bij het tweede deelaspect gaat het om opgaven die een beroep doen op het vermogen om hoofd- en bijzaken te onderscheiden: kan de lezer belangrijke en minder belangrijke informatie binnen een tekst onderscheiden? Onder het derde deelaspect vallen opgaven die een beroep doen op het vermogen om kernzinnen te identificeren. Het vierde deelaspect bevat opgaven die de lezer vragen detailinformatie op te zoeken in een tekst: *relevante details oppikken*. In het vijfde deelaspect wordt een beroep gedaan op het vermogen om de tekststructuur te begrijpen, zoals bijvoorbeeld ordeningen in tijd of oorzaak-gevolg relaties. Het zesde deelaspect bevat opgaven die vragen naar het begrijpen van verwijzingen tussen zinnen. Tenslotte doet het laatste deelaspect een beroep op het vermogen om de betekenis van signaalwoorden te doorzien.

Het vierde hoofdaspect, *Woordenschat en woordgebruik: woorden en zinnen begrijpen*, is weer onderverdeeld in twee deelaspecten: *zinsbegrip* en *woordbegrip*. Het gaat er bij beide deelaspecten om in hoeverre leerlingen in staat zijn om zins- en woordsbegrip uit de context af te kunnen leiden. Enerzijds gaat het om het vermogen om, met gebruik van compenserende strategieën, betekenis op zinsniveau af te leiden. Anderzijds gaat het om het vermogen om de betekenis van frequente en minder frequente woorden weer te geven of af te leiden op basis van woordsoort, woordvorm of spelling in een bekende context.

Tabel 5-6. Leerlingmodel leesvaardigheid Engels (CvTE, 2014)

Hoofdaspect	Deelaspect
1. Doelgericht lezen: Toepassen van leesstrategieën (te toetsen via deelaspecten bij 2 en 3.)	1.1 Oriënterend, globaal lezen; scannen, hoofdgedachte afleiden (3.1, 3.2) 1.2 Selectief (zoekend) lezen; relevante details oppikken uit de tekst (3.4) 1.3 Intensief lezen (2.2, 3.3, 3.5)
2. Tekstfuncties herkennen: tekstdoel en -soort herkennen	2.1 Het doel van de schrijver herkennen <i>Kan het doel van de schrijver herkennen: Informeren, Overtuigen, Amuseren, Instrueren</i> 2.2 Tekstinformatie gebruiken en conclusies trekken <i>Kan naar aanleiding van tekstuele informatie conclusies trekken die uitgaan boven de letterlijke betekenis van de tekst (vergelijken, classificeren, illustreren)</i>
3. Tekststructuur begrijpen: informatie uit teksten halen	3.1 Hoofdgedachte van de tekst afleiden <i>Kan de belangrijkste gedachte uit de tekst herkennen en/of verwoorden</i> 3.2 Hoofd- en bijzaken onderscheiden <i>Kan belangrijke en minder belangrijk informatie binnen de tekst onderscheiden</i> 3.3 Kernzinnen identificeren <i>Kan belangrijke zinnen identificeren (bijv. de zin met de belangrijkste informatie uit een alinea)</i> 3.4 Relevante details oppikken <i>Kan detailinformatie opzoeken in een tekst</i> 3.5 Begrijpen van de tekststructuur <i>Kan de structuur van een tekst doorzien (ordeningen in tijd, oorzaak)</i> 3.6 Verwijzingen tussen zinnen doorzien <i>Kan de betekenis van verwijswwoorden doorzien</i> 3.7 Betekenis van signaalwoorden doorzien <i>Kan de betekenis van signaalwoorden gebruikt in zinnen doorzien</i>
4. Woordenschat en woordgebruik: woorden en zinnen begrijpen	4.1 Zinsbegrip: zinnen begrijpen in hun context <i>Kan de betekenis van zinnen, in de context van een tekst over een bekend onderwerp weergeven of afleiden</i> 4.2 Woordbegrip: woorden begrijpen in hun context <i>Kan de betekenis van (frequente én infrequente) woorden, gebruikt in een zin over een bekend onderwerp, weergeven of afleiden op basis van woordsoort/vorm en spelling</i>

5.2.2 Operationalisatie leesvaardigheid Engels

Voor leesvaardigheid zijn de opgaven ook ontwikkeld door een constructiegroep van docenten uit het voortgezet onderwijs, verdeeld over twee constructiegroepen (vmbo en havo/vwo). De opgaven voor leesvaardigheid zijn eveneens door de vaststellingscommissie goedgekeurd.

Het leerlingmodel leesvaardigheid (14 deelaspecten) is veel uitgebreider dan het leerlingmodel voor schrijfvaardigheid (8 deelaspecten). Enerzijds heeft dit gezorgd voor een meer doelgerichte constructie: het construeren op signaalwoorden is bijvoorbeeld fijnmaziger dan het vergelijkbare deelaspect van schrijfvaardigheid dat het gehele gebruik van structuurwoorden bevraagt. Anderzijds geeft het minder variatie binnen het deelaspect. Ook is gebleken dat het lastig is om een duidelijk onderscheid te maken tussen sommige deelaspecten. Neem bijvoorbeeld de hoofdgedachte en een kernzin identificeren. In korte teksten kunnen deze deelvaardigheden overlappen. Het is belangrijk om hier goede afspraken over te maken: wat verstaan we onder een bepaalde deelvaardigheid? Daarnaast zou een pretest meer informatie kunnen geven over het nut van deze fijnmazigheid.

Het eerste hoofdaspect beoogt de leesstrategieën van een leerling te meten. In de huidige opzet van de toets zal dit een uitdaging zijn, omdat alleen de uitkomst van een leestaak kan worden gemeten en niet de aanpak. In de constructie is er gestart met de overige hoofdaspecten. Tijdens de constructie is er vervolgens telkens een inschatting gemaakt van de strategie die nodig was voor het oplossen van een opgave. Toen werd duidelijk dat bepaalde opgaven heel duidelijk naar één bepaalde strategie vragen, maar dat er ook opgaven zijn waarbij het strategiegebruik sterk van de leerling afhangt. Met deze kennis is er een aantal opgaven voor dit hoofdaspect geselecteerd die wél duidelijk naar een bepaalde strategie lijken te vragen (zie Tabel 5-7).

Ook bij de verschillende deelvaardigheden voor lezen bleek het ene opgavetype meer geschikt dan het andere. Het grootste deel van de opgaven bij lezen is een meerkeuze-opgave (zie Tabel 5-8).

Dit opgavetype past vooral bij de deelvaardigheden die een hoofdgedachte bevragen, conclusies trekken over de informatie uit de tekst of bijvoorbeeld het doel van de schrijver omdat er een antwoord buiten de tekst wordt verwacht, maar een open-opgave hier niet direct beoordeeld zou kunnen worden.

Een deelvaardigheid als kernzinnen identificeren vraagt juist naar een actie in de tekst zelf waardoor er in de tekst zelf gemarkeerd kan worden als opdracht (markeeropgave). Het begrijpen van tekststructuur kan ook op een actievere manier in de tekst bevraagd worden door bijvoorbeeld te laten slepen met de structuur van de tekst of een juist structuurwoord te kiezen in de tekst zelf (dropdown-opgave).

Tabel 5-7. Aantal vastgestelde opgaven en responsen per deelaspect van het leerlingmodel leesvaardigheid Engels

	vmbo-bb	vmbo-kb	vmbo-gt	havo	vwo	Totaal uniek
Deelaspect	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen
ENG-1.1	8 (8)	10 (10)	3 (6)	5 (5)	8 (8)	33 (36)
ENG-1.2	4 (13)	3 (8)	2 (7)	3 (11)	2 (5)	14 (44)
ENG-1.3	5 (8)	3 (12)	2 (8)	6 (18)	5 (11)	18 (48)
ENG-2.1	10 (10)	12 (12)	6 (6)	11 (11)	9 (9)	46 (46)
ENG-2.2	7 (7)	14 (14)	14 (17)	9 (12)	6 (9)	50 (59)
ENG-3.1	8 (8)	23 (23)	10 (10)	9 (9)	7 (11)	56 (60)
ENG-3.2	8 (8)	10 (10)	6 (8)	11 (11)	9 (9)	43 (45)
ENG-3.3	6 (9)	9 (12)	5 (5)	9 (14)	7 (10)	32 (46)
ENG-3.4	10 (19)	26 (61)	8 (20)	15 (35)	4 (13)	62 (147)
ENG-3.5	8 (8)	5 (8)	4 (9)	9 (11)	6 (9)	32 (45)
ENG-3.6	5 (8)	11 (11)	10 (20)	13 (14)	11 (12)	42 (56)
ENG-3.7	7 (7)	12 (12)	10 (10)	11 (11)	7 (9)	47 (49)
ENG-4.1	9 (9)	10 (12)	13 (13)	11 (11)	15 (15)	45 (47)
ENG-4.2	10 (13)	8 (11)	7 (9)	16 (24)	9 (14)	50 (71)
Totaal	105 (135)	156 (216)	100 (148)	138 (197)	105 (144)	570 (799)

Tabel 5-8. Aantal vastgestelde opgaven en responsen leesvaardigheid Engels per opgavetype

	vmbo-bb	vmbo-kb	vmbo-gt	havo	vwo	Totaal uniek
	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen
Alinea-opgave	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Correctie-opgave	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Dropdown-opgave	2 (6)	6 (17)	6 (14)	10 (10)	9 (11)	29 (51)
Hotspot-opgave	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	1 (1)
Kort open-opgave	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Markeer-opgave	24 (27)	23 (26)	26 (33)	25 (31)	25 (29)	109 (31)
Matrix-opgave	8 (28)	14 (48)	8 (30)	13 (47)	6 (24)	48 (173)
Meerkeuze-opgave	60 (60)	105 (105)	54 (54)	72 (72)	52 (52)	329 (329)
Multiple respons-opgave	1 (1)	0 (0)	0 (0)	0 (0)	0 (0)	1 (1)
Sleep-opgave	1 (4)	3 (12)	4 (15)	11 (30)	11 (26)	29 (86)
Sleep-opgave afbeelding	0 (0)	1 (4)	0 (0)	0 (0)	0 (0)	1 (4)
Volgorde-opgave	8 (8)	4 (4)	2 (2)	7 (7)	2 (2)	23 (23)
Totaal	105 (135)	156 (216)	100 (148)	138 (197)	105 (144)	570 (799)

5.3 Literatuur

College voor Toetsen en Examens (2014). *Publieksversie toetswijzer diagnostische tussentijdse toets voor Nederlands, Engels en wiskunde*. Utrecht: College voor Toetsen en Examens.

<https://www.pilotdt.nl/documenten/publicaties/2014/12/15/toetswijzer-dtt>

Deane, P. (2011). *Writing Assessment and Cognition*. Princeton, New Jersey: Educational Testing Service.

Deane, P., Odendahl, O., Quinlan, Th., Fowles, M., Welsh, C., Bivens-Tatum, J. (2008). *Cognitive Models of Writing: Writing Proficiency as a Complex Integrated Skill*. Research Report ETS RR-08-55. Princeton: Educational Testing Service.

Meijerink, H. (2008). *Over de drempels met taal en rekenen: hoofdrapport van de Expertgroep doorlopende leerlijnen taal en rekenen*. Enschede: Expertgroep Doorlopende Leerlijnen Taal en Rekenen.

Roelofs, E., & Schouwstra, S. (Red.). (2012). *Diagnostische tussentijdse toets: Verslag van de voorstudie*. Arnhem: Cito.

Van Hest, E., Beltman E. H., & Kleintjes, F. (2001). *Koppeling mvt-examens aan het Europese Referentiekader*. Arnhem: Citogroep, unit Voortgezet Onderwijs.

6 Beoordelen van open schrijftaken

Roelien Linthorst en Karen Keune⁵

6.1 Inleiding

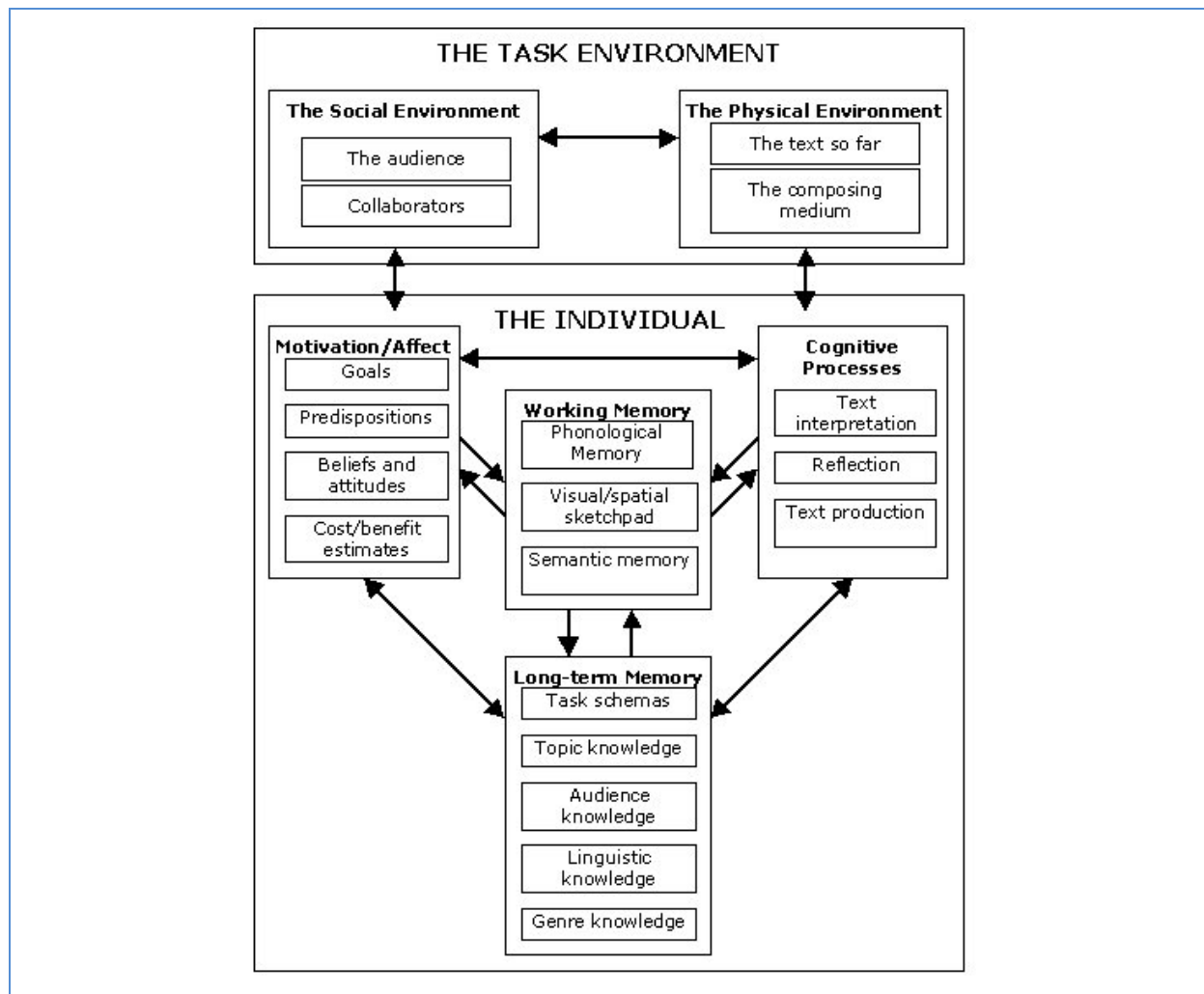
In de Toetswijzer DTT (CvTE, 2014), die de basis vormde voor de ontwikkeling van de DTT, wordt aangegeven dat er naast het gesloten, adaptieve en diagnostische deel van de DTT ook aanvullende open schrijftaken ontwikkeld zullen worden, zowel voor Engels als voor Nederlands. Hierover wordt in de Toetswijzer het volgende gezegd:

Bij de open schrijfopgave produceert de leerling op basis van een opdracht een tekst. De opgave wordt op de computer gemaakt. Automatische scoring van open schrijftaken bij de DTT is nog in ontwikkeling. De eerste jaren is de open schrijfopgave daarom facultatief. De leraar kan het schrijfvaardigheidsniveau van de leerlingen op basis van een beoordelingsformat beoordelen. Het is de bedoeling dat op termijn aspecten van de open schrijfopgave automatisch beoordeeld kunnen worden (CvTE, p. 11).

De keuze van de toetswijzercommissie om naast de kort open en gesloten opgaven ook (facultatieve) open schrijftaken aan te willen bieden, heeft te maken met de noodzaak om schrijfvaardigheid in de DTT op te delen in verschillende deelvaardigheden en deze afzonderlijk te toetsen om zo een goede, verfijnde diagnose te kunnen stellen. Bij het schrijven van een daadwerkelijke tekst zijn deze deelvaardigheden echter moeilijk apart te onderscheiden. Schrijven is een complexe vaardigheid waarbij een schrijver voortdurend bezig is met een ingewikkeld samenspel van verschillende deelactiviteiten, zoals ook blijkt uit het model van Hayes uit 1996 waarin het schrijfproces gerepresenteerd wordt (zie Figuur 6-1).

Bij open schrijftaken spelen al deze deelprocessen vrijwel tegelijkertijd een rol. Open schrijftaken sluiten om die reden ook goed aan bij wat er in de dagelijkse lespraktijk gebeurt. Desondanks kleven er aan open schrijftaken als toetsinstrument nadelen. Zoals uit het hierboven weergegeven model van Hayes (1996) blijkt, doet het schrijven van een tekst beroep op verschillende cognitieve processen. Die verschillende aspecten van het schrijven zijn in een beoordeling van een open schrijftaak vaak niet direct terug te zien: leerlingen kunnen eenzelfde score krijgen voor een schrijftaak, terwijl hun teksten op allerlei punten verschillen. Veelal is ook onduidelijk hoe verschillende aspecten meewegen bij de score. Daarnaast is de beoordelaarsovereenstemming bij open schrijftaken vaak gering (De Glopper & Willemsen, 2014). Dit kan deels verklaard worden doordat docenten verschillende aspecten van schrijfvaardigheid anders waarderen. De ene docent laat de rol van spellings- en formuleringsfouten bijvoorbeeld veel zwaarder meewegen in het eindoordeel dan de andere docent. Een ander nadeel van open schrijftaken is dat de afname tijdrovend is, waardoor leerlingen vaak maar één schrijftaak voorgelegd krijgen in een toets. Bouwer en Van den Bergh (2015) toonden echter aan dat op basis van een docentoordeel over één schrijftaak niet te oordelen valt over de algemene schrijfvaardigheid van een leerling, gezien de grote verschillen tussen oordelen van docenten én de grote verschillen tussen prestaties op verschillende schrijftaken van leerlingen. Om te kunnen generaliseren over genres heen, zou een schrijftoets volgens hen minimaal twaalf taken moeten bevatten.

⁵ Met dank aan Rick Godschalk, Hanna Trommel en Jennifer Leusink voor hun bijdrage aan de verschillende deelonderzoeken.



Figuur 6-1. Hayes' model van schrijfvaardigheid (1996)

In gesloten schrijftaken kunnen aan leerlingen opdrachten voorgelegd worden waarin slechts één schrijfaspect tegelijk aan de orde komt. De toetsuitslag levert hierdoor gerichte informatie op die de leerling kan gebruiken om zijn schrijfvaardigheid te verbeteren. Ook biedt het de mogelijkheid om de leerling in korte tijd verschillende contexten en schrijfdoelen voor te leggen, waardoor de toetsafname efficiënter verloopt en er meer recht gedaan wordt aan de diversiteit van schrijfvaardigheid dan bij een enkele open schrijftaak. Ook leent een gesloten vorm van schrijfvaardigheidstoetsing zich voor een adaptieve toetsafname terwijl dat bij open schrijftaken lastiger realiseerbaar is (Linthorst & Schuurs, 2014).

Kortom, open schrijftaken doen een beroep op alle voor het schrijven relevante cognitieve processen en sluiten aan bij wat er in de onderwijspraktijk gebeurt. Om die redenen maken open schrijftaken een valide indruk. Desondanks kleven er aan een open schrijftaak als toetsinstrument verschillende nadelen, onder andere door de verschillen die optreden in de beoordelingen van verschillende beoordelaars. Aan die beoordelingsproblematiek kan wellicht tegemoetgekomen worden indien het mogelijk zou zijn om schrijftaken automatisch te beoordelen. In de voorstudie van de DTT werd mede om bovenstaande redenen de suggestie gedaan om in de uiteindelijke DTT te kiezen voor een getrappt model, waarbij het resultaat op de (geautomatiseerd beoordeelde) open schrijfta(a)k(en) medebepalend zou zijn voor de gesloten adaptieve opgaven die de leerling vervolgens maakt om een verfijnde diagnose te kunnen stellen (Roelofs & Schouwstra, 2012). Om die reden is in het kader van de DTT verkend in hoeverre het mogelijk is om leerlingteksten geautomatiseerd te beoordelen. Van dit deelonderzoek en de onderzoeken die daarop volgden, wordt in dit hoofdstuk verslag gedaan.

6.2 Geautomatiseerd beoordelen van schrijfvaardigheid

Tijdens de eerste try-out in 2013 (Feenstra & Keune, 2013) is onderzocht in hoeverre het mogelijk is om open schrijftaken geautomatiseerd te beoordelen. Voor het Nederlands zijn hiervoor nog geen programma's beschikbaar. Wel is er voor het Nederlands de leesbaarheidspredictor T-scan (Kraf, Van der Sloot, Pander Maat, Van den Bosch, Van Gompel, 2013). Dit is een taaltechnologische applicatie die automatisch teksten analyseert door zinnen en woorden te ontleden en woordsoorten te benoemen. Ook bevat de applicatie verschillende frequentie- en woordenlijsten. Aan de hand van deze gegevens voorspelt T-scan tekstcomplexiteit op basis van meer dan 150 tekstenmerken.

Met behulp van deze maten voor tekstcomplexiteit kunnen alleen de lagere orde vaardigheden geanalyseerd worden. In het leerlingmodel Schrijfvaardigheid Nederlands (zie Hoofdstuk 4) zijn dit de aspecten *3 Linguïstische vaardigheden (woord- en zinsniveau)* en *4 Orthografische vaardigheden (spelling en interpunctie)*. Een schrijfproduct volledig beoordelen is niet mogelijk. Het doel van het onderzoek in de try-out (Feenstra & Keune, 2013) was daarom ook te bepalen of het voor bovengenoemde aspecten mogelijk is om op basis van tekstenmerken onderscheid te maken tussen kwalitatief goede en minder goede schrijfproducten. Indien dit het geval zou zijn, zou dit als basis kunnen dienen voor verdere ontwikkeling van geautomatiseerde schrijftaakbeoordeling.

Voor het onderzoek zijn bij leerlingen uit 2 vmbo-bb, 2 vmbo-kb, 2 vmbo-gt, 3 havo en 3 vwo twee stelopdrachten afgenomen. De 438 schrijfproducten van de leerlingen uit 2 vmbo en vergeleken met 253 schrijfproducten van leerlingen uit 3 havo en 3 vwo. Hiervoor zijn tekstenmerken uit T-scan geselecteerd die iets zeggen over twee deelaspecten van *3 Linguïstische vaardigheden (woordgebruik en zinsbouw)* en het deelaspect *2.3 samenhang* binnen de tekst. Voor het meten van de *Orthografische vaardigheden (spelling en interpunctie)* is gebruik gemaakt van de automatische spellingscorrector Valkuil.net (<http://valkuil.net/>).

Uit het onderzoek bleek dat het gebruik van bepaalde maten voor woordgebruik, zinsbouw en samenhang meer kenmerkend waren voor vmbo-leerlingen en andere maten juist voor havo/vwo-leerlingen. Nadere inspectie van individuele maten maakte echter duidelijk dat een aantal factoren de betrouwbaarheid van deze analyses sterk beïnvloedde. De specifieke schrijftaak bleek van grotere invloed op de tekstenmerken dan de leerweg/schoolsoort van de leerling. Verder werden teksten met spelfouten, niet bestaande woorden en afkortingen onterecht beoordeeld met een hoge woordcomplexiteit en werden teksten met grammaticaal incorrecte zinnen en zinnen met ontbrekende interpunctie onterecht beoordeeld met een hoge zinscomplexiteit.

Het geautomatiseerd beoordelen van schrijfproducten binnen de DTT bleek vooralsnog niet mogelijk te zijn. Analyse van tekstenmerken geeft interessante informatie over lagere orde vaardigheden, maar niet zodanig dat een leerling hierop beoordeeld kan worden. Verdere ontwikkeling hiervan voor de Nederlandse taal is een zeer complex, tijdrovend en kostbaar proces en lijkt daardoor voorlopig onhaalbaar. Het ontwikkelen van een tool voor diagnostische feedback op de lagere orde vaardigheden van schrijfproducten lijkt echter wel haalbaar en kan waardevol zijn voor leerling en docent.

6.3 Onderzoek naar beoordelingsmodellen

Zoals in de vorige paragraaf geschetst, is het nog niet mogelijk om teksten volledig en betrouwbaar geautomatiseerd te beoordelen. Het aanbieden van geautomatiseerd beoordeelbare open schrijftaken binnen DTT bleek daarmee binnen de pilot niet haalbaar. Wat wel haalbaar leek, was het ontwikkelen van open schrijftaken voorzien van beproefde beoordelingsmodellen waarmee de docent zelf de tekst van de leerling kan beoordelen. Hiervoor zijn twee deelonderzoeken verricht. In een eerste deelonderzoek, uitgevoerd in de tweede try-out, zijn drie verschillende beoordelingsmodellen onderzocht op hun bruikbaarheid bij zowel Nederlands (Linthorst & Keune, 2014), als Engels (Godschalk, Vrijs en Schouwstra, 2014).

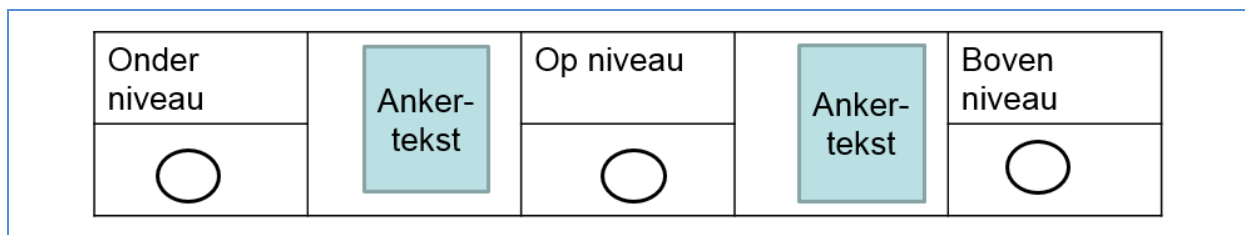
In een tweede deelonderzoek in 2016 zijn, op basis van de ervaringen opgedaan in het eerste deelonderzoek, nieuwe schrijftaken ontwikkeld en is een eerste verkenning gedaan naar de samenhang tussen deze open schrijftaken en de resultaten op de DTT. Beide deelonderzoeken zijn uitgevoerd bij zowel

Engels als Nederlands. Een beoogd neveneffect van beide deelonderzoeken was dat in de twee onderzoeken bruikbare schrijfproducten van leerlingen verzameld zouden worden, die ingezet zouden kunnen worden in de doorontwikkeling van het geautomatiseerd beoordelen van schrijfvaardigheid. Daarvoor is een corpus van leerlingteksten zeer wenselijk.

6.3.1 Onderzoek beoordelingsmodellen Nederlands

Het eerste deelonderzoek, naar de bruikbaarheid van drie verschillende beoordelingsmodellen, is uitgevoerd in het voorjaar van 2014, zowel bij Engels als bij Nederlands. Voor Nederlands zijn twee open schrijfp opdrachten voorgelegd aan de leerlingen (Linthorst & Keune, 2014). Het onderzoek Nederlands kende twee doelen: ten eerste dienden de verzamelde leerlingproducten als mogelijke input voor vervolgonderzoek naar geautomatiseerde beoordeling van schrijfvaardigheid, daarnaast werden de leerlingproducten gebruikt voor de beoordeling van drie potentieel geschikte beoordelingsmodellen op hun betrouwbaarheid en bruikbaarheid voor opname in de DTT. Deze drie beoordelingsmodellen waren:

1. Het *globale beoordelingsmodel* waarbij de beoordelaar gevraagd werd een oordeel over de tekst te vellen aan de hand van vier beoordelingspunten die aansluiten bij de hoofdaspecten van de DTT (*doel & publiek, structuur, woordgebruik & zinsbouw en spelling & interpunctie*)
2. De *beoordeling met behulp van anker teksten* waarbij ankeropstellen ingezet werden die een bepaalde grens markeerden. In dit onderzoek zijn er voor elk van de vier hoofdaspecten en voor vmbo-tl, havo en vwo ankers geselecteerd op de grens 'onder & op niveau' en de grens 'op & boven niveau'. Aan de beoordelaars werd vervolgens gevraagd de te beoordelen teksten te plaatsen naast deze ankeropstellen, zoals in Figuur 6-2 weergegeven wordt.



Figuur 6-2. Schematische weergave ankerbeoordeling

3. Het *analytische beoordelingsmodel* waarbij beoordelaars gevraagd werd een tekst te beoordelen aan de hand van meerdere beoordelingspunten die samen een zo nauwkeurig mogelijk beeld geven van het schrijfvaardigheidsniveau van een leerling. De beoordelingscriteria bij dit model kwamen overeen met de 12 (vmbo) en 13 (havo/vwo) deelaspecten die in het leerlingmodel schrijfvaardigheid onderscheiden worden.

In totaal zijn er 480 leerlingteksten verzameld die door 8 beoordelaars beoordeeld zijn aan de hand van het design dat weergegeven is in Tabel 6-1. Bijvoorbeeld, beoordelaar 1 heeft de teksten 1 tot en met 60 beoordeeld via het ankermodel en het analytische model, teksten 181 tot en met 240 met het globale model en het ankermodel en de teksten 241 tot en met 360 met het globale en het analytische model.

Tabel 6-1. Weergave design beoordeling leerlingteksten

Model	Tekst							
	1-60	61-120	121-180	181-240	241-360	301-360	361-420	421-480
Globaal	5	3	7	1	1	6	7	4
	4	6	2	8	5	2	3	8
Anker	1	3	2	4	5	8	7	6
	4	2	3	1	7	6	5	8
Analytisch	1	2	7	8	1	2	5	6
	5	6	3	4	7	8	3	4

De beoordeling vond plaats van globaal naar specifiek: eerst beoordeelde de beoordelaar de teksten volgens het globale model en vervolgens via het ankermodel of het analytische model, of de beoordelaar beoordeelde eerst via het ankermodel en daarna volgens het analytische model. Op die manier werd voorkomen dat de beoordelingen op het globale en het ankermodel beïnvloed zouden worden door de gedetailleerde beoordeling bij het analytische model. Na iedere beoordelingsronde werd de beoordelaar ook bevraagd naar zijn of haar ervaringen met het werken met het specifieke beoordelingsmodel. Dit gebeurde mondeling aan de hand van een voorafgaand opgestelde vragenlijst. Op die manier werd aanvullende informatie verzameld over de kwaliteit, bruikbaarheid en efficiëntie van de modellen.

Een uitgebreid verslag van de resultaten van het onderzoek is terug te vinden in Linthorst en Keune (2014). De interbeoordelaarsbetrouwbaarheid van alle modellen was redelijk goed, waarbij de interbeoordelaarsovereenstemming bij het ankermodel op de meeste punten iets hoger bleek te liggen dan bij het globale model. Ook werd er bij dit model meer spreiding gevonden in de beoordelingen zelf. Tussen het globale model en het analytische model werden geen opvallende verschillen gevonden. Hoewel het werken met de ankerteksten er met betrekking tot de (interbeoordelaars)betrouwbaarheid iets beter uitsprong, had deze beoordelingswijze niet de directe voorkeur van de beoordelaars. Zij vonden het werken met de ankerteksten tijdrovender, vermoedelijk ook veroorzaakt doordat zij voor ieder niveau en voor ieder hoofdaspect nieuwe ankerteksten aangeboden kregen. Daarnaast plaatsten enkele beoordelaars vraagtekens bij de keuze van enkele specifieke ankerteksten, die zij niet passend vonden voor het te beoordelen niveau. De beoordeling aan de hand van het globale model verliep het snelst en was volgens de beoordelaars het meest efficiënt. Het analytische model deed volgens de beoordelaars het meest recht aan de prestaties van de leerling, omdat de tekst op veel verschillende aspecten beoordeeld werd.

6.3.2 Onderzoek beoordelingsmodellen Engels

Bij Engels is in 2014 ook een onderzoek naar de geschiktheid van verschillende beoordelingsmodellen uitgevoerd, maar dit onderzoek had een iets andere insteek (Godschalk, Vrijs en Schouwstra, 2014). De verzamelde leerlingproducten Engels zijn beoordeeld aan de hand van een beoordelingsmodel waarmee niet alleen de globale kwaliteit van het schrijfproduct beoordeeld werd, maar ook de communicatieve effectiviteit. Bij communicatieve effectiviteit gaat het erom of de leerling daadwerkelijk het beoogde schrijfdoel bereikt. Communicatieve effectiviteit is niet opgenomen in het leerlingmodel Engels, omdat het lastig toetsbaar is in gesloten opgaven. De toetswijzercommissie Engels achtte het echter wel belangrijk voor open schrijftaken, vandaar dat onderzocht is of het mogelijk is om communicatieve effectiviteit apart te beoordelen bij open schrijftaken.

In het onderzoek zijn drie schrijftaken afgenomen bij vmbo- en havo/vwo-leerlingen. Dit leverde 571 schrijfproducten op. Zowel de beoordeling van de globale kwaliteit van ieder schrijfproduct als de communicatieve effectiviteit gebeurde holistisch. De beoordelaar werd gevraagd een rapportcijfer voor het schrijfproduct te geven en een classificatie van de leerweg/schoolsoort die paste bij het schrijfproduct. Een groot deel van de leerlingteksten (60%) is door twee beoordelaars beoordeeld, het resterende deel door één beoordelaar.

Ook over dit deelonderzoek is uitgebreid gerapporteerd in het onderzoeksrapport DTT van 2014 (Godschalk, Vrijs en Schouwstra, 2014). Uit de resultaten bleek dat de beoordelaars er niet goed in slaagden om een onderscheid te maken tussen communicatieve effectiviteit en globale kwaliteit van de schrijfproducten met behulp van de niveau-inschatting en de cijfermatige beoordeling.

6.4 Ontwikkeling van open schrijftaken en eerste verkenning samenhang

De ervaringen die opgedaan waren met het eerdere onderzoek naar geautomatiseerde beoordeling van schrijfvaardigheid (paragraaf 6.2) en met het deelonderzoek naar beoordelingsmodellen (paragraaf 6.3), waren aanleiding voor een vervolgonderzoek in het schooljaar 2016-2017. Dit kleinschalige onderzoek had twee doelen: enerzijds diende het om enkele open schrijftaken voor Engels en Nederlands te ontwikkelen in het kader van de ontwikkeling van de DTT, anderzijds wilden we op kleinschalige wijze verkennen óf en welke samenhang er was tussen de resultaten van leerlingen op de DTT en een of meerdere open schrijftaken.

6.4.1 Ontwikkeling van open schrijftaken

Met het oog op bovenstaande doelen hebben de constructiegroepen Engels en Nederlands medio 2016 enkele open schrijftaken ontwikkeld. Belangrijke uitgangspunten bij de constructie van deze taken waren voor beide talen:

- Er komen aparte schrijftaken voor vmbo en voor havo/vwo. Gezien het niveauverschil tussen vmbo leerjaar 2 en havo/vwo leerjaar 3 werd het niet realistisch geacht om bij deze leerlingen altijd precies dezelfde schrijftaken af te nemen. Sommige schrijftaken leenden zich wel voor afname bij zowel vmbo- als havo/vwo-leerlingen en voor onderlinge vergelijkingen, maar in andere gevallen is er, ook op advies van de vaststellingscommissies, gedifferentieerd tussen deze schoolsoorten.
- Iedere schrijftaak moet binnen een lesuur van ongeveer 50 minuten af te nemen zijn, om de tijdsbelasting van de deelnemende scholen niet te hoog te laten zijn.
- De schrijftaak heeft een duidelijk communicatief doel, is zo authentiek mogelijk en sluit zo veel mogelijk aan bij de belevingswereld van de leerling.
- Er worden schrijftaken voor verschillende tekstdoelen (bijvoorbeeld informeren en overtuigen) en verschillende tekstgenres (bijvoorbeeld e-mail, artikel, instructie) ontwikkeld om het domein schrijfvaardigheid zo breed mogelijk te representeren.
- De 'overschrijfbaarheid' van de schrijftaak wordt zo veel mogelijk ingeperkt. Voor de vergelijkbaarheid van de uiteindelijke schrijfproducten van leerlingen is enige sturing in de schrijftaak wenselijk. Bovendien hebben leerlingen achtergrondinformatie over het onderwerp nodig om een inhoudelijk goede tekst te kunnen schrijven. Tegelijkertijd vergroot deze sturing de 'overschrijfbaarheid': leerlingen kunnen vaak formuleringen uit de casus of uit bronmateriaal meenemen in hun eigen tekst, waarmee de eigen tekstproductie verminderd wordt. Om dit probleem te ondervangen hebben we bij de ontwikkeling van de open schrijftaken gezocht naar input die op een andere wijze verstrekt werd, bijvoorbeeld door audio- of videomateriaal of door afbeeldingen. Bovendien is er bij Engels voor gekozen om de opdracht in het Nederlands te verstrekken, zodat leerlingen Engelse begrippen niet letterlijk uit de casus konden overnemen.

Net als bij de reguliere DTT zijn de ontwikkelde schrijftaken voorgelegd aan de vaststellingscommissies. Ook zijn de schrijftaken voorgelegd aan leerplanontwikkelaars Nederlands en Engels van SLO. Zowel bij Nederlands als bij Engels zijn er uiteindelijk drie schrijftaken aangewezen die kleinschalig gepretest zouden worden (zie Tabel 6-2).

Tabel 6-2. Geselecteerde schrijftaken Nederlands en Engels

Vak	Schrijftaak	Leerweg/schoolsoort
Nederlands	Gescheiden gymles	vmbo
	Tekenbeet	vmbo & havo/vwo
	Voedingsles	havo/vwo
Engels	Homework	vmbo
	Fantastic Beasts	vmbo & havo/vwo
	Olympic gymnast	havo/vwo

Op basis van de ervaringen die opgedaan zijn in het eerste deelonderzoek (zie paragraaf 6.3) is ervoor gekozen om de schrijftaken te voorzien van een globaal beoordelingsmodel, waarbij de beoordelaars de taken moeten beoordelen aan de hand van de vier hoofdaspecten uit het leerlingmodel schrijfvaardigheid. Hoewel uit het eerste deelonderzoek het model met de anker teksten het meest veelbelovend leek te zijn wanneer het gaat om betrouwbaarheid en interbeoordelaarsovereenstemming, bleek dit model ook arbeidsintensief te zijn. We konden niet te veel tijd vragen voor de beoordeling van de leerlingteksten in dit tweede deelonderzoek. We wilden graag ervaren docenten inschakelen voor de beoordeling van de leerlingteksten en er werd al een groot beroep gedaan op de tijdsinvestering van de deelnemende scholen, aangezien de scholen de open schrijftaken afnamen naast de reguliere DTT. Om die reden, en vanwege het feit dat er voor het gebruik van ankerbeoordeling een extra afname nodig zou zijn om mogelijk

geschikte ankeropstellen te verzamelen, is er gekozen voor een globaal beoordelingsmodel. In Figuur 6-3 is weergegeven hoe dit beoordelingsmodel eruitzag voor hoofdaspect 1 bij Nederlands.

1. Afstemming op doel en publiek (retorische vaardigheden)							
Is de leerling in staat een doelgerichte tekst te schrijven die afgestemd is op het gevraagde publiek?							
Aspecten die hierbij onder andere een rol spelen:							
Kan de leerling inschatten welke informatie wel en niet in de tekst moet worden opgenomen?							
Kan de leerling de juiste toon kiezen en vasthouden?							
Kan de leerling een passend schrijfdoel hanteren en daaraan vasthouden?							
...vergeleken met wat een leerling van deze leerweg (vmbo-bb, -kb, -gt/tl, havo of vwo) zou moeten kunnen							
heel zwak	zwak	matig	voldoende	goed	heel goed		niet beoordeelbaar
1	2	3	4	5	6		0

Figuur 6-3. Globaal beoordelingsmodel voor hoofdaspect 1 bij Nederlands

Zoals in de figuur te zien is, is het beoordelingsmodel voorzien van een zespuntsschaal variërend van heel zwak tot heel goed, waarbij de beoordelaar moet aangeven hoe goed de leerling op het hoofdaspect presteert vergeleken met wat een leerling van deze leerweg/schoolsoort zou moeten kunnen. Door weergave van de deelaspecten bij het hoofdaspect is geprobeerd inzichtelijk te maken hoe het hoofdaspect binnen de DTT gedefinieerd wordt.

6.4.2 Verkenning samenhang open schrijftaken en DTT

Naast de ontwikkeling van enkele open schrijftaken op zich, was het ook de wens om een eerste verkenning uit te voeren naar de samenhang tussen tekst(en) die leerlingen zelf schrijven en de resultaten op de adaptieve DTT. Bovendien wilden we zien hoe de schrijftaken in praktijk uitpakten en ervaring opdoen met de afname van open schrijftaken in Facet. Een probleem bij de afname van de open schrijftaken was dat meervoudig beoordelen (het beoordelen van een product van leerlingen op meerdere aspecten) binnen Facet nog niet mogelijk was en we de schrijfproducten wel op de vier verschillende hoofdaspecten wilden laten beoordelen. Om dit probleem te ondervangen is er uiteindelijk gekozen voor een work-around-oplossing, waarbij leerlingen na het schrijven van hun tekst nog driemaal door een (leeg) scherm moesten klikken, zodat het leerlingproduct door de beoordelaar uiteindelijk in vier aparte schermen (voor elk van de vier aspecten één scherm) beoordeeld kon worden.

Om de open schrijftaken te kunnen piloten en de samenhang met de gesloten DTT te kunnen verkennen, zijn in het voorjaar van 2017 de scholen die de DTT afnamen benaderd met de vraag of ze naast de reguliere DTT ook een open schrijftaak Engels en/of Nederlands af wilden nemen bij één of meerdere klassen. Eén van de vereisten voor deelname was dat er van de deelnemende scholen verwacht werd dat ze de schrijfproducten van hun leerlingen ook zelf zouden nakijken aan de hand van het bijgeleverde beoordelingsmodel. Daarnaast zouden de schrijfproducten beoordeeld worden door een tweede onafhankelijke beoordelaar, te weten een Cito-stagiaire van de universiteit Nijmegen (voor Engels) en Cito-stagiaires van de universiteit Groningen (voor Nederlands). Aangezien de beoordeling van open schrijftaken vaak verschilt afhankelijk van wie de beoordeling uitvoert, was een beoordeling door twee personen van een schrijfproduct een minimumvoorwaarde voor dit deelonderzoek.

De eigen beoordeling door de school bleek echter een hoge drempel voor deelname. De respons voor deelname aan dit onderzoek was laag. Hierbij speelde ook de grote tijdsinvestering die scholen al gedaan hadden voor de reguliere DTT een rol, waardoor afname van een extra opdracht voor scholen vaak als te belastend ervaren werd. In Tabel 6-3 is weergegeven hoeveel scholen zich aangemeld hebben voor de afname van de open schrijftaken en hoeveel afnames er daadwerkelijk plaatsgevonden hebben.

Tabel 6-3. Deelname open schrijftaak voorjaar 2017

	Aantal aangemelde scholen vmbo	Aantal aangemelde scholen havo/vwo	Daadwerkelijke deelname
Nederlands	2	1	2 vmbo-scholen met in totaal 175 leerlingen en 1 havo/vwo-school met 31 leerlingen
Engels	1	1	1 havo/vwo-school met 27 leerlingen

Uiteindelijk hebben drie scholen een open schrijftaak voor Nederlands afgenomen en heeft één school deelgenomen aan de afname van de schrijftaak Engels. Helaas zagen de deelnemende scholen Nederlands uiteindelijk af van de eigen beoordeling van de open schrijftaken, waardoor er niet gerapporteerd kan worden over de resultaten bij Nederlands. Om die reden zal hieronder enkel verslag gedaan worden van de resultaten bij Engels.

De open schrijftaak Engels is afgenomen bij 27 leerlingen uit havo 3. De opdracht 'Fantastic Beasts' bestond uit het schrijven van een YouTube-bericht op basis van een video. Hierbij moest de leerling aangeven of hij of zij de film die beschreven werd in de video zou willen bezoeken, waarbij minimaal drie argumenten moesten worden gegeven. Deze instructies zijn in het Nederlands gegeven om letterlijk kopiëren in de leerlingtekst te voorkomen. De video werd in het Engels aangeboden.

De 27 leerlingproducten zijn door twee beoordelaars beoordeeld: de docent Engels van de leerlingen en een Cito-stagiaire van de universiteit Nijmegen. Op basis van hun beoordelingen is er eerst gekeken naar de mate van overeenstemming tussen de beoordelaars. Helaas bleek er onvoldoende overeenstemming tussen de beoordelaars te zijn. De lage overeenstemming tussen de beoordelaars maakt dat de beoordelingen instabiel zijn en dat ze daarom niet gebruikt kunnen worden voor verdere analyse. Een verkenning van de samenhang tussen de open schrijftaak en de resultaten op de DTT zelf is met deze data daarom helaas niet goed mogelijk.

Bij de tegenvallende beoordelaarsovereenstemming van de open schrijftaak Engels speelt vermoedelijk mee dat één van de beoordelaars geen docent was. In het eerste deelonderzoek naar de verschillende beoordelingsmodellen zagen we wel een redelijk goede beoordelaarsovereenstemming, maar bij dat deelonderzoek waren enkel docenten met leservaring betrokken die ook een gezamenlijke scholingsbijeenkomst hadden gevolgd. Dat was nu niet het geval.

6.5 Conclusies en aanbevelingen

Het bleek niet haalbaar om binnen de driejarige DTT-pilot schrijftaken aan te bieden die op een betrouwbare manier geautomatiseerd beoordeeld kunnen worden. De twee deelonderzoeken die we verricht hebben, hebben wel bruikbare inzichten opgeleverd over de afname van open schrijftaken en de beoordeling daarvan.

Wat betreft de *beoordeling* van open schrijftaken door beoordelaars leerde het eerste deelonderzoek dat ankerbeoordeling perspectieven biedt, maar tegelijkertijd tijdrovend is. Zeker wanneer teksten van een bepaalde schoolsoort of leerweg op meerdere aspecten (met verschillende ankers) beoordeeld moeten worden. Het opzetten van de ankerbeoordeling vraagt ook een flinke tijdsinvestering omdat er idealiter eerst een afname van de schrijftaak plaatsvindt voor de selectie van ankers alvorens de ankers daadwerkelijk ingezet kunnen worden. Een globale beoordeling aan de hand van enkele beoordelingsaspecten is daarom vaak praktischer, maar in het tweede deelonderzoek zagen we dat ook hier de interbeoordelaarsovereenstemming te wensen overliet. Het ideale beoordelingsmodel hebben we in deze twee deelonderzoeken dan ook niet gevonden. Wellicht bieden recente ontwikkelingen met paarsgewijze vergelijkingen (o.a. Van Daal, Lesterhuis, Coertjens, Donche, & De Maeyer., 2016) mogelijkheden om op vrij efficiënte wijze toch tot een betrouwbaar oordeel te kunnen komen. Paarsgewijs vergelijken vraagt echter nog steeds een behoorlijke tijdsinvestering, levert uiteindelijk een rangordening van leerlingteksten op en biedt daarom wellicht minder aanknopingspunten voor een formatieve vorm van schrijfvaardigheidsevaluatie.

Wat betreft de *schrijftaak* zelf hebben we in het tweede deelonderzoek gezien dat het aanbieden van bronmateriaal via kijk-/luisterfragmenten duidelijk voordelen heeft ten opzichte van een schriftelijke casus. De overschrijfbaarheid wordt hierdoor verminderd en bovendien bleek uit de afgelegde schoolbezoeken dat leerlingen audio- of videomateriaal aansprekender vinden. Dit biedt in de toekomstperspectieven voor verdere integratie van taalvaardigheden bij de toetsing van taalvaardigheid.

Er zal in dat geval ook gekeken moeten worden hoe de instructie over de schrijftaken aan de leerlingen gegeven kan worden. Hoewel de deelnemende scholen bij de afname van de schrijftaken een handleiding aangeleverd kregen waarin de procedure voor de toetsafname geschetst werd en waarin aangegeven werd welke instructie leerlingen moesten krijgen, bleek in de praktijk dat docenten deze instructie nauwelijks uitvoerden. De instructie beperkte zich veelal tot de instructie dat leerlingen konden inloggen met hun inloggegevens en aan de slag konden. Het verstrekken van instructie op andere wijze, bijvoorbeeld door een audio- of videobericht in de afname-omgeving lijkt daarom wenselijk. Vermoedelijk krijgen leerlingen daardoor een betere en meer gelijke instructie.

Het lijkt hoe dan ook wenselijk om de beoordeling van open schrijftaken te voorzien van een training van beoordelaars. Ervaring en training van beoordelaars leek in deze deelonderzoeken effect te hebben, omdat de overeenstemming tussen de beoordelaars in het eerste deelonderzoek beduidend hoger lag dan in het tweede deelonderzoek. In het eerste deelonderzoek was er een gezamenlijke startbijeenkomst voor de beoordelaars georganiseerd, waarbij de globale beoordelingsprincipes besproken werden. In het tweede deelonderzoek hadden de beoordelaars geen specifieke instructie vooraf ontvangen.

De gewenste vergelijking tussen open schrijftaken en de resultaten op de DTT bleek in het tweede deelonderzoek niet mogelijk. Om dit goed uit te kunnen voeren, zou een uitgebreider onderzoek opgezet moeten worden, met afname van meer verschillende open schrijftaken bij dezelfde leerlingen, gezien de taakafhankelijkheid van schrijfvaardigheid (o.a. Brouwer & Van den Bergh, 2015; Feenstra & Keune, 2013). Ook ligt een meer kwalitatieve vergelijking daarbij eerder voor de hand, omdat de resultaten op een open schrijftaak niet zonder meer vergelijkbaar zijn met de resultaten op de DTT. Gezien de gevoerde discussie over open en gesloten toetsing van schrijfvaardigheid (zie o.a. De Gloppe & Willemsen, 2014; Pulles, Den Ouden, Herrlitz, & Van den Bergh, 2013) is vervolgonderzoek naar open en gesloten vormen van schrijfvaardigheidstoetsing in de toekomst aanbevelenswaardig.

6.6 Literatuur

- Bouwer, R. & Van den Bergh, H. (2015). Toetsen van schrijfvaardigheid: hoeveel beoordelaars, hoeveel taken? *Levende Talen Tijdschrift*, 16 (3), 3-12.
- College voor Toetsen en Examens (2014). *Publieksversie toetswijzer diagnostische tussentijdse toets voor Nederlands, Engels en wiskunde*. Utrecht: College voor Toetsen en Examens.
<https://www.pilotdt.nl/documenten/publicaties/2014/12/15/toetswijzer-dtt>
- De Gloppe, K., & Willemsen, A. (2014). Kunnen gesloten toetsen bijdragen aan de toetsing van schrijfvaardigheid? *Levende Talen Tijdschrift*, 15(1), 31-38.
- Feenstra, H. & Keune, K. (2013). Geautomatiseerde beoordeling schrijfvaardigheid Nederlands. In S. Schouwstra (Red.), *Diagnostische tussentijdse toets: Onderzoek 2013* (pp. 63 - 91). Arnhem: Cito.
- Godschalk, R., Vrijs, W. & Schouwstra, S. (2014). Beoordeling open vragen schrijfvaardigheid Engels. In S. Schouwstra (Red.), *De diagnostisch tussentijdse toets: onderzoek 2014* (pp. 82-90). Arnhem: Cito.
- Hayes, J.R. (1996). A New Framework for Understanding Cognition and Affect in Writing. In: C.M. Levy, S. Randell (Red.). *The Science of Writing. Theories, Methods, Individual Differences and Applications*. Mahwah, New Jersey: Lawrence Erlbaum.

- Kraf, R., van der Sloot, K., Pander Maat, H., van den Bosch, A., & van Gompel, M. (2013). *T-Scan*, <http://language.link.let.uu.nl/tscan>.
- Linthorst, R. & Keune, K. (2014). Beoordeling open vragen schrijfvaardigheid Nederlands. In S. Schouwstra (Red.), *De diagnostisch tussentijdse toets: onderzoek 2014* (pp. 71-82). Arnhem: Cito.
- Linthorst, T.R. & Schuurs, U. (2014). Diagnostische toetsing bij het schoolvak Nederlands. *28ste Conferentie Onderwijs Nederlands*, 28, 59-63.
- Pulles, T., Den Ouden, H., Herrlitz, W., & Van den Bergh, H. (2013). Kan een meerkeuzetoets bijdragen aan het meten van schriftelijke taalvaardigheid? *Levende Talen Tijdschrift*, 14(2) 31-41.
- Roelofs, E., & Schouwstra, S. (Red.). (2012). *Diagnostische tussentijdse toets: Verslag van de voorstudie*. Arnhem: Cito.
- Van Daal, T., Lesterhuis, M., Coertjens, L., Donche, V. & De Maeyer, S. (2016). Validity of comparative judgement to assess academic writing: examining implications of its holistic character and building on a shared consensus. *Assessment in education: principles, policy & practice*. <http://dx.doi.org/10.1080/0969594X.2016.1253542>.

7 DTT wiskunde

Irene van Stiphout

7.1 Leerlingmodel

De meerwaarde van de DTT wiskunde is gezocht in het bieden van een verklaring voor achterblijvende prestaties van leerlingen. Het idee hiervoor kwam voort uit de constatering in onderzoek (Bruin-Muurling, 2010; Kraemer, 2011; Van Stiphout, 2011; Roorda, 2012) dat het niveau van leerlingen achterblijft ten opzichte van de verwachtingen. Uit deze onderzoeken kwam een beeld naar voren dat de beheersing van leerlingen kwetsbaar is. Dat wil zeggen dat leerlingen een redelijke beheersing hebben van standaardopgaven, maar dat ze weinig flexibel daarbuiten zijn. Ook bleek dat leerlingen dezelfde wiskundige concepten in verschillende vakken niet herkennen en kennis over wiskundige concepten die opgedaan is in het ene vak niet of nauwelijks gebruiken in het andere vak (Roorda, 2012). Daarnaast gebruiken leerlingen vaak weinig geavanceerde strategieën en blijven ze daarin hangen waardoor ze niet toekomen aan verder generaliseren of abstraheren (Kraemer, 2011). De hoop was dat een kader dat zou aansluiten op recente ontwikkelingen in het onderwijs betekenisvol zou kunnen zijn.

In het actieplan Beter presteren (Van Bijsterveldt-Vliegenthart, 2011) is ingezet op een hoger niveau van leerlingen. Deze combinatie van de behoefte aan een hoger niveau enerzijds en anderzijds de inzichten in waarom dat hoger niveau niet wordt bereikt, heeft geleid tot een tweedimensionaal verklarend leerlingmodel, dat bestaat uit een dimensie van vakdomeinen en een zogeheten vakdidactische dimensie (zie Tabel 7-1). De rationale achter deze tweede dimensie is dat docent en vanuit hun eigen toetsing weten welke onderwerpen (de domeinen) leerlingen wel of niet beheersen, maar minder zicht hebben op mogelijke redenen *waarom* het niveau niet is bereikt.

In de vakdidactische dimensie van het leerlingmodel is gekozen voor een driedeling waarin de onderstaande ontwikkelingen zijn verweven:

- Het herkennen van onderliggende wiskundige structuren (Devlin, 2012);
- Het komen tot niveauverhoging in termen van objectvorming en kunnen omgaan met meervoudige betekenis (bijv. Sfard, 1991);
- Het zien van samenhang tussen wiskundige concepten binnen en tussen de domeinen van de tussendoelen (NRC & MLSC, 2001).

Deze driedeling is voor het eerst beschreven door Bruin-Muurling (2010) en gebruikt door Van Stiphout (2011) voor het analyseren van leerlingresultaten. Het gebruik ervan als onderdeel van een leerlingmodel was niet eerder gedaan. Voor een meer uitgebreide beschrijving van het leerlingmodel, zie de voorstudie (Roelofs & Schouwstra, 2012) en de toetswijzer (CvTE, 2014).

Tabel 7-1. Leerlingmodel wiskunde (CvTE, 2014)

Vakdomeinen	Vakdidactische aspecten		
	Structuur	Meerduidigheid	Samenhang
Domein B: Getallen (en variabelen)	B1	B2	B3
Domein C: Verhoudingen	C1	C2	C3
Domein D: Meten en meetkunde	D1	D2	D3
Domein E: Verbanden en formules	E1	E2	E3
Domein F: Informatieverwerking en onzekerheid (alleen havo/vwo)	F1	F2	F3

Al vroeg in het proces van het ontwerp van DTT is onderkend dat het leerlingmodel vernieuwend was. De verwachting dat een diagnose op deze driedeling inderdaad verklarend zou zijn, was ambitieus. In 2012 is daarom een verkennend onderzoek gedaan onder drie klassen op een school waarbij de hoofdvraag was in hoeverre de uitwerking van een leerling gerelateerd kan worden aan het onderliggende aspect (structuur zien, objectvorming, samenhang zien) dat de opgave beoogt te meten. Een beperkt aantal leerlingen heeft een setje opgaven gekregen. Op basis van de geschreven antwoorden van leerlingen zijn aanvullende interviews gehouden met leerlingen over hun strategieën. De fouten die leerlingen maakten, leken inderdaad herleidbaar tot de vakdidactische aspecten.

In de loop van 2013 werden verschillende problemen met het leerlingmodel zichtbaar. Zo bleek het onder toetsdeskundigen niet mogelijk een goede interbeoordelaarsbetrouwbaarheid te realiseren bij het indelen van de ontwikkelde opgaven in *structuur*, *meerdere* of *samenhang*. Tevens waren de opgaven die door constructiegroep-leden werden ontwikkeld soms 'te plat' om in de driedeling te passen. Daarnaast waren er zorgen over de haalbaarheid van de meerwaarde van de driedeling. Tot slot kwam de vakdidactische driedeling niet herkenbaar naar voren in de statistische analyses van leerlingresponsen in de pretesten.

De toetsdeskundigen wiskunde hebben verschillende malen bestaande opgaven ingedeeld volgens deze driedeling om tot een beschrijving te komen van de indeling in de drie vakdidactische aspecten en de interbeoordelaarsbetrouwbaarheid van de indeling te vergroten. Eind 2015 en begin 2016 is gekeken of inhoudelijke perspectieven op het leerlingmodel te combineren waren met psychometrische perspectieven met het doel om te komen tot een empirisch onderbouwd leerlingmodel. Helaas leverde deze combinatie geen empirische onderbouwing van het bestaande leerlingmodel op noch een aanzet tot een alternatief leerlingmodel.

Hierop zijn verschillende alternatieven onderzocht waarbij het streven was tot een model te komen dat positieve punten van bestaande kaders zou combineren, tegemoet zou komen aan de beperkingen en kritische geluiden, en werkbaar zou zijn vanaf de constructiegroepen aan het begin van het traject tot aan de rapportage aan het einde van het traject. Een andere wens was dat het alternatieve leerlingmodel op een vrij eenvoudige wijze te koppelen zou zijn aan door scholen gebruikte indelingen als RTTI (Reproductie, Training – toepassing niveau 1, Transfer – toepassing niveau 2, Inzicht) en OBIT (Onthouden, Begrijpen, Integreren en Toepassen). De hoop was dat op die manier de rapportages meer betekenis zouden krijgen doordat docenten ze zouden kunnen koppelen aan hun eigen praktijk. De alternatieve driedeling die hieruit is voortgekomen is Kennis – Inzicht – Toepassen (KIT). Deze indeling is vergelijkbaar met het leerlingmodel van TIMSS (Trends in International Mathematics and Science Study), waarin Knowing, Applying en Reasoning de drie categorieën zijn (Mullis & Martin, 2014). Hoewel dit model een aantal voordelen lijkt te hebben, was er binnen de driejarige pilot DTT geen tijd en ruimte om het model verder uit te werken.

Aantallen opgaven

Tabel 7-2 laat het aantal opgaven zien dat is geconstrueerd en goedgekeurd na afname. Er zijn gemiddeld per deelaspect in een leerweg/schoolsoort 12 opgaven die samen 18 responsen opleveren. Voor vmbo zijn er iets meer opgaven per deelaspect (gemiddeld 14 opgaven die 20 responsen opleveren) dan in havo/vwo (gemiddeld 10 opgaven en 15 responsen). Van alle opgaven is 73% (70% van de responsen) gericht op één leerweg/schoolsoort (uniek), de overige 27% van de opgaven is geschikt voor twee of drie leerwegen/schoolsoorten.

Tabel 7-2. Aantal goedgekeurde opgaven en responsen na de adaptieve afname 2017 per deelaspect van het leerlingmodel

	vmbo-bb	vmbo-kb	vmbo-gt	Havo	Vwo	Totaal uniek
	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen
B1	26 (28)	22 (24)	15 (24)	10 (21)	11 (17)	75 (97)
B2	15 (32)	15 (34)	14 (19)	10 (15)	11 (19)	59 (106)
B3	14 (18)	12 (16)	13 (15)	8 (22)	9 (24)	40 (61)
C1	20 (25)	18 (23)	18 (22)	10 (10)	9 (10)	63 (71)
C2	12 (20)	12 (20)	18 (27)	9 (11)	8 (10)	32 (43)
C3	11 (18)	13 (17)	16 (16)	12 (20)	7 (15)	45 (61)
D1	17 (23)	21 (22)	15 (16)	10 (14)	10 (14)	54 (66)
D2	15 (32)	13 (24)	14 (20)	8 (14)	8 (14)	45 (85)
D3	10 (14)	12 (16)	11 (14)	10 (12)	13 (15)	34 (45)
E1	12 (14)	16 (24)	15 (21)	12 (16)	14 (17)	55 (76)
E2	8 (8)	12 (13)	11 (12)	12 (20)	9 (10)	33 (43)
E3	11 (17)	9 (15)	10 (19)	5 (11)	5 (11)	34 (61)
F1				12 (16)	9 (13)	14 (18)
F2				10 (16)	11 (18)	15 (22)
F3				9 (13)	7 (11)	11 (15)
Totaal	171 (249)	175 (248)	170 (225)	147 (231)	141 (218)	609 (870)

7.2 Operationalisatie

In Tabel 7-3 is het totale aantal goedgekeurde opgaven weergegeven per opgavetype. De opgaven zijn geconstrueerd door vier constructiegroepen, twee voor vmbo en twee voor havo/vwo.

De constructiegroepen bestonden ieder uit drie docenten die op dat moment werkzaam waren in het onderwijstype waarvoor zij opgaven construeerden. Aan de hand van de twee dimensies, de inhoudelijke dimensie (de vakdomeinen) en de vakdidactische dimensie, is een beschrijving gemaakt per deelaspect aan de hand waarvan constructiegroepsleden opgaven hebben geconstrueerd. Nagenoeg alle opgaven zijn voorgelegd aan de vaststellingscommissie (zie Hoofdstuk 2). Na de pretesten zijn opgaven die opvielen in de analyses opnieuw voorgelegd aan de vaststellingscommissie. Als gevolg hiervan zijn opgaven behouden, aangepast of afgekeurd.

Bij sommige opgaven duurde het wat langer voor ze naar de vaststellingscommissie konden omdat er, vooral in het begin, technische problemen waren bij de opgaven. De wensen van de constructiegroepen en toetsdeskundigen wat betreft de digitale opgaven liepen niet altijd synchroon met de technische mogelijkheden. De digitale mogelijkheden zijn gedurende de constructie flink toegenomen met de komst van onder andere de Digitale Wiskunde Omgeving, Geogebra en het computeralgebrasysteem Maxima dat is ingezet in de automatische beoordeling.

Verschillende beschikbare opgavetypen (zie Hoofdstuk 2) zijn speciaal ontwikkeld voor de talen en niet gebruikt voor het ontwikkelen van opgaven voor wiskunde. Dit betreft de typen alinea-opgave, correctie-opgave en markeer-opgave. Het opgavetype multiple response is in de loop der tijd minder gebruikt bij wiskunde. Uit de analyse van de pretesten bleek dat dit opgavetype problematisch is omdat leerlingen hierin niet noodzakelijkerwijs een keuze voor een antwoord hoeven te maken. Geen keuze maken is ook een antwoord maar hoeft geen geschikte afspiegeling te zijn van de vaardigheid van de leerling.

De opgavetypen meerkeuze en kort-open komen verreweg het vaakst voor. Vanaf de start van de constructie waren dit de opgavetypen die het meest pasten bij de operationalisatie van het leerlingmodel en de toenmalige technische mogelijkheden. De opgavetypen DWO en het gebruik van computeralgebrasysteem waren mogelijk vanaf medio 2015. In de loop van 2016 werd het mogelijk om Geogebra-opgaven te construeren en op te nemen in de toetsen. Die ontwikkelingen worden uitgebreider beschreven in Hoofdstuk 8.

Tabel 7-3. Aantal goedgekeurde opgaven en responsen na de adaptieve afname 2017 per opgavetype

	vmbo-bb	vmbo-gt	vmbo-kb	havo	vwo	Totaal uniek
	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen
Categoriseer-opgave	0 (0)	1 (1)	0 (0)	0 (0)	0 (0)	1 (1)
Combinatie-opgave	5 (8)	5 (10)	6 (13)	3 (5)	5 (9)	16 (30)
Dropdown-opgave	16 (26)	9 (14)	12 (18)	12 (33)	8 (26)	44 (89)
DWO-opgave	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	2 (2)
Geogebra-opgave	0 (0)	3 (3)	1 (1)	2 (2)	2 (2)	6 (6)
Hotspot-opgave	2 (2)	2 (2)	2 (2)	0 (0)	1 (1)	7 (7)
Korte open-opgave	77 (97)	74 (88)	79 (105)	55 (64)	54 (61)	261 (316)
Matrix-opgave	12 (46)	13 (42)	10 (33)	18 (38)	15 (38)	50 (140)
Meerkeuze-opgave	42 (42)	54 (54)	47 (48)	37 (37)	37 (37)	162 (163)
Multiple response-opgave	2 (2)	1 (1)	3 (3)	5 (11)	6 (12)	13 (24)
Sleep-opgave	6 (8)	3 (5)	5 (9)	9 (28)	5 (16)	23 (50)
Sleep-opgave afbeelding	5 (14)	1 (1)	3 (9)	3 (10)	4 (12)	11 (29)
Volgorde-opgave	3 (3)	3 (3)	6 (6)	2 (2)	3 (3)	13 (13)
Totaal	171 (249)	170 (225)	175 (248)	147 (231)	141 (218)	609 (870)

In Tabel 7-4 staan de aantallen afgekeurde opgaven en responsen per deelaspect. In het algemeen kan gesteld worden dat de geconstrueerde opgaven aan de moeilijke kant waren voor de niveaus waarvoor ze waren bedoeld. In de loop van de constructie is gepoogd hierop bij te sturen. De noodzaak tot het construeren van makkelijkere opgaven in combinatie met het verkennen van de uitbreiding van de technische mogelijkheden zorgde voor een grote druk op de constructiegroepen.

In Tabel 7-5 staan de aantallen afgekeurde opgaven en responsen per opgavetype. Opvallend is dat er geen Geogebra-opgaven zijn afgevalen, ondanks dat lang niet alle leerlingen gewend zijn aan dit type opgaven (zie ook Hoofdstuk 8).

Tabel 7-4. Aantal afgekeurde opgaven en responsen na de adaptieve afname 2017 per deelaspect van het leerlingmodel wiskunde

	vmbo-bb	vmbo-gt	havo	vmbo-kb	vwo	Totaal uniek
Deelaspect	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen
B1	4 (4)	6 (6)	3 (7)	4 (4)	2 (6)	18 (22)
B2	5 (5)	9 (11)	5 (13)	6 (9)	4 (9)	25 (38)
B3	2 (6)	6 (13)	6 (17)	4 (11)	4 (15)	14 (32)
C1	1 (1)	3 (3)	5 (9)	6 (6)	3 (7)	14 (18)
C2	3 (3)	4 (4)	4 (5)	4 (4)	3 (4)	8 (9)
C3	6 (6)	4 (5)	3 (4)	8 (8)	4 (5)	16 (18)
D1	6 (6)	6 (6)	5 (5)	6 (6)	4 (4)	18 (18)
D2	2 (2)	9 (14)	3 (3)	2 (5)	2 (2)	15 (23)
D3	5 (5)	6 (6)	5 (6)	5 (5)	3 (4)	14 (15)
E1	2 (2)	5 (10)	4 (9)	4 (5)	1 (1)	14 (24)
E2	4 (4)	4 (4)	2 (2)	4 (4)	4 (5)	10 (11)
E3	4 (8)	5 (12)	9 (17)	5 (10)	8 (11)	24 (43)
F1			2 (4)		2 (4)	2 (4)
F2			3 (3)		0 (0)	3 (3)
F3			5 (8)		4 (7)	5 (8)
Totaal	44 (52)	67 (94)	64 (112)	58 (77)	48 (84)	200 (286)

Tabel 7-5. Aantal afgekeurde opgaven en responsen na de adaptieve afname 2017 per opgavetype wiskunde

	vmbo-bb	vmbo-gt	Havo	vmbo-kb	vwo	Totaal uniek
	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen	Aantal opgaven en responsen
Categoriseer-opgave	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Combinatie-opgave	3 (7)	2 (3)	2 (2)	3 (6)	1 (1)	8 (13)
Dropdown-opgave	1 (1)	3 (4)	3 (14)	2 (5)	2 (10)	8 (23)
DWO-opgave	0 (0)	1 (1)	0 (0)	1 (1)	0 (0)	1 (1)
Geogebra-opgave	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Hotspot-opgave	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Korte open-opgave	17 (17)	26 (30)	34 (52)	30 (34)	26 (41)	100 (123)
Matrix-opgave	1 (5)	7 (28)	10 (25)	3 (12)	5 (17)	18 (57)
Meerkeuze-opgave	18 (18)	23 (23)	10 (10)	16 (16)	8 (8)	50 (50)
Multiple response-opgave	1 (1)	1 (1)	3 (6)	1 (1)	4 (4)	6 (9)
Sleep-opgave	0 (0)	1 (1)	0 (0)	0 (0)	0 (0)	1 (1)
Sleep-opgave afbeelding	1 (1)	1 (1)	2 (3)	1 (1)	2 (3)	3 (4)
Volgorde-opgave	2 (2)	2 (2)	0 (0)	1 (1)	0 (0)	5 (5)
Totaal	44 (52)	67 (94)	64 (112)	58 (77)	48 (84)	200 (286)

7.3 Literatuur

- Bruin-Muurling, G. (2010). *The development of proficiency in the fraction domain: Affordances and constraints in the curriculum* (Dissertatie). Eindhoven: Eindhoven University of Technology. Also available at: <https://pure.tue.nl/ws/files/2840528/692951.pdf>
- College voor Toetsen en Examens (2014). *Publieksversie toetswijzer diagnostische tussentijdse toets voor Nederlands, Engels en wiskunde*. Utrecht: College voor Toetsen en Examens. <https://www.pilotdt.nl/documenten/publicaties/2014/12/15/toetswijzer-dtt>
- Devlin, K. (2012). *Introduction to Mathematical Thinking*. Petaluma, CA: Devlin.
- Kraemer, J.-M. (2011). *Oplossingsmethoden voor aftrekken tot 100* (Dissertatie). Arnhem: Cito. An English summary can be found in the digital version of the thesis at www.alexandria.tue.nl/extra2/721544.pdf
- Mullis, I. V., & Martin, M. O. (2014). *TIMSS Advanced 2015 Assessment Frameworks*. International Association for the Evaluation of Educational Achievement. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement.
- National Research Council, & Mathematics Learning Study Committee. (2001). *Adding it up: Helping children learn mathematics*. National Academies Press.
- Roelofs, E., & Schouwstra, S. (Red.). (2012). *Diagnostische tussentijdse toets: Verslag van de voorstudie*. Arnhem: Cito.
- Roorda, G. (2012). *Ontwikkeling in verandering: ontwikkeling van wiskundige bekwaamheid van leerlingen met betrekking tot het concept afgeleide*. University Library Groningen.
- Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics*, 22(1), 1–36.
- Van Bijsterveldt-Vliegthart, M. (2011, 5 mei). *Actieplan beter presteren* [bijlage Kamerbrief]. Geraadpleegd van <https://www.rijksoverheid.nl/documenten/kamerstukken/2011/05/23/actieplan-vo-beter-presteren>.
- Van Stiphout, I. M. (2011). *The development of algebraic proficiency* (Dissertatie). Technische Universiteit Eindhoven). Geraadpleegd op: <http://alexandria.tue.nl/extra2/719774.pdf>.

8 Automatische beoordeling wiskunde

Paul Drijvers en Joke Hofstee

8.1 Probleemstelling

De DTT is een adaptieve toets: de toets past zich aan aan het niveau van de leerling om de zwakke en sterke punten van de leerling efficiënt in kaart te brengen. Om dit te kunnen realiseren is het nodig dat de leerlingresponsen automatisch worden beoordeeld. Automatische beoordeling is dus een voorwaarde voor adaptiviteit.

Hoewel het onderwerp internationaal sterk in de belangstelling staat (zie bijvoorbeeld Fife, 2011; Stacey & Wiliam, 2013; Williamson, Mislevy & Bejar, 2006) en er veel ontwikkelingen zijn op dit gebied, moeten we constateren dat kant-en-klare oplossingen voor automatische beoordeling bij wiskundetoetsen niet voorhanden waren. De realisatie van een adequate automatische beoordeling was dan ook een van de grote uitdagingen van de DTT wiskunde. Deze uitdaging concentreerde zich op drie kernpunten: (1) de automatische beoordeling van opgaven waarin leerlingen constructieruimte hebben, (2) de 'intelligentie' van automatische beoordeling, en (3) de beoordeling van tussenstappen. We lichten deze drie kernpunten nader toe.

- 1) *Automatische beoordeling van opgaven waarin leerlingen constructieruimte hebben*
Bij traditionele schriftelijke wiskundetoetsen kunnen leerlingen in hun uitwerkingen eenvoudig tekst, formules, grafieken en berekeningen combineren en integreren, en op kladpapier schetsjes tekenen of proefberekeningen maken. Dergelijke ruimte om 'wiskunde te doen' is in een digitale toetsomgeving veelal beperkt, terwijl die juist zeer relevant is bij diagnostische toetsing. De uitdaging is dus om ook bij digitale wiskundetoetsen opgaven te ontwerpen waarin de leerling iets kan maken, en constructieruimte ervaart, zoals dat ook op papier aan de orde is. Dit vraagt om de beschikbaarheid van wiskundig gereedschap en stelt specifieke eisen aan automatische beoordeling.
- 2) *De 'intelligentie' van automatische beoordeling*
Net als bij andere vakken is de adequate beoordeling van responsen bij wiskunde een subtiële aangelegenheid. Bij sommige opgaven zullen de antwoorden x^2-9 en $(x-3)(x+3)$ beide goed moeten worden gerekend, terwijl bij opgaven waarin de vraag is om uit te werken of te ontbinden juist alleen het eerste dan wel het tweede antwoord als correct dient te worden beschouwd. Voor een menselijke beoordelaar zijn dit soort beslissingen helder. Bij automatische beoordeling is het zaak om mechanismen voor deze subtiliteit te implementeren, samen met mogelijkheden voor de opgaveconstructeur om deze per opgave aan te sturen.
- 3) *De beoordeling van tussenstappen*
Veel toetsopgaven bij wiskunde vragen om een stapsgewijze oplossing. In elke stap van dit proces kan een leerling een fout maken, waarmee vervolgens wordt doorgerekend. Een menselijke beoordelaar, zoals een wiskundedocent, zal in dergelijke gevallen de uitwerking veelal met een deelscore belonen. Met name voor diagnostische doeleinden kunnen dergelijke beoordelingen van belang zijn. Bij de automatische beoordeling van digitale responsen staat de beoordeling van tussenstappen echter nog in de kinderschoenen, terwijl het wel een vraag is van docenten uit het veld.

8.2 Werkwijze

Het werk aan de realisatie van de automatische beoordeling van wiskunde-opgaven is uitgevoerd in nauwe samenwerking tussen toetsdeskundigen wiskunde van Cito, softwareontwikkelaars en architecten van Cito en van ketenpartners zoals CvTE en DUO die verantwoordelijk zijn voor de ontwikkeling van de afspeelomgeving Facet. In deze continue ontwikkelingen zijn de volgende mijlpalen te identificeren:

2014: Het Cito-interne onderzoek ‘Digitale diagnostische toetsing van wiskunde: onderzoek van een nieuwe omgeving’

Dit betreft een onderzoek in het kader van de interne onderzoeksagenda van de unit CTE van Cito. In dit onderzoek is een variant op de DTT ontwikkeld binnen de Digitale Wiskunde Omgeving van de Universiteit Utrecht, die op dat moment meer mogelijkheden bood dan de keten QuestifyBuilder - Facet. Deze variant is afgenomen bij 578 leerlingen van klas 3 havo/vwo. De conclusie is dat een omgeving zoals de Digitale Wiskunde Omgeving het door de beschikbaarheid van wiskundige middelen zoals een formule-editor, een grafiekenvenster en een meetkundevenster mogelijk maakt gevarieerde en rijke opgaven te ontwerpen. Deze bieden een grote constructieruimte aan leerlingen en maken daarmee een meer genuanceerde diagnose haalbaar (zie Drijvers, Visser, Straat & Schouwstra, 2014).

2015: Internationale expertmeeting Automatische beoordeling

Dit was een tweedaagse conferentie die Cito in samenwerking met CvTE heeft georganiseerd in Arnhem. Zes toonaangevende nationale en internationale experts hebben de stand van zaken op het terrein van automatische beoordeling van digitale reken- en wiskundetoetsen gepresenteerd aan een vijftigtal deelnemers. De nadruk daarbij lag op de beoordeling van tussenstappen. De conclusies zijn dat er behoefte is aan de inbouw van tools voor grafieken en meetkunde in Facet en dat de noodzaak van beoordeling van tussenstappen groter is voor een diagnostische toets zoals de DTT dan voor een summatieve toets zoals de Rekentoets VO (Drijvers & van Reeuwijk, 2015).

2015-2016: Uitbreiding QuestifyBuilder en Facet

De belangrijkste uitbreidingen van QuestifyBuilder en Facet vonden plaats in de periode 2015-2016 en betreffen de ontwikkeling van een formule-editor (native in Facet), de integratie van een module van de Digitale Wiskunde Omgeving voor het tekenen van grafieken, de integratie van het meetkundepakket Geogebra, en het gebruik van het computeralgebrasysteem Maxima voor de automatische beoordeling van formules en algebraïsche expressies. Deze uitbreidingen hebben ook geleid tot aanpassingen van de mogelijkheden voor opgaveconstructeurs om deze nieuwe middelen binnen QuestifyBuilder, de auteursomgeving, in te zetten.

2016: Interactieonderzoek

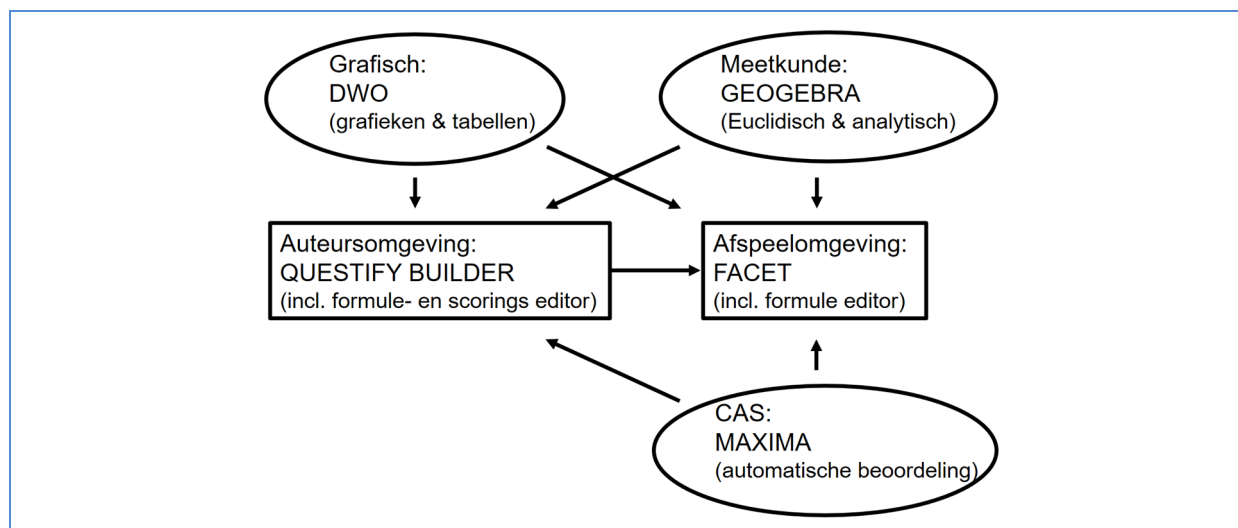
Dit was een onderzoek naar de interacties van leerlingen met de nieuwe middelen binnen Facet. Aan de hand van een gedetailleerde observatie van 19 leerlingen uit klas 3 van havo, vwo en/of vmbo-gt die werken met de DTT, zijn de interacties van de leerlingen met Facet geanalyseerd, waarbij met name gekeken is naar de nieuwe mogelijkheden van de toetsomgeving. Dit heeft geleid tot ontwerprichtlijnen voor opgaven waarin de Digitale Wiskunde Omgeving of Geogebra worden gebruikt (Groenheiden, 2016).

2016-2017: Deelname Erasmus+ project Advise-Me

In een consortium met de Open Universiteit, de Universiteit Utrecht, de Universiteit Paris-Est Créteil en de Universiteit van Saarland is een Erasmus+ onderzoeksvoorstel ingediend met de titel “Automatic Diagnostics with Intermediate Steps in Mathematics Education”. Dit project is gehonoreerd en in het najaar van 2016 van start gegaan. Het richt zich met name op de automatische beoordeling van tussenstappen. Hoewel dit onderzoek niet onderdeel is van de pilot DTT en de resultaten pas na afloop van het DTT-project beschikbaar zullen zijn, is het onderzoek vanuit een behoefte van de DTT ontstaan.

8.3 Resultaten

De resultaten van de ontwikkeling van automatische beoordeling worden besproken aan de hand van de drie in 3.1.1 onderscheiden kernpunten. Als leidraad daarbij kan het overzicht van de architectuur van QuestifyBuilder en Facet dienen, zoals afgebeeld is in Figuur 8-1.

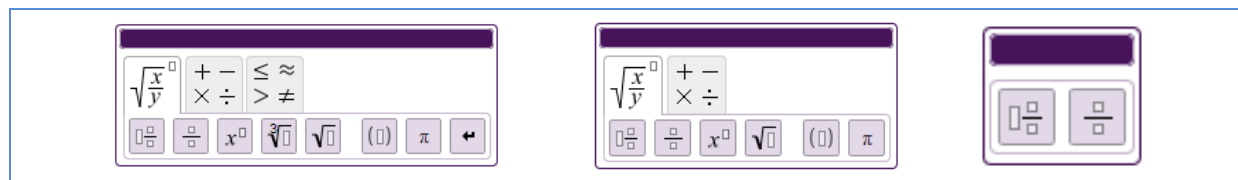


Figuur 8-1. Overzicht van wiskundige middelen in de architectuur van QuestifyBuilder en Facet

Het eerste kernpunt betreft de **automatische beoordeling van opgaven waarin leerlingen constructieruimte hebben**. In dit kader is het volgende gerealiseerd.

Formules invoeren

Binnen Facet is een formule-editor ingebouwd, waarmee leerlingen wiskundige formules en expressies kunnen invoeren. Het formulevenster kan uit één of meerdere regels bestaan. De ingevoerde formules worden in een standaard format (MathML) geïnterpreteerd en zijn daarmee ook geschikt voor automatische beoordeling. De opgaveconstructeur kan bij het ontwerpen van een opgave de volledige formule-editor beschikbaar stellen, maar de constructeur kan ook kiezen voor een beperkte of een minimale variant, die meer geschikt is voor bijvoorbeeld de DTT in het vmbo of de Rekentoets VO (zie Figuur 8-2).



Figuur 8-2. De formule-editor in volledige, beperkte of minimale uitvoering

Grafieken tekenen

De grafische module van de Digitale Wiskunde Omgeving (Universiteit Utrecht) is geschikt gemaakt voor gebruik in de keten QuestifyBuilder - Facet (zie linksboven in Figuur 8-1). Hierdoor kunnen opgaven worden ontworpen waarin leerlingen bijvoorbeeld punten en grafieken tekenen. In de vorm van zogeheten 'Custom Interactions' kunnen deze activiteiten worden ingebed in toetsopgaven en is de respons van leerlingen beschikbaar voor automatische beoordeling.

In het linker scherm van Figuur 8-3 is een voorbeeldopgave afgebeeld. De vraag is om door middel van twee punten de somgrafiek van de twee gegeven grafieken te tekenen. In het rechter scherm heeft de leerling twee punten aangeklikt en de lijn erdoor getrokken. De automatische beoordeling zorgt ervoor dat elk tweetal punten dat op die lijn ligt als correct wordt herkend. Dit geeft de leerling flexibiliteit en maakt dat verschillende manieren van oplossen als correct kunnen worden herkend.



Figuur 8-3. Fictieve opgave die gebruik maakt van de grafische module van de Digitale Wiskunde Omgeving (DWO)

Meetkundige constructies uitvoeren

Door de inbedding van een pakket voor dynamische meetkunde, Geogebra, kunnen leerlingen in Facet meetkundige constructies maken (zie rechtsboven in Figuur 8-1). Denk bijvoorbeeld aan het construeren van cirkels of hoogtelijnen. Ook deze constructies zijn geschikt voor automatische beoordeling. Geogebra is een vrij uitgebreid pakket, waarvan de mogelijkheden groter zijn dan in veel gevallen nodig is voor de DTT. Net als bij de formule-editor is de menubalk in te perken. Afhankelijk van de opgave kan de opgaveconstructeur kiezen welke knoppen voor de leerling beschikbaar zijn. In Figuur 8-4 staan twee varianten. Het ligt voor de hand om te streven naar een beperkt aantal varianten van de menubalk, zodat leerlingen daaraan gewend kunnen raken.

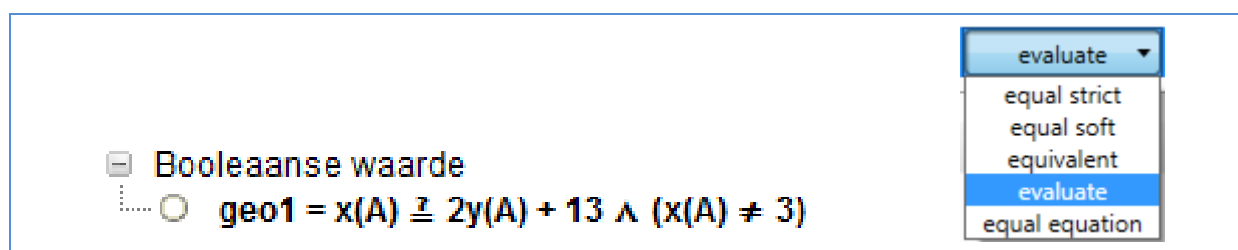


Figuur 8-4. Twee varianten van de menubalk van Geogebra

Het tweede kernpunt betreft de **'intelligentie' van automatische beoordeling**. Met intelligentie bedoelen we hier dat opgaven die de kandidaten door de constructieruimte op een groot aantal verschillende manieren correct kunnen beantwoorden ook als zodanig worden beoordeeld. Het gaat er dus om dat alle correcte responsen als zodanig worden herkend. Dit vindt op twee manieren plaats: door het gebruik van Boolse variabelen in Geogebra (zie Kader 8-1) en door de inzet van een computeralgebrasysteem (CAS), in dit geval het open source programma Maxima (zie rechtsonder in Figuur 8-1). Het ingezette computeralgebrasysteem kan onder andere nagaan of een expressie exact gelijk dan wel algebraïsch equivalent is aan de sleutel, en of een algebraïsche uitdrukking voldoet aan een gespecificeerd criterium (zie Kader 8-1). Om deze beoordelingswijze (met uitzondering van de laatste) in te stellen, kan de toetsconstructeur gebruik maken van de betreffende opties in de scoringsmodule van QuestifyBuilder, afgebeeld in Figuur 8-5 (rechts).

Kader 8-1. Gerealiseerde automatische beoordeling

- Met de instelling *equal strict* gaat het CAS na of een expressie exact gelijk is aan de door de toetsconstructeur ingevoerde sleutel. Deze instelling beoordeelt dus zeer streng: als de sleutel gelijk is aan $x+3$, dan wordt de respons $3+x$ als incorrect beoordeeld.
- In de instelling *equivalent* gaat het CAS na of de sleutel en de respons algebraïsch equivalent zijn. Deze beoordeling is minder streng: als de sleutel gelijk is aan $x+3$, dan wordt ook de respons $3+x$ als correct beoordeeld, net als de responsen $(x^2-9)/(x-3)$ en $x+1+1+1$.
- De instelling *equal soft*, die niet standaard in Maxima aanwezig is maar waarvoor een script is ontwikkeld, houdt qua strengheid het midden tussen de twee eerdergenoemde opties: als de sleutel gelijk is aan $x+3$, dan wordt de respons $3+x$ als correct beoordeeld, maar de minder vereenvoudigde vormen $(x^2-9)/(x-3)$ en $x+1+1+1$ niet.
- In de instelling *evaluate* gaat het CAS na of een algebraïsche uitdrukking voldoet aan een door de opgaveconstructeur opgegeven criterium. Deze optie is met name nuttig als gevraagd wordt om expressies te genereren die aan bepaalde voorwaarden voldoen. Stel bijvoorbeeld dat de vraag is om via een vergelijking een parabool te tekenen die door het punt (3, 9) gaat. In principe zijn er oneindig veel parabolen waarvoor dit geldt, en evenveel bijbehorende formules. Om de respons nu te beoordelen, wordt nagegaan of gegeven uitdrukking inderdaad kwadratisch is en of de waarde voor $x=3$ gelijk is aan 9.
- De instelling *dependency* heeft betrekking op een aantal verschillende numerieke invoervakken of verschillende coördinaten van punten. Nagegaan wordt of de tweede invoer in een bepaalde relatie staat tot de eerste. Deze optie kan ten eerste worden gebruikt om na te gaan of met een eventuele foutief antwoord op correcte wijze wordt doorgerekend. Een tweede toepassing is om na te gaan of een punt in het vlak aan bepaalde voorwaarden voldoet. In de opgave afgebeeld in Figuur 8-3 worden de coördinaten van de twee door de leerling ingevoerde punten afgelezen. De eerste coördinaat van het eerste punt, zeg a , mag elke waarde hebben. De tweede coördinaat moet er echter voor zorgen dat het punt op de gevraagde lijn met vergelijking $y = 5 - x/2$ ligt. Daarom voert de opgaveontwerper met *dependency* de voorwaarde $b = 5 - a/2$ in. Voor de coördinaten van het tweede punt, (c, d) , zijn dan de aanvullende afhankelijkheden dat c niet gelijk is aan a en dat $d = 5 - c/2$.
- Met de instelling *equal equation* gaat het CAS na of twee opvolgende vergelijkingen equivalent zijn, dat wil zeggen dezelfde oplossingsverzamelingen hebben.
- Specifiek voor Geogebra is het gebruik van Boolse variabelen. Dit zijn variabelen die de toetsconstructeur in Geogebra definieert en die de waarden 0 of 1 hebben. Als de kandidaat bijvoorbeeld een punt A moet bepalen op de lijn met vergelijking $x = 2y + 13$ en met x -coördinaat ongelijk aan 3, dan kan de Boolse variabele in Figuur 8-5 (links) beoordelen of dit correct is uitgevoerd. De waarde van deze variabele, 0 of 1 voor niet waar of waar, komt binnen bij Facet en kan automatisch worden beoordeeld.



Figuur 8-5. Gereedschap voor de toetsconstructeur: Boolse variabele in Geogebra (links) en keuzemenu in Questify (rechts)

Het derde kernpunt betreft *de beoordeling van tussenstappen*. Dit is het lastigste punt van de drie, dat in het kader van DTT slechts in beperkte mate is gerealiseerd. Een eerste manier om dit te doen is om foutieve alternatieven bij meerkeuze opgaven dan wel foutieve antwoorden bij open opgaven te interpreteren in termen van een fout bij een tussenstap. Als in een opgave $3 \cdot 4^2$ berekend moet worden, dan wijst het antwoord 144, of de keuze voor 144 als alternatief, erop dat het kwadrateren pas na het vermenigvuldigen heeft plaatsgevonden en niet ervoor. Deze aanpak is in de DTT weinig gebruikt. Een tweede manier is om expliciet om het noteren van tussenstappen te vragen en gebruik te maken van de optie *dependency*. Als bijvoorbeeld gevraagd wordt hoeveel geld er op een spaarrekening staat na

2 jaar, waarvan het eerste jaar met 3% en het tweede jaar met 2% rente, dan kan als tussenstap om het bedrag na 1 jaar worden gevraagd. Als dat niet correct is, kan met dependency alsnog worden nagegaan of het eindantwoord wel 1,02 keer zo groot is als het tussenantwoord. Het nadeel aan deze methode is dat de deelstappen al sterk voorgestructureerd moeten worden, terwijl het bedenken van de sequentie van tussenstappen juist een kenmerk is van het probleemoplossend vermogen van de kandidaat. Om deze reden is van deze methode weinig gebruik gemaakt. Een derde aanpak, die nog niet is gerealiseerd, betreft het gebruik van meerregelige antwoordvakken en van zogeheten 'domain reasoners' om de tussenstappen van leerlingen te beoordelen (Drijvers, ingediend). Deze aanpak is het onderwerp van het Erasmus+ project Advise-Me.

8.4 Toekomstperspectieven

Er is binnen de DTT belangrijke voortgang geboekt ten aanzien van het automatisch beoordelen van wiskunde. Het is zinvol om de resultaten van DTT in te zetten bij andere digitale toetsen rekenen en wiskunde, waarin automatische beoordeling een rol speelt. Denk hierbij aan de Rekentoets VO en aan de computereexamens wiskunde van vmbo, maar ook aan de Adaptieve Centrale Eindtoets. Ten derde moeten de mogelijkheden verder worden ontwikkeld, zoals dat bijvoorbeeld momenteel plaatsvindt in het Erasmus+ project Advise-Me, waarin door middel van 'domain reasoners' en 'student modeling' wordt gewerkt aan de beoordeling van tussenstappen (zie <http://advise-me.ou.nl>).

8.5 Literatuur

- Drijvers, P. (ingediend). Digital assessment of mathematics: opportunities, issues and criteria. *Mesure et évaluation en education*.
- Drijvers, P., & van Reeuwijk, M. (2015). *Automatische beoordeling van wiskunde*. Rapportage expertmeeting Arnhem, 16-18 februari 2015. Arnhem/Utrecht: Cito/CvTE.
- Drijvers, P., Visser, S., Straat, H., & Schouwstra, S. (2014). Digitale diagnostische toetsing van wiskunde. Onderzoek van een nieuwe omgeving. In S. Schouwstra (Red.), *Diagnostische tussentijdse toets: Onderzoek 2014* (pp.53-70). Arnhem: Cito.
- Fife, J. H. (2011). *Automated scoring of CBAL mathematics tasks with m-rater*. Research Memorandum. Princeton, NJ: ETS. <http://www.ets.org/Media/Research/pdf/RM-11-12.pdf>.
- Groenheiden, M. (2016). *Intuïtieve Interacties van leerlingen met digitale hulpmiddelen tijdens de Diagnostische Tussentijdse Toets Wiskunde*. Intern rapport. Arnhem: Cito.
- Stacey, K. & Wilam, D. (2013). Technology and Assessment in Mathematics. In M. A. Clements, A. Bishop, C. Keitel, J. Kilpatrick, & F. Leung (Eds.), *Third International Handbook of Mathematics Education* (pp. 721-751). New York/Berlin: Springer.
- Williamson, D. M., Mislevy, R. J., & Bejar, I. (Eds.) (2006). *Automated scoring of complex tasks in computer-based testing*. Mahwah, NJ: Lawrence Erlbaum Associates.

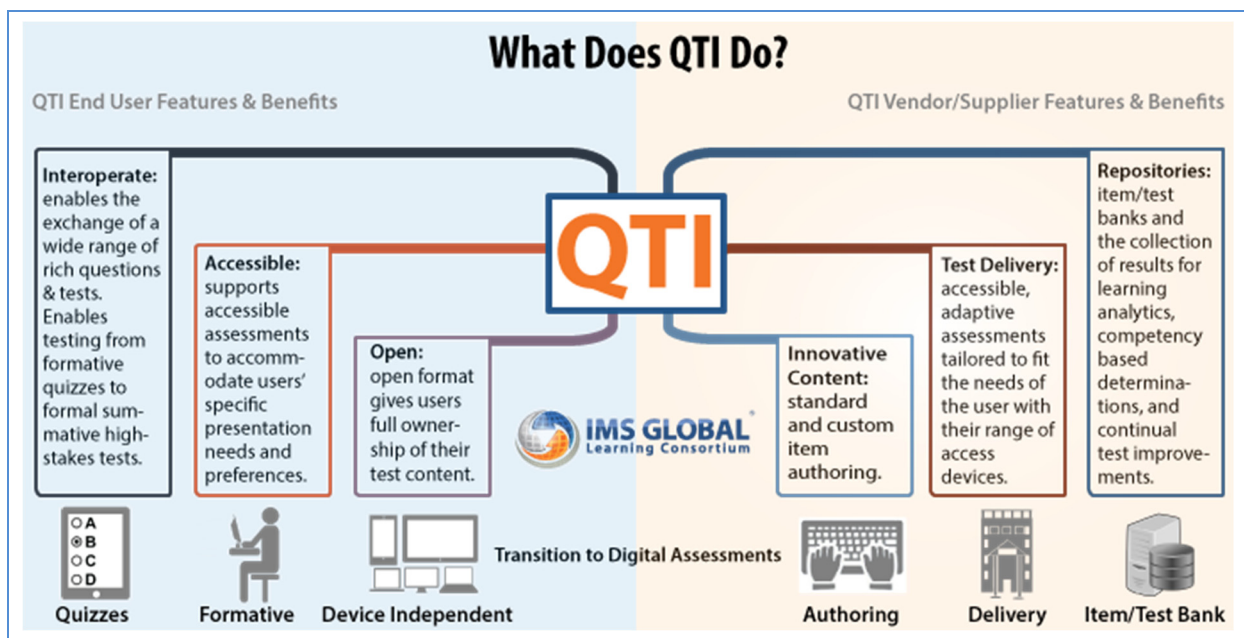
9 Uitwisseling van informatie tussen de auteursomgeving en Facet

Arjan Aarnink

9.1 Inleiding

In toets- en examenketens is er veelal sprake van meerdere stakeholders, meerdere leveranciers en meerdere systemen. Om uitwisseling van informatie en de koppeling van systemen zo goed mogelijk te laten verlopen, wordt gebruikt gemaakt van open standaarden. Open standaarden zijn goed gedocumenteerd, er zitten geen restricties aan het gebruik en ze zijn gratis te gebruiken. In de Facet-keten (waar ook de DTT in was opgenomen) wordt gebruikt gemaakt van Question & Test Interoperability (QTI) (<https://www.imsglobal.org/activity/qtiapip>). Deze standaard wordt beheerd door IMS Global Learning Consortium (<https://www.imsglobal.org/>).

QTI is een standaard om toetsmateriaal uit te wisselen. QTI beschrijft de toets, de items, de opmaak, de scoring en het kandidaatresultaat. In dit hoofdstuk worden de termen item en interactie gebruikt zoals ze worden gebruikt in de QTI-standaard (zie ook Hoofdstuk 3). In Figuur 9-1 staan globaal de QTI kenmerken beschreven.



Figuur 9-1. QTI kenmerken (bron: <http://www.imsglobal.org/activity/qtiapip>)

9.2 Dutch Exam Profile

Het exacte gebruik van de QTI-standaard voor Facet is beschreven in het 'Dutch Exam Profile' (DEP). Het DEP is een toepassingsprofiel met extensies van de internationale standaard voor de uitwisseling van toetsmateriaal IMS QTI (zie ook Edustandaard, 2017). Ten tijde van het DTT-project is versie 2.1 gebruikt van de DEP.

De extensies van het DEP zijn toegevoegd om functionaliteiten te ondersteunen die (nog) niet door QTI worden ondersteund. Men kan hierbij denken aan speciaal gedrag en presentatie van items (zie ook sectie 9.4). Maar ook bijzondere vormen van scoring kunnen in het DEP worden beschreven. Het streven was om deze extensies zo beperkt mogelijk te houden. Ook zet Cito zich in om de betreffende functionaliteiten ondersteund te krijgen in nieuwe versies van QTI. Met de komst van QTI2.2 en QTI3 zullen veel functionaliteiten die nu nog in het DEP zijn gespecificeerd, in QTI beschikbaar komen. Het doel is dan ook

om het DEP in de toekomst grotendeels te ontmantelen zodat een betere interoperabiliteit van toetsmateriaal wordt gerealiseerd.

Het DEP wordt beheerd door Edustandaard, de organisatie die zich bezighoudt met onderwijsstandaarden in het Nederlandse onderwijs. Meer informatie over de DEP-standaard is te vinden op de website van Edustandaard (https://www.edustandaard.nl/standaard_afspraken/dutch-exam-profile-dep/).

9.3 Toepassing van standaarden in de keten

De QTI-standaard wordt grofweg op twee plaatsen in de Facet-keten gebruikt:

- 1) *Uitwisseling tussen het Cito-auteurssysteem (Questify Builder) en Facet*
Deze uitwisseling is te vinden in het linker groene blokje in het schema in Figuur 9-2. Het betreft een DEP-package met daarin alle informatie over de items en toetsen. De DEP-package wordt aangeboden als zipbestand. Het zipbestand wordt getekend en versleuteld ('signed and encrypt'), zodat alleen de partijen de inhoud kunnen lezen die daartoe gemachtigd zijn. Daarnaast wordt er een Facet-manifest meegestuurd. In dit manifest staat beschreven voor wie de toets bestemd is. Zo worden onder meer de doelgroep (bijvoorbeeld schoolsoort/leerweg) en afnameperiode (de periode waarbinnen de toets afgenomen mag worden) beschreven.
- 2) *Uitwisseling van leerlingresultaten tussen Facet en Cito-analyse tools*
Deze uitwisseling is te vinden in het rechter groene blokje van het schema in Figuur 9-2. In deze stap worden de leerlingantwoorden en uitkomsten uitgewisseld tussen Facet en Cito. Deze uitwisseling wordt in de facet-keten 'datadump' genoemd en is gebaseerd op het QTI-schema voor 'Result Reporting'. Meer informatie is te vinden op de IMS website (<http://www.imsglobal.org>).

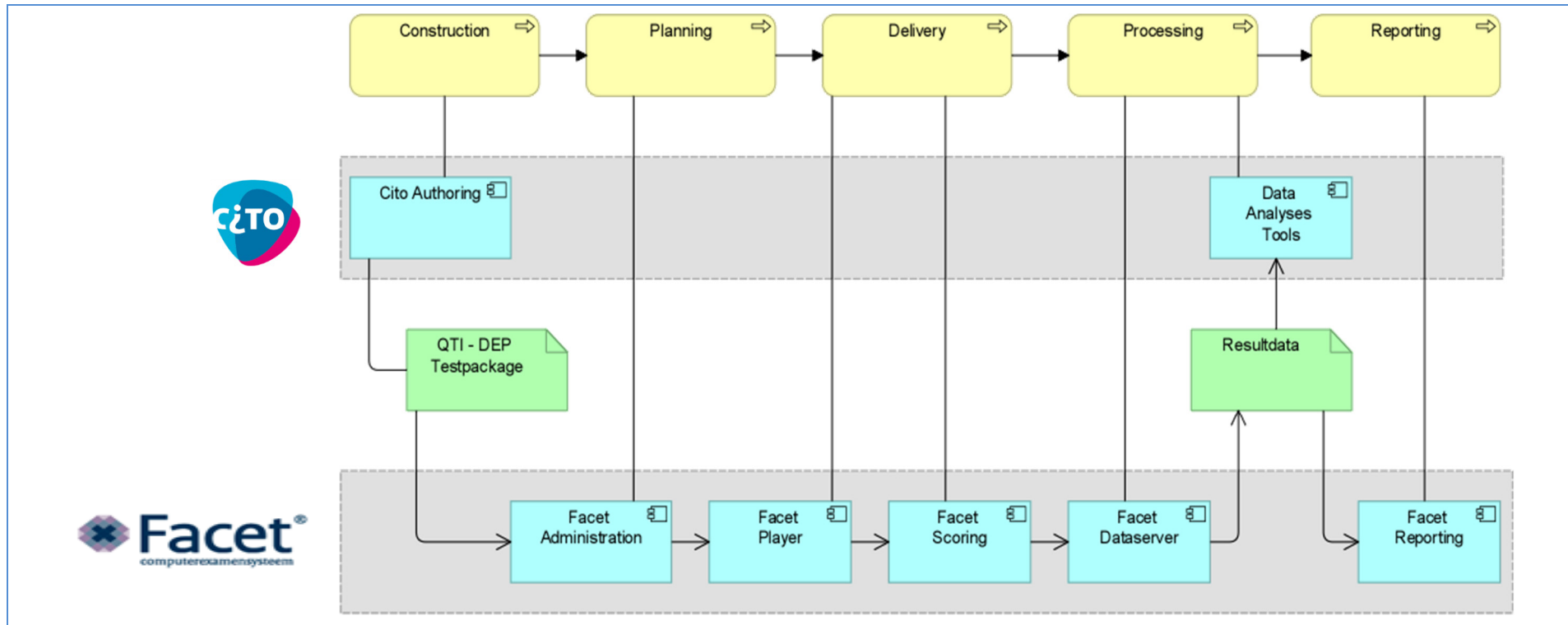
9.4 Noodzakelijke afwijkingen en extensies

De DTT is een innovatieve toets. Hoewel het streven was om zoveel mogelijk bij de QTI-standaard aan te sluiten, bleek dat op enkele punten niet mogelijk. Onderstaand een overzicht van aanpassingen aan QTI die ten behoeve van de DTT zijn doorgevoerd:

Gebruikmaken van bestaande uitbreidingsmogelijkheden van QTI

In enkele gevallen was bij DTT een afwijkend gedrag van bestaande itemtypes gewenst. QTI biedt de mogelijkheid om ander gedrag af te dwingen. Enkele voorbeelden:

- Om een stuk tekst te verdelen in alinea's is gebruik gemaakt van een bestaand QTI itemtype 'hottext interaction' (markeer-opgave). Met behulp van een *class-name* wordt van het standaardgedrag afgeweken. In een markeer-opgave ('hottext interaction') kunnen tekstdelen gemarkeerd (geselecteerd) worden. In een alinea-opgave verschijnt voor de gemarkeerde zin bovendien een alinea-einde.
- Om afbeeldingen te verdelen in categorieën, is gebruik gemaakt van een bestaand QTI itemtype 'graphic GapMatch interaction' (Sleep-opgave afbeelding). Met behulp van een 'class-name' wordt van het standaardgedrag afgeweken. In een sleep-opgave afbeelding ('graphic GapMatch interaction') kan een element (antwoordmogelijkheid) naar een plek op een afbeelding gesleept worden. In een categoriseer-opgave kan een leerling meerdere elementen naar een plek op een afbeelding slepen.



Figuur 9-2. Weergave van informatie-uitwisseling tussen de Cito-auteursomgeving en Facet

Extensies op QTI

De QTI-standaard bood voor enkele itemtypes geen plaats die voor de DTT noodzakelijk waren. Om de items toch aan te bieden zijn extensies aan de QTI toegevoegd. Hiervoor is de standaardwijze voor uitbreiden ('extend') gebruikt zoals reeds beschikbaar in de Facet-keten: Dit wordt gedaan door de extensies op te nemen in het DEP. Enkele voorbeelden:

- Om de DTT adaptief af te kunnen nemen, is een extensie gemaakt. In deze extensie wordt de mogelijkheid geboden om een adaptieve module en sturbestand mee te leveren met de toets.
- Om wiskundige tekens en afbeeldingen op te nemen in bijvoorbeeld 'inline choice interacties' (dropdown-opgaven), is een extensie gemaakt voor dit type opgaven.
- Het aanbieden van een formule-editor is standaard niet mogelijk. Via een extensie is dit mogelijk gemaakt. Tevens kan de opgaveconstructeur aangeven welke knoppen van de formule-editor aan de kandidaat beschikbaar worden gesteld.

'Custom Interactions'

Geavanceerde interacties kunnen worden aangeboden met behulp van QTI CI's ('Custom Interactions'). Het is een semigestructureerde manier om op JavaScript en HTML5 gebaseerde interacties aan te bieden aan de kandidaat. Cito levert deze CI's. Er is een afspraak gemaakt tussen Cito en DUO hoe de resultaten van de CI's in Facet moeten worden opgeslagen.

Enkele voorbeelden:

- Interacties gebaseerd op Digitale Wiskundeomgeving (DWO)
 - Meer informatie is te vinden op de website van het Freudenthal Instituut (<http://www.fisme.uu.nl/wisweb/welcome.html>)
- Interacties gebaseerd op Geogebra.
 - Meer informatie is te vinden op de website van Geogebra (<https://www.geogebra.org/>)

'Custom operators'

Om wiskundige items goed te kunnen nakijken, maakt Facet gebruik van een Computer Algebra System (CAS). In de QTI-package geeft Cito score-regels mee aan Facet hoe de betreffende wiskundige items moeten worden gescoord. Deze regels worden verpakt in 'custom operators'. Dit is een gestructureerde manier om informatie mee te geven in QTI.

Enkele voorbeelden:

- Custom operator om te bepalen of twee formules *equal* zijn.
- Custom operator om te bepalen of formules *equivalent* zijn.

9.5 Toekomstige ontwikkelingen

In de vorige paragraaf staat beschreven op welke punten van de QTI-standaard is afgeweken voor de DTT. In deze paragraaf zal worden ingegaan op ontwikkelingen die sinds de ontwikkeling van de DTT spelen met betrekking tot standaarden.

Extensies op QTI

Tijdens de start van de DTT was QTI versie 2.1 beschikbaar. Het was noodzakelijk om extensies op QTI aan te brengen (via DEP) om essentiële functionaliteit voor de DTT beschikbaar te maken. Ondertussen is QTI2.2 beschikbaar. Deze versie voorziet in veel functionaliteiten die een groot aantal DEP-extensies overbodig maakt. Men kan hierbij denken aan de ondersteuning van mathML voor de presentatie van wiskundige formules. Met de komst van QTI3 (verwacht in 2018) zouden vrijwel alle DEP-extensies overbodig worden. Cito is betrokken bij de ontwikkeling van QTI en heeft de kennis van onder andere DTT ingebracht bij de ontwikkeling van QTI3.

'Custom Interactions'

Custom Interactions (CI's) zijn erg krachtig. Echter was een afspraak tussen Cito en DUO noodzakelijk om goede werking te garanderen. Met de komst van een IMS-standaard voor CI's (verwacht in 2018) zullen bilaterale afspraken tot het verleden behoren en worden CI's interoperabel.

Computergestuurd Adaptief Toetsen (CAT)

Cito heeft veel kennis opgedaan met betrekking tot het integreren van adaptieve logica in een QTI-omgeving. Deze kennis is gedeeld met IMS. In 2016 is Cito samen met enkele andere IMS-leden een werkgroep gestart om een standaard voor CAT te ontwikkelen. Deze standaard is sinds eind 2017 in concept beschikbaar. Verwacht wordt dat deze standaard in 2018 definitief zal worden. Hiermee zal de noodzaak grotendeels verdwijnen om bilaterale afspraken te maken tussen leveranciers van CAT-modules en afnameomgevingen.

9.6 Literatuur

Edustandaard. (2017). *Dutch Exam Profile (DEP)*. Geraadpleegd op november 2017, op https://www.edustandaard.nl/standaard_afspraken/dutch-exam-profile-dep/dutch-exam-profile-versie-4-0/.

IMS Global Learning Consortium (2017). *Question and Test Interoperability and Accessible Portable Item Protocol Background*. Geraadpleegd op november 2017, op <https://www.imsglobal.org/activity/qtiapip>.

10 Dataverwerking en itemanalyses

Jesse Koops

10.1 Inleiding

De reguliere ICT keten van ontwikkeling, afname, dataopslag en analyse voor de DTT was gecompliceerd (zie het vorige Hoofdstuk 9). Een groot deel van deze keten was tijdens de ontwikkeling van de DTT tussen 2012 en 2017 ook nog niet beschikbaar of er ontbraken nog belangrijke functionaliteiten. Een andere kwestie was dat de beoogde veranderingen of extra functionaliteiten in het standaard ontwikkel- en afnameproces van toetsen opgenomen moesten worden, hetgeen veel tijd kostte. Tegelijkertijd moesten de resultaten van de veranderingen en extra functionaliteiten voor de DTT direct beschikbaar zijn voor de analyses en verdere ontwikkeling van de DTT. Dit gold bijvoorbeeld voor hercodering bij sleutelfouten, koppeling van opgaven en responsen aan de inhoudelijke aspecten en het automatisch beoordelen van wiskundige formules. Deze functionaliteiten moesten direct beschikbaar zijn voor de DTT-analyses, terwijl deze nog niet in het standaardproces waren geïmplementeerd.

Tijdens de opbouw van de DTT is daarom gewerkt met een speciaal voor de DTT-pretestfase ontwikkeld en gradueel gegroeid systeem bestaande uit een relationele database en meer of minder specifiek voor de DTT ontwikkelde of aangepaste tools en libraries (een library is een verzameling van functies die programma's aan kunnen roepen). Deze konden vrij gemakkelijk van jaar tot jaar worden aangepast. Hierdoor konden ze voorzien in de steeds veranderende behoeftes en het software landschap van de DTT. Een nadeel van deze benadering was dat het door de veranderingen die elk jaar zijn geïmplementeerd (resultierend in een gewijzigde datastructuur) moeilijk was om data van verschillende jaren samen te voegen voor één analyse.

Hieronder volgt een lijst van tools die (gedeeltelijk) voor de DTT werden ontwikkeld of aangepast:

- Relationele databases (verschillend per jaar)
- Scripts om QTI-items (opgaven), DUO/Facet data en andere databronnen uit te lezen
- scripts om data te exporteren uit de database naar het format dat noodzakelijk was voor de kalibratie en blokindelingen (zie ook Hoofdstuk 12)
- QTI-scoring engine (2015-2017)
- Tijdelijke scoring systemen (2013-2015)
- Sleutelcorrectie tools (2013-2015)
- CAS systeem voor scoring van open wiskunde antwoorden, gebaseerd op SymPy (www.sympy.org/en/index.html)
- Online viewer voor QTI-items (opgaven)
- Dashboards en andere dataviewers
- Online toets en itemanalyse applicaties
- 3DC standaardbepalingsapplicatie (Cito, 2018)

Hierbij moet worden opgemerkt dat het ontwikkelen van deze tools noodzakelijk was om de DTT verder te ontwikkelen, maar dat ze niet tot het op te leveren eindproduct behoren. Oplevering zou ook niet nuttig zijn om verschillende redenen. Een aantal tools zijn inmiddels achterhaald door de keten en sommige tools zijn specifiek aangepast op de eenmalig gebruikte datastructuur voor een bepaald jaar of voor de systemen die op Cito gebruikt worden.

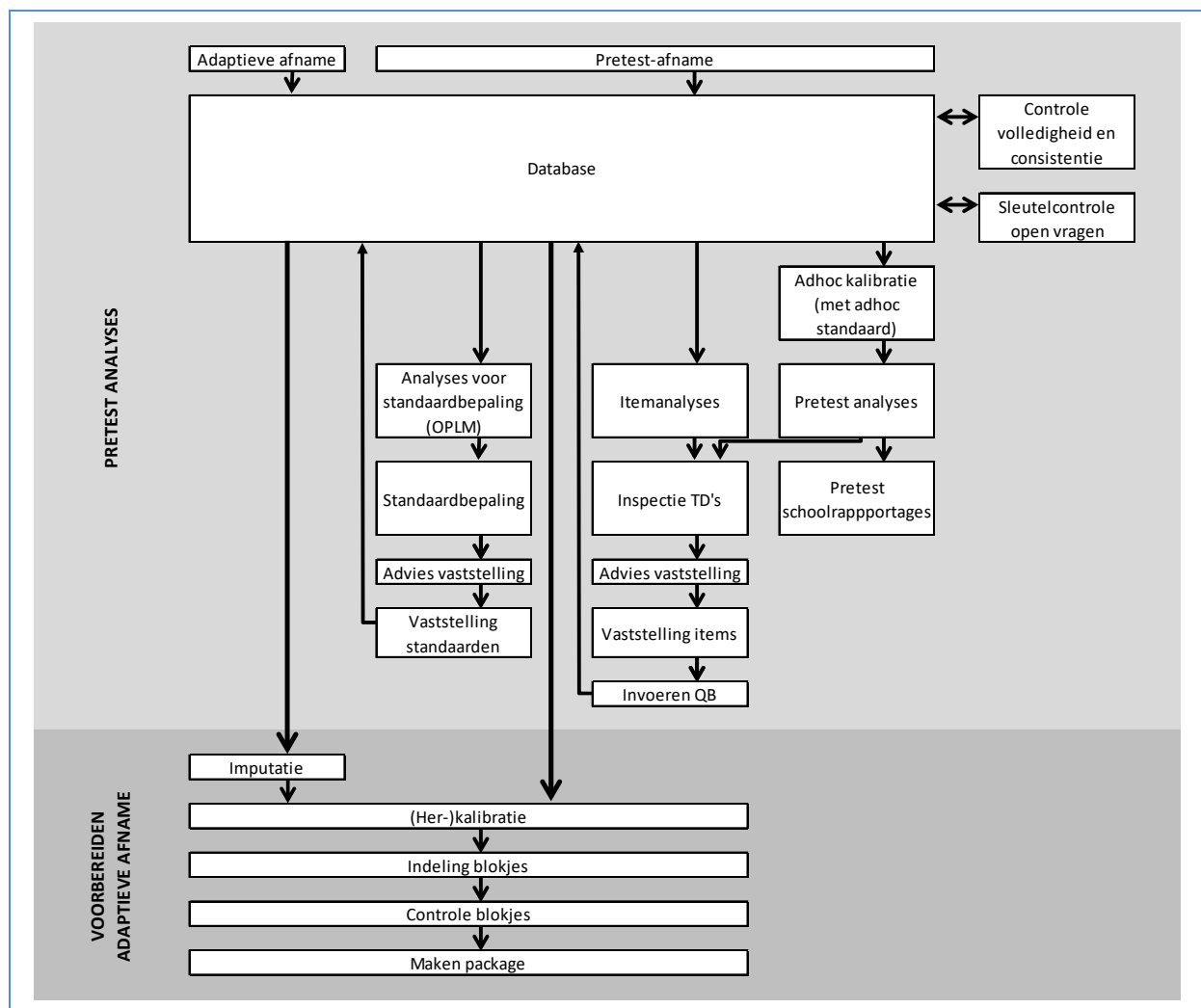
Waar mogelijk generaliseert en standaardiseert Cito interne DTT-applicaties zodat deze door eenieder gebruikt kunnen worden en het onderwijs ten goede komen. Uit bovenstaand rijtje is de 3DC standaardbepalingsapplicatie generiek gemaakt en gepubliceerd als gratis software (Cito, 2018). Verder

vormen de online toets en itemanalyse⁶ applicaties de inspiratie voor een grafische gebruikersinterface voor het (reeds bestaande) toets en itemanalyse pakket Dexter (Maris, Bechger, Koops en Partchev, 2017).

10.2 Dataverwerking

Na ontvangst van de data werden deze ingelezen uit de Xml-bestanden die door DUO werden opgeleverd op basis van de item- en toetsdefinities in de Questify packages. De data kwamen terecht in een relationele database. Vervolgens werden de data gecontroleerd op volledigheid op basis van lijsten met deelnemende scholen en vakken.

Als de data compleet waren, werden er enkele automatische checks uitgevoerd op de sleutels en scoring en volgde een eerste snelle exploratieve analyse om structurele problemen op te sporen die kunnen wijzen op een technische fout bij de afname, de scoring of het inlezen van de bestanden. Ter illustratie hieronder het stroomschema van de dataverwerking in 2016 (Figuur 10-1).



Figuur 10-1. Stroomschema van de dataverwerking DTT in 2016 (Cito, 2016)

10.2.1 Dataverwerking per jaar

Hieronder volgt een korte beschrijving van de belangrijkste elementen van de dataverwerking in elk jaar van de DTT-ontwikkeling.

⁶ Feitelijk is sprake van itemresponsanalyse, maar dit wordt doorgaans itemanalyse genoemd. Om die reden hanteren we in dit hoofdstuk de term itemanalyses

2013

In het eerste jaar van de ontwikkeling van de DTT werd nog gebruik gemaakt van de oude afname omgeving ExamenTester van Cito. ExamenTester is ontwikkeld voor gebruik bij examens en is niet toegesneden op diagnostische toetsen. De DTT vraagt bijvoorbeeld om complexere opgavetypen, scoring en metadatering dan in ExamenTester mogelijk is. Voor de eerste try-out betekende dit dat slechts 45% van alle opgaven automatisch via ExamenTester konden worden beoordeeld.

Om de overige opgaven op de voor de DTT noodzakelijke wijze te kunnen verwerken, zijn de volgende maatregelen genomen:

- Voor de beoordeling van de niet-automatisch na te kijken, open antwoorden zijn beoordelingsinstructies opgesteld door de toetsdeskundigen.
- Voor de handmatige beoordeling van open antwoorden zijn beoordelaars geworven die in een periode van 4 weken alle open responsen hebben beoordeeld, inclusief de schrijfoopdrachten voor de talen. In totaal zijn ruim 850 uren door beoordelaars besteed aan de handmatige beoordeling van de open responsen.
- Voor de beoordeling van alle (niet-numerieke) open opgaven zijn twee tijdelijke beoordelingstools ontwikkeld.
 - Een tool voor de open tekstopgaven
 - Een tool voor de wiskunde opgaven gemaakt met de wiskunde toolbox (Flash)
- Alle opgaven met meerdere responsen en alle 'gesloten' opgaven met Flash-inhoud⁷ zijn via een database automatisch beoordeeld. Hiervoor zijn specifieke routines ontwikkeld.

2014

In 2014 werd voor het eerst gewerkt met Questify Builder als auteursomgeving en Facet als afname-omgeving. Alle opgaven waren in theorie automatisch scorebaar, met uitzondering van de stelopdrachten voor Nederlands en Engels. De beoordeling gebeurde nog wel buiten de Questify-Facet keten omdat in Questify Builder nog niet alle mogelijke opgavetypen van sleutels konden worden voorzien en in Facet niet alle opgavetypen automatisch gescoord konden worden. Ook nieuw in 2014 waren enkele opgaven waarbij een formule-editor werd gebruikt. In dit jaar werden wat voorzichtige experimenten gedaan met automatische scoring van formule-antwoorden.

Er waren nog geheel geen voorzieningen voor codering in Questify en Facet voor opgaven met meerdere interacties. Codering van afzonderlijke responsen gebeurde in een database. Toetsdeskundigen konden sleutels corrigeren met een voor de DTT ontwikkelde sleutelcorrectie-tool.

2015

De verdere ontwikkeling van Facet en Questify zorgde ervoor dat in 2015 opgaven in Questify gecodeerd konden worden op interactieniveau. Omdat nog niet alle opgaven automatisch gescoord konden worden en er in Facet geen herscoring mogelijk was (hetgeen bij een pretest met open opgaven essentieel is) werd geen gebruik gemaakt van de automatische scoring en codering in Facet. Het was wel een grote vooruitgang dat de coderingen op interactieniveau automatisch konden worden uitgelezen uit de QTI-itemdefinities die uit Questify Builder werden geëxporteerd.

Voor de scoring van de antwoorden op kort open-opgaven met invoer van formules is in 2015 een eigen bescheiden library ontwikkeld gebaseerd op de open source library SymPy.

2016

De dataverwerking in 2016 vond plaats onder vergelijkbare randvoorwaarden en omstandigheden als in 2015. In 2016 kon gewerkt worden met een geavanceerder databasesysteem (PostgreSQL in plaats van SQLite) en met een interne server waarop onlineapplicaties konden worden ontwikkeld en geïnstalleerd. Hierdoor kon de itemanalyses en sleutelcorrectie aanmerkelijk soepeler verlopen (zie voor een beschrijving de volgende sectie).

⁷ In ExamenTester was het alleen mogelijk om nieuwere vormen van opgaven te maken via Adobe Flash. Adobe Flash is een computerprogramma waarmee animaties, webvideo's en webapplicaties gemaakt kunnen worden.

2017

De Questify-Facet keten was inmiddels praktisch compleet. De dataverwerking was vergelijkbaar met die in 2016. Kort open-opgaven met invoer van formules konden inmiddels automatisch beoordeeld worden in Facet en ook herscoring was mogelijk in Facet, zij het met enkele beperkingen. Omdat het intern ontwikkelde systeem niet onderhevig was aan dezelfde beperkingen als Facet werd voor de meeste doeleinden de eigen ontwikkelde beoordelingsmodule gebruikt voor de analyses. De Facet-beoordeling werd gebruikt ter controle.

10.2.2 Leerpunten in de DTT-keten

De ad hoc ontwikkelde tools dienden om bepaalde problemen in de DTT-keten te omzeilen. Sommige van deze problemen waren van tijdelijke aard en inherent aan de snelle ontwikkeling en het stadium waarin het project zich bevond, enkele andere deden zich tot het eind van het project voor. Terugkijkend kunnen (naast omstandigheden zoals tijdsdruk) twee globale oorzaken voor de ondervonden problemen worden aangewezen.

Modulariteit

Modulariteit in software refereert aan de mate waarin ‘modules’ van elkaar gescheiden zijn. Een module is een op zichzelf staand deel van een programma of library dat gespecialiseerd is in één taak of een set van nauw gerelateerde taken. Een module kan zodoende worden ingezet om diezelfde taak te vervullen binnen verschillende applicaties. Een hoge mate van modulariteit zorgt dat een aanpassing in één module niet leidt tot aanpassingen in andere modules; daarnaast kunnen modules los van elkaar worden gebruikt.

Een voorbeeld waarin modulariteit binnen de DTT-software-keten onvoldoende gerealiseerd werd, is de herziening van de beoordeling (‘herscoring’) van leerlingantwoorden. Een modulaire oplossing (die binnen Cito gebruikt werd om rond de beperkingen van de DTT-keten te werken) kan bestaan uit een module met een functie die als input de beoordelingsregel en een leerlingantwoord verwacht en op basis daarvan een beoordeling (‘score’) voor de opgave genereert. Een dergelijke module kan op meerdere plaatsen worden gebruikt, zowel bij scoring in een toetsstelsel als scoring achteraf met dezelfde of een aangepaste sleutel.

De beoordeling en herziening van de beoordeling (herscoring) in Facet kon niet worden losgekoppeld van de afname software en de uiteindelijke implementatie vereist dat een leerling virtueel de complete toets opnieuw doorloopt. Daarom is herscoring in Facet weliswaar inmiddels mogelijk maar nog niet voor een adaptieve afname (immers, dit zou het toetspad beïnvloeden). Een consequentie is dat herscoring van zaai-opgaven in de adaptieve afname ook in het laatste jaar van de ontwikkeling nog steeds bij Cito plaats vond in plaats van via Facet. In een modulaire oplossing zou het daadwerkelijk doorlopen van de toets functioneel gescheiden zijn van de scoring van de opgaven.

Data integriteit en identificatievariabelen

Variabelen die gebruikt worden om data-elementen te identificeren (‘identifiers’, bijvoorbeeld een itemcode of toets-id) moeten aan minimaal twee voorwaarden voldoen:

1. *Uniek* voor een element binnen de scope waarin ze worden gebruikt (d.w.z. dezelfde identificatiecode wordt niet voor verschillende elementen gebruikt en hetzelfde element heeft altijd dezelfde identificatiecode)
2. *Stabiel* (zolang een element niet verandert, doet de identificatiecode dat ook niet)

Met name de stabiliteit liet te wensen over. Zo werd in verschillende jaren en in verschillende systemen een verschillende identificatie voor de conceptcoderingen gebruikt, veranderden identifiers van antwoordopties bij updates van Questify en waren de identificatiecodes van de opgaven (opgave-id’s) niet consistent met betrekking tot gebruik van hoofdletters of kleine letters en mintekens of onderstreping (‘underscores’). Daarnaast bleek het niet mogelijk om geheel uit te sluiten dat opgaven in de auteursomgeving (Questify Builder) werden aangepast (het zogenoemde bevroren) en was het in veel gevallen ook niet mogelijk om te achterhalen of en door wie een opgave was aangepast.

10.3 Itemanalyses en sleutelcorrectie

Om de kwaliteit van de opgaven te kunnen evalueren zijn beschrijvende itemanalyses uitgevoerd die informatie geven over de (statistische) kenmerken van de ontwikkelde opgaven.

Voor opgave-evaluatie en sleutelcorrectie is het bijzonder belangrijk dat een opgave kan worden beoordeeld door een inhoudelijk deskundige in de context van de data en de psychometrische gegevens. In de standaard DTT-keten was het zo dat de opgave en de sleutel te zien waren in de auteursomgeving (Questify Builder), de scoring gedaan werd in Facet en de responsgegevens beschikbaar waren in een Microsoft Server database. Psychometrische gegevens zijn normaal gesproken het resultaat van analyse met software als SPSS of R en in het geval van de DTT een Fortran routine. Analyseresultaten worden normaal gesproken opgeleverd als tekst-, CSF- en Excelbestanden of afbeeldingen.

Een dergelijke versnippering van de data die een toetsdeskundige nodig heeft om een opgave te beoordelen is niet bevorderlijk. Om deze reden werd intern een onlineapplicatie ontwikkeld waarin deze gegevens zoveel mogelijk waren gecombineerd.

Om de kwaliteit van de opgaven te kunnen evalueren zijn beschrijvende itemanalyses uitgevoerd die informatie geven over de (statistische) kenmerken van de ontwikkelde opgaven. Voor elk vak waren binnen elke schoolsoort en leerweg verschillende toetsversies met meerdere secties (één sectie per hoofdaspect) zodat alle opgaven gepretest zouden worden, zie Tabel 10-1 voor de belangrijkste psychometrische maten die werden gebruikt voor de beoordeling van de kwaliteit van een opgave.

Tabel 10-1. Beschrijving van de belangrijkste psychometrische maten om de kwaliteit van de opgaven te evalueren

Maat	Beschrijving
P-waarde	P-waarde (proportie goed) van de conceptrespons. Indien het een polytome respons is dan wordt hier het gemiddelde weergegeven.
Rit, Rir	Correlatie van de conceptrespons met de somscore van de toets en de correlatie van de conceptrespons met de toets minus de betreffende conceptrespons. Rit en Rir zijn maten voor het onderscheidend vermogen van een opgave. In het geval van een adaptieve toets is de correlatie met het blok waarin de opgave voorkwam gebruikt, deze is echter alleen informatief in de eerste blokken die alle leerlingen hebben gekregen (in de Hoofdstukken 12 en 13 wordt de blokindeling uitgebreid besproken).
Cronbach's alpha, Cronbach's alpha - opgave	Cronbach's alpha is een ondergrens voor de betrouwbaarheid van de toetsscore. Ook bij deze maat geldt dat ze alleen te interpreteren is in de pretest of de eerste blokken in de adaptieve toets.
P_{onder}	De kans om de conceptrespons correct te beantwoorden voor de leerlingen die onder niveau zitten na de ad-hoc-kalibratie (20% van de pretestleerlingen).
P_{op}	De kans om de conceptrespons correct te beantwoorden voor de leerlingen die op niveau zitten na de ad-hoc-kalibratie (60% van de pretestleerlingen).
P_{boven}	De kans om de conceptrespons correct te beantwoorden voor de leerlingen die boven niveau zitten na de ad-hoc-kalibratie (20% van de pretestleerlingen).

In de online applicatie konden de toetsdeskundigen de uitkomsten van de toets- en itemanalyses zien (zie Figuur 10-2). De toetsdeskundigen konden per opgave commentaar schrijven. Indien gewenst kon elke opgave getoond worden (zie Figuur 10-3) en een Excel-bestand gedownload worden met de uitkomsten.

Alle opgaven die mogelijk problematisch waren werden gemarkeerd in een Excel-bestand en door de toetsdeskundigen inhoudelijk geïnspecteerd en besproken met de vaststellingscommissie (zie ook Hoofdstuk 2). De vaststellingscommissie beoordeelde de voorgelegde opgaven en besloot of de besproken opgaven werden goed- of afgekeurd.

Figuur 10-2. Schermafdruk van de toets- en itemanalyses

Afname Pre-test		Vak Engels Nederlands Wiskunde															
♦ Sectie	♦ populatie	♦ hoofd-attribuut	♦ items in sectie	♦ Item	♦ Item type	♦ aant. resp.	♦ concept	♦ gem. score	♦ min. score	♦ max. score	♦ riT	♦ riR	♦ cr a	♦ cr a - item	♦ cr a40	♦ advies	♦ commentaar
EN-BB_H1	BB	1	56	ITM-dtt-en-vm-0228-SC	gapMatch	276	CONCEPTRESPONSE_ENG-1	0.971	0	1	0.218	0.194	0.778	0.776	0.715		
							CONCEPTRESPONSE2_ENG-1	0.822	0	1	0.290	0.236	0.778	0.775	0.715		
							CONCEPTRESPONSE3_ENG-1	0.696	0	1	0.388	0.327	0.778	0.772	0.715		
							CONCEPTRESPONSE4_ENG-1	0.982	0	1	0.200	0.180	0.778	0.777	0.715		
							CONCEPTRESPONSE5_ENG-1	0.754	0	1	0.399	0.342	0.778	0.771	0.715		
							CONCEPTRESPONSE6_ENG-1	0.699	0	1	0.413	0.354	0.778	0.771	0.715		
EN-BB_H1	BB	1	56	ITM-dtt-en-vm-0377-SC	gapMatch	277	CONCEPTRESPONSE_ENG-1	0.386	0	1	0.232	0.161	0.780	0.779	0.717		
							CONCEPTRESPONSE2_ENG-1	0.542	0	1	0.269	0.197	0.780	0.778	0.717		
							CONCEPTRESPONSE3_ENG-1	0.545	0	1	0.245	0.173	0.780	0.779	0.717		

Figuur 10-3. Schermafdruk van de toets- en itemanalyses waarbij ingezoomd is op een opgave

Sectie	populatie	hoofd-attribuut	items in sectie	item	taal	item type	aant. resp.	gem. score	min. score	max. score	riT	riR	cr a	cr a - item	cr a40	besluit VC	commentaar			
EN-GLTL-b1	GLTL	ENG-1	36	ITM-dtt-en-vm-0605-SC	ja	gapMatch	2311	79	1	ENG-1-2	0.177	0	1	0.383	0.321	0.785	0.779	0.802	Afgekeurd	Item is te ingewikkeld
EN-KB-b1	KB	ENG-1	33	ITM-dtt-en-vm-0605-SC	ja	gapMatch	1203	77	1	ENG-1-2	0.099	0	1	0.344	0.288	0.743	0.736	0.778	Afgekeurd	Item is te ingewikkeld

Wys Anna, can't make it tonight Grandma fell and broke her hip so have to go to the hospital with my dad. I'll email mr Gerards we can't hand in the assignment tomorrow.

Hope she's ok.
No problem at all. See u tomorrow at school? xx

Yeah, sure!

Je stuurt een e-mail naar Je docent naar aanleiding van een bericht.

Sleep DRIE passende zinnen naar het lege e-mailvak.

Dear Mr Gerards

I had to go to the hospital with my dad, because my grandmother broke her hip.

Kind regards

Anna

Anna and I have decided it is impossible to hand in the assignment tomorrow. I hope you understand the situation and will give us some more time to finish it. You will get the assignment later, no worries. Unfortunately, Anne and I could not finish the assignment in time.

Antwoord

RESPONSE

Anna and I have decided it is impossible to hand in the assignment tomorrow. 1261

I had to go to the hospital with my dad, because my grandmother broke her hip. 1437

I hope you understand the situation and will give us some more time to finish it. 107

You will get the assignment later, no worries. 47

Unfortunately, Anne and I could not finish the assignment in time. 655

Anna and I have decided it is impossible to hand in the assignment tomorrow. 491

I had to go to the hospital with my dad, because my grandmother broke her hip. 1545

I hope you understand the situation and will give us some more time to finish it. 718

You will get the assignment later, no worries. 234

Unfortunately, Anne and I could not finish the assignment in time. 520

Sleutels

If([and([E,01] in RESPONSE, [0,02] in RESPONSE, [C,03] in RESPONSE)):
CONCEPTRESPONSE_ENG-1 = 1
CONCEPTRESPONSE_ENG-1-2 = 1

10.4 Literatuur

Cito (2016). *Diagnostische tussentijdse toets: Verslag pretest 2016*. Arnhem: Cito.

Cito (2018). *Computerized standardsetting using 3DC*. Geraadpleegd op <http://www.cito.com/our-expertise/implementation/3dc>

Maris, G., Bechger, T., Koops, J. & Partchev, I. (2017). *Dexter: Data Management and Analysis of Tests*. R package version 0.6.1.

R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

11 Standaardbepaling

Sanneke Schouwstra, Jesse Koops en Karen Keune

11.1 Doel van standaardbepaling

In de DTT worden diagnoses gesteld voor schrijfvaardigheid Nederlands, schrijfvaardigheid Engels en wiskunde. Hierbij wordt per aspect uit het leerlingmodel (kennisaspecten of deelvaardigheden) bepaald in welke beheersingscategorie de leerling valt: onder, op of boven niveau. Het doel van de standaardbepaling was om per hoofdaspect de grenzen te bepalen tussen onder en op niveau en tussen op en boven niveau voor elke leerweg/schoolsoort. Voor elk vak heeft daarvoor per schoolsoort en leerweg (vmbo-bb, vmbo-kb, vmbo-gt, havo en vwo) een standaardbepaling plaatsgevonden waarin de standaarden voor de hoofdaspecten werden gezet: in totaal hebben er dus na elke pretest vijftien standaardbepalingen met experts plaatsgevonden. Voor deze standaardbepaling is een variatie van de *Data-Driven Direct Consensus Methode* (3DC) gebruikt (Feskens, Keuning, Van Til, & Verheyen, 2014; Keuning, Straat & Feskens, 2017).

11.2 De 3DC methode

De 3DC methode (Feskens, Keuning, Van Til, & Verheyen, 2014) is een nieuwe methode, die elementen uit de Angoff (Angoff, 1971), de Bookmark (Mitzel, Lewis, Patz, & Green, 2001) en de Direct Consensus Methode (Sireci, Hambleton, & Pitoniak, 2004) combineert. In 2013, toen het globale plan voor de standaardbepaling voor de DTT werd uitgewerkt (Keune & Schouwstra, 2014) was deze methode al succesvol toegepast bij de standaardbepaling van de toetsen MBO-COE Nederlands en MBO-COE Rekenen en voor het bepalen van de prestatiestandaarden voor het Europees Referentie Kader (Feskens, Keuning, Van Til, & Verheyen, 2014). De 3DC methode is gebaseerd op het idee dat de opgaven in een toets kunnen worden onderverdeeld in clusters. Deze methode is daarom uitstekend geschikt voor toetsen die eenduidig te onderscheiden clusters van opgaven hebben, zoals de DTT. De opgaven van de DTT zijn geclusterd per aspect van het leerlingmodel. Per cluster van opgaven kan vervolgens bepaald worden hoeveel responsen correct moeten zijn om in de beoogde beheersingscategorie te vallen.

Bij de 3 DC standaardbepaling wordt gewerkt met een team van experts op het vakgebied. Om tot goed gefundeerde standaarden te komen worden de experts voorzien van zoveel mogelijk verschillende informatiebronnen. Voorafgaand aan de standaardbepalingssessie voor de DTT hebben de experts de tussendoelen van SLO (2012) en de toetswijzer (CvTE, 2014) gekregen. In de tussendoelen van SLO staat beschreven wat de leerlingen halverwege het voortgezet onderwijs moeten kunnen en in de toetswijzer van het CvTE staat beschreven welke aspecten gediagnosticeerd moeten worden. Tijdens de sessie krijgen de experts de clusters met opgaven te zien en gebruiken ze ook hun eigen ervaring als docent bij hun oordeel. Bovendien krijgen de experts tijdens de standaardbepaling empirische feedback over de relatieve moeilijkheid van de opgaven, over de oordelen van de andere experts en eventueel over de impact van de oordelen op het aantal leerlingen dat in een bepaalde niveaugroep valt (onder, op of boven-niveau).

11.3 De experts

Het streven van CvTE was om voor elke standaardbepaling 10 tot 12 experts te werven. Gemiddeld deden uiteindelijk negen docenten mee per sessie, zie Tabel 11-1. Er is vooral geworven onder de scholen die meededen in de pilot, omdat het belangrijk was dat de inhoudelijk experts mensen waren die recente ervaring hadden met het onderwijzen van de doelgroep.

In 2015 hebben 93 docenten meegedaan (van de 107 aanmeldingen) en in 2016 82 docenten (van de 90 aanmeldingen). Vijftig procent van de docenten die zich in 2016 hadden aangemeld, hadden al in 2015 meegedaan aan de standaardbepalingen. Om te zorgen voor consistentie in de oordelen zijn de grenzen bij de verschillende leerwegen/schoolsoorten deels door dezelfde groep experts bepaald. In 2015 deed 40% van de docenten mee aan meerdere standaardbepalingen (bijvoorbeeld van havo en van vwo) en in 2016 47% (zie Tabel 11-2).

Tabel 11-1. Aantal experts per standaardbepaling

		Aantal experts per standaardbepaling				
		vmbo-bb	vmbo-kb	vmbo-gt	havo	vwo
2015	Schrijfvaardigheid Engels	8	10	9	12	11
	Schrijfvaardigheid Nederlands	7	9	7	8	10
	Wiskunde	8	10	11	8	10
2016	Schrijfvaardigheid Engels	10	11	10	8	7
	Schrijfvaardigheid Nederlands	7	8	12	12	11
	Wiskunde	5	8	9	11	12

Tabel 11-2. Aantal experts dat heeft deelgenomen aan de standaardbepalingen

		Aangemeld	Deelgenomen	Meerdere sessies
2015	Schrijfvaardigheid Engels	34	32	50%
	Schrijfvaardigheid Nederlands	36	29	38%
	Wiskunde	39	34	32%
2016	Schrijfvaardigheid Engels	27	22	77%
	Schrijfvaardigheid Nederlands	35	33	36%
	Wiskunde	32	30	37%

11.4 De opgaven

De 3DC is een toetsgeoriënteerde standaardbepaling omdat de experts hun oordelen over de grenzen baseren op de inhoud van de toets. Bij de DTT moesten de experts een oordeel vormen over het aantal responsen dat een grensleerling correct zou moeten hebben bij elk cluster opgaven. In beide pretesten waren echter in totaal meer opgaven afgenomen⁸ in elke leerweg/schoolsoort dan de experts op een dag zouden kunnen bekijken, daarom werd een selectie van opgaven gebruikt.

In 2015 zijn tien tot twaalf opgaven per cluster geselecteerd. In dat jaar was de standaardbepaling gericht op de diagnoses die gerapporteerd zouden worden bij de eerste (beperkt) adaptieve afname in 2016. In de beperkt adaptieve afname zou bij de talen uitsluitend op de hoofdaspecten gerapporteerd worden, terwijl bij wiskunde op alle deelaspecten van slechts twee domeinen gerapporteerd zou worden. Bij de talen zijn daarom in 2015 grenzen bepaald tussen onder en op niveau en tussen op en boven niveau op vier clusters (één cluster per hoofdaspect) van ongeveer twaalf opgaven. Bij wiskunde werden grenzen gezet op zes clusters (per deelaspect van domein *D Meten en meetkunde* en domein *E Verbanden en formules*) van ieder ongeveer tien opgaven.

In 2016 was de standaardbepaling gericht op de diagnoses van de volledig adaptieve afname van 2017. Bij de talen zijn in 2016, net zoals bij wiskunde, grenzen bepaald per deelaspect. Deze grenzen zijn vervolgens samengenomen tot een standaard per hoofdaspect. Bij Engels waren er acht clusters, bij Nederlands twaalf (vmbo) tot dertien (havo/vwo) en bij wiskunde waren er zes (vmbo) tot negen (havo/vwo) clusters. Omdat er bij de talen meer clusters beoordeeld moesten worden dan in 2015 is als richtlijn een aantal van minstens twaalf responsen per cluster genomen (in plaats van de twaalf opgaven in 2015). Bij de talen waren uiteindelijk gemiddeld veertien responsen per cluster en bij wiskunde, net zoals in 2015, zestien

⁸ Om alle ontwikkelde opgaven te kunnen pretesten is met een design gewerkt, waardoor elke leerling alleen een gedeelte van de opgaven hoefde te maken (Cito, 2015; Cito, 2016).

responsen. In Tabel 11-3 is het gemiddelde aantal opgaven en responsen gebruikt voor de standaardbepalingen te zien

Om een grotere variëteit aan soorten opgaven, contexten en/of genres in de standaardbepaling mee te nemen zijn zoveel mogelijk opgaven gebruikt met een beperkt aantal responsen. Complexe⁹ en problematische¹⁰ opgaven zijn niet gebruikt. Indien er na uitsluiting van de complexe en problematische opgaven nog te veel opgaven waren, zijn eerst de opgaven geselecteerd die ook in aangrenzende leerwegen/schoolsoorten zijn gebruikt en vervolgens een random selectie van de resterende opgaven. Door opgaven te gebruiken die in aangrenzende leerwegen/schoolsoorten zijn afgenomen kan gecontroleerd worden of de standaarden in de aangrenzende leerwegen/schoolsoorten consistent zijn met elkaar (of men bijvoorbeeld niet hogere eisen stelt aan vmbo-bb dan aan vmbo-kb).

Tabel 11-3. Gemiddeld aantal opgaven en responsen in de standaardbepalingen

		Schrijfvaardigheid Engels		Schrijfvaardigheid Nederlands		Wiskunde	
		Opgaven	Responsen	Opgaven	Responsen	Opgaven	Responsen
2015	Gemiddeld per cluster	11	23	13	54	10	16
2016	Gemiddeld per cluster	11	14	5	14	10	16

11.5 De taak van de experts

De experts kregen per post informatie over de standaardbepaling, de tussendoelen en de toetswijzer. Voorafgaand aan de standaardbepaling kregen de experts een korte training waarin uitleg gegeven werd over de DTT, de standaardbepalingsprocedure en de computerapplicaties die tijdens de standaardbepaling werden gebruikt. De vijftien standaardbepalingen vonden plaats in een periode van twee weken, nadat alle gegevens van de pretest verwerkt waren, de toets- en itemanalyses waren uitgevoerd en eventuele sleutelcorrecties waren doorgevoerd. Elke standaardbepaling duurde één dag. Per dag vonden simultaan maximaal twee standaardbepalingen van twee verschillende vakken plaats, zodat de experts aan meerdere standaardbepalingen konden deelnemen en inval mogelijk was in geval van ziekte.

Tijdens de standaardbepaling kregen de experts per aspect (cluster) de opgaven te zien in een previewer, zodat zij de opgaven precies zo zagen als de leerlingen. Vervolgens vormden de experts een oordeel over het aantal responsen dat de 'grensleerling' correct zou moeten hebben¹¹. Elke expert moest per cluster aangeven hoeveel responsen de leerling naar verwachting correct zou moeten hebben op het cluster als zijn/haar vaardigheid zich precies op de grens onder/op niveau van het aspect zou bevinden. Hun oordeel over het aantal responsen dat de 'grensleerling' correct zou moeten hebben, konden ze aangeven op een digitaal standaardbepalingsformulier (zie Figuur 11-1). Tijdens de standaardbepaling werd een computerapplicatie gebruikt, die de hele standaardbepalingssessie ondersteunde (de 3DC applicatie; Koops, 2016; Cito, 2018).

11.5.1 Het standaardbepalingsformulier

De pretestgegevens werden gebruikt om het digitale formulier te maken, waarop de experts hun oordeel over het aantal correcte responsen konden aangeven. Met behulp van de pretestgegevens is de moeilijkheid bepaald van de opgaven die meegingen in de standaardbepaling. De moeilijkheid is bepaald door alle vmbo-opgaven op één vaardigheidsschaal te leggen en de havo/vwo-opgaven op één vaardigheidsschaal te leggen met behulp van het One Parameter Logistic Model (Verhelst, Glas, & Verstralen, 1995). Voor elke leerweg/schoolsoort is met behulp van simulaties voor elk mogelijk aantal

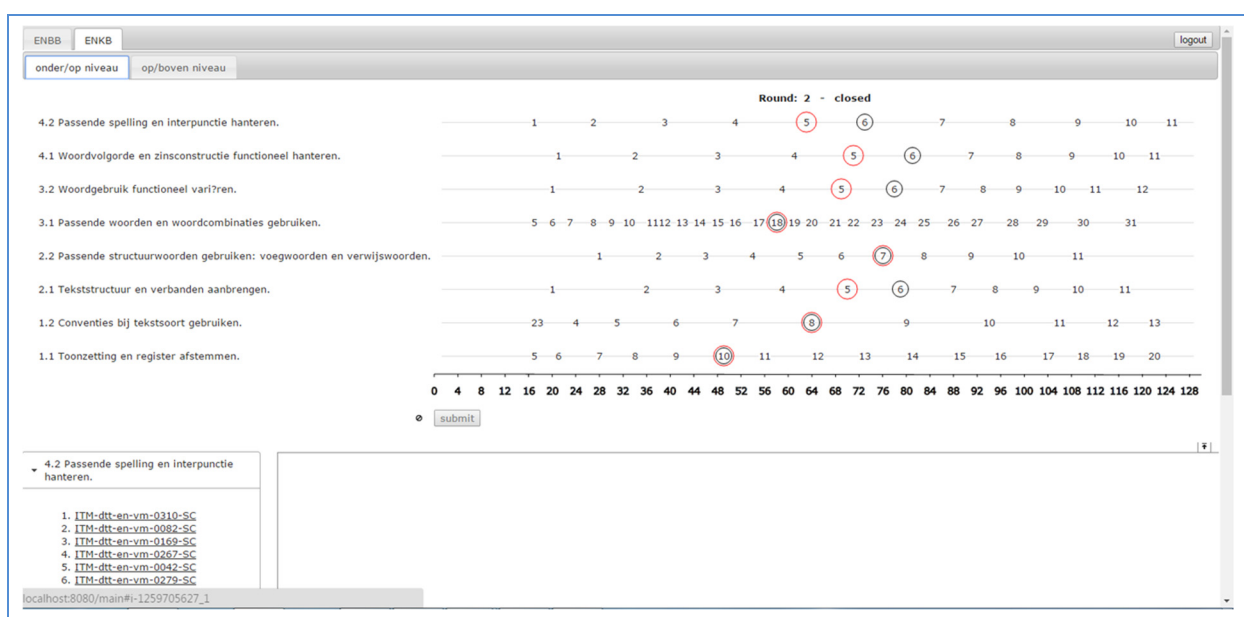
⁹ Opgaven die verschillende aspecten meten of opgaven met polytome responsen.

¹⁰ Opgaven waarbij meer dan de helft van de conceptresponsen extreem makkelijk of extreem moeilijk zijn of een zeer geringe samenhang vertonen met de overige responsen.

¹¹ In de standaard 3DC methode worden de experts gevraagd om een grensscore aan te geven in plaats van het aantal correcte responsen.

correcte antwoorden op de volledige toets uitgerekend wat het verwachte aantal correcte antwoorden is op een deelaspect (cluster). Deze uitkomsten zijn weergegeven op het digitale standaardbepalingsformulier (zie Figuur 11-1). In de standaardbepaling voor de DTT wordt het One Parameter Logistic Model alleen gebruikt voor de weergave op dit formulier. Het model wordt niet gebruikt voor het uitdrukken van de standaard.

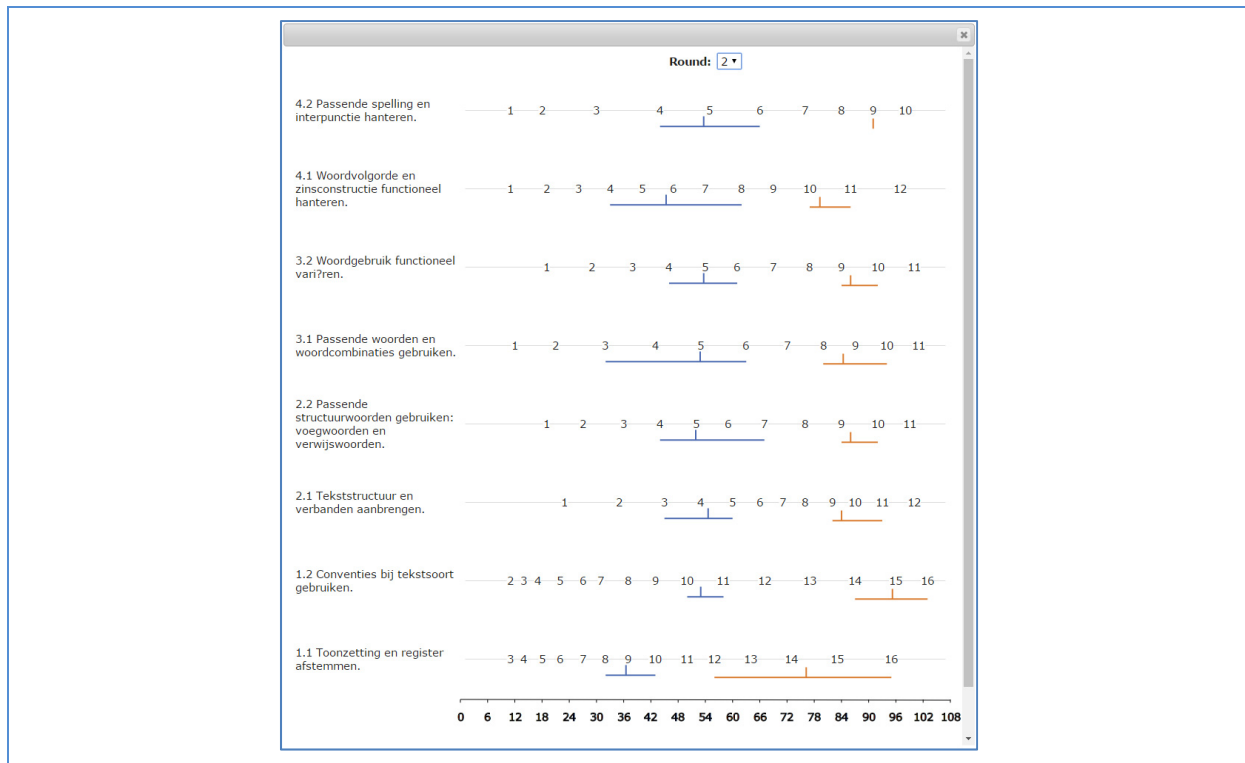
In Figuur 11-1 is een voorbeeld van het digitale standaardbepalingsformulier te zien met fictieve oordelen (in de cirkels) van een expert. Bij aanvang kregen de experts dus het formulier zonder oordelen (cirkels). Op de horizontale lijnen van het formulier is het aantal correcte responsen voor elk deelaspect te zien. Het aantal correcte responsen voor een deelaspect is gerelateerd aan het totaal aantal correcte responsen. Bijvoorbeeld, op het formulier in Figuur 11-1 is te zien dat acht correcte responsen bij *Conventies* te verwachten zouden zijn bij een leerling die in totaal ongeveer 64 correcte responsen heeft en vijf correcte responsen bij *Spelling*. Op dit formulier zien de experts dus de relatieve moeilijkheid van het cluster in verhouding tot de andere clusters. De visuele relatie tussen de aspecten helpt de experts een realistisch oordeel te maken.



Figuur 11-1. Een schermafdruk van een standaardbepalingsformulier met fictieve oordelen van een expert (zichtbaar als de zwarte en rode cirkels)

11.5.2 Oordelen van de experts

Per te zetten grens vonden twee beoordelingsrondes plaats. De experts begonnen met de onder/op-grens. Op grond van de opgaven bepaalde elke expert ze het aantal responsen dat de 'grensleerling' correct zou moeten hebben en kon het oordeel aangeven op het formulier (zwarte cirkels). Na de eerste ronde, waarin experts hun oordelen individueel hadden bepaald, vond een groepsdiscussie plaats. De groepsleider gaf normatieve feedback met behulp van de computerapplicatie. De groepsleider toonde de oordelen van alle experts en besprak de onderlinge overeenstemming en variatie van de oordelen. Ook kregen de experts een feedbackformulier te zien dat speciaal voor de DTT is ontwikkeld, zoals weergegeven in Figuur 11-2. Op het formulier konden ze het bereik en de modus van de oordelen zien. Vervolgens werden minstens twee experts uitgenodigd om hun oordeel toe te lichten en argumenten te delen met de groep en ontstond een discussie. De discussie diende te waarborgen dat de experts op grond van dezelfde inhoudelijke gronden tot een oordeel kwamen, de experts hoefden niet noodzakelijkerwijs tot hetzelfde oordeel te komen. Na afloop van de discussie kregen de experts de gelegenheid om een definitief oordeel te geven (de rode cirkels in Figuur 11-1). Nadat het definitieve oordeel over de onder/op-grens was gegeven, gingen de experts aan de slag met een oordeel over de op/bovengrens op dezelfde wijze.



Figuur 11-2. Feedbackformulier van de expertoordelen, met in blauw de oordelen over de onder/op-grens en in rood de oordelen over de op/boven-grens

Bij de DTT is altijd eerst de ondergrens bepaald en vervolgens de bovengrens. De ervaring was echter dat de vakexperts in de loop van de dag coulanter werden. Hun eerste oordeel leek doorgaans erg streng. Een andere opzet waarbij bijvoorbeeld de eerste standaard nogmaals beoordeeld wordt aan het eind van de dag zou mogelijk geschikter zijn. Ook zou onderzocht kunnen worden of het zinvol is om de bovengrens eerst te zetten en daarna de ondergrens. Dit zou wellicht in een minder hoge standaard resulteren.

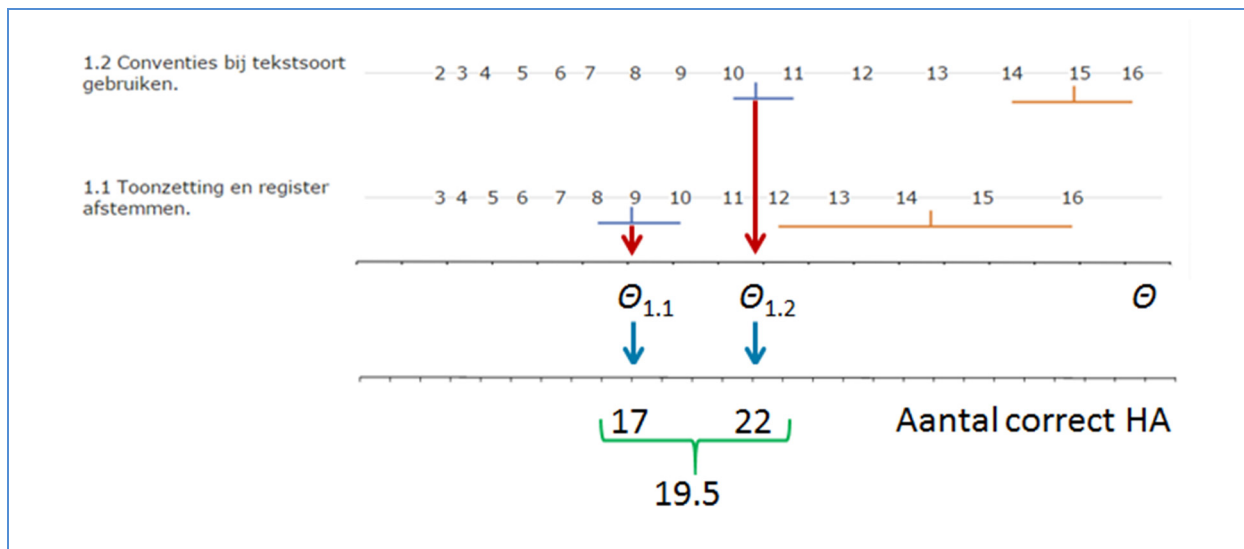
11.6 Vaststelling van de standaard

Na de standaardbepalingssessies werden de standaarden uitgerekend op grond van de oordelen en werd een rapport geschreven. Voor de eerste, beperkt adaptieve afname¹² (2015-2016) werden de oordelen van de experts op de (deel-) aspecten als standaard genomen. Voor de volledig adaptieve afname (2016-2017) zijn de deelaspecten als clusters ingezet om vervolgens de standaard voor de hoofdaspecten te berekenen. Om rekening te houden met verschillen in de relatieve moeilijkheid van de clusters, zijn de oordelen niet gemiddeld maar is interpolatie gebruikt om een standaard te verkrijgen.

Hiervoor is het gemiddelde clusteroordeel van de experts vertaald in een vaardigheidsschatting met lineaire interpolatie ($\theta_{1.1}$ en $\theta_{1.2}$ in Figuur 11-3). De vaardigheidsschattingen zijn vervolgens vertaald naar een aantal correcte responsen op een hoofdaspect. In de illustratie in Figuur 11-3 correspondeert de vaardigheidsschatting $\theta_{1.1}$ met 17 correcte responsen op het hoofdaspect en $\theta_{1.2}$ met 22 correcte responsen. Het gemiddelde aantal correcte responsen van de deelaspecten behorend bij het hoofdaspect leverde vervolgens de standaard voor het hoofdaspect (in de illustratie in Figuur 11-3 is dat gemiddelde 19,5).

¹² Bij de talen gaf de eerste adaptieve afname alleen diagnoses op hoofdaspecten en bij wiskunde alleen diagnoses van de drie wiskundige deelaspecten in twee domeinen.

Figuur 11-3. Illustratie van het samennemen van de oordelen tot een standaard



Na elke pretest is een rapport over de standaardbepaling geschreven voor de vaststellingscommissies van het CvTE. Het rapport bevat de oordelen van de experts en de standaarden, evenals impactinformatie. Impactinformatie toont het aantal pretestleerlingen dat in elke categorie valt (onder, op of boven niveau) na toepassing van de standaarden. De commissie gaf een advies over de standaarden. Op basis van de impact kon bijvoorbeeld een iets hogere of lagere standaard worden aanbevolen. Nadat het CvTE een definitieve, formele beslissing had genomen over de standaarden, konden de standaarden worden toegepast op de pretestleerlingen en konden de opgaven worden gekalibreerd.

11.7 Conclusie

In twee opeenvolgende jaren zijn opgaven gepretest en hebben standaardbepalingen plaatsgevonden (zie de pretestverslagen Cito, 2015 en Cito, 2016). Door de herziene opdracht van OCW in 2014 (zie Hoofdstuk 1) was het niet mogelijk om alle opgaven in één grote pretest af te nemen en met slechts één standaardbepaling te werken. Er zou nader onderzoek verricht kunnen worden naar de mogelijke impact van die gescheiden standaardbepalingen op de kalibratie.

Voor de standaardbepalingen is de 3DC methode succesvol toegepast. Net zoals bij andere criterium-gerelateerde methodes spelen vakexperts een centrale rol in de standaardbepalingen met de 3DC methode. Zij baseren hun oordelen over wat verwacht zou mogen worden van een grensleerling op de opgaven, hun ervaring als docent, empirische feedback en op de inhoudelijke beschrijving van wat een leerling zou moeten kunnen in de tussendoelen (SLO, 2012).

Het hele proces van standaardbepaling tot aan het definitieve besluit nam ongeveer vier weken in beslag. Een norm-gerelateerde methode, zoals is gebruikt bij de ad-hoc-kalibratie (zie Hoofdstuk 10) zou sneller werken maar is minder geschikt. Bij een norm-gerelateerde methode worden standaarden vastgesteld op grond van de *relatieve* positie van leerlingen (Cizek & Bunch, 2007). Bijvoorbeeld, er zou sprake zijn van een norm-gerelateerde methode als de 20% laagst presterende leerlingen als onder niveau geclassificeerd zouden worden. Die norm-gerelateerde methode is minder geschikt wanneer we willen weten hoe een leerling het doet ten opzichte van de doelen die gesteld zijn, onafhankelijk van hoe andere leerlingen het doen. In dat geval kan men beter een (absoluut) criterium-gerelateerde methode gebruiken, zoals de 3DC methode. Juist bij een diagnostische toets waarbij we voor elke leerling de individuele leerbehoeftes willen bepalen, onafhankelijk van hoe andere leerlingen presteren, is een criterium-gerelateerde methode passend.

Om rekening te houden met de andere structuur van de DTT in vergelijking met bijvoorbeeld een centraal examen, is de 3DC methode op een paar punten aangepast voor de DTT:

- De clusters zijn gebaseerd op leerlingmodellen.
- Het IRT-model is alleen gebruikt voor het standaardbepalingsformulier, niet om de standaard uit te drukken.
- Voor de DTT zijn de experts gevraagd om een aantal correcte responsen aan te geven, niet een grensscore.
- De clusters zijn niet gecombineerd tot één standaard maar twee tot vier clusteroordelen zijn gecombineerd tot een standaard voor elk aspect.
- In plaats van een gemiddelde score uit te rekenen, is interpolatie gebruikt om een standaard uit te rekenen.

Na enkele adaptieve afnames door marktpartijen zou geëvalueerd kunnen worden hoe de werkelijke impact is van de standaarden. Indien blijkt dat aanzienlijk meer of minder leerlingen onder, op of boven niveau zitten dan verwacht, zou de standaardbepaling herhaald kunnen worden.

11.8 Literatuur

- Angoff, W.H. (1971). Scale, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 508-600). Washington, DC: American Council on Education.
- Cito (2015). *Diagnostische tussentijdse toets: Verslag pretest 2015*. Arnhem: Cito.
- Cito (2016). *Diagnostische tussentijdse toets: Verslag pretest 2016*. Arnhem: Cito.
- Cito (2018). *Computerized standardsetting using 3DC*. Geraadpleegd op <http://www.cito.com/our-expertise/implementation/3dc>
- Cizek, G. J., & Bunch, M. B. (2007). *Standard Setting*. London: Sage Publications, Inc.
- College voor Toetsen en Examens (2014). *Publieksversie toetswijzer diagnostische tussentijdse toets voor Nederlands, Engels en wiskunde*. Utrecht: College voor Toetsen en Examens. <https://www.pilotdtt.nl/documenten/publicaties/2014/12/15/toetswijzer-ddt>
- Feskens, R., Keuning, J., Van Til, A & Verheyen, R. (2014). *Prestatiestandaarden voor ERK in het examenjaar: Een internationaal ijkingsonderzoek*. Arnhem: Cito.
- Keune, K., & Schouwstra, S. (2014). Standaardbepaling door experts na de eerste pretest van de DTT. In S. Schouwstra (Red.), *De diagnostisch tussentijdse toets: onderzoek 2014* (pp. 47-52). Arnhem: Cito.
- Keuning, J. Straat, J. H., & Feskens, R. C. W. (2017). The Data-Driven Direct Consensus (3DC) procedure: A new approach to standard setting. In S. Blömeke & J. Gustafsson (Eds.), *Standard Setting in education: The Nordic countries in an international perspective* (pp. 263-278). Cham, Switzerland: Springer International Publishing.
- Koops, J. (2016). *Software and manual 3DC*. <http://www.cito.com/our-expertise/implementation/3dc>
- Mitzel, H.C., Lewis, D.M., Patz, R.J., & Green, D.R. (2001). The bookmark procedure: Psychological perspectives. In G.J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 249-281). Mahwah, NJ: Erlbaum.
- Sireci, S.G., Hambleton, R.K., & Pitoniak, M.J. (2004). Setting passing scores on licensure exams using direct consensus. *CLEAR Exam Review*, 15, 21-25.
- SLO: Nationaal expertisecentrum leerplanontwikkeling (april 2012). *Concept-tussendoelen kernvakken onderbouw vo*. Geraadpleegd op <http://www.slo.nl/nieuws/00291/>.
- Verhelst, N. D., Glas, C. A. W., & Verstralen, H. H. F. M. (1995). *OPLM: One parameters logistic model. Computer program and manual*. Arnhem, the Netherlands: Cito.

12 Psychometrische benadering binnen de DTT

Daniel van der Palm en Herbert Hoijtink

12.1 Model

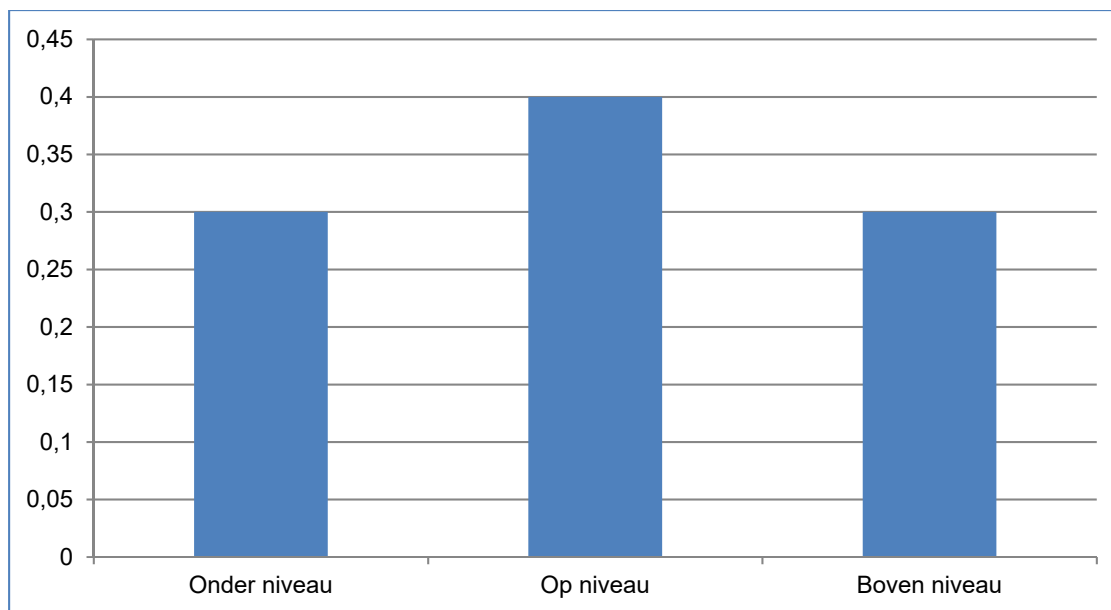
Het psychometrische model voor de DTT is een latenteklassemodel (LK-model) (Goodman, 1974; Lazarsfeld, 1950) dat gecombineerd is met zogenaamde diagnostische hypothesen (Roelofs & Schouwstra, 2012; Hoijtink & Sies, 2013; Sies, 2014). Bij de DTT heeft het LK-model een gefixeerd aantal latente klassen, namelijk drie, en wordt gebruik gemaakt van prior gewichten. Gedurende de afgelopen decennia is het gebruik van het latenteklassemodel gegroeid wat betreft het aantal toepassingen in de praktijk, maar zeker ook wat betreft het aantal wetenschappelijke publicaties. In eerste instantie was het latenteklassemodel bekend als 'Latent Structure Analysis' en binnen de econometrie wordt het ook wel het 'finite mixture'-model genoemd.

Het LK-model wordt gebruikt in situaties waarbij de te meten concept(en) niet direct observeerbaar zijn. Mensen kunnen bijvoorbeeld tot bepaalde latente klassen behoren wat hun persoonlijkheid betreft of politieke overtuiging. Zo kan men bijvoorbeeld niet direct observeren of een persoon introvert/extravert is of bijvoorbeeld boven gemiddeld open staat voor nieuwe ervaringen. Bij een LK-model gebruiken we om die reden indicator variabelen om (indirect) in te schatten tot welke latente klasse elk persoon behoort; dit laatste wordt ook wel classificatie genoemd. In het geval van de DTT worden er drie diagnostische hypothesen (Roelofs & Schouwstra, 2012) gesteld met betrekking tot de beheersing van een leerling: de leerling zit onder niveau (<), op niveau (=) of boven niveau (>). De DTT levert dus geen score op of een zak/slaag resultaat; het doel is om tot een diagnose te komen over de mate waarin een leerling (deel-) aspecten beheerst om, waar nodig, extra tijd en aandacht te kunnen geven aan materie waar een leerling mee worstelt of extra uitdaging te bieden bij materie die de leerling al goed beheerst.

12.2 Ad-hoc-kalibratie, kalibratie en herkalibratie

Er kunnen drie groepen leerlingen onderscheiden worden: de onder-niveaugroep, de op-niveaugroep en de boven-niveaugroep. Het doel van een kalibratie is om voor elk van de drie niveaugroepen de succesansen voor items en testlets te bepalen. Een item is de opgave of het opgave-onderdeel dat tot één onafhankelijk beoordeelde respons leidt. Bij een testlet geeft een leerling antwoord op een meerdere opgave-onderdelen die bij elkaar horen en afhankelijk van elkaar zijn. De eerste kalibratie die plaats kan vinden is een ad-hoc-kalibratie en bestaat uit een simpeler proces dan de standaardkalibratie. In deze ad-hoc-kalibratie wordt een relatieve norm toegepast. Een relatieve norm heeft als voordeel dat deze zeer eenvoudig en snel is toe te passen. De ad-hoc-kalibratie stelt ons in staat om na elke pretest de scholen feedback te geven over de prestaties van hun leerlingen op de pretest in vergelijking met andere pretestscholen (beschreven in de pretestverslagen, Cito, 2015; Cito, 2016). De ad-hoc-kalibratie is ook gebruikt bij de toets- en itemanalyses (zie Hoofdstuk 10) om een indicatie te geven van het onderscheidend vermogen.

Bij een ad-hoc-kalibratie wordt er voor elke leerling per aspect een inschatting gemaakt tot welke niveaugroep hij of zij behoort op basis van het percentiel waar de betreffende leerling in valt wat zijn of haar aantal correcte antwoorden (somscore) betreft (zie bijvoorbeeld Figuur 12-1). De leerlingen die in het 20^{ste} percentiel vallen of lager worden ingedeeld in de onder-niveaugroep, leerlingen die boven het 20^{ste} percentiel zitten maar lager dan het 81^{ste} percentiel worden ingedeeld in de op-niveaugroep en de resterende leerlingen in de boven-niveaugroep. Nu de groepen gedefinieerd zijn kunnen de succesansen voor de items en testlets bepaald worden. Deze laatste stap van de kalibratie gebeurt op dezelfde wijze als bij een standaardkalibratie. Om die reden wordt nu het gehele proces van een standaardkalibratie uiteengezet.



Figuur 12-1. Voorbeeld van proportionele verdeling van leerlingen bij een ad-hoc-kalibratie op basis van aantal correcte antwoorden

Bij een standaardkalibratie wordt een (absolute) criterium-gerelateerde methode gebruikt. Met behulp van experts en de data-driven direct consensus methode worden voor elk vak de standaarden bepaald (zie Hoofdstuk 11). Per aspect zijn er twee standaarden; dit zijn de aantallen correcte responsen die volgens de experts horen bij (a) een leerling wiens beheersing van een aspect nét op niveau is en (b) een leerling wiens beheersing nét boven niveau is. Nadat de standaarden bepaald zijn wordt er voor elke pretestleerling per aspect vastgelegd of hij of zij onder, op of boven niveau zit. Het essentiële verschil tussen een ad-hoc-kalibratie en een standaardkalibratie is dus of de grenswaarden van de drie niveaus gebaseerd zijn op relatieve percentielen of op absolute standaarden. Bij relatieve percentielen (de ad-hoc-kalibratie) hangt de norm af van de prestaties van andere leerlingen. Absolute standaarden die zijn gezet met behulp van vakexperts zijn gebaseerd op de tussendoelen, de toetsinhoud en empirische feedback (zie Hoofdstuk 11).

Na het bepalen van de drie groepen leerlingen moeten de parameters van het LK-model worden geschat. Er zijn drie sets aan parameters: (1) de succesansen voor losse items, (2) de succesansen voor testlets en (3) de prior-modelkansen. De succesansen voor de losse items kunnen simpelweg worden geobserveerd binnen de drie groepen leerlingen (de onder-niveaugroep, de op-niveaugroep en de boven-niveaugroep). De succeskans voor item j binnen de boven-niveaugroep, bijvoorbeeld, wordt als volgt berekend,

$$P_j(>) = \frac{N_j(>)}{N_j(>) + M_j(>)}$$

waarbij $N_j(>)$ staat voor het aantal leerlingen in de boven-niveaugroep dat item j correct beantwoord heeft en $M_j(>)$ het aantal boven-niveauleerlingen dat item j incorrect beantwoord heeft.

Bij een testlet geeft een leerling antwoord op een set opgaven of onderdelen van opgaven die bij elkaar horen en afhankelijk van elkaar zijn. Het is om die reden belangrijk dat er rekening gehouden wordt met het feit dat de verschillende antwoorden op deze set opgaven onderling een statistische afhankelijkheid kunnen vertonen. De oplossing voor dit probleem is dat we succesansen voor testlets op het niveau van responspatronen modelleren. Een testlet die drie responsen oplevert heeft $2^3 = 8$ mogelijke responspatronen en voor elk van de drie niveaugroepen moeten we de kans schatten dat elk van de 8 mogelijke responspatronen gegeven wordt door een leerling. Het aantal mogelijke responspatronen neemt snel toe naargelang er meer responsen in een testlet zitten en lang niet alle mogelijke responspatronen zullen daadwerkelijk geobserveerd worden bij een afname. Om die reden gebruiken we voor testlets een

tweede-orde log-lineair model om alsnog voor elk mogelijk responspatroon een kans te kunnen schatten. Zonder deze informatie zou het niet mogelijk zijn om de gehele reeks aan mogelijke responspatronen mee te nemen in de simulatieprocedure. In het geval dat er meer dan vier responsen zijn die eigenlijk tot één testlet behoren, worden deze responsen niet als testlet opgevat maar verwerkt als losse items.

Naast de twee sets aan succesansen hebben we tot slot ook nog de priors nodig om de niveaugroep van een leerling in te schatten op basis van zijn of haar responspatroon. De priors zijn drie gewichten die ons in staat stellen om tendensen in de diagnoses bij te sturen. Een tendens kan bijvoorbeeld zijn dat minder onder-niveauleerlingen correct gediagnosticeerd worden dan op-niveauleerlingen. In eerste instantie gebruiken we zogenaamde niet-informatieve priors; de drie gewichten zijn dan allen gelijk aan $1/3$ en sturen dus nog niet bij. De verdere details omtrent priors komen in sectie 12.4 aan bod.

12.2.1 Herkalibratie

Het resultaat van een kalibratie is een verzameling van geschatte succesansen voor de items en geschatte kansen voor de responspatronen van de testlets. Wanneer echter nieuwe afnamedata beschikbaar komen voor eerder gekalibreerde items en testlets, dan willen we in principe de kalibratie opnieuw uitvoeren op basis van alle beschikbare data om de invloed van steekproeffluctuaties te verminderen (standaardfouten van modelparameters nemen af naarmate de steekproefgrootte toeneemt). Het is dan van belang om eerst te onderzoeken of de resultaten van een kalibratie op basis van de nieuwe data (item- en testletparameters) niet te veel afwijken van de eerder geschatte parameters. Omdat we te maken hebben met steekproeven ligt het in de lijn der verwachting dat er kleine verschillen zijn in de geschatte kansen tussen verschillende afnames, maar er kan ook sprake zijn van een systematische verschuiving, wat ook wel 'parameter drift' wordt genoemd (zie ook Maas, 2017). Indien er sprake is van een verschuiving, dan moet er nader bekeken worden hoe we omgaan met deze discrepantie. Als de resultaten redelijkerwijs voldoende overeenkomen, dan kan er een herkalibratie gedaan worden op basis van alle beschikbare data. Vervolgens wordt de output van deze herkalibratie gebruikt bij het samenstellen van nieuwe toetsen.

12.2.2 Imputatie

Vanwege de adaptieve opzet van de DTT krijgen niet alle leerlingen alle opgaven voorgelegd en is sprake van missende waarden. Binnen de groep onder-niveauleerlingen zullen bepaalde opgaven bijvoorbeeld niet of weinig gemaakt zijn omdat deze opgaven later in de toets voorkomen en alleen worden aangeboden aan leerlingen die waarschijnlijk op-niveau zitten. Een andere oorzaak voor missende waarden die voor komt is dat sommige leerlingen niet alle opgaven voorgelegd hebben gekregen omdat de diagnose al zeker genoeg was (zie Hoofdstuk 13). Als de missende waarden niet zouden worden aangepakt, dan zouden er een aantal opgaven kunnen zijn waarvoor er binnen één of meer van de drie groepen geen of een zeer laag aantal responsen geobserveerd is. Oftewel, de succeskans van die niveaugroep kan dan niet geschat worden of de schatting is zeer instabiel, omdat deze gebaseerd is op een klein aantal observaties.

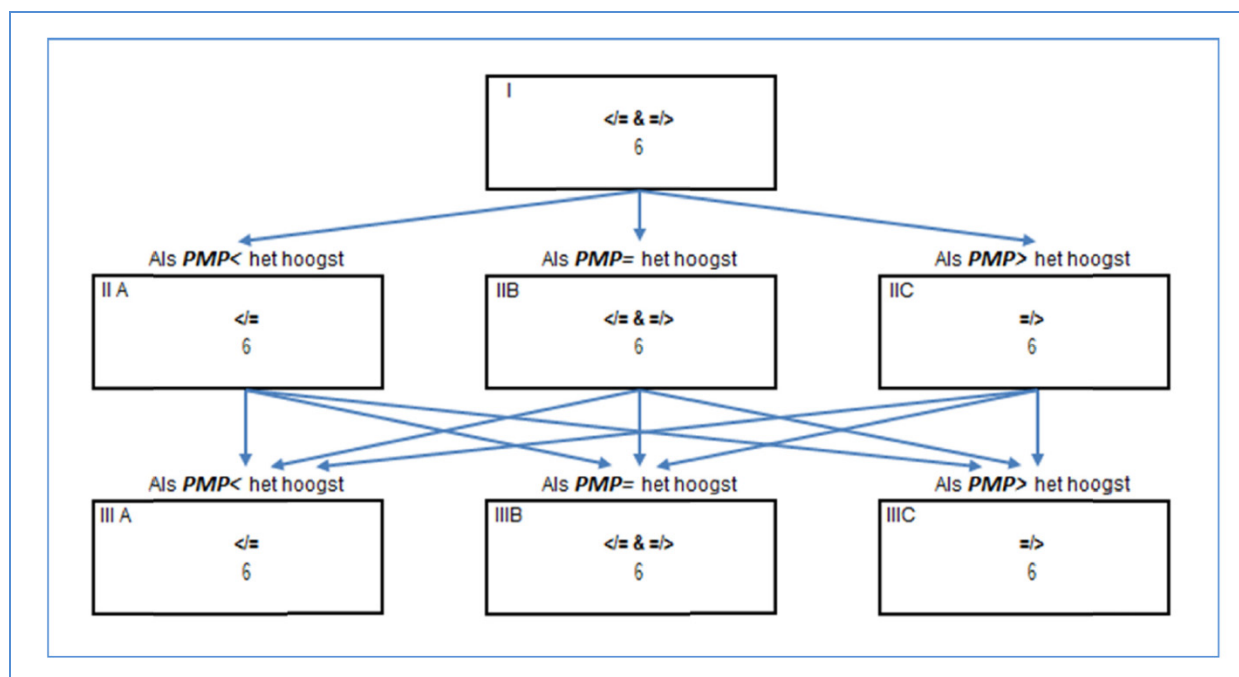
Met behulp van imputatie is het mogelijk om alsnog responsen te schatten voor leerlingen met missende waarden en op deze manier kan de stabiliteit van de geschatte kansen verhoogd worden. In het geval van de DTT wordt er een LK-model gebruikt om de simultane kansverdeling voor alle responsen te schatten. Aan de hand van een dergelijke geschatte kansverdeling kunnen er responsen gegenereerd worden voor leerlingen met missende waarden.

Het aantal latente klassen dat gebruikt wordt voor de imputatie wordt iteratief bepaald aan de hand van een informatiecriterium. Een informatiecriterium maakt het mogelijk om een model te kiezen waarbij modelfit wordt afgewogen tegen het aantal parameters. Voor elke dataset wordt er eerst een LK-model met twee klassen geschat en vervolgens wordt in elke iteratie het aantal klassen opgehoogd, totdat de verbetering in de modelfit niet meer opweegt tegen de toename in het aantal parameters.

Als het LK-model voor imputatie geschat is dan wordt voor iedere leerling met minimaal één missende waarde berekend wat diens kans is om tot elk van de drie groepen te behoren en wordt de leerling willekeurig ingedeeld op basis van die kansen. Vervolgens worden simpelweg de conditionele responskansen gebruikt om voor iedere leerling responsen te genereren voor items met missende waarden.

12.3 Optimalisatie van de blokindeling door middel van simulatie

Met de drie sets aan parameters is het LK-model in principe volledig gedefinieerd. Wat echter nog niet vast staat is welke opgaven we precies voorleggen aan de leerlingen. In theorie is het mogelijk om iedere leerling alle opgaven voor te leggen, ongeacht zijn of haar beheersingsniveau. In de praktijk zou dit echter een zeer tijdrovend proces zijn en een te grote belasting vormen voor de leerling. Daarom wordt gebruik gemaakt van een adaptief design (zie ook Hoofdstuk 13). Een adaptief design bij de DTT bestaat uit 2, 3 of 4 lagen¹³ aan opgaven met in elke laag een apart blok aan opgaven voor leerlingen die waarschijnlijk onder niveau zitten, leerlingen die waarschijnlijk op niveau zitten en leerlingen die waarschijnlijk boven niveau zitten.



Figuur 12-2. De adaptieve procedure met blokken (Hojtink & Sies, 2014)

Zoals te zien is in Figuur 12-2 maken alle leerlingen voor één aspect eerst allemaal hetzelfde blok aan opgaven. Deze eerste set opgaven zou dus goed onderscheid moeten kunnen maken tussen leerlingen onder en leerlingen op niveau, evenals tussen leerlingen op en leerlingen boven niveau. Na het doorlopen van het eerste blok wordt er een inschatting worden gemaakt van de niveaugroep. Deze berekening is als volgt:

De priors en de succesansen voor elk van de drie niveaugroepen zijn geschat en dus is het mogelijk om voor elke niveaugroep te berekenen wat de kans is dat een bepaald responspatroon wordt geobserveerd (Hojtink, Béland, & Vermeulen, 2012); namelijk, de prior behorende bij een niveaugroep vermenigvuldigd met de kans dat voor item 1 de gegeven respons zich voordoet en zo verder de productsom over alle items die beantwoord zijn door een leerling. Als we deze drie kansen elk delen door de som van deze drie kansen dan komen we tot de 'posterior membership probabilities' (PMP). Deze PMP's geven aan hoe waarschijnlijk het is dat een leerling gegeven zijn of haar responspatroon tot een specifieke niveaugroep behoort.

Na het eerste blok wordt elke leerling voorlopig geplaatst in de niveaugroep waarvoor hij of zij de grootste PMP heeft (die het meest waarschijnlijk is). Vervolgens krijgt de leerling in laag 2 een blok opgaven dat hoort bij deze niveaugroep. Bijvoorbeeld, als de leerling waarschijnlijk onder niveau zit krijgt de leerling opgaven die een goed onderscheid maken tussen onder en op niveau (linker blok in Figuur 12-2). Als de

¹³ In 2016 was sprake van 2 lagen en in 2017 sprake van 3 lagen zonder doortoetsen (wiskunde) of met doortoetsen (talen)

leerling waarschijnlijk boven niveau zit, krijgt de leerling juist opgaven die een goed onderscheid maken tussen op- en boven-niveau (rechter blok in Figuur 12-2). Aan het eind van laag 2 (en laag 3 als er 4 lagen zijn) wordt dit proces herhaald. Aan het eind van de toets zijn de PMP's hetgeen gerapporteerd wordt: de toetsuitkomst. Aan het eind van de toets weten we hoe waarschijnlijk het is dat de leerling gegeven zijn of haar responspatroon bij de onder-niveaugroep hoort ($PMP <$), bij de op-niveaugroep hoort ($PMP =$) en bij de boven-niveaugroep ($PMP >$) hoort. De leerling wordt geplaatst in de niveaugroep die het meest waarschijnlijk is (de hoogste PMP heeft). Bijvoorbeeld, als de leerling het meest waarschijnlijk bij de op-niveaugroep hoort, krijgt de leerling de diagnose op niveau.

Hiermee is in de technische zin een toetsafname voltooid voor één hoofdaspect. Het is echter dan nog niet duidelijk op welke wijze we opgaven zouden moeten verdelen over de verschillende blokken. Om die vraag te kunnen beantwoorden maken we gebruik van een simulatieprocedure.

12.3.1 Simulatie

Door middel van simulatie is het mogelijk om te zoeken naar een zo goed mogelijk verdeling van de opgaven en responsen over de blokken. Allereerst is het nodig om te definiëren welke criteria we hanteren om de kwaliteit van een indeling te beoordelen. Er zijn twee criteria waarmee rekening is gehouden: (1) de grootte van de kansen op correcte diagnoses gegeven het ware niveau van iedere leerling en (2) de balans tussen de drie kansen op correcte diagnoses.

In de Tabel 12-1 worden de 9 mogelijke uitkomsten weergegeven van de toets, waarbij we onderscheid maken tussen de werkelijke niveaugroep van een leerling en de ingeschatte niveaugroep aan het eind van de toetsafname. In het hypothetische voorbeeld krijgt 88% van de onder-niveauleerlingen de juiste diagnose onder niveau, de rest krijgt een verkeerde diagnose: 9% de verkeerde diagnose op niveau en 3% de diagnose boven niveau. Van de op-niveauleerlingen wordt 87% correct gediagnosticeerd en van de boven-niveauleerlingen 91%.

Tabel 12-1. Voorbeeld van classificatietabel

Ware niveau	Geschat niveau		
	Onder	Op	boven
Onder	0,88	0,09	0,03
Op	0,07	0,87	0,06
Boven	0,01	0,08	0,91

Het eerste criterium, de kans op correcte diagnoses, houdt dus in dat we de som van de diagonaal van deze 3-bij-3 tabel proberen te maximaliseren, aangezien dit de uitkomsten zijn waarbij de ware en geschatte niveaugroep overeenkomen. De simulatie is ontwikkeld om zo goed mogelijk aan de twee criteria te voldoen en werkt als volgt:

- 1) De opgaven worden op willekeurige wijze verdeeld over de blokken.
 - a. In het geval van bijvoorbeeld 3 lagen (zonder doortoetsen) zijn er 7 blokken, zoals weergegeven in Figuur 12-2. Als er 7 blokken moeten komen, worden de opgaven over 5 blokken verdeeld, omdat het middelste op-niveaublok in laag 2 en 3 wordt gevormd door opgaven uit onder-niveaublok en uit het boven-niveaublok. Hiervoor wordt per deelaspect willekeurig de helft van de opgaven genomen uit het onder-niveaublok en de helft uit het 'boven niveau' blok.
 - b. Het aantal opgaven dat aan een blok toegewezen mag worden wordt beperkt door de beschikbare toetsduur. Bij aanvang is bekend wat de maximale toetsduur is die beschikbaar is per hoofdaspect. Bijvoorbeeld, als er drie uur beschikbaar is voor de hele afname en er zijn vier hoofdaspecten en zaai-opgaven, dan is er per hoofdaspect 2160 seconden beschikbaar ($10800 \text{ seconden} / 5$). De beschikbare toetsduur per hoofdaspect wordt gedeeld door het aantal blokken dat een leerling moet maken (lagen) om tot de beschikbare tijd per blok te komen. Vervolgens delen we de beschikbare tijd per te maken blok door de beschikbare antwoordtijd per opgave om tot een maximaal aantal opgaven per blok te komen. Bijvoorbeeld, bij een toetsduur per hoofdaspect van 2100 seconden, en

een design met 3 lagen met doortoetsen krijgt een leerling vier blokken: $2100/(4 \text{ blokken}) = 525$ seconden per blok, $525/(50 \text{ seconden antwoordtijd benodigd per opgave}) = 10$ opgaven, dus mag een leerling per blok maximaal 10 opgaven aangereikt krijgen. Het kan dus betekenen dat niet alle beschikbare opgaven daadwerkelijk in de toets voorkomen. De benodigde antwoordtijd per opgave werd uitgerekend op grond van de waargenomen responstijden tijdens de voorgaande afname.

- 2) Voor de blokindeling worden 10.000 gesimuleerde leerlingen aangemaakt die onder niveau zitten, 10.000 leerlingen die op niveau zitten en 10.000 leerlingen die boven niveau zitten. Deze gesimuleerde leerlingen doorlopen de toets met de gecreëerde blokindeling. Er worden antwoorden op de opgaven gegenereerd die de gesimuleerde leerlingen volgens de blokindeling krijgen, op basis van de succesansen conditioneel op de niveaugroep. Aan het einde van elke laag wordt de niveaugroep van de gesimuleerde leerling geschat. Aan het eind van de toets is de diagnose bekend en blijkt in welke van de negen mogelijke uitkomstgroepen de gesimuleerde leerling is terecht gekomen (in de classificatietabel, zie Tabel 12-1); de frequentie van de betreffende uitkomst in de 3-bij-3 tabel wordt met één opgehoogd.
- 3) Deze procedure van blokindelingen maken en simuleren loopt door totdat er 100 blokindelingen gevonden zijn die voldoen aan de eisen. Er zijn drie restricties:
 - a. De tijdsrestrictie
 - b. De eis dat ieder blok minimaal één respons moet hebben voor elk deelaspect
 - c. De eis dat de aantallen responsen voldoende evenredig verdeeld zijn over de blokken
- 4) Van deze 100 indelingen wordt uitgerekend wat de som is van de drie kansen op correcte diagnoses en worden de 10 indelingen aangehouden die de hoogste som van de diagonaal hebben.
- 5) Tot slot worden de priors geoptimaliseerd om de verschillen in grootte van de kansen op correcte diagnoses te minimaliseren (zie Sectie 12.4).

Deze tien beste oplossingen worden teruggekoppeld en toetsdeskundigen bepalen in samenspraak met de afdeling psychometrie welke oplossing de voorkeur heeft (zie verder Hoofdstuk 13).

Opgavegroepen in de blokindeling

Soms moet bij de optimalisaties rekening gehouden worden met de aanwezigheid van opgavegroepen. Dit zijn opgaven die op verschillende schermen gepresenteerd worden en die bij elkaar horen, bijvoorbeeld drie opgaven die over dezelfde tekst gaan. Dit betekent dat dat deze opgaven verplicht tegelijkertijd aangeboden moeten worden in een specifieke volgorde. Het zoekalgoritme om een optimale verdeling van opgaven te vinden houdt rekening met opgavegroepen door van tevoren dergelijke opgaven samen te bundelen in een tijdelijke structuur. Als we twee opgaven hebben met ieder één respons en deze opgaven behoren tot dezelfde opgavegroep, dan maakt de simulatie er tijdelijk één opgave van met twee responsen. Deze kwestie speelt overigens alleen in het deel van de simulatie waar bekeken wordt of willekeurige blokindelingen voldoen aan de verschillende restricties. Als er eenmaal een blokindeling is gevonden die voldoet aan de restricties dan worden deze bundels weer teruggebracht naar hun originele vorm, maar dan hebben we wel de garantie dat de betreffende opgaven direct achter elkaar zijn geplaatst in hetzelfde blok.

12.3.2 Uitstapkansen

In het adaptieve ontwerp is er ook rekening mee gehouden dat een leerling niet alle blokken opgaven hoeft te maken als de diagnose al heel zeker is, bijvoorbeeld als de leerling de opgaven consistent heel goed of heel slecht maakt (zie verder Hoofdstuk 13). De leerling mag als het ware uit de toets stappen.

Een uitstapkans is een criterium waarmee per niveau aan het einde van elke laag bekeken kan worden of een leerling kan uitstappen omdat er voldoende zekerheid is omtrent diens niveau. Deze uitstapkans zou berekend kunnen worden voor elke laag apart, maar bij de technische opzet van de adaptieve module is gekozen voor één set aan uitstapkansen per hoofdaspect (en één set voor elk deelaspect). De set uitstapkansen wordt berekend op basis van de opgaven in de eerste laag, omdat dit de meest conservatieve waarden oplevert. Deze uitstapkansen worden vervolgens na elke laag toegepast.

Bij het bepalen van de uitstapkansen maken we allereerst de aanname dat items en testlets in een diagnostische tussentijdse toets -toets onderscheidend vermogen hebben. De mate van onderscheidend vermogen varieert van item tot item en van testlet tot testlet. Desalniettemin geldt dat elke additionele

opgave die we de leerlingen voorleggen, de zekerheid omtrent het ingeschatte niveau verhoogt (tenzij de succesansen dus gelijk zijn voor de verschillende niveaus, maar dit kunnen we bij de kalibratie detecteren en komt hooguit voor een zeer beperkt aantal items voor).

Onder het DTT-model geldt dat hoe groter het aantal opgaven is dat een leerling maakt, des te groter de kans is dat er een correcte diagnose plaatsvindt (i.e., het geschatte niveau is gelijk aan het ware niveau van een leerling). Theoretisch gezien is het zo dat de kans op correcte diagnose naar één gaat als het aantal opgaven naar oneindig gaat, aangezien de invloed van toeval onder die omstandigheden meer en meer wegvalt. In de praktijk hebben we natuurlijk een beperkt aantal opgaven en is de invloed van toeval groter. We komen dus niet zo snel tot een toets die perfect onderscheid maakt.

Ook al maakt de toets geen perfect onderscheid, het observeren van een lichtelijk afwijkend responspatroon (gegeven het niveau van een leerling) is waarschijnlijker dan het observeren van een zeer afwijkend responspatroon. Oftewel, de kans dat een leerling uit de onder-niveaugroep één heel moeilijke opgave toevallig toch correct beantwoord (door gokken of door een moment van extra inzicht) is groter dan de kans dat diezelfde leerling tien heel moeilijke opgaven correct weet te beantwoorden. Dit betekent ook in het algemeen, als er een verkeerde diagnose gemaakt wordt, dat het verschil tussen de waarschijnlijkheid (PMP) van de ingeschatte niveaugroep en de waarschijnlijkheid (PMP) van de ware niveaugroep doorgaans relatief klein is. Bijvoorbeeld, als een onder-niveauleerling de verkeerde diagnose op niveau krijgt is het waarschijnlijker dat de toetsuitkomst van die leerling [$PMP_{<} = 0,3$; $PMP_{=} = 0,5$; $PMP_{>} = 0,2$] is, dan dat de toetsuitkomst [$PMP_{<} = 0,1$; $PMP_{=} = 0,2$; $PMP_{>} = 0,7$] is. We kunnen dan ook concluderen dat er gemiddeld genomen een hogere proportie aan incorrecte diagnoses voor komt in de verzameling van PMP-vectoren die weinig afwijken van een uniforme verdeling [$1/3, 1/3, 1/3$] en dat we minder fouten zullen aantreffen onder de leerlingen met PMP-vectoren die sterk neigen naar één niveau, bijvoorbeeld [$0,7, 0,2, 0,1$].

De regel is geïmplementeerd dat er alleen vroegtijdig uitgestapt mag worden als het percentage correcte diagnoses in de groep uitstappers 95% of hoger is (zie Hoofdstuk 13) en uitstappen mag op zijn vroegst na het volledig doorlopen van de eerste laag. Om een optimale blokindeling te vinden in de simulatieprocedure, wordt voor elk van de drie niveaugroepen uitgerekend wanneer voldaan kan worden aan deze 95% eis. De gesimuleerde leerlingen—die bijvoorbeeld als onder niveau worden ingeschat—hebben in ieder geval een PMP voor dit niveau groter dan $1/3$ en een PMP kan oplopen tot (praktisch) 1,0. Zoals eerder genoemd is het de verwachting dat er meer foute diagnoses voorkomen onder de leerlingen wiens PMP voor ‘onder niveau’ maar nét de grootste is. Om die reden worden de leerlingen die ingeschat worden op ‘onder niveau’ geordend op basis van hun $PMP(<)$ van klein naar groot. Het is in principe mogelijk dat het aantal foute diagnoses direct al lager is dan 5%, als bijvoorbeeld de eerste laag opgaven bevat die zeer goed onderscheid kunnen maken tussen onder- en op-niveau en tussen op-niveau en boven-niveau. In dat geval wordt de uitstapkans gelijkgesteld aan ,01 en mag dus vrijwel iedereen uitstappen na de eerste laag als ze gediagnosticeerd zijn als onder niveau. Het is ook mogelijk dat de eis van 95% helemaal niet gehaald kan worden door minder of minder goede opgaven; de uitstapkans wordt dan op ,99 gezet. Voor de gevallen tussen deze twee extremen in wordt in de simulatieprocedure de minimale PMP steeds wat opgehoogd totdat de groep met PMP's boven een specifieke PMP-waarde minimaal voor 95% bestaat uit correct gediagnosticeerde leerlingen. Wanneer er voldaan is aan de 95% eis wordt de uitstapkans gelijkgesteld aan deze PMP.

12.4 Optimalisatie van de priors

Zoals eerder al kort genoemd zijn priors de gewichten die gebruikt kunnen worden om tendensen in de kansen op correcte diagnoses bij te sturen. We hebben bij de DTT drie niveaus en dus drie prior gewichten. Als we eenmaal een bepaalde blokindeling gevonden hebben in de simulatieprocedure dan weten we in ieder geval dat de som van de kansen op correcte diagnoses het hoogst was voor deze indeling (binnen de set van onderzochte blokindelingen). Dit garandeert echter niet dat de kans op correcte diagnoses voor de drie niveaus even groot zijn. De verschillen tussen deze kansen zijn meestal niet bijzonder groot omdat er doorgaans vele alternatieve indelingen zijn die deze afwijking niet vertonen en in het zoekalgoritme verdringen deze alternatieve indelingen eventuele indelingen die mankementen vertonen.

Doordat we echter te maken hebben met discrete eenheden die we moeten verdelen (items en testlets) lukt het nooit om de drie kansen op correcte diagnoses exact gelijk te krijgen. Om dit alsnog zo goed mogelijk te bewerkstelligen gebruiken we priors. Voor de 10 beste oplossingen worden de volgende stappen gezet om de priors te optimaliseren en de kansen op correcte classificatie zo gelijk mogelijk te krijgen:

- (1) neem van een indeling de kansen op correcte classificatie, $pr_{<}$, $pr_{=}$ en $pr_{>}$, en bereken de vector $\mathbf{m}$ als volgt: $\mathbf{m} = \left(pr_{>}, \frac{pr_{>}}{pr_{=}/pr_{<}}, pr_{<} \right)$
- (2) Normaliseer vector $\mathbf{m}$ door elk van de drie elementen te delen door de som van $\mathbf{m}$
- (3) Vermenigvuldig de genormaliseerde vector $\mathbf{m}$ met de oude priors en normaliseer het resultaat opnieuw om tot een nieuwe set aan priors te komen
- (4) herbekeken $pr_{<}$, $pr_{=}$ en $pr_{>}$, gegeven de nieuwe priors en ga weer naar stap 1 zolang de aanpassing van de priors in een iteratie groter is dan de ingestelde ondergrens.

12.5 Literatuur

- Cito (2015). *Diagnostische tussentijdse toets: Verslag pretest 2015*. Arnhem: Cito.
- Cito (2016). *Diagnostische tussentijdse toets: Verslag pretest 2016*. Arnhem: Cito.
- Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215-231.
- Hojtink, H., & Sies, A. (2013). De psychometrie van de diagnostisch tussentijdse toets. In S. Schouwstra (Red.), *De diagnostisch tussentijdse toets: onderzoek 2013* (pp. 13-62). Arnhem: Cito.
- Hojtink, H., & Sies, A. (2014). Adaptieve procedure met itemblokken. In S. Schouwstra (Red.), *De diagnostisch tussentijdse toets: onderzoek 2014* (pp. 33-40). Arnhem: Cito.
- Hojtink, H., Béland, S., & Vermeulen, J. A. (2014). Cognitive diagnostic assessment via Bayesian evaluation of informative diagnostic hypotheses. *Psychological Methods*, 19(1), 21-38.
- Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, & J. A. Clausen (Eds.), *Measurement and prediction* (pp. 361-412). Princeton: Princeton University Press.
- Maas, L. (2017). *Potential impact of item parameter drift on diagnostic accuracy in educational testing*. [ongepubliceerde master thesis]. Utrecht: Utrecht University.
- Roelofs, E., & Schouwstra, S. (Red.). (2012). *Diagnostische tussentijdse toets: Verslag van de voorstudie*. Arnhem: Cito.
- Sies, A.M.A. (2014). *Cognitive diagnostic testing using calibrated hypotheses* [ongepubliceerde master thesis]. Utrecht: Utrecht University.

13 Adaptiviteit

Sanneke Schouwstra, Peter van Os en Joke Hofstee

De DTT heeft een formatieve functie. De DTT dient derhalve meer fijnmazige informatie te geven dan een algemene vaardigheidsscore, zodat het leerproces afgestemd kan worden op de individuele leerbehoefte. Om die reden is de DTT gebaseerd op leerlingmodellen die alle kennis- en vaardigheidsaspecten beschrijven, die noodzakelijk zijn voor een prestatie op het vereiste niveau (zie Hoofdstuk 1). De DTT beoogt al deze kennis- en vaardigheidsaspecten te diagnosticeren (zie de Hoofdstukken 4, 5 en 7). Voor elk van de eerste vaardigheden die geoperationaliseerd zijn (schrijfvaardigheid Nederlands, schrijfvaardigheid Engels en wiskunde), moesten acht tot dertien deelaspecten gediagnosticeerd kunnen worden. Zoveel diagnoses kunnen alleen geleverd worden als de toets adaptief is. Met een computergestuurde adaptieve toets (CAT) kan effectiever getoetst worden doordat na elke beantwoorde opgave nagegaan wordt hoeveel we al weten en welke opgaven nog voorgelegd moeten worden om een accurate diagnose te kunnen stellen.

Hoe de adaptiviteit eruitziet, is afgestemd op de functie van de toets. Adaptiviteit kan ingezet worden voor verschillende doeleinden. Binnen onderwijskundige toetsen wordt adaptiviteit vaak gebruikt om nauwkeuriger te meten over de hele vaardigheidsverdeling, of om ook in de uiteindes van de vaardigheidsverdeling nauwkeurig te kunnen meten. Bij psychologische meetinstrumenten gebruikt men adaptiviteit met name om het aantal vragen, en daarmee de toetstijd, te reduceren (zie onder andere Hol, 2006). Bij de DTT is de primaire reden voor de adaptiviteit om in dezelfde toetstijd als bij een lineaire afname méér accurate diagnoses te kunnen stellen. In de DTT willen we voor elk aspect van een leerlingmodel kunnen schatten of de leerling onder, op of boven niveau zit (zie Roelofs & Schouwstra, 2012; Sies, 2014). Hierbij wordt geen model gebruikt dat gebaseerd is op de itemresponstheorie, zoals doorgaans het geval is bij computergestuurde adaptieve toetsen (Eggen, 2012), maar het model beschreven in Hoofdstuk 12.

13.1 Gradaties van adaptiviteit

In de DTT zijn verschillende gradaties van adaptiviteit gecombineerd (zie Figuur 13-1). Allereerst, is de DTT ontwikkeld voor vijf verschillende leerwegen/schoolsoorten (jaar 2 van vmbo-bb, vmbo-kb en vmbo-gt en jaar 3 van havo en vwo) en voor elke schoolsoort/leerweg is een aparte versie. Tussen de versies van aangrenzende leerwegen/schoolsoorten is overlap in opgaven gecreëerd¹⁴, omdat oorspronkelijk de opdracht was om ook een opstroomindicatie te geven (zie Roelofs & Schouwstra, 2012). In de herziene opdracht is deze eis komen te vervallen, mede omdat dit zou vereisen dat ook in de tweede klas havo (waar de DTT niet afgenomen wordt) gepretest zou moeten worden (zie Hoofdstuk 1).

Ten tweede, zijn er naast de verschillende versies twee stadia in de meting. Schattingen na de voorstudie en de eerste try-out lieten zien dat vele uren toetsen nodig zouden zijn voor het diagnosticeren van een volledig leerlingmodel. Om die reden is besloten om in twee stadia te werken: In de eerste fase worden alle hoofdaspecten gediagnosticeerd. In het tweede stadium, het doortoetsstadium, worden deelaspecten alleen nader gediagnosticeerd als dit nodig is en er nog toetstijd is (Hojtink & Sies, 2013). In sectie 13.4.2 wordt het doortoetsalgoritme besproken waarmee bepaald wordt in welke volgorde de deelaspecten nader gediagnosticeerd worden.

Tenslotte, vindt de meest verfijnde adaptiviteit plaats bij de diagnostisering van de hoofdaspecten. Om sneller tot de gewenste diagnoses te komen wordt in deze eerste fase een adaptieve procedure met opgave-blokken gebruikt.

¹⁴ Tussen 3 havo en 2 vmbo-gt is geen overlap om dat deze niet aangrenzend zijn door het verschillende leerjaar



Figuur 13-1. Gradaties van adaptiviteit in de DTT

13.2 Adaptieve procedure met blokken

Bij leesvaardigheid, maar ook bij schrijfvaardigheid Nederlands zijn er soms meerdere opgaven bij één tekst. Deze opgaven dienen natuurlijk bij elkaar te blijven en samen aangeboden te worden. In de DTT zijn dus bestaande blokken van opgaven (testlets). Dit is echter niet de voornaamste reden om met blokken opgaven te werken. Op grond van een adaptieve afname op opgaveniveau is herkalibratie, het opnieuw bepalen van de parameters van opgaven, niet mogelijk zonder een nieuwe pretest (Zwitsers & Maris, 2015), terwijl herkalibratie altijd nodig is na verloop van tijd. Wanneer men met een blokkenstructuur werkt is herkalibratie op grond van de adaptieve afname wel mogelijk.

Herkalibratie is noodzakelijk als de afnamegroep op een andere wijze reageert op de opgaven dan tijdens de pretest (of eerdere afnames) is gebeurd, waardoor de parameters van de opgave veranderen (zogenaamde itemparameter drift, zie ook Glas, 2000). Na een paar jaar zal de afnamegroep op een andere wijze op de opgaven reageren doordat leerlingen hebben geoefend voor de meting, het onderwijs veranderd is of doordat de opgaven bekend zijn geworden (Maas, 2017). Men kan al eerder een andere reactie verwachten als de wijze van afnemen verandert of als de pretestgroep heel klein was of niet representatief was voor de leerlingpopulatie. Na de herziene opdracht van het Ministerie van OCW (zie Hoofdstuk 1), waarin met een vrijwillige groep van scholen werd gewerkt, werd herkalibratie noodzakelijk in verband met de kleinere pretestgroep (500-1000 observaties per opgave).

Het werken met blokken van opgaven ondervangt ook andere nadelen van adaptiviteit op opgaveniveau (zie onder andere Hendrickson, 2007). Wanneer men werkt met blokken is de kwaliteit van de toets beter te controleren door experts. Toetsdeskundigen kunnen de inhoud van de blokken controleren om contexteffecten te voorkomen, de inhoud van de opgaven te balanceren en de juiste volgorde te bepalen. Ook de mate waarin de opgaven gezien zijn door leerlingen ('item exposure') is eenvoudig te controleren en adaptiviteit met blokken is technisch minder veeleisend.

Simulatieonderzoek heeft laten zien dat de adaptieve procedure met blokken goed werkt wanneer de spreiding van de succesansen als leidraad wordt genomen en niet de moeilijkheid van de opgaven. Het informatieverlies ten opzichte van de adaptieve procedure op opgaveniveau is gering (Hojtink & Sies, 2014).

13.2.1 Blokkenontwerp

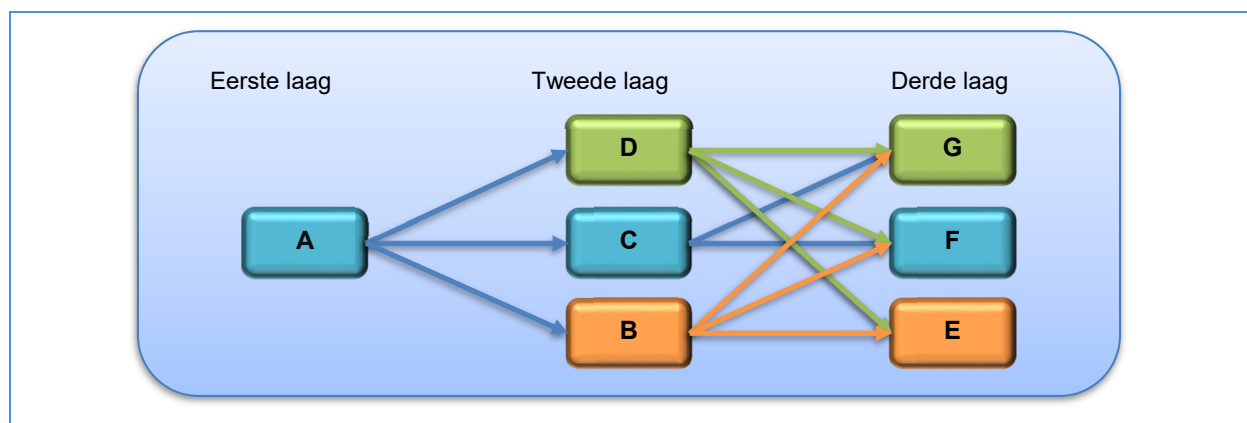
In Figuur 13-2 staat is het blokkenontwerp getekend. Het blokje in de eerste laag (A) wordt aan alle leerlingen aangeboden. Hierna wordt één blokje uit de tweede laag aangeboden:

- Blokje B wordt aangeboden aan de leerlingen die op grond van de eerste laag (blokje A) de grootste kans op 'onder niveau' hadden,
- Blokje C wordt aangeboden aan de leerlingen die op grond van de eerste laag (blokje A) de grootste kans op 'op niveau' hadden en
- Blokje D wordt aangeboden aan de leerlingen die op grond van de eerste laag (blokje A) de grootste kans op 'boven niveau' hadden.

Na het tweede blokje wordt één blokje uit de laatste (derde) laag aangeboden:

- Blokje E wordt aangeboden aan de leerlingen die op grond van de tweede laag de grootste kans op 'onder niveau' hadden,
- Blokje F wordt aangeboden aan de leerlingen die op grond van de tweede laag de grootste kans op 'op niveau' hadden en
- Blokje G wordt aangeboden aan de leerlingen die op grond van de tweede laag de grootste kans op 'boven niveau' hadden.

Nadat de derde laag is doorlopen wordt begonnen aan het tweede doortoetsstadium. Voor elk deelaspect is een doortoetsblokje met opgaven.



Figuur 13-2. Blokkenontwerp

In de eerste, beperkte, adaptieve afname van 2016 zijn per hoofdaspect alleen de eerste twee lagen gemaakt en is niet doorgetoetst (Cito, 2015). In de laatste afname van de pilot in 2017 is het gehele adaptieve ontwerp met drie lagen en doortoetsen geïmplementeerd (Cito, 2016).

13.2.2 Creëren blokken

De blokken zijn eerst automatisch samengesteld. De opgaven behorende bij een hoofdaspect zijn 1000 keer aselekt over de blokken verdeeld (zie Hoofdstuk 12; Cito, 2016). Vervolgens is met behulp van gesimuleerde data nagegaan hoe accuraat de diagnoses zijn bij elk van deze 1000 oplossingen. De tien oplossingen met de meest accurate diagnose op hoofdaspectniveau zijn voorgelegd aan psychometrici. De psychometrici bestudeerden voor een adaptieve afname van schrijfvaardigheid Nederlands, schrijfvaardigheid Engels en wiskunde in totaal 600 oplossingen (3 vakken x 4 hoofdaspecten x 5 leerwegen/schoolsoorten x 10 oplossingen). Hierbij werd niet alleen gelet op de accuratesse van de diagnose van het hoofdaspect, maar ook de accuratesse van de deelaspect-diagnoses en werd gelet op de verdeling van de responsen behorend bij de verschillende aspecten over de blokken.

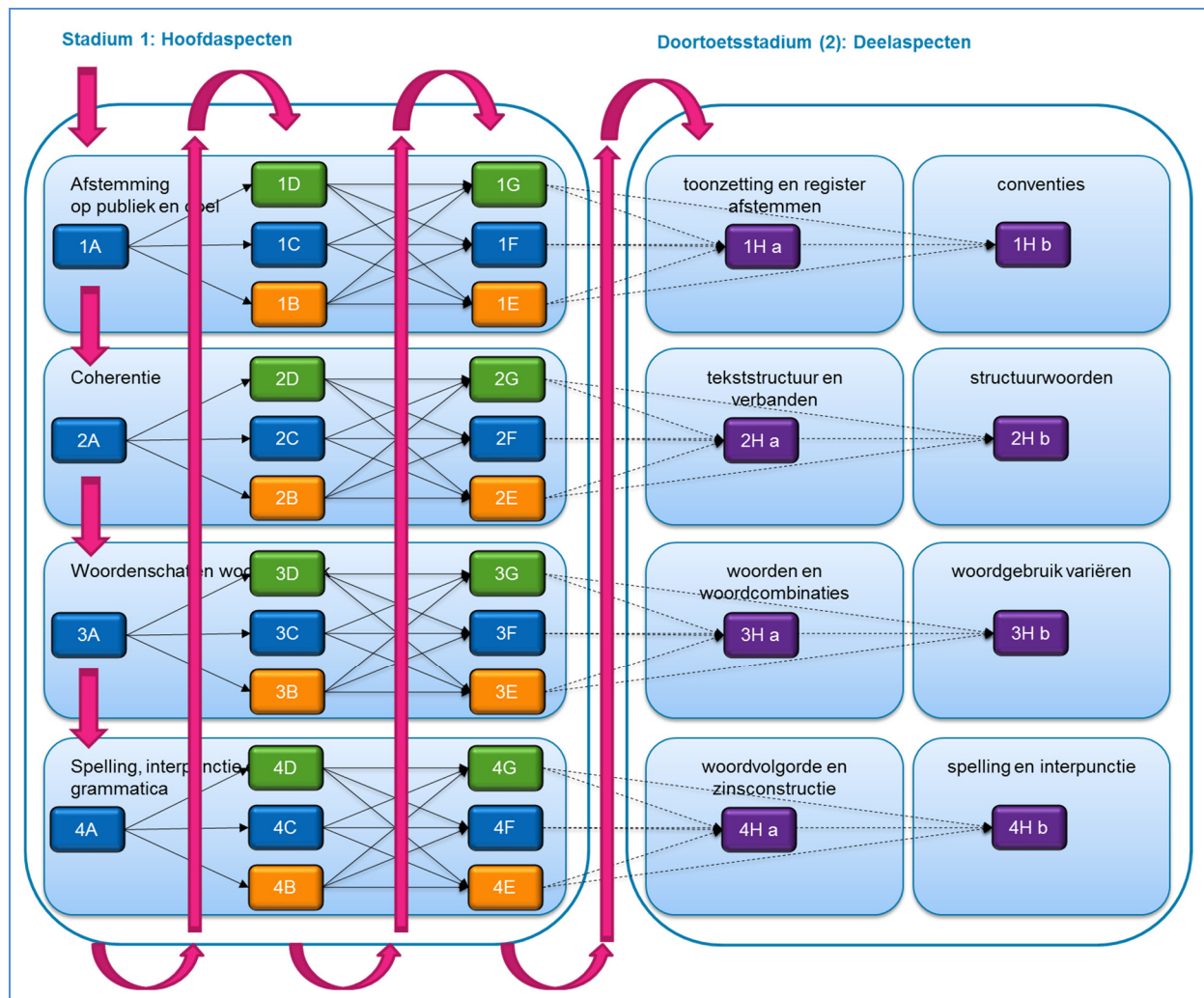
Vervolgens bestudeerden de toetsdeskundigen de geadviseerde oplossingen (20 per vak).

De toetsdeskundigen letten hierbij op potentiële contexteffecten, volgorde effecten of een ongebalanceerde

inhoud. Indien gewenst konden de blokindeling dan handmatig aangepast worden, waarbij natuurlijk de accuratesse met hulp van simulatie gecontroleerd werd. Bijvoorbeeld, bij schrijfvaardigheid Engels vmbo-bb waren veel opgaven die het formuleren van een briefaanhef toetsen in één blokje terecht gekomen na de automatisch samengestelde blokindeling (voor de afname 2017), ook al werden verschillende opgavetypen gebruikt. De oplossing is vervolgens handmatig aangepast. Deze aanhef-opgaven zijn handmatig gelijk verdeeld over de blokken van hoofdaspect 1, *Afstemming op publiek en doel*, en binnen de blokken ver uit elkaar geplaatst.

13.3 Adaptief verloop

In de eerste beperkte adaptieve afname van 2016 maakten de leerlingen per hoofdaspect de opgaven. Ze begonnen met de blokjes van het eerste hoofdaspect. Hierna maakten ze de blokjes van het tweede hoofdaspect, daarna van het derde en tenslotte de blokjes van het vierde hoofdaspect (Cito, 2016). De vaststellingscommissie heeft bepaald dat in het adaptieve prototype van 2017 de leerlingen per laag door de meting heen zouden lopen, omdat hiermee de meting gevarieerder is voor leerlingen (zie Figuur 13-3). In deze variant maken de leerlingen eerst allemaal het eerste blokje van elk hoofdaspect en aansluitend een aantal zaai-opgaven (zie sectie 13.5). Vervolgens maken ze van elk hoofdaspect een tweede blokje en tenslotte van elk hoofdaspect een derde blokje (zie Figuur 13-3). Nadat het eerste stadium geheel is doorlopen beginnen ze aan de doortoetsblokjes. De volgorde waarin dat gebeurt wordt bepaald door het doortoetsalgoritme (zie sectie 13.4.2).



Figuur 13-3. Volgorde van de adaptieve toets (roze pijlen) bij schrijfvaardigheid Engels in 2017

Aan het verloop per laag (zoals in 2017) kleven wel twee nadelen. Allereerst zijn leerlingen veel langer met exact dezelfde opgaven bezig. Het eerste blokje wordt namelijk door alle leerlingen gemaakt. In de opzet per laag krijgen alle leerlingen dus eerst voor elk hoofdaspect eenzelfde blokje. Pas nadat ze de eerste vier blokjes hebben gemaakt van de eerste laag, vindt de eerste adaptieve beslissing plaats (blokjes B, C of D van het eerste hoofdaspect, zie Figuur 13-3). Het maakt afkijken dus iets gemakkelijker.

Een tweede nadeel is dat het langer duurt voordat de eerste diagnose gesteld kan worden. In een verloop per hoofdaspect kan na drie blokjes de eerste diagnose (voor het eerste hoofdaspect) gesteld worden. In een verloop per laag kan pas na negen blokjes de eerste diagnose gesteld worden. Dit levert een risico op als niet de volledige toetstijd gebruikt wordt, omdat leerlingen dan geen volledige rapportage kunnen krijgen.

13.4 Adaptieve regels

Er zijn drie adaptieve regels geïmplementeerd: Een stopregel, een regel voor doortoetsvolgorde en een regel voor de bruikbaarheid van de uitkomsten.

13.4.1 Regel voor stoppen of overslaan blokken

Het is mogelijk dat de diagnose al heel zeker is voordat de derde laag is bereikt, bijvoorbeeld als de leerling de opgaven consistent heel goed of heel slecht maakt. Wanneer de diagnoses van de hoofdaspecten en bijbehorende deelaspecten zeker genoeg zijn, krijgt de leerling geen blokken meer met opgaven over die hoofd- en deelaspecten. In de eerste adaptieve afname (2017) is gewerkt met een vast criterium van 0,95. Het is echter beter om het criterium te laten afhangen van het gewenste percentage correcte diagnoses. Om die reden is ook een procedure uitgewerkt om te bepalen bij welke waarschijnlijkheid de gewenste accuraatheid (percentage correcte diagnoses) is bereikt.

Met behulp van simulaties is een grenswaarde per aspect bepaald. Per aspect is uitgerekend bij welke waarschijnlijkheid ('posterior model probability', zie Hoofdstuk 12) 95% van de leerlingen een correcte diagnose krijgt na het eerste blokje opgaven. Na elk antwoord wordt de waarschijnlijkheid van alle diagnoses bijgewerkt. Als de waarschijnlijkheden van het hoofdaspect en van de bijbehorende deelaspecten hoger zijn dan de grenswaardes, krijgt de leerling geen blokken met opgaven meer voorgelegd voor die hoofd- en deelaspecten.

Doorgaans ligt de waarschijnlijkheid waarbij 95% van de leerlingen correct gediagnosticeerd wordt (de grenswaarde) onder de ,95. We zien dat voor veel aspecten 95% van onder-niveauleerlingen en 95% van de boven-niveauleerlingen correct gediagnosticeerd worden bij een waarschijnlijkheid die (ver) onder de ,90 ligt. Voor de op-niveauleerlingen ligt de grenswaarde waarbij 95% van de op-niveauleerlingen correct gediagnosticeerd worden veelal tussen de ,90 en ,95 (zie Bijlage 17.2 voor alle grenswaardes).

13.4.2 Regel voor doortoetsvolgorde

Nadat in het eerste stadium alle hoofdaspecten zijn gediagnosticeerd, begint het doortoetsstadium. Indien de geplande toetstijd (3 uur) nog niet voorbij is en de afname nog niet stop is gezet door de docent of afnamecoördinator, kan doorgedaan worden met het nader toetsen van deelaspecten. Zoals hiervoor beschreven, hoeft een deelaspect niet verder getoetst te worden als de diagnose al zeker genoeg is. Dit is het geval als de waarschijnlijkheid van een van de diagnoses van het deelaspect al hoger is dan de grenswaarde, waarbij 95% van de leerlingen correct gediagnosticeerd wordt. Alle overige deelaspecten komen in aanmerking voor doortoetsen.

Vanuit het idee dat alle leerlingen, ook de excellente leerlingen, hun leeropbrengsten kunnen maximaliseren door relatieve verbeterpunten aan te pakken is een algoritme uitgewerkt waarbij de nadruk ligt op de relatieve verbeterpunten (Hojtink & Schouwstra, 2014). Er is sprake van een indicatie wanneer er een aanzienlijke kans is dat de 'beheersing van de leerling van het deelaspect relatief verbetering behoeft. Hiervoor wordt gekeken naar de geschatte 'posterior model probabilities' na het eerste stadium (dus nadat de leerling drie blokken opgaven heeft gemaakt of heeft mogen overslaan voor elk hoofdaspect).

- Als de meeste deelaspecten naar schatting boven niveau zijn zal als eerste het deelaspect doorgetoetst worden dan het minst waarschijnlijk boven niveau is (de kleinste PMP>).

- Als de meeste deelaspecten naar schatting op niveau zijn zal als eerste het deelaspect doorgetoetst worden dan het meest waarschijnlijk onder niveau is (de grootste PMP<).
- Als de meeste deelaspecten naar schatting onder niveau zijn zal als eerste het deelaspect doorgetoetst worden dan het minst waarschijnlijk onder niveau is (de kleinste PMP<).

Het tweede stadium stopt wanneer de beschikbare afnametijd op is of de diagnoses voor alle deelaspecten voldoende zeker zijn. Natuurlijk zijn alternatieve algoritmes denkbaar. Bijvoorbeeld een algoritme waarbij niet alleen gekeken wordt naar de (relatieve) verbeterpunten, maar ook naar de sterke punten. In dit geval zou ook sprake zijn van een indicatie voor doortoetsen als de beheersing van de leerling van het deelaspect relatief sterk is.

13.4.3 Regel voor bruikbaarheid van uitkomsten

In de adaptieve module wordt na elke beantwoorde opgave de waarschijnlijkheid van alle diagnoses (de 'posterior model probabilities') voor alle aspecten en deelaspecten bijgewerkt. Dus na één respons levert de adaptieve module al een schatting op voor het betreffende aspect. Om te voorkomen dat diagnoses worden gerapporteerd die op veel te weinig responsen zijn gebaseerd is een regel geformuleerd voor de bruikbaarheid van de schattingen.

Bij het evalueren van de blokindelingen wordt onder meer beoordeeld hoe accuraat de diagnoses zijn (zie sectie 13.2.2). Op dit moment is een vuistregel gehanteerd dat indien minder dan 60% van de onder, op of boven-niveauleerlingen correct gediagnosticeerd worden de uitkomsten voor geen enkele leerling bruikbaar zijn voor rapportage. Indien dit het geval is kan in het sturbestand van de adaptieve toets aangegeven worden dat de uitkomsten van het betreffende aspect in die blokindeling niet bruikbaar zijn.

Voor alle aspecten waarbij de simulaties laten zien dat meer dan 60% van de leerlingen correct gediagnosticeerd worden, zijn de volgende drie regels geïmplementeerd in de adaptieve module:

- Als de waarschijnlijkheid van de diagnose boven de grenswaarde ligt, is de toetsuitkomst bruikbaar en kan die gerapporteerd worden;
- Als de waarschijnlijkheid van de diagnose onder de grenswaarde ligt én de leerling heeft alle drie lagen uit het eerste stadium doorlopen, dan is de diagnose van het betreffende hoofdaspect bruikbaar en kan die gerapporteerd worden;
- Als de waarschijnlijkheid van de diagnose onder de grenswaarde ligt én de leerling heeft het hele doortoetsblok gemaakt, dan is de diagnose van het betreffende deelaspect bruikbaar en kan die gerapporteerd worden.

Als de leerling dus heel vroeg stopt met de afname is het mogelijk dat de leerling voor geen enkel aspect voldoende responsen heeft gegeven. In dit geval zal de leerling helemaal geen rapportage krijgen.

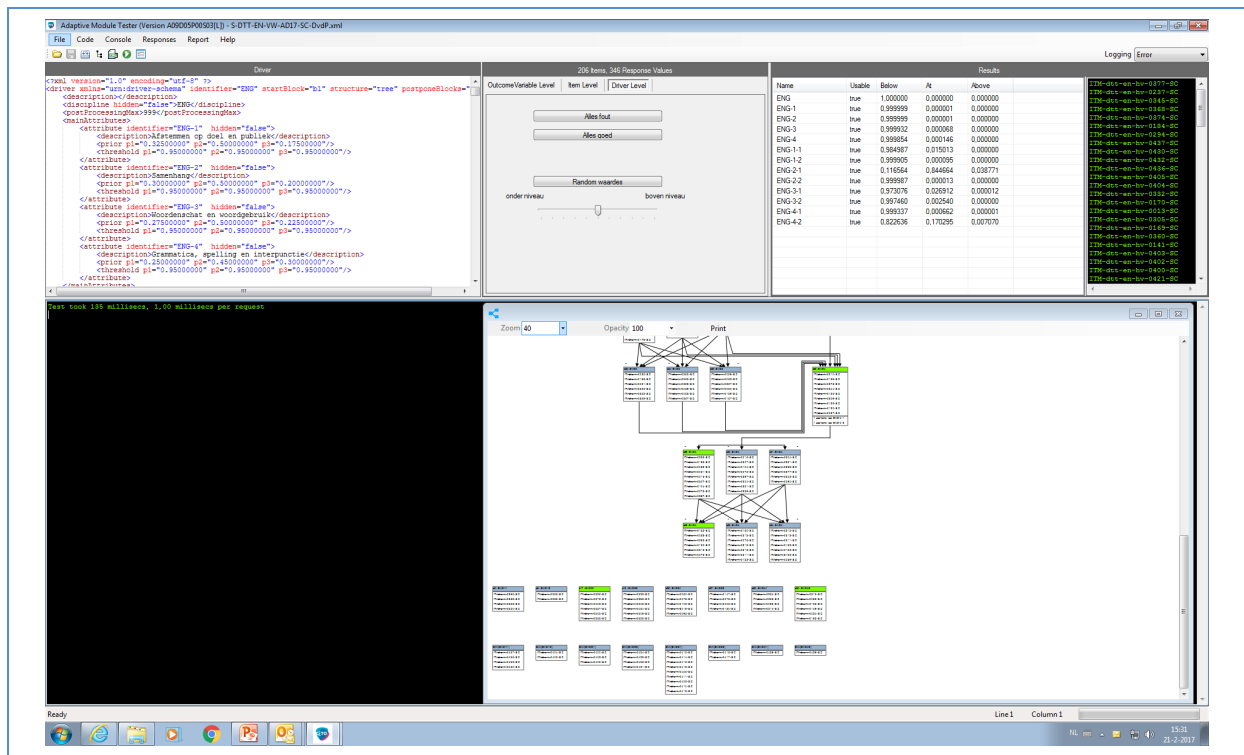
13.5 Zaaïen

In het adaptieve ontwerp is ook rekening gehouden met de mogelijkheid om nieuwe opgaven te zaaïen. Door nieuwe opgaven in de adaptieve toets op te nemen kunnen de kenmerken van deze opgaven met behulp van de adaptieve afname bepaald worden en kunnen de opgaven (indien goed bevonden) gekalibreerd worden (zie Hoofdstuk 12). Het voordeel van zaaïen is dat er geen grootschalige pretest nodig is om nieuwe opgaven te kalibreren. Door middel van zaaïen komen er jaarlijks nieuwe opgaven bij, die gebruikt kunnen worden voor verversing en onderhoud van de opgavebank.

Om te zorgen dat alle leerlingen, ongeacht hun niveau, de zaaïopgaven maken worden de zaaïopgaven aangeboden na het eerste blokje. Van tevoren moet in een bestand (zie volgende paragraaf) per deelaspect aangegeven worden hoeveel zaaïopgaven een leerling krijgt. De hoeveelheid zaaïopgaven die opgenomen kan worden zal meestal afhangen van de beschikbare toetstijd en het verwachte aantal observaties. Het zaaïen mag niet te veel afnametijd in beslag nemen en om de opgaven na afname te kalibreren zijn minstens 500 observaties nodig. In de adaptieve afname van 2017 kregen alle leerlingen dezelfde zaaï-opgaven, maar het is ook mogelijk dat elke leerling slechts een aselect geselecteerd aantal opgaven krijgt.

13.6 Adaptieve architectuur

Voor de DTT is een adaptieve module geprogrammeerd. Die module wordt meegeleverd met een toets. Ook wordt een zogenaamd stuurbestand meegeleverd, waarin alle instellingen voor de adaptieve afname staan. Voor het aanmaken van deze stuurbestanden, het uittesten van de stuurbestanden (zie Figuur 13-4) en voor het visualiseren van het adaptieve ontwerp zijn aparte tools geprogrammeerd. De module en tools zijn opgeleverd en voorzien van documentatie, zodat andere partijen deze ook kunnen gebruiken.

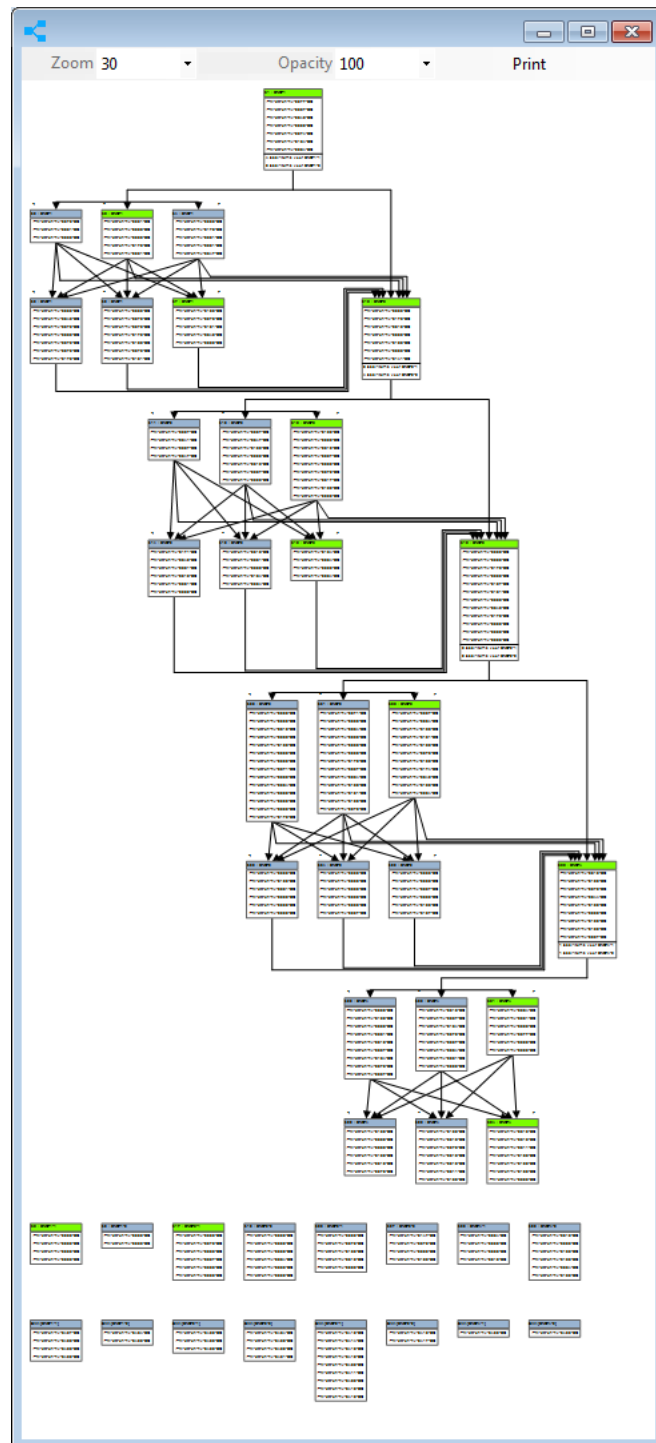


Figuur 13-4. Schermafprint van de tool om de stuurbestanden te testen

Een adaptieve toets bestaat dus uit een package met de opgaven, de adaptieve module en een stuurbestand. In het stuurbestand staan de instellingen van de adaptieve toets:

1. De parameters van de responsen of van combinatie van responsen bij testlets ($P<$; $P=$ en $P>$).
2. De startwaardes ('prior model probabilities')
3. Instellingen voor de adaptieve regels
 - a. De grenswaardes voor stoppen of overslaan
 - b. De instelling voor de bruikbaarheid (in principe wel of niet bruikbaar)
4. Het adaptieve ontwerp voor de gehele toets, dat wil zeggen de bloksamenstelling, de volgorde van de hoofdaspecten en de zaai-opgaven (zie Figuur 13-5 voor een voorbeeld).

In de adaptieve module zelf wordt na elke opgave de waarschijnlijkheid van de diagnoses (de 'posterior model probabilities') uitgerekend. Er wordt ook bepaald of blokken overgeslagen mogen worden en of de toetsuitkomst bruikbaar is. Na het eerste stadium wordt met behulp van het doortoetsalgoritme bepaald in welke volgorde de deelaspecten doorgetoetst gaan worden.



Figuur 13-5. Visualisatie van een adaptieve toets met in het groen het adaptieve pad van een gesimuleerde leerling, die bijna alles correct heeft beantwoord

13.8 Conclusie

Bij de DTT is adaptiviteit gebruikt om meer diagnoses te kunnen stellen in de gegeven afnametijd. Bij de DTT zijn verschillende gradaties van adaptiviteit gebruikt. Allereerst zijn verschillende versies voor elke leerweg/schoolsoort. Binnen elke versie doorloopt een leerling twee stadia. In het eerste stadium wordt een diagnose gesteld voor alle hoofdaspecten en worden opgaven gezaaid. Hierbinnen vindt adaptiviteit plaats op het niveau van opgaveblokjes. In het tweede stadium, het doortoetstadium, worden diagnoses voor deelaspecten gesteld. Het verloop van de toets wordt gereguleerd door drie adaptieve regels.

In de DTT is de adaptiviteit gebaseerd op het model van Sies en Hoijtink (2014). Het is een geheel nieuwe benadering waarbij verschillende aspecten nog nader onderzoek verdienen. De voor- en nadelen van deze benadering zouden onderzocht kunnen worden door een vergelijking met IRT-gebaseerde benaderingen, zoals bijvoorbeeld de benadering van Eggen en Straetmans (2000). Ook zou de impact van de onzekerheid over de parameters op de accuratesse van de diagnoses nader onderzocht moeten worden, zoals Veldkamp (2012) adviseert voor elk operationele computergestuurde adaptieve toets. Ook naar de adaptieve regels zou nader onderzoek gedaan kunnen worden, zo zouden alternatieve algoritmes voor het doortoetsen ontwikkeld kunnen worden.

Tijdens de ontwikkeling is de adaptiviteit gegroeid. In de eerste adaptieve afname van 2016 was alleen sprake van het eerste stadium waarin een diagnose werd gesteld voor de hoofdaspecten. Op dat moment werd ook alleen met twee lagen van blokken gewerkt. Het is denkbaar dat in de toekomst de adaptiviteit nog iets verder groeit. Echter een volledig adaptieve afname beperkt de mogelijkheden voor herkalibratie en kwaliteitscontrole. Daarnaast concluderen Zwitser en Maris (2015) dat het toevoegen van additionele stadia in een multistagetest niet noodzakelijkerwijs leidt tot een hogere efficiëntie. Nader onderzoek is nodig om een ontwerp te vinden dat die efficiëntie maximaliseert. Een mogelijkheid om de efficiëntie te verhogen, is het gebruik van de samenhang tussen aspecten of een mogelijke hiërarchische structuur voor de adaptiviteit, zoals bijvoorbeeld Gierl en Zhou (2008) doen.

13.9 Literatuur

- Cito (2015). *Diagnostische tussentijdse toets: Verslag pretest 2015*. Arnhem: Cito.
- Cito (2016). *Diagnostische tussentijdse toets: Verslag pretest 2016*. Arnhem: Cito.
- Eggen, T. J. H. M., & Straetmans, G. J. J. M. (2000). Computerized adaptive testing for classifying examinees into three categories. *Educational and Psychological Measurement*, 60, 713- 734.
- Eggen, T. J. H. M. (2012). Computerized Adaptive Testing Item Selection in Computerized Adaptive Learning Systems. In T. J. H. M. Eggen & B. P. Veldkamp (Eds.), *Psychometrics in Practice at RCEC* (pp. 11-21). Enschede: RCEC, Cito/Universiteit van Twente.
- Gierl, M. J., & Zhou, J. (2008). Computer adaptive attribute testing. A new approach to cognitive diagnostic assessment. *Zeitschrift fur Psychologie*, 216, 29-39.
- Glas, C. A. W. (2000). Item calibration and parameter drift. In W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: theory and practice* (pp. 183–199). Dordrecht: Kluwer Academic Publishers.
- Hendrickson, A. (2007). An NCME instructional module on multistage testing. *Educational Measurement: Issues and Practice*, 26(2), 44-52.
- Hoijtink, H., & Schouwstra, S. (2014). Interpretatie van het initiële profiel op deelattribuutniveau. In S. Schouwstra (Red.), *De diagnostische tussentijdse toets: onderzoek 2014* (pp. 41-46). Arnhem: Cito.
- Hoijtink, H., & Sies, A. (2013). De psychometrie van de diagnostische tussentijdse toets. In S. Schouwstra (Red.), *De diagnostische tussentijdse toets: onderzoek 2013* (pp. 13-62). Arnhem: Cito.

- Hoijtink, H., & Sies, A. (2014). Adaptieve procedure met itemblokken. In S. Schouwstra (Red.), *De diagnostische tussentijdse toets: onderzoek 2014* (pp. 33-40). Arnhem: Cito.
- Hol, A. M. (2006). *A CAT with personality and attitude* (Dissertatie). Amsterdam: Universiteit van Amsterdam.
- Maas, L. (2017). *Potential impact of item parameter drift on diagnostic accuracy in educational testing*. [ongepubliceerde master thesis]. Utrecht University, Utrecht.
- Roelofs, E., & Schouwstra, S. (Red.). (2012). *Diagnostische tussentijdse toets: Verslag van de voorstudie*. Arnhem: Cito.
- Sies, A.M.A, (2014). *Cognitive diagnostic testing using calibrated hypotheses* [ongepubliceerde master thesis]. Utrecht University, Utrecht.
- Veldkamp, B. P. (2012). Ensuring the future of computerized adaptive testing. In T. J. H. M. Eggen & B. P. Veldkamp (Eds.), *Psychometrics in Practice at RCEC* (pp. 35-46). Enschede: RCEC, Cito/Universiteit van Twente.
- Zwitser, R. J., & Maris, G. K. J. (2015). Conditional statistical inference with multistage testing designs. *Psychometrika*, 80(1), 65-84.

14 De creatie van een nieuwe rapportagevorm voor de DTT

Sanneke Schouwstra en Patrick de Klein

De DTT is primair bedoeld om docenten en leerlingen te informeren over de verbeter- en sterktepunten van de leerling. De informatie die de DTT oplevert moet de docent in staat stellen het leerproces van de leerlingen bij te sturen op grond van hun leerbehoefte. Om deze formatieve functie te kunnen vervullen, diende zorgvuldig aandacht besteed te worden aan de rapportages. Vanaf het begin van de DTT-ontwikkeling is het belang van de rapportage daarom benadrukt en is het rapportagemodel toegevoegd aan het conceptuele raamwerk van de toets (zie Hoofdstuk 1).

14.1 Stadia in de ontwikkeling van de rapportagevorm

Er zijn vier stadia te onderscheiden in de ontwikkeling van de DTT-rapportages. Tijdens de voorstudie is de blauwdruk voor de rapportage uitgewerkt. Allereerst is een literatuuronderzoek verricht naar de principes en het belang van feedback en rapportages. Uit dit literatuuronderzoek zijn enkele ontwerpprincipes gedestilleerd, die gehanteerd zijn bij het creëren van prototypische rapportages (zie het voorbeeld in Figuur 14-1). Voor elk vak zijn de prototypische rapportages in veldraadplegingen voorgelegd aan docenten.

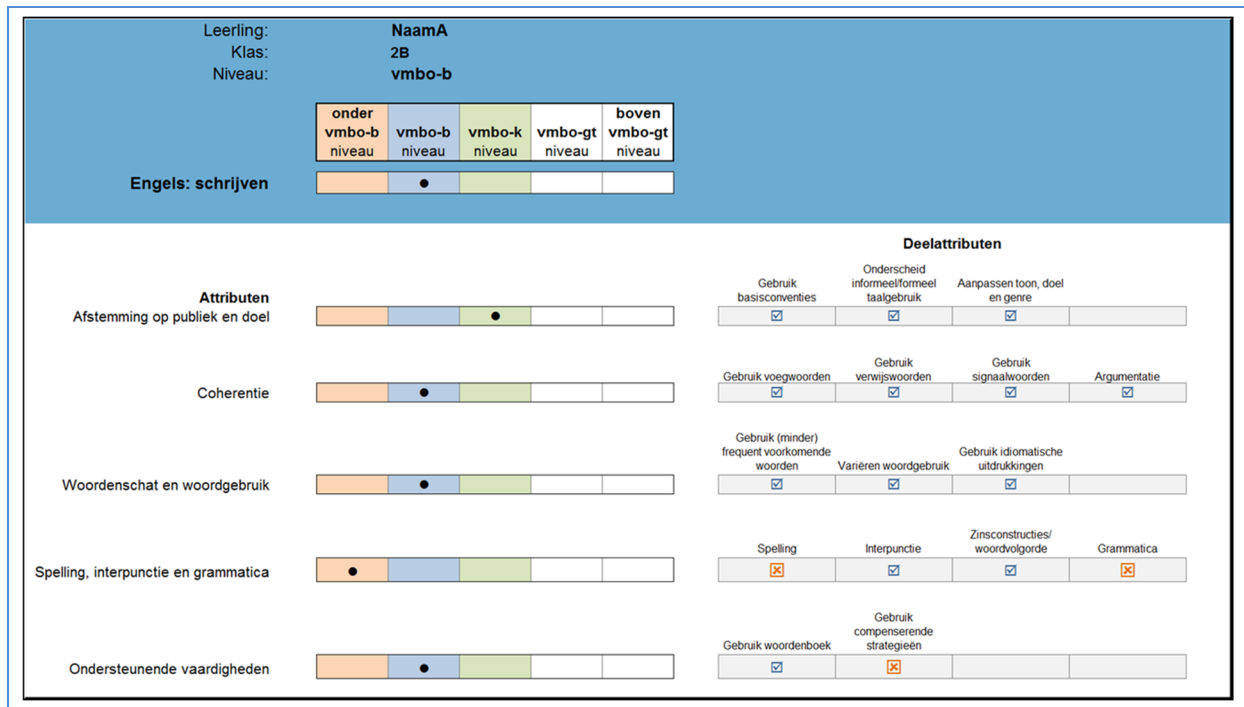
Uit de veldraadplegingen kwam naar voren dat enkele docenten een aanduiding van de betrouwbaarheid van de diagnoses minder bruikbaar of lastig te interpreteren vonden (Roelofs & Schouwstra, 2012). Voor een eerlijk en juist gebruik van de resultaten is het echter cruciaal dat de gebruiker maximaal geïnformeerd wordt over de mate van zekerheid. Om die reden is in het tweede stadium opnieuw naar de visualisatie gekeken en is de zogenaamde bolletjes-weergave bedacht (zie Figuur 14-2).

In de bolletjes-weergave wordt per aspect de meest waarschijnlijke diagnose aangegeven met een gekleurd bolletje. Bijvoorbeeld, in Figuur 14-2 is te zien dat de leerling waarschijnlijk *Coherentie* op niveau beheerst. Naast een bolletje bij de meest waarschijnlijke diagnose, zijn de andere twee diagnoses ook getoond. Hoe waarschijnlijker de diagnose hoe groter en zichtbaarder het bolletje is. Bijvoorbeeld, diagnose voor *Woordenschat en woordgebruik* is op niveau (blauw bolletje), maar op grond van de antwoorden van de leerling zou de leerling ook onder niveau kunnen zitten. Het formaat van het oranje bolletje geeft aan dat dit redelijk waarschijnlijk is.

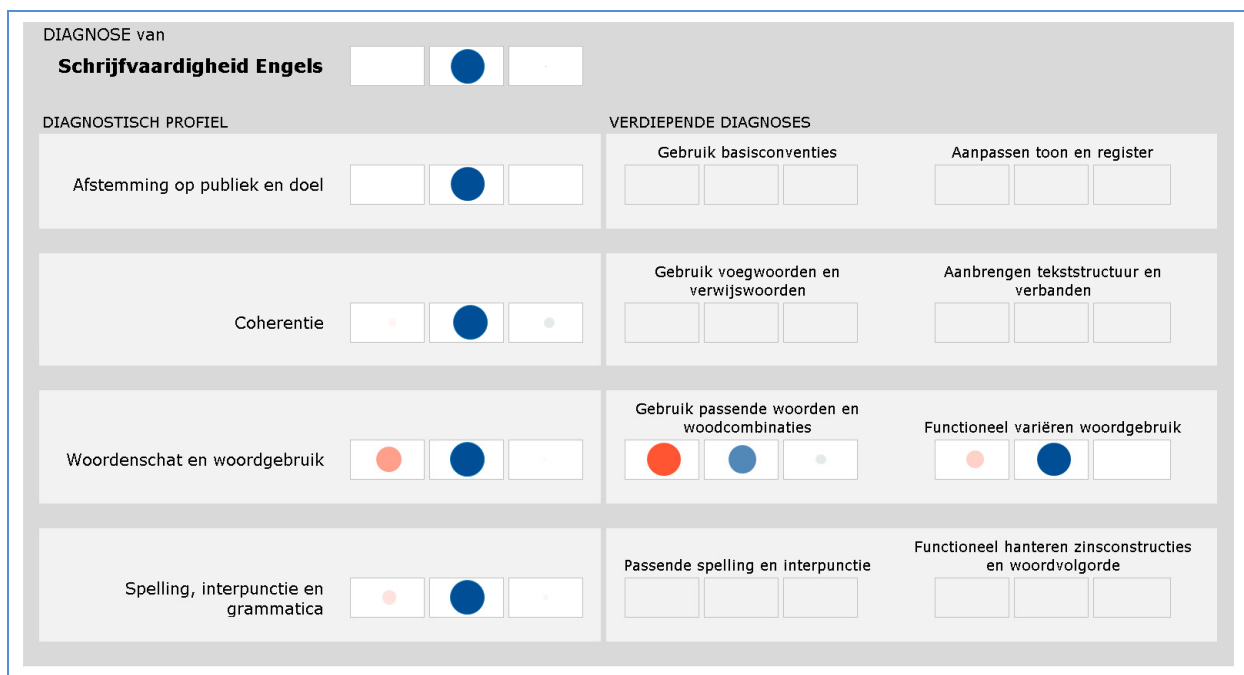
In dit tweede stadium zijn voor alle drie de vakken voorbeelden gemaakt van de leerlingrapportages om aan de toetswijzercommissies voor te leggen voor feedback. Voor elk vak zijn op basis van de leerlingmodellen en gesimuleerde gegevens twee leerlingrapportages gemaakt en twee groepsrapportages:

- Een leerlingrapportage van een leerling met een verbeterpunt
- Een leerlingrapportage van een leerling met een sterk punt
- Een groepsrapportage van een vmbo-klas
- Een groepsrapportage van een vwo-klas met een specifiek verbeterpunt

Na de feedback van de toetswijzerscommissies is een zogenaamde mock-up gemaakt in het derde stadium. De mock-up is bedoeld om een eerste indruk te geven hoe de rapportage van de DTT voor docenten eruit zou kunnen zien en zou kunnen werken. Mede op grond van deze mock-up heeft de DUO tenslotte onder regie van het CvTE de definitieve rapportage gerealiseerd in het laatste stadium. Cito heeft in dit laatste stadium voorbeeldopgaven gemaakt ter illustratie van de gerapporteerde diagnoses.



Figuur 14-1. Voorbeeld van een prototypische leerlingrapportage uit de voorstudie (eerste stadium)



Figuur 14-2. Voorbeeld van een visualisatie met bolletjes (tweede stadium)

14.2 Basis van de DTT-rapportage

In de Memorie van Toelichting bij het wetsvoorstel voor invoering van de DTT (2013) waren drie niveaus van rapportage voorzien: leerling, school en landelijk. In de voorstudie is uiteengezet dat groepsrapportages zeer wenselijk zijn omdat de primaire gebruiker de docent is. Tijdens de ontwikkeling waren daarom vier rapportages voorzien in plaats van drie: op leerlingniveau, op schoolniveau, op nationaal

niveau en bovendien op groepsniveau. Na de herziene opdracht van het Ministerie van OCW in 2014 (zie Hoofdstuk 1) is het vierde niveau, het nationale niveau, komen te vervallen.

Aan alle rapportage-ontwerpen lagen de volgende ontwerpprincipes ten grondslag (zie ook Roelofs & Schouwstra, 2012):

- In één oogopslag een helder beeld geven van de leerbehoefte van de leerling
- Communiceren door middel van beelden (data visualisatie)
- Uitkomsten liefst grafisch, in getallen en in tekst
- Alle informatie, beelden en kleuren moeten doelgericht gebruikt worden. Dus geen kleuren voor de mooiigheid.
- Bevorderen intuïtief correcte interpretatie toetsuitkomsten
- Visualisatie van de (on-)zekerheid van de toetsuitkomsten
- Nadruk op opvallende uitkomsten
- Eenvoudig te gebruiken voor docenten (bijvoorbeeld eenvoudige navigatie)

Natuurlijk werd hierbij gezorgd dat het geheel overzichtelijk was, de gegevens op een zinvolle wijze gegroepeerd waren, geen statistisch jargon gebruikt werd en weinig tekst op het scherm kwam. Een andere wens was dat voorbeeldopgaven gecreëerd werden bij de gerapporteerde toetsuitkomsten.

De rapportage moest voor elke leerling meer fijnmazige informatie geven dan een algemene prestatie- of vaardigheidsscore. De gerapporteerde uitkomsten zijn gebaseerd op leerlingmodellen met hoofd- en deelaspecten: de DTT diagnosticeert de kennis- en vaardigheidsaspecten die noodzakelijk zijn om op beoogd niveau Engels te schrijven/Nederlands te schrijven/wiskunde te bedrijven. In de leerlingrapportage diende dit hele leerlingmodel zichtbaar te zijn. Voor elk aspect zijn er drie mogelijke uitkomsten die op de rapportage staan: de leerling beheerst waarschijnlijk het aspect onder niveau, op niveau of boven niveau.

14.3 De mock-up

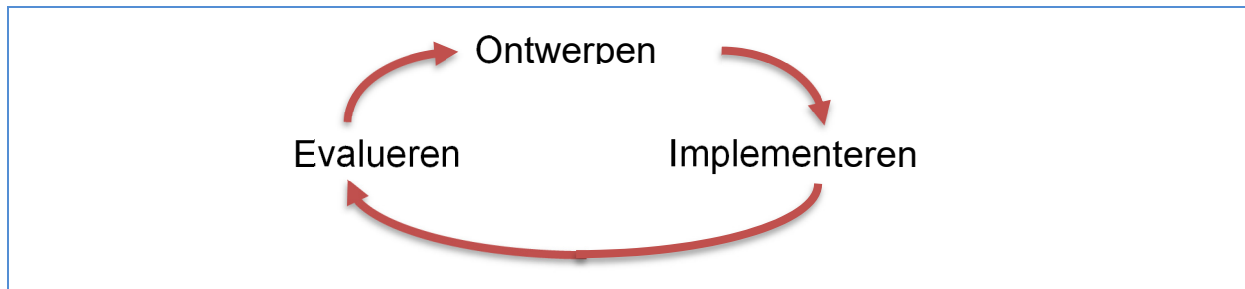
Na de voorstudie en de statische visualisaties van de rapportages is een mock-up gemaakt. De mock-up is ontwikkeld voor drie doelen. Allereerst is de mock-up is gemaakt als illustratie van een werkende rapportage, die gebaseerd is op een leerlingmodel, in presentaties en trainingen, bijvoorbeeld voor de onderwijsadviseurs die de pilotscholen hebben bijgestaan in de implementatie van de DTT¹⁵. De mock-up was ook nodig om feedback te krijgen van scholen en docenten, zodat de specificaties voor de uiteindelijke rapportage verbeterd konden worden. Het CvTE heeft twee wijzen feedback gekregen op de rapportage in de mock-up: via een online enquête en via de klankbordgroep van het CvTE. De mock-up fungeerde ten derde als specificatie voor de realisatie in Facet door DUO.

14.3.1 Wijze ontwikkelen van de mock-up

De mock-up was een werkend HTML5 prototype van de rapportage module die op een agile wijze is ontwikkeld. Agile ontwikkelen is een snelle manier die heel wendbaar en flexibel is. In korte tijdsperiodes (iteraties genoemd, zie Figuur 14-3) wordt een hele cyclus doorlopen van ontwerpen, implementeren en evalueren. Na elke iteratie is er een nieuwe versie van het product, in ons geval de mock-up.

De ontwerpprincipes uit de voorstudie en de statische visualisaties, zoals in Figuur 14-2, vormden de basis voor de mock-up. Het werkende prototype maakte gebruik van data die voor eerder psychometrisch onderzoek was gesimuleerd (Hojtink & Sies, 2014). De gesimuleerde data is aangeleverd in een Excel-bestand en zijn van daaruit omgezet in Json voor gebruik in Angular JS, een open source webapplicatie framework.

¹⁵ De adviseurs zijn benaderd via Edventure, een branchevereniging van onderwijsadviesbureaus



Figuur 14-3. De cyclus van één iteratie

Er waren rapportages voorzien op vier niveaus: leerlingniveau, groepsniveau, schoolniveau en landelijk niveau. Tijdens de voorstudie (Roelofs & Schouwstra, 2012) waren de volgende gebruikers gespecificeerd:

- Gegevens op leerlingniveau: docent, ouder, leerling en mentor
- Gegevens op groepsniveau: docent, vaksectie, schoolmanagement
- Gegevens op schoolniveau: schoolmanagement
- Gegevens op landelijk niveau: beleidsmakers

Bij het ontwerpen van de mock-up lag de focus op de primaire gebruiker: de docent (zie Hoofdstuk 1). Om deze reden zijn in de mock-up een leerlingrapportage en een groepsrapportage ontwikkeld. Bij aanvang van de ontwikkeling van de mock-up zijn gebruikersverhalen voor de docent uitgewerkt als basis voor de verdere ontwikkeling (zie een voorbeeld in Figuur 14 4). Deze gebruikersverhalen waren de basis voor de verder ontwikkeling. De gebruikersdoelen waren de basis voor de wijze waarop de interacties en schermen in de rapportagemodule zijn uitgewerkt.

Wie is de gebruiker?	De docent
In welke situatie wil de gebruiker de rapportage gebruiken?	Tienminutengesprek
Wat wil de gebruiker bereiken met de rapportage?	Een leerlingrapportage snel vinden en afdrukken, zodat ik het aan de ouders kan laten zien tijdens het gesprek.
Wat betekent dit voor het ontwerp van de rapportage?	Een lijst waaruit een specifieke leerling gekozen kan worden
	Af te drukken, bij voorkeur op 1 pagina
	Eenvoudig te gebruiken

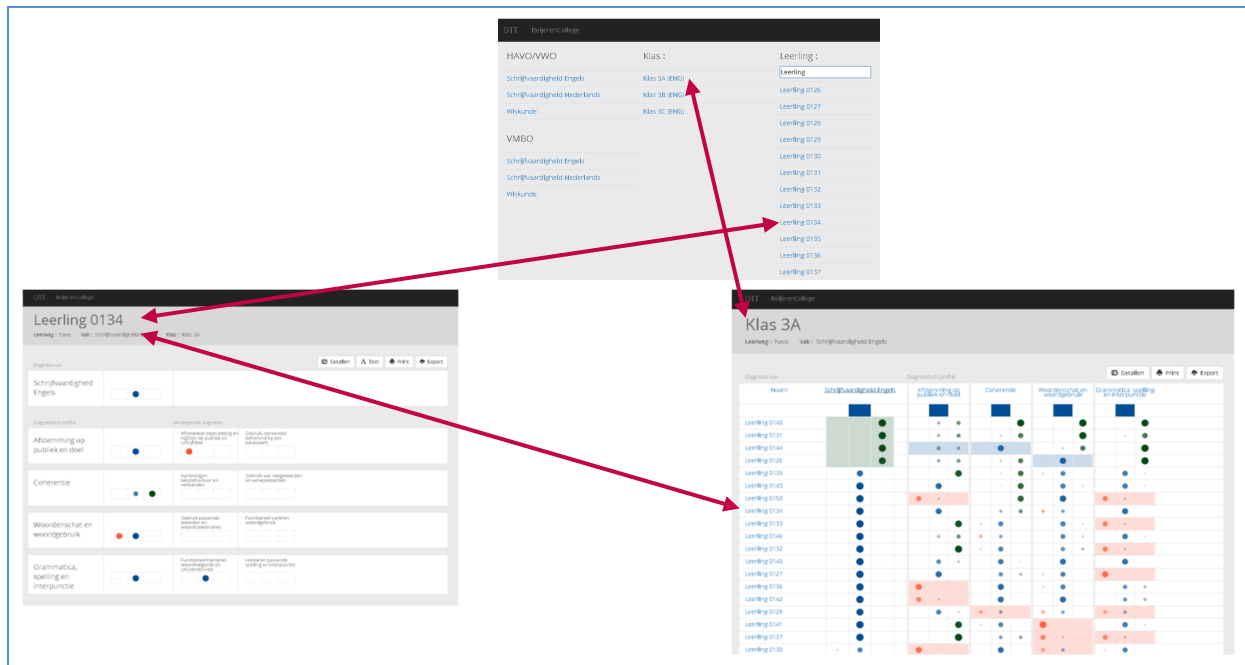
Figuur 14-4. Voorbeeld gebruikersverhaal

14.3.2 Ontwerp van de mock-up

In het interactieontwerp van de mock-up is een eenvoudige navigatie tussen het startscherm, de leerlingrapportage en de groepsrapportage gerealiseerd zoals weergegeven in Figuur 14-5. In het visuele ontwerp van de mock-up zijn de belangrijke kenmerken van de statische visualisaties (zie Figuur 14-2) overgenomen. Allereerst moesten voor één leerling de onderlinge structuur van hoofd- en deelaspecten en alle diagnoses op één scherm te zien zijn. De tekst op de rapportages was tot een minimum beperkt.

Een zeer belangrijk aspect dat overgenomen is in het visuele ontwerp, was de wijze waarop de zekerheid gevisualiseerd was in bolletjes. In de eerste vorm van de leerlingrapportage werd alleen de meest waarschijnlijke diagnose weergegeven met een bolletje (de standaardweergave). In de tweede vorm werden ook de andere twee diagnoses weergegeven: hoe waarschijnlijker de alternatieve diagnoses waren,

hoe groter en zichtbaarder de bolletjes van die alternatieve diagnoses (zie Figuur 14-6). Er werd ook een vorm gemaakt waarin de grootte en zichtbaarheid van alle bolletjes varieerden als functie van de waarschijnlijkheid van de diagnose (de 'posterior model probabilities'). In deze vorm kon ook de meest waarschijnlijke diagnose een klein bolletje hebben. Daarnaast was het idee dat er ook een getalsmatige en tekstuele weergave van de diagnoses zou komen.



Figuur 14-5. Interactieontwerp



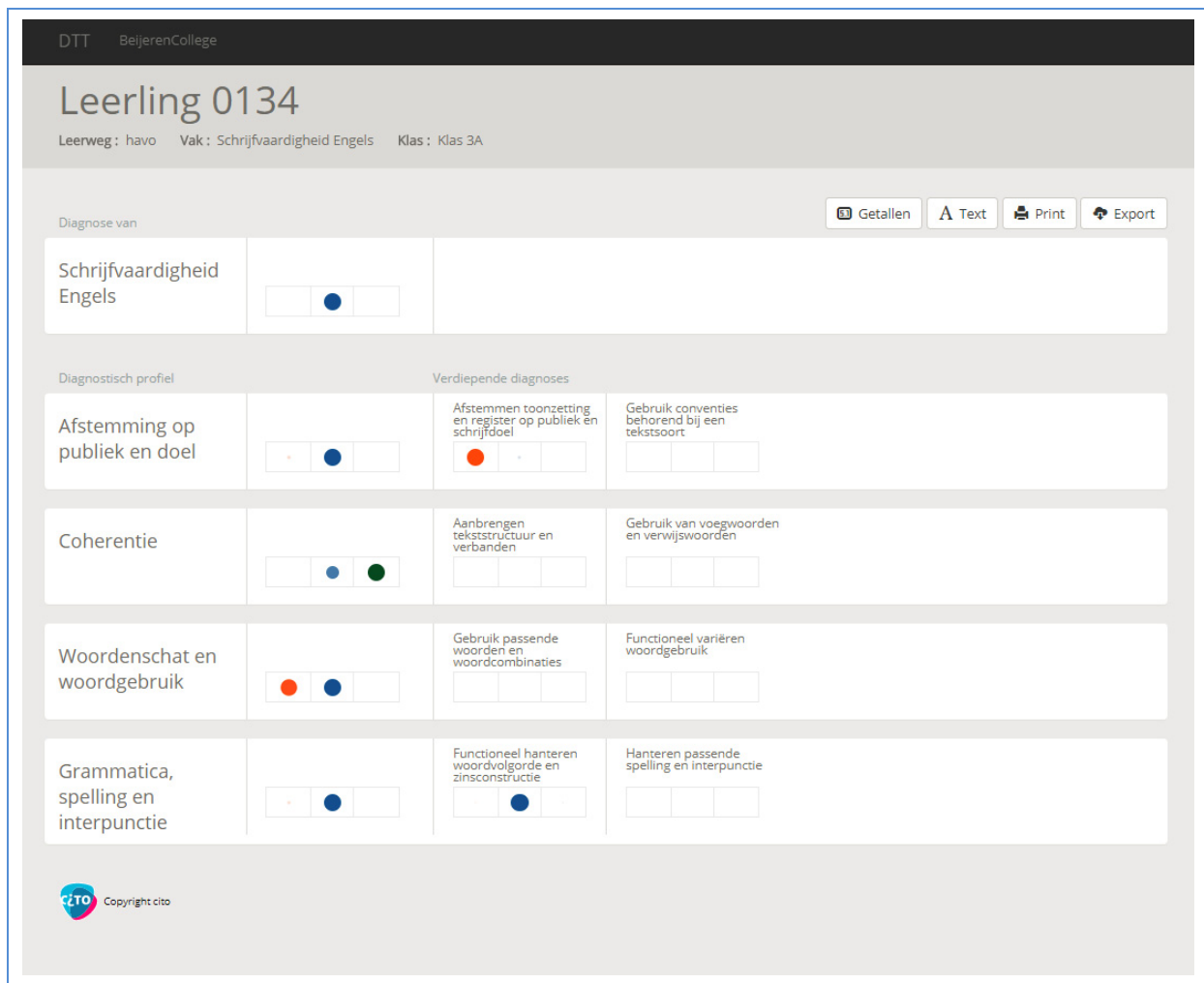
Figuur 14-6. De drie vormen van weergave

Andere aspecten van het uiteindelijke visuele ontwerp waren (zie Figuur 14-7):

- Wit gebied voor de inhoud
- Duidelijke interactiviteit door knoppen
- Kleuren alleen gebruiken voor het aangeven van de betekenis van de toetsuitkomsten
- Het signaleren van belangrijke uitkomsten moet een gevolg zijn van de kleuren

Bij de groepsrapportage zijn ook de markeringen en mogelijkheden voor sorteren uitgewerkt. Voor elke klas (groep) werden de meest voorkomende diagnoses in de klas(-sen) per aspect gemarkeerd, (zie Figuur 14-8 boven aan elke kolom). Ook werden per leerling de opvallende uitkomsten gemarkeerd: dat wil zeggen leerlingen die overall extra aandacht (remediërend of verdiepend) nodig hadden en bij elke leerling de aspecten die extra aandacht nodig hadden.

Wat sortering betreft moest bij de groepsrapportage op naam en op de diagnose van elk hoofdaspect gesorteerd kunnen worden. Verder moest er bij de sortering op de overall diagnose voor het vak een handige groepering komen. Dit werd gerealiseerd door een algoritme waarbij binnen de sortering op de overall diagnose (onder, op of boven niveau), gesorteerd wordt op het deelaspect met binnen de groep met de meeste diagnoses onder niveau, vervolgens op het deelaspect met de op een na meeste diagnoses onder niveau enzovoort (zie de sortering in Figuur 14-8).



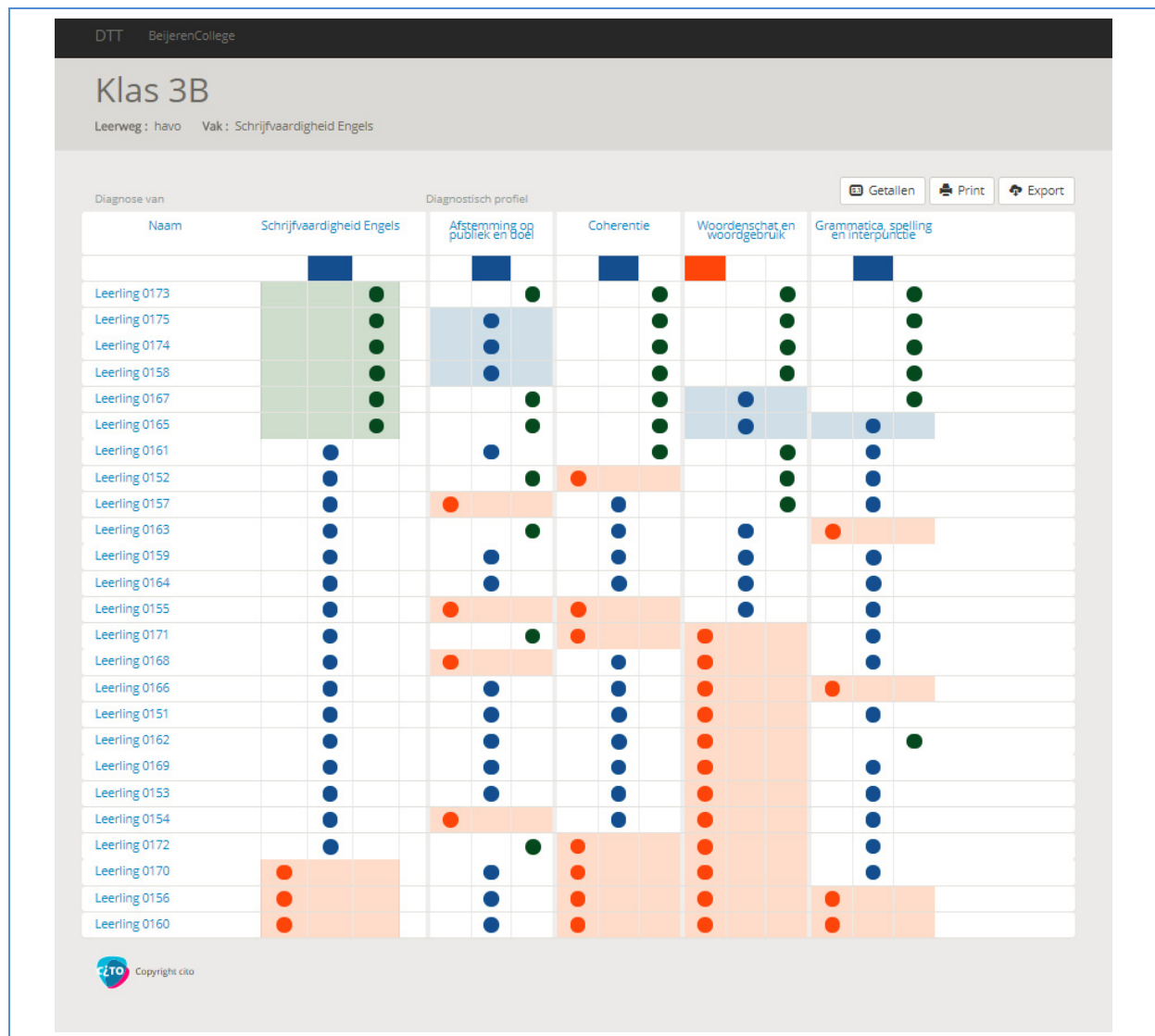
Figuur 14-7. Leerlingrapportage in de mock-up

14.4 Het definitieve product

De uiteindelijke mock-up is gebruikt om feedback van de klankbordgroep, scholen en docenten te krijgen. Er is een online enquête afgenomen onder scholen. Uit de enquête bleek dat de docenten en schoolleiders de visualisatie goed te begrijpen vonden en dat ze vonden dat de leerling- en klasrapportages waardevolle en bruikbare informatie opleverden. Opvallend was dat de meerderheid van de docenten en schoolleiders vooral de weergave wilden zien waarin de onzekerheid gevisualiseerd werd. De grafische weergave met bolletjes werd ook verkozen boven een getalsmatige of tekstuele weergave. De tekstuele weergave werd het minst noodzakelijk geacht. In Bijlage 17.4 worden de resultaten van de enquête beschreven.

Na de feedback is onder regie van het CvTE de rapportage in Facet gerealiseerd waarbij de mock-up als specificatie is gebruikt. Er is besloten om geen tekstuele beschrijving te realiseren omdat uit de feedback bleek dat deze weinig geraadpleegd zou worden. Op grond van de feedback is de markering van de groepsrapportage ook iets gewijzigd. Omdat de toetsuitkomsten van de DTT (zonder scores) geheel anders zijn dan van overige toetsen in Facet moesten hiervoor speciale aanpassingen gerealiseerd worden in Facet.

De visualisatie van de toetsuitkomsten in de rapportage van Facet leek sterk op de visualisatie in de mock-up (zie Figuur 14-9). De interacties waren uiteindelijk wel anders dan in het ontwerp. Zo was het niet mogelijk dat een docent ter plekke in een klaslokaal de rapportages bekeek via een tablet of computer (zoals beschreven in de gebruikersverhalen). Ook was het helaas niet mogelijk om alle rapportages van een afname tegelijkertijd te downloaden.



Figuur 14-8. Groepsrapportage in de mock-up in de standaard weergave

Naast de leerling- en groepsrapportages wilden docenten ook zicht op de gemaakte opgaven. Doorgaans wordt inzage in de opgaven gebruikt voor drie doeleinden. Allereerst is het belangrijk om de leerlingen van tevoren een idee te kunnen geven van het soort opgaven dat ze zullen krijgen en leerlingen eventueel te laten oefenen met nieuwe digitale opgavetypen en digitale hulpmiddelen. Ook voor andere partijen was het relevant om het soort opgaven te zien zodat deze andere partijen, zoals uitgeverijen, kunnen aansluiten bij de ontwikkelingen en innovaties. Om deze reden zijn oefenopgaven gemaakt en op een website (<https://oefenen.facet.onl>) geplaatst. Ten tweede is het belangrijk dat elke leerling de gelegenheid krijgt om te controleren of er geen fouten in de beoordeling zijn gesloten. Daarom kon bij gerede twijfel aan de toetsuitkomst de gemaakte toets onder supervisie ingekeken worden. Een derde doel waarvoor inzage doorgaans gebruikt wordt door docenten, is om meer grip te krijgen op de verbeterpunten en sterke punten die naar voren komen in de toetsuitkomst, zodat het onderwijsleerproces aangepast kan worden. Om docenten dergelijke aanknopingspunten te bieden is besloten om voorbeeldopgaven aan te bieden bij de toetsuitkomst. Als een docent of leerling ziet dat een leerling een bepaald verbeterpunt heeft kunnen ze samen naar het soort opgaven kijken waarmee de leerling bij de DTT moeite had.

Per deelaspect zijn twee voorbeeldopgaven ontwikkeld met een korte omschrijving voor welke diagnose de opgave illustratief is en wat de sleutel is. Deze voorbeeldopgaven konden nog niet direct via de leerlingrapportage ingezien worden, maar waren voor docenten en leerlingen in te zien via een website. Idealiter zouden docenten en leerlingen direct via de rapportage de voorbeeldopgaven te zien krijgen in

plaats van via een website. Ook zou in aanvulling op de voorbeeldopgaven, een voorbeeld van een lesdoel en een les getoond kunnen worden die ingaan op de gerapporteerde verbeter- of sterktepunten.

Leerlingrapportage 'Naam Leerling'

Weergave: genuanceerd

Vak: DTT Engels schrijfvaardigheid **Niveau:** VO-vwo **Afnamegroep:**

Diagnose van

DTT Engels schrijfvaardigheid

Hoofdaspecten

Afstemmen op doel en publiek	
Samenhang	
Woordenschat en woordgebruik	
Grammatica, spelling en interpunctie	

Deelaspecten

Toonzetting en register afstemmen		Conventies bij tekstsoort gebruiken	
Tekststructuur en verbanden aanbrengen		Passende structuurwoorden gebruiken: voegwoorden en verwijswwoorden	
Passende woorden en woordcombinaties gebruiken		Woordgebruik functioneel variëren	
Woordvolgorde en zinsconstructie functioneel hanteren		Passende spelling en interpunctie hanteren	

Legenda

Onder niveau	
Op niveau	
Boven niveau	

Figuur 14-9. Voorbeeld van een definitieve leerlingrapportage zoals die kan worden afgedrukt

14.5 Literatuur

Hojtink, H., & Sies, A. (2014). Adaptieve procedure met itemblokken. In S. Schouwstra (Red.), *De diagnostisch tussentijdse toets: onderzoek 2014* (pp. 33-40). Arnhem: Cito.

Roelofs, E., & Schouwstra, S. (Red.). (2012). *Diagnostische tussentijdse toets: Verslag van de voorstudie*. Arnhem: Cito.

15 Resultaten van de DTT-afname 2017

Sanneke Schouwstra, Jesse Koops en Daniel van der Palm

In februari 2017 is de volledig adaptieve afname geweest. Na de afname zijn de uitkomsten geanalyseerd. Hierbij is gekeken naar de toetspaden, de toetsduur, toetsuitkomsten en de eigenschappen van de (zaai-) opgaven. Met alle goedgekeurde opgaven zijn nieuwe blokindelingen voor de adaptieve toetsen geconstrueerd.

In 2017 hebben 125 schoollocaties meegedaan aan de afnames. In totaal zijn er 30490 afnames gestart, dat is minder dan in 2016. In 2017 zijn ook extra pretest afnames geweest in vmbo-bb (zie Tabel 15-1), omdat in 2016 te weinig observaties waren gerealiseerd (<500) voor een betrouwbare kalibratie (Cito, 2016). De gegevens van de extra pretestafnames zijn samengevoegd met de pretestafnames in 2016 voor herkalibratie.

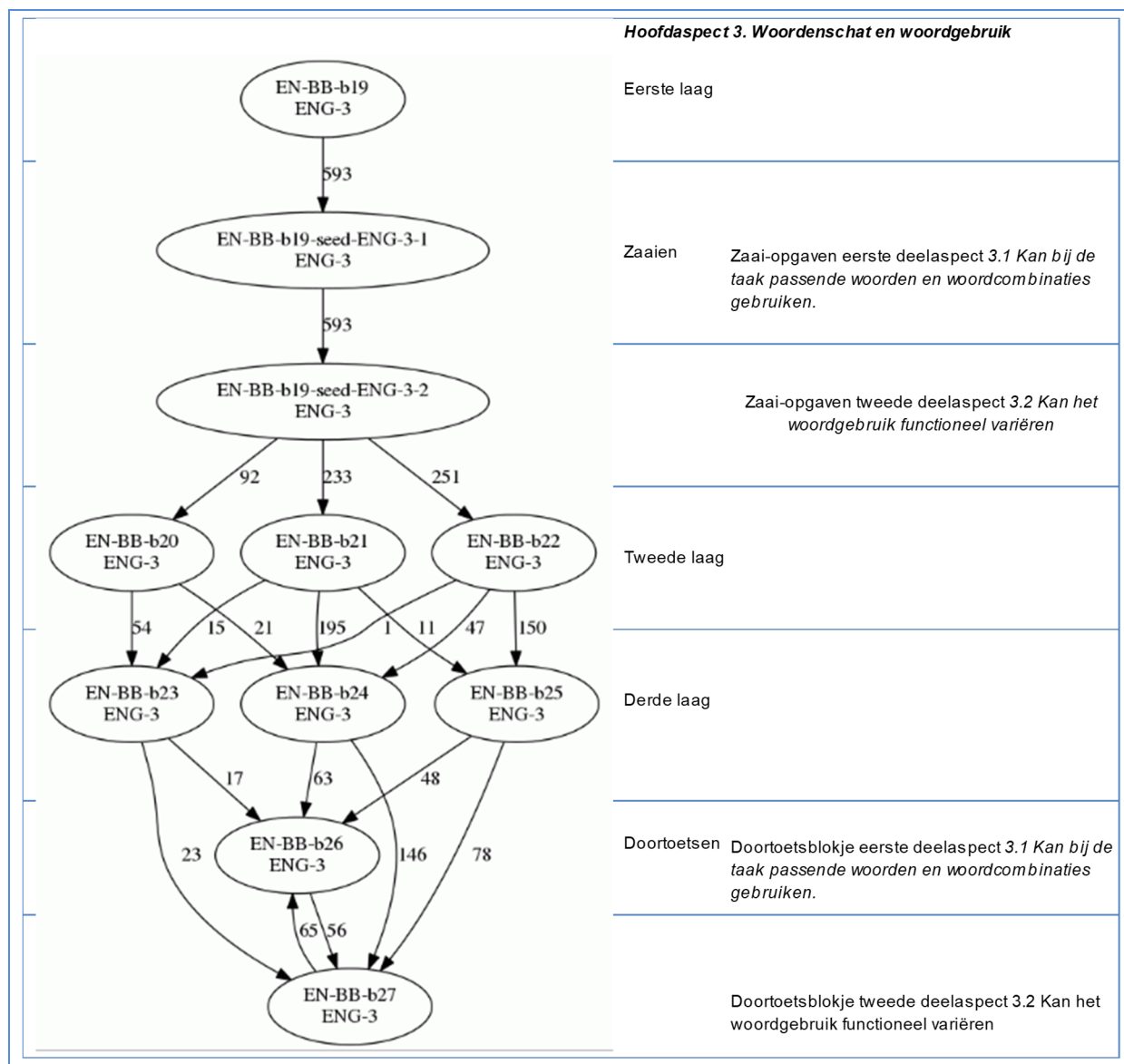
Tabel 15-1. Aantal afnames per vak en per toets

		Schrijfvaardigheid Engels			Schrijfvaardigheid Nederlands			Wiskunde		
		2015	2016	2017	2015	2016	2017	2015	2016	2017
pretest	vmbo-bb	505	358	213	747	383	273	618	343	316
	vmbo-kb	793	607		787	581		1117	753	
	vmbo-gt	1321	1475		1791	1397		1651	1586	
	havo	1269	1036		1321	820		1777	1314	
	vwo	1448	1150		1774	922		1839	1122	
	Totaal	5336	4626	213	6420	4103	273	7002	5118	316
adaptief	vmbo-bb		808	598		988	809		919	729
	vmbo-kb		1266	1208		1362	1325		1574	1365
	vmbo-gt		2364	2319		3404	2955		2753	2396
	havo		2265	2681		1956	2366		3056	3211
	vwo		2110	2390		2255	2607		2013	2729
	Totaal		8813	9196		9965	10062		10315	10430
Totaal		5336	13439	9409	6420	14068	10335	7002	15433	10746

Opmerking: In elk jaar hebben sommige leerlingen aan meerdere toetsen (pretest en adaptief) en/of aan meerdere vakken meegedaan.

15.1 Doorlopen paden in de adaptieve afname

Door de adaptiviteit kan een leerling op heel veel verschillende manieren door de toets gaan. Het totaal aantal mogelijke paden per toets is enorm groot. Alleen het eerste stadium (zonder doortoetsen) kent al 6561 verschillende paden per toets. Om toch zicht te krijgen op de doorlopen paden is per hoofdaspect het aantal leerlingen dat een bepaald pad tussen de lagen heeft gehad gevisualiseerd, zie bijvoorbeeld Figuur 15-1.



Figuur 15-1. Voorbeeld van een visualisatie van de paden bij hoofdaspect 3. Woordenschat en woordgebruik, Schrijfvaardigheid Engels in vmbo-bb

In het voorbeeld zien we dat in vmbo-bb 593 leerlingen na de eerste laag van 3. *Woordenschat en woordgebruik* bij Schrijfvaardigheid de zaai-opgaven voor het eerste deelaspect 3.1 *Kan bij de taak passende woorden en woordcombinaties gebruiken* hebben gemaakt. Diezelfde 593 leerlingen hebben vervolgens de zaai-opgaven van het tweede deelaspect 3.2 *Kan het woordgebruik functioneel variëren* gemaakt.

Na de zaai-opgaven heeft de eerste doorverwijzing plaatsgevonden: in de tweede laag hebben 92 leerlingen het onder-niveaublokje gekregen, 233 het op-niveaublokje en 251 leerlingen het boven-niveaublokje. Daarna begonnen leerlingen aan de derde laag.

- Na het onder-niveaublokje zijn 54 leerlingen in de derde laag ook naar het onder-niveaublokje gegaan en 21 leerlingen naar op-niveau. Er zijn geen leerlingen naar boven-niveau gegaan.
- Van de leerlingen die in de tweede laag het op-niveaublokje hebben gemaakt zijn 15 leerlingen in de derde laag naar het onder-niveaublokje gegaan, 195 naar op-niveau en 11 leerlingen naar het boven-niveaublokje.

- Na het boven-niveaublokje is 1 leerling naar het onder-niveaublokje in de derde laag gegaan, 47 naar het op-niveaublokje en 150 naar het boven-niveaublokje.

Na de derde laag is het doortoetsen begonnen:

- 128 leerlingen zijn begonnen aan het doortoetsblokje *3.1 Kan bij de taak passende woorden en woordcombinaties gebruiken*: 17 vanuit het onder-niveaublokje, 63 vanuit het op-niveaublokje en 48 vanuit het boven-niveaublokje. Hierna hebben 56 leerlingen het doortoetsblokje gemaakt van het tweede deelaspect.
- 247 leerlingen zijn als eerste begonnen met het doortoetsblokje voor deelaspect *3.2 Kan het woordgebruik functioneel variëren*, waarvan 65 vervolgens het doortoetsblokje voor het eerste deelaspect hebben gemaakt.

Voor alle drie de vakken, voor elke leerweg/schoolsoort en voor elk hoofdaspect is een dergelijke visualisatie gemaakt (in totaal 60, zie Bijlage 17.3). We zien dat het vrij zeldzaam is dat leerlingen vanuit een onder-niveaublokje in de tweede laag naar het boven-niveaublokje in de derde laag gaan en vice versa dat leerlingen vanuit een boven-niveaublokje in de tweede laag naar het onder-niveaublokje in de derde laag gaan. Alle overige paden komen met regelmaat voor.

15.2 Toetsduur en antwoordtijd

De gemiddelde antwoordtijd per opgave was bij schrijfvaardigheid Engels ongeveer een halve minuut, bij schrijfvaardigheid Nederlands beduidend langer: bijna anderhalve minuut. Bij wiskunde varieerde de antwoordtijd behoorlijk per schoolsoort en leerweg. De gemiddelde antwoordtijd in vmbo was ongeveer een minuut, in havo anderhalve minuut en in vwo zelfs bijna twee minuten (zie Tabel 15-2).

Uit de antwoordtijden kunnen we afleiden dat een aantal schoollocaties gebruik heeft gemaakt van de mogelijkheid om de leerlingen een pauze te geven. Omdat geen directe informatie gegeven wordt over pauzeren in de aangeleverde gegevens, moeten we dat uit de antwoordtijden afleiden. Indien in een afname de maximale antwoordtijd in een toets gemiddeld over alle leerlingen ver boven de gemiddelde antwoordtijd per opgave ligt (bijvoorbeeld boven de 20 minuten) kunnen we aannemen dat sprake is van een pauze. In 7% van de afnames is de gemiddelde maximum antwoordtijd boven de 20 minuten.

Alhoewel het de bedoeling was dat leerlingen drie uur zouden werken aan de DTT, zien we dat gemiddeld veel minder tijd aan de DTT is besteed (zie Tabel 15-2). Als we corrigeren voor een pauze (totale toetstijd minus de maximale antwoordtijd in een toets gemiddeld over alle leerlingen) zien we dat bij Engels de leerlingen ruim een uur aan de adaptieve DTT hebben besteed, bij Nederlands en wiskunde 1,5 tot 2 uur.

Tabel 15-2. Gemiddelde toetsduur en antwoordtijd per vak en leerweg/schoolsoort

vak	Schoolsoort/ leerweg	N	Gemiddeld aantal opgaven	Gemiddelde toetsduur	Gecorrigeerde gemiddelde toetsduur	Gemiddelde antwoordtijd per opgave
Engels	vmbo-bb	598	139	00:58:51	00:54:57	00:00:25
	vmbo-kb	1208	151	01:11:33	01:07:29	00:00:28
	vmbo-gt	2319	149	01:09:37	01:06:35	00:00:28
	havo	2681	161	01:21:02	01:17:27	00:00:30
	vwo	2390	143	01:19:00	01:14:00	00:00:33
Nederlands	vmbo-bb	809	81	01:39:02	01:29:00	00:01:13
	vmbo-kb	1325	93	02:02:46	01:50:01	00:01:19
	vmbo-gt	2955	90	02:10:26	02:00:36	00:01:27
	havo	2366	91	02:11:07	02:04:50	00:01:26
	vwo	2607	98	02:26:30	02:18:29	00:01:30
Wiskunde	vmbo-bb	729	107	01:31:03	01:21:30	00:00:51
	vmbo-kb	1365	105	01:40:49	01:31:19	00:00:58
	vmbo-gt	2396	97	01:42:03	01:34:30	00:01:03
	havo	3211	71	01:45:27	01:38:10	00:01:29
	vwo	2729	74	02:21:03	02:09:53	00:01:54

15.3 Toetsuitkomsten

Voor de vakken is per hoofdaspect het percentage leerlingen berekend dat een diagnose onder niveau, een diagnose op niveau en een diagnose boven niveau krijgt (zie Tabel 15-3, Tabel 15-4 en Tabel 15-5).

We hebben de toetsuitkomsten vergeleken met de verwachte impact tijdens de standaardbepalingen.

De verwachte impact is het percentage pretestleerlingen dat onder, op of boven niveau valt als de standaard wordt toegepast (zie ook Hoofdstuk 11).

Tabel 15-3. Toetsuitkomsten schrijfvaardigheid Engels van de adaptieve afname 2017 vergeleken met de impact van de standaarden tijdens de pretest 2016

Toetsuitkomsten 2017				Impact standaarden pretestdata 2016			
		% onder-niveau	% op-niveau	% boven-niveau	% onder-niveau	% op-niveau	% boven-niveau
Schrijfvaardigheid Engels							
vmbo-bb	1. Afstemming op publiek en doel	22,2%	55,4%	22,4%	13,5%	50,3%	36,2%
	2. Coherentie	26,8%	54,2%	19,0%	28,3%	43,1%	28,6%
	3. Woordenschat en woordgebruik	19,7%	45,5%	34,7%	12,9%	46,9%	40,2%
	4. Grammatica, spelling en interpunctie	28,0%	45,0%	27,0%	24,2%	50,1%	25,6%
vmbo-kb	1. Afstemming op publiek en doel	20,1%	50,1%	29,8%	16,4%	44,4%	39,1%
	2. Coherentie	25,0%	49,4%	25,6%	19,9%	42,4%	37,7%
	3. Woordenschat en woordgebruik	18,3%	54,8%	26,9%	13,6%	50,9%	35,4%
	4. Grammatica, spelling en interpunctie	15,2%	58,7%	26,0%	15,1%	67,1%	17,8%
vmbo-gt	1. Afstemming op publiek en doel	9,3%	60,1%	30,6%	12,8%	66,6%	20,6%
	2. Coherentie	12,8%	47,4%	39,8%	15,1%	52,3%	32,6%
	3. Woordenschat en woordgebruik	11,8%	50,1%	38,1%	11,1%	62,1%	26,8%
	4. Grammatica, spelling en interpunctie	16,1%	64,5%	19,4%	13,2%	65,5%	21,3%
havo	1. Afstemming op publiek en doel	16,8%	62,8%	20,4%	13,1%	66,7%	20,2%
	2. Coherentie	23,3%	54,3%	22,4%	26,7%	56,6%	16,8%
	3. Woordenschat en woordgebruik	21,5%	54,9%	23,7%	23,6%	56,9%	19,5%
	4. Grammatica, spelling en interpunctie	16,0%	59,4%	24,6%	19,1%	61,7%	19,2%
vwo	1. Afstemming op publiek en doel	22,0%	66,9%	11,2%	13,6%	64,0%	22,4%
	2. Coherentie	9,0%	62,8%	28,2%	24,4%	54,9%	20,7%
	3. Woordenschat en woordgebruik	23,8%	52,1%	24,0%	34,6%	40,4%	24,9%
	4. Grammatica, spelling en interpunctie	24,4%	53,5%	22,2%	36,1%	46,8%	17,2%
Gemiddeld		19,1%	55,1%	25,8%	19,4%	54,5%	26,1%

Opmerking. De rood gemarkeerde percentages zijn 10% lager dan in de pretest, de groen gemarkeerde percentages zijn 10% hoger dan in de pretest.

Over leerwegen/schoolsoorten en aspecten heen zien we dat bij schrijfvaardigheid Engels (zie Tabel 15-3) conform de verwachting rond de 55% van de leerlingen op-niveau gediagnosticeerd wordt, ruim een kwart (26%) boven niveau 19% onder niveau. Het meest opvallende is dat in vwo bij drie van de vier hoofdaspecten minder leerlingen onder-niveau vallen dan verwacht werd tijdens de pretest.

Tabel 15-4. Toetsuitkomsten schrijfvaardigheid Nederlands van de adaptieve afname 2017 vergeleken met de impact van de standaarden tijdens de pretest 2016

Schrijfvaardigheid Nederlands		Toetsuitkomsten 2017			Impact standaarden pretest data 2016		
		% onder-niveau	% op-niveau	% boven-niveau	% onder-niveau	% op-niveau	% boven-niveau
vmbo-bb	1 Retorische vaardigheden	11,0%	51,5%	37,5%	12,5%	62,7%	24,7%
	2 Tekststructurele vaardigheden	15,9%	57,7%	26,4%	24,1%	58,4%	17,5%
	3 Linguïstische vaardigheden	21,1%	61,8%	17,1%	24,7%	49,8%	25,4%
	4 Orthografische vaardigheden	23,7%	62,7%	13,6%	36,7%	41,7%	21,6%
vmbo-kb	1 Retorische vaardigheden	10,1%	48,8%	41,1%	17,6%	41,4%	41,0%
	2 Tekststructurele vaardigheden	16,2%	50,9%	32,9%	17,9%	50,3%	31,8%
	3 Linguïstische vaardigheden	9,2%	64,2%	26,6%	14,8%	54,7%	30,5%
	4 Orthografische vaardigheden	19,3%	66,3%	14,4%	14,5%	59,4%	26,1%
vmbo-gt	1 Retorische vaardigheden	4,5%	45,1%	50,4%	9,1%	47,7%	43,2%
	2 Tekststructurele vaardigheden	7,2%	64,5%	28,3%	28,6%	54,9%	16,5%
	3 Linguïstische vaardigheden	7,9%	68,8%	23,3%	11,2%	62,8%	26,0%
	4 Orthografische vaardigheden	26,5%	46,2%	27,4%	20,0%	64,7%	15,3%
havo	1 Retorische vaardigheden	12,7%	59,8%	27,4%	12,9%	62,2%	24,9%
	2 Tekststructurele vaardigheden	18,5%	60,5%	20,9%	26,8%	48,9%	24,3%
	3 Linguïstische vaardigheden	17,3%	67,6%	15,0%	35,0%	50,3%	14,7%
	4 Orthografische vaardigheden	37,2%	59,8%	3,0%	47,9%	38,8%	13,3%
vwo	1 Retorische vaardigheden	1,4%	70,1%	28,6%	15,2%	69,7%	15,2%
	2 Tekststructurele vaardigheden	13,5%	70,3%	16,2%	12,7%	70,5%	16,8%
	3 Linguïstische vaardigheden	26,9%	61,2%	11,9%	18,4%	69,3%	12,3%
	4 Orthografische vaardigheden	23,9%	69,1%	7,0%	25,3%	62,7%	12,0%
Gemiddeld		16,2%	60,3%	23,5%	21,3%	56,0%	22,7%

Opmerking. De rood gemarkeerde percentages zijn 10% lager dan in de pretest, de groen gemarkeerde percentages zijn 10% hoger dan in de pretest.

Bij schrijfvaardigheid Nederlands is doorgaans 60% op niveau, wat minder dan een kwart boven niveau en 16% onder niveau (zie Tabel 15-4). De percentages variëren echter aanzienlijk over de aspecten binnen de leerwegen/schoolsoorten. Het meest opvallende is dat bij havo meer leerlingen op niveau zitten dan verwacht (en minder onder niveau). Bij vmbo-bb zitten ook meer leerlingen op niveau bij hoofdaspect 3 *Linguïstische vaardigheden* en 4 *Orthografische vaardigheden*. Bij vier aspecten (in verschillende leerwegen/schoolsoorten) zitten meer leerlingen boven niveau dan verwacht.

Tabel 15-5. Toetsuitkomsten wiskunde van de adaptieve afname 2017 vergeleken met de impact van de standaarden tijdens de pretest 2016 (domeinen B, C en F) en de pretest 2015 (domeinen D en E)

Wiskunde		Toetsuitkomsten 2017			Impact standaarden pretest 2016*		
		% onder-niveau	% op-niveau	% boven-niveau	% onder-niveau	% op-niveau	% boven-niveau
vmbo-bb	B: Getallen (havo/vwo en variabelen)	43,2%	46,6%	10,2%	33,7%	53,2%	13,2%
	C: Verhoudingen	56,0%	37,3%	6,8%	35,8%	45,7%	18,5%
	D: Meten en meetkunde	32,7%	54,6%	12,7%	30,4%	54,0%	15,5%
	E: Verbanden en formules	34,8%	53,8%	11,4%	27,1%	56,5%	16,4%
vmbo-kb	B: Getallen (havo/vwo en variabelen)	41,2%	46,3%	12,5%	36,6%	45,7%	17,7%
	C: Verhoudingen	46,3%	43,7%	10,0%	42,0%	40,9%	17,1%
	D: Meten en meetkunde	47,1%	49,4%	3,5%	37,5%	56,6%	5,9%
	E: Verbanden en formules	37,4%	53,4%	9,2%	22,9%	65,4%	11,7%
vmbo-gt	B: Getallen (havo/vwo en variabelen)	35,6%	51,9%	12,5%	32,4%	53,7%	13,9%
	C: Verhoudingen	36,2%	47,7%	16,1%	30,3%	53,3%	16,4%
	D: Meten en meetkunde	33,8%	56,8%	9,4%	31,2%	62,6%	6,3%
	E: Verbanden en formules	27,9%	57,0%	15,0%	23,4%	70,7%	5,9%
havo	B: Getallen (havo/vwo en variabelen)	48,0%	46,3%	5,6%	18,2%	66,4%	15,4%
	C: Verhoudingen	45,2%	42,5%	12,4%	17,1%	66,1%	16,8%
	D: Meten en meetkunde	37,4%	49,3%	13,3%	39,1%	49,3%	11,6%
	E: Verbanden en formules	46,6%	40,6%	12,8%	28,2%	65,8%	6,0%
	F: Informatieverwerking en onzekerheid (havo/vwo)	54,8%	39,9%	5,4%	16,4%	67,6%	16,0%
vwo	B: Getallen (havo/vwo en variabelen)	42,2%	51,7%	6,2%	11,4%	78,0%	10,6%
	C: Verhoudingen	45,7%	38,1%	16,1%	17,2%	64,2%	18,5%
	D: Meten en meetkunde	36,8%	53,7%	9,5%	31,8%	58,4%	9,8%
	E: Verbanden en formules	47,5%	46,8%	5,7%	38,3%	57,7%	4,0%
	F: Informatieverwerking en onzekerheid (havo/vwo)	53,9%	41,4%	4,7%	14,9%	69,7%	15,4%
Gemiddeld		41,6%	48,2%	10,2%	27,3%	60,1%	12,6%

Opmerking. De rood gemarkeerde percentages zijn 10% lager dan in de pretest, de groen gemarkeerde percentages zijn 10% hoger dan in de pretest.

Bij wiskunde (Tabel 15-5) zien we dat doorgaans *minder dan de helft* van de leerlingen de diagnose op niveau krijgen en meer dan veertig procent de diagnose onder niveau. De toetsuitkomsten zijn dus veel lager dan verwacht. We zien dat vooral bij havo en vwo een veel hoger percentage leerlingen onder niveau valt.

Vraag die nog niet onderzocht is, is of de gekozen aanpak van adaptiviteit de DTT wiskunde niet te moeilijk maakt. Tijdens de pretesten kregen de leerlingen opgaven uit twee (vmbo) of drie (havo/vwo) domeinen. Bij de adaptieve afname krijgen de leerlingen opgaven uit alle vier (vmbo) en alle vijf (havo/vwo) domeinen. Bovendien kregen bij de pretesten de leerlingen alle opgaven per domein aangeboden, dus bijvoorbeeld eerst alle opgaven van domein *D: Meten en meetkunde* en daarna alle opgaven van domein *E: Verbanden en formules*. Tijdens de adaptieve afname zijn de opgaven door elkaar aangeboden conform het advies van de vaststellingscommissie (zie Hoofdstuk 13). Die afwisseling zou bij wiskunde misschien de toets als geheel te moeilijk kunnen maken.

15.5 Iteanalyses en (her-)kalibratie

In 2017 is een beperkt aantal opgaven gezaaid. De zaai-opgaven zijn geanalyseerd en beoordeeld zoals bij de pretest is gebeurd (zie Hoofdstuk 4, Hoofdstuk 5 en Hoofdstuk 7). De zaai-opgaven die goedgekeurd zijn door de vaststellingscommissie, zijn gekalibreerd aan de hand van de toetsuitkomst (zie Hoofdstuk 12). Voor schrijfvaardigheid Engels zijn 62 zaai-opgaven goedgekeurd van de 75, voor schrijfvaardigheid Nederlands 17 van de 23 en bij wiskunde zijn 75 van de 96 zaai-opgaven goedgekeurd.

Alle andere opgaven, die eerder zijn gepretest, zijn geherkalibreerd. Om te zorgen dat de parameters gebaseerd zijn op voldoende observaties, zijn voor elke opgave de pretestgegevens samengenomen met de gegevens van één adaptieve afname. Bij de gegevens van de adaptieve afname zijn eerst gegevens geïmputeerd (zie Hoofdstuk 12). Hierna zijn de pretestgegevens samengevoegd met de geïmputeerde, adaptieve gegevens. De opgaven die in 2015 zijn gepretest zijn samengenomen met de adaptieve afname van 2016. De opgaven die in 2016 zijn gepretest zijn samengenomen met de adaptieve afname van 2017. Op grond van de beheersingscategorie (onder, op of boven niveau) zijn tenslotte de opgaven gekalibreerd. Bij de pretestleerlingen is per hoofdaspect de diagnose bepaald door het toepassen van de standaard (zie Hoofdstuk 11) en bij de leerlingen van de adaptieve afnames is de diagnose bepaald tijdens de afname (de toetsuitkomst).

15.6 Blokindeling voor oplevering 2018

Schrijfvaardigheid Nederlands

Met behulp van simulaties is de blokindeling gezocht (zie Hoofdstuk 12) waarbij zoveel mogelijk diagnoses correct zijn. De simulaties laten zien dat de beste blokindelingen bij schrijfvaardigheid Nederlands tot accurate diagnoses leiden (zie Tabel 15-6). De simulaties laten zien dat in doorsnee het percentage correcte diagnoses rond de 90% ligt. Bij slechts één diagnose is het percentage lager dan 70%. In havo bij hoofdaspect 1. *Doel en publiek* krijgt 67,3% van de op-niveauleerlingen de correcte diagnose, terwijl van de onder-niveauleerlingen 81,6% en van de boven-niveauleerlingen 79,7% een correcte diagnose krijgt.

Bij de deelaspecten is het percentage correcte diagnoses in doorsnee goed: 76%, maar enigszins lage percentages (lager dan 70%) komen vaak voor, met name in vmbo-gt (15 van de 36 deelaspect-diagnoses) en vwo (12 van de 39 diagnoses), zie Bijlage 17.2. In havo en vwo zijn drie deelaspecten voor het doortoetsen onvoldoende accuraat voor rapportage, dat wil zeggen dat bij een of meer van de diagnoses (onder, op of boven-niveau) het percentage correcte diagnoses onder 60% ligt. Dit is in havo het geval bij deelaspect 2.1 *Tekstelementen kiezen* en bij vwo bij de deelaspecten 1.1 *Voorkennis en informatiebehoefte bij de lezers inschatten* en 2.4 *Een standpunt weergeven en van passende argumenten voorzien* (zie Bijlage 17.2).

Schrijfvaardigheid Engels

De simulaties laten zien dat de beste blokindeling bij schrijfvaardigheid Engels tot zeer accurate diagnoses van de hoofdaspecten leiden. Hoe goed de oplossingen zijn kan gezien worden in Tabel 15-7. Bijvoorbeeld bij de diagnose van hoofdaspect 3. *Woordenschat en woordgebruik* in vmbo-bb zien we dat 96,4% van de onder-niveauleerlingen correct gediagnosticeerd worden, 96,6% van de op-niveauleerlingen en 92,3% van de boven-niveauleerlingen.

In doorsnee ligt bij schrijfvaardigheid Engels het percentage correcte diagnoses rond de 93%. In alle gevallen ligt het percentage correcte diagnoses bij de hoofdaspecten ruim boven de 85% (zie Tabel 15-7). Bij de deelaspecten ligt het percentage correcte diagnoses iets lager, omdat die diagnoses op minder responsen zijn gebaseerd. Maar ook hier ligt het percentage gemiddeld hoog: 83,4% correcte diagnoses *voordat* het doortoetsen begint (zie Bijlage 17.2 voor alle uitkomsten op deelaspect-niveau). Alleen in vwo bij deelaspect 1.1 *Kan toonzetting en register afstemmen op publiek en schrijfdoel* en de diagnose op niveau bij deelaspect 4.2 *Kan passende spelling en interpunctie hanteren* is het percentage correcte classificaties aan de lage kant (rond de 63%, zie Bijlage 17.2).

Tabel 15-6. Percentages correcte diagnoses schrijfvaardigheid Nederlands, aantal opgaven en responsen in de simulaties

Leerweg/ schoolsoort	Hoofdaspect	Onder niveau	Op niveau	Boven niveau	Aantal opgaven	Aantal responsen
vmbo-bb	1. Doel en publiek	95,0%	94,9%	94,9%	43	160
	2. Structuur	93,3%	87,3%	84,8%	51	129
	3. Woord- en zinsniveau	95,5%	95,5%	95,4%	47	157
	4. Spelling en interpunctie	94,9%	94,4%	94,4%	40	213
vmbo-kb	1. Doel en publiek	93,6%	93,5%	93,5%	44	190
	2. Structuur	90,6%	90,6%	90,6%	56	118
	3. Woord- en zinsniveau	97,0%	96,8%	96,8%	50	192
	4. Spelling en interpunctie	96,2%	95,7%	95,7%	37	212
vmbo-gt	1. Doel en publiek	89,0%	89,0%	89,0%	43	170
	2. Structuur	95,2%	94,5%	94,7%	64	127
	3. Woord- en zinsniveau	94,7%	94,5%	94,5%	43	163
	4. Spelling en interpunctie	85,9%	78,9%	73,6%	41	212
havo	1. Doel en publiek	81,6%	67,3%	79,7%	42	126
	2. Structuur	80,5%	76,7%	72,7%	67	160
	3. Woord- en zinsniveau	92,1%	92,8%	93,3%	43	174
	4. Spelling en interpunctie	95,0%	95,0%	95,0%	36	204
vwo	1. Doel en publiek	96,4%	82,9%	71,9%	50	106
	2. Structuur	76,6%	74,4%	71,4%	67	164
	3. Woord- en zinsniveau	91,6%	91,6%	91,6%	51	168
	4. Spelling en interpunctie	97,8%	97,6%	97,6%	34	176

Tabel 15-7. Percentages correcte diagnoses schrijfvaardigheid Engels, aantal opgaven en responsen in de simulaties

School- soort/ Leerweg	Hoofdaspect	Onder niveau	Op niveau	Boven niveau	Aantal opgaven	Aantal responsen
vmbo-bb	1. Afstemming op publiek en doel	90,3%	90,3%	90,3%	51	107
	2. Coherentie	94,1%	94,1%	94,1%	61	80
	3. Woordenschat en woordgebruik	96,4%	96,5%	96,6%	58	129
	4. Spelling, interpunctie en grammatica	92,9%	92,8%	92,8%	65	70
vmbo-kb	1. Afstemming op publiek en doel	89,0%	89,0%	89,0%	52	124
	2. Coherentie	95,4%	95,4%	95,4%	65	99
	3. Woordenschat en woordgebruik	97,1%	97,1%	97,1%	68	147
	4. Spelling, interpunctie en grammatica	96,0%	96,0%	96,0%	70	109
vmbo-gt	1. Afstemming op publiek en doel	95,6%	95,6%	95,6%	43	134
	2. Coherentie	96,1%	96,1%	96,2%	73	96
	3. Woordenschat en woordgebruik	96,0%	95,9%	96,0%	60	131
	4. Spelling, interpunctie en grammatica	93,6%	93,6%	93,6%	70	99
havo	1. Afstemming op publiek en doel	93,2%	93,2%	93,2%	48	149
	2. Coherentie	91,3%	91,3%	91,3%	54	84
	3. Woordenschat en woordgebruik	94,6%	94,6%	94,7%	78	130
	4. Spelling, interpunctie en grammatica	93,3%	93,3%	93,3%	73	109
vwo	1. Afstemming op publiek en doel	88,2%	88,2%	88,2%	39	100
	2. Coherentie	92,5%	92,5%	92,5%	51	92
	3. Woordenschat en woordgebruik	91,7%	91,7%	91,7%	72	114
	4. Spelling, interpunctie en grammatica	87,6%	87,6%	87,6%	52	82

Wiskunde

De beste blokindelingen bij wiskunde leiden tot redelijk accurate diagnoses (zie Tabel 15-8) wanneer alle opgaven in het eerste stadium worden gebruikt (dus zonder doortoetsen). In doorsnee ligt het percentage correcte classificaties rond de 82%. De accuratesse ligt lager dan bij de talen, vermoedelijk ook omdat het aantal responsen in veel gevallen aanzienlijk lager is.

In vmbo is de accuratesse in doorsnee 84%. Bij havo en vwo ligt de accuratesse iets lager (80%), dan bij vmbo (zie Tabel 15-8). Op deelaspect-niveau ligt het percentage correct voor de op-niveaudiagnose in vrijwel alle gevallen veel te laag (<60% correct). In doorsnee is het percentage correct voor de onder-niveaudiagnose (over de vijf leerwegen/schoolsoorten en drie deelaspecten heen) 70,6%, bij de op-niveau slechts 50,4% en bij de boven-niveau 74,1%.

Ook vorig jaar lag de accuratesse bij de deelaspecten te laag (<60%). Daarom zijn in 2017 alle opgaven gebruikt voor de diagnose op hoofdaspect-niveau (het eerste stadium) en waren er geen doortoetsblokjes voor de deelaspecten. Er werd ook niet gerapporteerd op deelaspect-niveau. Aangezien ook dit jaar in geen enkel domein binnen de leerwegen/schoolsoorten alle drie deelaspecten voldoende accuraat worden gediagnosticeerd, zijn weer alle opgaven opgenomen in het eerste stadium en zijn er geen doortoetsblokjes. Aangezien de diagnoses op deelaspect-niveau onvoldoende accuraat zijn, blijft rapportage van de deelaspect-diagnoses af te raden, zie Bijlage 17.2.

Tabel 15-8. Percentages correcte diagnoses wiskunde, aantal opgaven en responsen in de simulaties

Leerweg/ schoolsoort	Hoofdaspect	Onder niveau	Op niveau	Boven niveau	Aantal opgaven	Aantal responsen
vmbo-bb	B. Getallen	87.8%	87.8%	87.8%	70	90
	C. Verhoudingen	86.2%	86.2%	86.2%	56	76
	D. Meten en meetkunde	83.5%	83.5%	83.6%	49	77
	E. Verbanden en formules	79.5%	79.5%	79.5%	43	49
	F. Informatieverwerking en onzekerheid	76.7%	76.7%	76.7%	41	56
vmbo-kb	B. Getallen	80.5%	80.5%	80.5%	57	84
	C. Verhoudingen	81.4%	81.4%	81.4%	46	70
	D. Meten en meetkunde	83.6%	83.6%	83.6%	55	70
	E. Verbanden en formules	88.6%	88.6%	88.7%	41	63
	F. Informatieverwerking en onzekerheid	76.7%	76.7%	76.7%	41	56
vmbo-gt	B. Getallen	82.6%	82.6%	82.6%	53	70
	C. Verhoudingen	83.3%	83.3%	83.3%	67	77
	D. Meten en meetkunde	83.8%	83.8%	83.8%	52	63
	E. Verbanden en formules	87.2%	87.2%	87.2%	46	61
	F. Informatieverwerking en onzekerheid	76.7%	76.7%	76.7%	41	56
havo	B. Getallen en variabelen	79.0%	79.0%	79.0%	34	69
	C. Verhoudingen	80.0%	80.0%	80.0%	38	47
	D. Meten en meetkunde	77.4%	77.3%	77.4%	37	49
	E. Verbanden en formules	82.0%	82.0%	82.0%	36	56
	F. Informatieverwerking en onzekerheid	76.7%	76.7%	76.7%	41	56
vwo	B. Getallen en variabelen	77.9%	77.9%	77.9%	38	67
	C. Verhoudingen	77.4%	77.4%	77.4%	34	46
	D. Meten en meetkunde	84.7%	84.7%	84.8%	37	47
	E. Verbanden en formules	85.3%	85.2%	85.3%	42	48
	F. Informatieverwerking en onzekerheid	75.4%	75.4%	75.4%	36	48

15.7 Literatuur

Cito (2016). *Diagnostische tussentijdse toets: Verslag pretest 2016*. Arnhem: Cito.

16 Nawoord

Sanneke Schouwstra

Onder regie van het van CvTE is in de periode 2012-2017 samen met het onderwijsveld het prototype van de DTT ontwikkeld voor drie vakken. Voor vijf leerwegen/schoolsoorten zijn volledige adaptieve toetsen gemaakt voor schrijfvaardigheid Nederlands, schrijfvaardigheid Engels en wiskunde, die digitaal afgenomen zijn. Daarnaast zijn opgavebanken ontwikkeld voor leesvaardigheid Nederlands en leesvaardigheid Engels. In dit rapport is de inhoudelijke ontwikkeling van de DTT door Cito beschreven. Cito was verantwoordelijk voor de toetsconstructie, de psychometrische analyses, vormgeving van het rapportagemodel en de ontwikkeling van een adaptieve module.

Om aan nieuwe politieke vereisten en veranderende wensen van het onderwijsveld te voldoen, moest de toets sneller, op een meer flexibele en interactieve manier samen met scholen ontwikkeld worden. Onderzoek en innovaties op verschillende gebieden – toetsinhoud, psychometrie en technologie – hebben de volledige ontwikkeling van dit nieuwe instrument omgeven. Gedurende de ontwikkeling zijn de uitkomsten van het onderzoek en de opgedane kennis gedeeld met onder andere scholen, stakeholders, marktpartijen en onderwijsadviseurs via verslagen, workshops en presentaties.

De DTT is gebaseerd op nieuwe leerlingmodellen, die beschrijven welke kennis- en vaardigheidsaspecten noodzakelijk zijn voor een prestatie op niveau. Voor de DTT zijn nieuwe, meer authentieke, opgavevormen ontwikkeld die allen automatisch beoordeeld worden. Bij talen kunnen leerlingen bijvoorbeeld woorden selecteren en verbeteren. Bij wiskunde kunnen leerlingen in opgaven formules invoeren, grafieken tekenen en meetkundige constructies maken die automatisch beoordeeld worden. Om efficiënt te kunnen diagnosticeren is de toets adaptief, waarvoor een adaptieve module is ontwikkeld. De vorm van adaptiviteit die uitgewerkt is, was erop gericht zoveel mogelijk accurate diagnoses te kunnen stellen in de beschikbare afnametijd. De toets levert geen traditionele score, zoals vaak bij summatieve toetsen het geval is, maar geeft dankzij een nieuw psychometrisch model per aspect aan hoe waarschijnlijk het is dat de leerling onder niveau zit, op niveau zit en boven niveau zit. Voor deze toetsuitkomsten is een nieuwe rapportagevorm ontwikkeld.

Tijdens het onderzoek en de ontwikkeling van het prototype zijn verschillende mogelijkheden geopperd om de DTT verder te optimaliseren en ontwikkelen. Zo is het onder andere bij de talen wenselijk dat open schrijftaken in de toets worden opgenomen en een nakijkhulp ontwikkeld wordt, bijvoorbeeld met behulp van automatische beoordeling. Met dergelijke open schrijftaken kan ook de relatie tussen de diagnoses op deelaspecten en het vermogen om goed te schrijven onderzocht worden.

Bij wiskunde is verder onderzoek mogelijk voor een meer verfijnd leerlingmodel dat accuraat gediagnosticeerd kan worden en goed bruikbaar is voor docenten, ongeacht de lesmethode die ze hanteren. Ook zou bij wiskunde nader onderzocht kunnen worden of een andere adaptieve opbouw (per domein) de toets eenvoudiger maakt voor leerlingen. Op dit moment lijkt de vorm van de toets en de opgaven erg moeilijk voor de leerlingen.

Ook de adaptiviteit en rapportages zouden verder ontwikkeld kunnen worden om nog meer in te spelen op de behoeftes van scholen en docenten. De voorbeeldopgaven zouden bijvoorbeeld direct via de leerlingrapportage gekoppeld kunnen zijn. In aanvulling daarop zou een voorbeeld getoond kunnen worden van een lesdoel en een les dat ingaat op de gerapporteerde verbeter- of sterktepunten. Ook zouden alternatieve adaptieve algoritmes ontwikkeld kunnen worden. Bijvoorbeeld, in het ontwikkelde doortoetsalgoritme ligt nadruk op verbeterpunten omdat verwacht werd dat aanpak van deze punten tot hogere prestaties zouden leiden. Het is echter denkbaar dat docenten het wenselijk vinden dat in het doortoetsen gelijke nadruk op de verbeterpunten en sterke punten ligt (Hoijsink & Schouwstra, 2014). Hiervoor zou een alternatief doortoetsalgoritme ontwikkeld moeten worden.

De doorontwikkeling van de adaptiviteit kan ook gericht zijn op het reduceren van de toetsduur. Uit de evaluatie van het CvTE (te verschijnen) bleek dit een duidelijke wens van scholen. Al tijdens de voorstudie is geopperd dat de beheersing van het ene hoofdaspect implicaties kan hebben voor de beheersing van een ander hoofdaspect, bijvoorbeeld als de hoofdaspecten hiërarchisch geordend zijn (Roelofs &

Schouwstra, 2012, p. 109). Het zou bijvoorbeeld zo kunnen zijn dat om structuur in een tekst aan te kunnen brengen de leerling de linguïstische vaardigheden goed moet beheersen. In dat geval is het al heel waarschijnlijk dat een leerling de structuur niet beheerst, indien is gebleken dat die leerling de linguïstische vaardigheden niet beheerst. Dit zou tot een reductie kunnen leiden in het aantal af te nemen opgaven over structuur voor leerlingen die de linguïstische vaardigheden niet beheersen. Hiervoor zou verder wetenschappelijk onderzoek nodig zijn naar de onderlinge samenhang van de aspecten uit de leerlingmodellen.

Een andere manier om het aantal opgaven te verminderen is door verversing van opgaven. Uit onderzoek van Hoijtink en Sies (2013; 2014b) is gebleken dat het onderscheidend vermogen van de opgaven in sterke mate bepalend is voor het aantal benodigde opgaven. Hoe groter het onderscheidend vermogen des te minder opgaven nodig zijn. Wanneer men opgaven verversst kan men ervoor kiezen om de minder onderscheidende opgaven te vervangen. Op die wijze ontstaat na verloop van tijd een opgavebank met een groter onderscheidend vermogen en zullen minder opgaven nodig zijn voor een accurate diagnose.

Men kan er ook aan denken om meer lagen in de adaptieve procedure in te bouwen of zelfs adaptiviteit op opgaveniveau te implementeren. Eerder onderzoek (Hoijtink & Sies, 2014) heeft echter laten zien dat de blokkenprocedure bijna net zo accuraat was als adaptiviteit op opgaveniveau. De winst in tijd bij gelijkblijvende accuraatheid zal dus gering zijn. Natuurlijk kan de toetsduur ook teruggedrongen worden als men minder accurate diagnoses toestaat, maar dat vermindert de bruikbaarheid van de DTT aanzienlijk door de vele misdiagnoses.

Een andere mogelijkheid die niet zozeer de toetsduur maar wel de toetsdruk vermindert, is het spreiden van de toetsing over twee dagen, bijvoorbeeld de doortoetsing zou op de tweede dag plaats kunnen vinden. Men kan ook denken aan een opzet waarbij de docent op grond van de resultaten kiest of doorgetoetst gaat worden en bij wie. Een heldere scheiding tussen het eerste stadium en het doortoetsstadium heeft ook tot voordeel dat duidelijk is wanneer voldoende opgaven gemaakt zijn voor een rapportage van de hoofdaspect-diagnoses. Een dergelijke duidelijke indicatie is belangrijk om te zorgen dat scholen leerlingen niet te vroeg laten ophouden met de toets, nog voordat voldoende opgaven zijn beantwoord om een diagnose te kunnen geven.

Natuurlijk zijn er ook andere mogelijkheden om te zorgen dat duidelijk is wanneer voldoende opgaven beantwoord zijn voor een diagnose van de hoofdaspecten. Via een informatiescherm tijdens de toets zouden leerlingen bijvoorbeeld geïnformeerd kunnen worden op het moment dat ze voldoende opgaven voor een diagnose hebben gemaakt. Desalniettemin blijft het cruciaal dat scholen en docenten goed geïnformeerd worden over de benodigde toetstijd en de mogelijke consequenties (namelijk geen rapportage) wanneer te weinig tijd gegeven wordt voor de DTT.

Een voor de hand liggende groeimogelijkheid is de uitbreiding van de toets. De opgaven leesvaardigheid Engels en leesvaardigheid Nederlands kunnen gepretest worden. Ook zouden opgaven ontwikkeld kunnen worden voor een diagnostische meting van luistervaardigheid Engels en luistervaardigheid Nederlands en eventueel rekenen. Die nieuwe vaardigheden zullen ook om de ontwikkeling van nieuwe opgavetypen vragen en om een uitbreiding van de automatische beoordeling. Andere uitbreiding waaraan scholen mogelijk behoefte zouden kunnen hebben zijn een remediëringstoets om na te gaan of de leerling baat heeft gehad bij remediëring, als die is ingezet, en speciale toetsversies voor leerlingen met een beperking.

We kunnen concluderen dat de DTT een geheel nieuw product is met mooie ontwikkel- en groeimogelijkheden. Het is een innovatieve toets voor alle leerlingen aan het einde van de onderbouw, die gebaseerd is op leerlingmodellen, die digitaal en adaptief afgenomen wordt, met nieuwe opgavetypen, een nieuw onderliggend psychometrisch model met een nieuwe rapportagevorm. Het is een toets die laat zien waar de sterkte en verbeterpunten van leerlingen liggen zodat de docent gerichter actie kan ondernemen voor verbetering en maatwerk, bijvoorbeeld met behulp van de lessenseries die door SLO (2017) ontwikkeld zijn. De scholen reageerden positief op de nieuwe opgavetypen en nieuwe rapportagevorm en ervoeren de DTT als een stimulans voor formatief evalueren (CvTE, te verschijnen).

In juni 2016 is het politieke besluit genomen om aan het einde van de pilotperiode de DTT over te dragen aan marktpartijen. In het kader daarvan zijn er verschillende kennisdelingssessies voor marktpartijen

geweest. Eind mei 2017 hebben twee marktpartijen aangekondigd de DTT te gaan aanbieden voor scholen, en in juni is begonnen met de oplevering van het prototype aan het Ministerie van OCW, dat heeft zorg gedragen voor de openbaarmaking en daarmee overdracht aan de markt.

16.1 Literatuur

College voor Toetsen en Examens (2017). *Slo-materiaal*. Geraadpleegd op december 2017, op <https://www.pilotdtt.nl/slo-materiaal>.

College voor Toetsen en Examens (te verschijnen). *Eindrapport project DTT*. Utrecht: College voor Toetsen en Examens.

Hojtink, H., & S. Schouwstra (2014). Interpretatie van het initiële profiel op deelattribuutniveau. In S. Schouwstra (Red.), *De diagnostische tussentijdse toets: onderzoek 2014* (pp. 41-46). Arnhem: Cito.

Hojtink, H., & Sies, A. (2013). De psychometrie van de diagnostisch tussentijdse toets. In S. Schouwstra (Red.), *De diagnostisch tussentijdse toets: onderzoek 2013* (pp. 13-62). Arnhem: Cito.

Hojtink, H., & Sies, A. (2014). Adaptieve procedure met itemblokken. In S. Schouwstra (Red.), *De diagnostisch tussentijdse toets: onderzoek 2014* (pp. 33-40). Arnhem: Cito.

Hojtink, H., & Sies, A. (2014b). Een reconstructie van de succesansen in de eerste try-out. In S. Schouwstra (Red.), *De diagnostisch tussentijdse toets: onderzoek 2014* (pp. 11-32). Arnhem: Cito.

Roelofs, E., & Schouwstra, S. (Red.). (2012). *Diagnostische tussentijdse toets: Verslag van de voorstudie*. Arnhem: Cito.

17 Bijlagen

17.1 Bijlage Voorbeelden van opgavetypen

Alinea-opgave

Een opgavetype waarbij de leerling een tekst in alinea's onderverdeelt. De alinea'scheidingen worden direct zichtbaar in de tekst. Geschikt om de leerling zelf tekststructuur te laten aanbrengen.

Soort interactie: Klikken (meerkeuze)

DTT Nederlands schrijfvaardigheid havo-vwo voorbeeldopgaven

Vraag 7 van 15



I van 3 alinea'sgrenzen geplaatst.

Niveau: havo

Het einde van de gloeilamp

Ruim honderd jaar lang werden alle huizen en kantoren in Nederland verlicht met behulp van de gloeilamp. De lampen waren niet duur en waren gemakkelijk in gebruik. Ging er een lamp kapot, dan draaide je er gewoon een nieuw peertje in. Aan het tijdperk van de gloeilamp is echter een einde gekomen.

Sinds 1 september 2011 mogen winkels geen houdere gloeilampen van 60 watt of meer inkopen. De handel in matte gloeilampen is vanaf 1 september 2009 al verboden. Hierdoor zijn er steeds minder gloeilampen in de winkel te koop. En sinds 1 september 2012 mogen winkels hun bestaande voorraden nog wel verkopen, maar kunnen ze geen nieuwe gloeilampen meer inkopen. Vanaf dat moment is in de hele Europese Unie de handel in gloeilampen verboden. De gloeilamp is natuurlijk een grote slokop waar het om energie gaat. Maar 8% van de stroom die een gloeilamp krijgt, wordt omgezet in licht; de rest wordt omgezet in warmte. Bij spaarlampen en ledlampen krijg je veel meer licht voor de stroom die je erin stopt. Toch zijn veel mensen ook kritisch over de spaarlamp: daar zit kwik in en de lampen gaan veel minder lang mee dan wordt beloofd. Met al die weggegooid spaarlampen komt er dus veel van dat gevaarlijke, giftige kwik in ons milieu. Dat nadeel had die goede oude gloeilamp dan weer niet...

Je schrijft een nieuwsbericht over het einde van de gloeilamp.

Verdeel de tekst in vier alinea's. Klik de zinnen aan waarmee de nieuwe alinea's moeten beginnen.

1 2 3 4 5 6 7 8 9 10 11 12 ...

Inleveren Volgende >

Categoriseer-opgave

De leerling sleept gegeven elementen (tekst, afbeeldingen) in vakken per categorie. Geschikt om elementen te sorteren, denk bijvoorbeeld aan formeel/informeel bij tekstsoorten.

Soort interactie: Slepen

DTT wiskunde havo-vwo voorbeeldopgaven **Vraag 1 van 12**

Welke grafieken, tabellen en formules horen bij elkaar?

grafiek 1

grafiek 2

grafiek 3

$y = 2x + 3$

$y = -x + 2$

$y = -0,5x + 3$

x	-3	-2	-1	0	1	2	3	4
y	4,5	4	3,5	3	2,5	2	1,5	1

x	-3	-2	-1	0	1	2	3	4
y	-3	-1	1	3	5	7	9	11

x	-3	-2	-1	0	1	2	3	4
y	5	4	3	2	1	0	-1	-2

<https://oefenen.facet.nl/facet/speler/afname/item/fe2cce939667fb078f2830e057b383fbbac2b04a04017a005f2ac641fb7827c>

Inleveren Volgende >

Combinatie-opgave

Een opgavetype waarbij meerdere soorten interacties (bijvoorbeeld een meerkeuzevraag en een invulvraag) op één scherm worden gecombineerd.

Soort interactie: Diversen

Correctie-opgave

De leerling selecteert en verbetert onderdelen van een tekst. Geschikt om grammatica en spelling te toetsen.

Soort interactie: Klikken en invullen

DTT Engels schrijfvaardigheid Voorbeeldopgaven

Vraag 8 van 18



Niveau: VMBO
Aspect: Spelling, interpunctie en grammatica

1 van 1 selecties gemaakt

Our mum likes to ask really  boaring questions.

Klik op het woord dat verkeerd gespeld is en verbeter het.

1

boaring

x

...

2

3

4

5

6

7

8

9

10

11

12

...

Inleveren

Volgende >

Dropdown-opgave

Een opgavetype waarbij de leerling een antwoord kiest uit een keuzelijst. De leerling ziet daarna direct het effect van zijn keuze in de tekst. Dit opgavetype wordt veel gebruikt als woordenschat-vraag waarbij de leerling het meest geschikte woord invoegt in de tekst.

Soort interactie: Klikken (meerkeuze)

DTT Engels schrijfvaardigheid Voorbeeldopgaven
Vraag 15 van 18

Niveau: Havo/Vwo
Aspect: Woordenschat en woordgebruik

Vul het juiste woord in.

I'm really sorry, but I will have to call our date. I've got too much homework.

back
for
off
on

... 7 8 9 10 11 12 13 14 15 16 17 18
Inleveren Volgende >

DTT Nederlands schrijfvaardigheid vmbo voorbeeldopgaven
Vraag 6 van 12

Niveau: KB

Je hebt een bericht geschreven over een opmerkelijke gebeurtenis. In de tekst ontbreekt een aantal woorden.

Kies het juiste woord om een logische tekst te krijgen.

Eeneige tweelingbroers getrouwd met eeneige tweelingzussen

In India zijn afgelopen donderdag twee tweelingbroers getrouwd met twee tweelingzussen. De twee broers waren hier blij mee, [] ze al van jongs af aan met tweelingzussen wilden trouwen. Ze hebben vijf jaar gezocht naar de perfecte vrouwelijke tweeling en hebben oproepen gedaan in kranten en op websites.

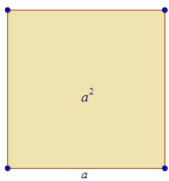
Volgens een van de bruidegoms was het een perfecte dag, [] hij het jammer vond dat er in het publiek niet veel tweelingen zaten. [] is er nog meer goed nieuws te melden: de vrouwen zijn alle [] een tweeling.

Bovendien
Daardoor
Daarom

1 2 3 4 5 6 7 8 9 10 11 12
Inleveren Volgende >

DTT wiskunde havo-vwo voorbeeldopgaven
Vraag 9 van 12

Het vierkant heeft zijde a en oppervlakte a^2 .



Kies bij de gaten in de tekst steeds de juiste optie.

Als alle zijden van het vierkant twee keer zo groot worden, wordt de oppervlakte [] keer zo groot.

Om de oppervlakte van het vierkant te verdubbelen, moeten alle zijden vermenigvuldigd worden met [].

Als de oppervlakte van het vierkant drie keer zo groot is geworden, dan hebben de zijden een lengte van [].

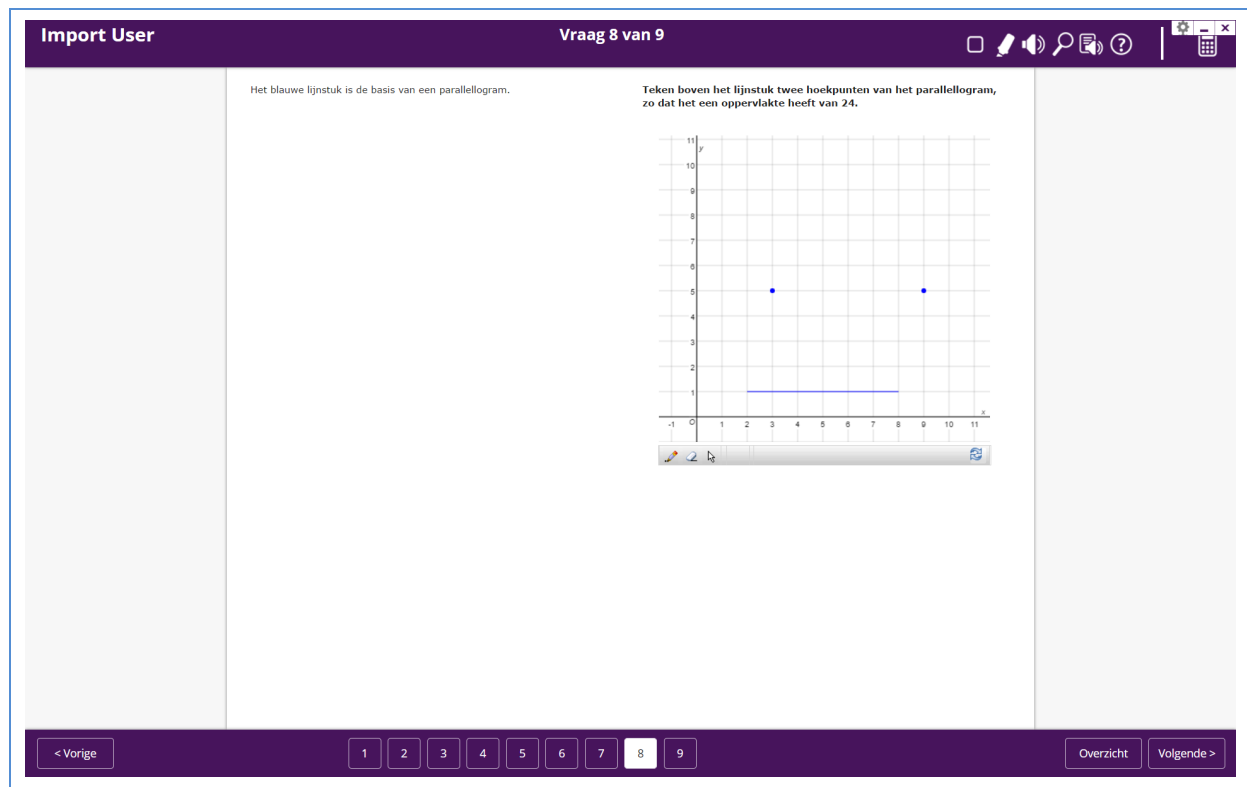
$1\frac{1}{2} \cdot a$
 $\sqrt{3} \cdot a$
 $3a$

1 2 3 4 5 6 7 8 9 10 11 12
Inleveren Volgende >

DWO-opgave

De leerling kan zelf grafiekpunten of grafieken tekenen als antwoord.

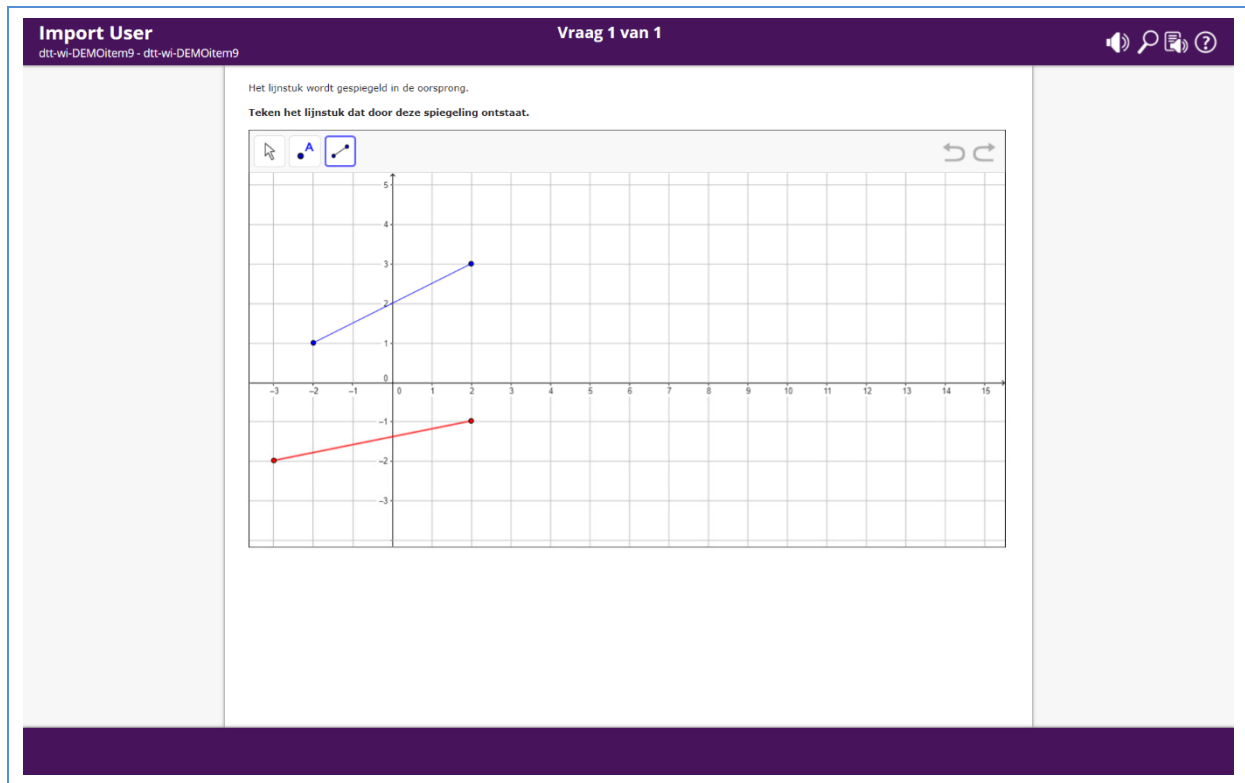
Soort interactie: Diversen



Geogebra-opgave

De leerling kan diverse wiskundige interacties uitvoeren als antwoord. Bijvoorbeeld het correcte meetkundige figuur tekenen of aanpassen.

Soort interactie: Diversen



Hotspot-opgave

De leerling klikt het juiste antwoord aan op een afbeelding.

Soort interactie: Klikken

DTT Nederlands schrijfvaardigheid havo-vwo voorbeeldopgaven **Vraag 3 van 15**

Niveau: havo/vwo

Voor Nederlands moet je binnenkort een informatief verslag over dode talen inleveren. Met dit verslag wil je de lezers informeren over dode talen in de wereld. Je hebt gebrainstormd over deelonderwerpen voor jouw tekst. Je hebt heel veel ideeën, maar niet alles is geschikt.

Klik de elementen aan die in jouw informatieve verslag passen.

1 2 **3** 4 5 6 7 8 9 10 11 12 ...

Inleveren Volgende >

DTT wiskunde vmbo voorbeeldopgaven **Vraag 10 van 11**

Klik op alle breuken die kleiner zijn dan $\frac{5}{7}$.

1 2 3 4 5 6 7 8 9 10 11

Inleveren Volgende >

Kort open-opgave

De leerling typt zelf een antwoord in. Dit opgavetype wordt bij talen gebruikt om spelling te toetsen. Bij wiskunde-opgaven kunnen de leerlingen input geven in de vorm van getallen en formules, via een speciale formule-editor.

Soort interactie: Invullen

Import User

Vraag 9 van 18

Niveau: VMBO
Aspect: Spelling, interpunctie en grammatica

Luister naar het fragment en spel het woord op de juiste manier.

00:00 / 00:02

In that dress she looks so

... 3 4 5 6 7 8 9 10 11 12 13 ...

Inleveren Volgende >

DTT wiskunde havo-vwo voorbeeldopgaven

Vraag 3 van 12

Suzanne wil een stacaravan huren. Camping 'Rusthof' rekent een vast bedrag van € 45,- en een prijs per nacht van € 3,70.

Stel een formule op voor de huurprijs h in euro van een stacaravan op camping Rusthof bij een verblijf van n nachten.

huurprijs $h =$

$\sqrt{\frac{x}{y}}$
 $+$
 $-$
 $\leq$
 $\approx$

$\frac{x}{y}$
 $\times$
 $\div$
 $>$
 $\neq$

$\frac{1}{x}$
 x^0
 $\sqrt[n]{x}$
 $\sqrt{x}$
 $()$
 π

1 2 3 4 5 6 7 8 9 10 11 12

Inleveren Volgende >

DTT wiskunde vmbo voorbeeldopgaven

Vraag 9 van 11

Marlise geeft haar hond elke dag dezelfde hoeveelheid brokken. Ze koopt een zak brokken van 500 gram. Na 5 dagen heeft ze nog 250 gram en na 10 dagen is de zak leeg.



Marlise maakt bij deze gegevens een formule voor de hoeveelheid brokken in de zak.

Welke getallen horen in de lege vakken?

Inhoud zak in gram = x aantal dagen +

▲ Rekenmachine

×

sin

cos

tan

a^b/c

$\times 10^y$

$\sqrt{}$

x^2

$\wedge$

x^{-1}

π

INV

←

→

(

)

AC

7

8

9

DEL

C

4

5

6

×

÷

1

2

3

+

-

0

,

(-)

ANS

=

1

2

3

4

5

6

7

8

9

10

11

Inleveren

Volgende >

Markeer-opgave

De leerling selecteert actieve onderdelen van een tekst. Geschikt om antwoorden te zoeken in de tekst, bijvoorbeeld een citaat of fout gespelde woorden.

Soort interactie: Klikken

DTT Engels schrijfvaardigheid Voorbeeldopgaven

Vraag 4 van 18



Niveau: VMBO
Aspect: Coherentie

Klik op het zinsdeel waar 'there' naar verwijst.

I try to go to **the zoo** whenever I have a day off. The animals **there** always make me feel better about myself.



1 2 3 4 5 6 7 8 9 10 11 12 ...

Inleveren Volgende >

DTT Nederlands schrijfvaardigheid vmbo voorbeeldopgaven

Vraag 7 van 12



Niveau: GT

De sportdag

Vorige week hadden we sportdag op school. Die was erg geslaagd. Je kon tijdens de sportdag allerlei sporten beoefenen zoals handbal, honkbal, voetbal, netbal en minivoetbal. Je zelfs kon een geweldige wandeling met een quiz maken!
Voor de brugklassers was het de eerste sportdag. **Ging het voor deze leerlingen dus vooral om samenwerking met de klas en de docenten.**
Hierdoor leerden de docenten de leerlingen ook beter kennen.
Aan het einde van de dag was er ook nog een leuke afsluiter voor alle leerlingen. Het was voor jouw vast ook een sportieve en fijne dag!

Je schrijft een stukje over de sportdag voor de website van de school. Drie zinnen zijn niet goed geformuleerd.
Klik de drie zinnen aan die niet goed zijn geformuleerd.

1 2 3 4 5 6 7 8 9 10 11 12

Inleveren Volgende >

Matrix-opgave

De leerling kiest uit (meestal) twee mogelijkheden, zoals 'wel' of 'niet'. Geschikt voor het bevragen van details.

Soort interactie: Klikken (meerkeuze)

DTT Engels schrijfvaardigheid Voorbeeldopgaven
Vraag 6 van 18

Niveau: VMBO
Aspect: Woordenschat en woordgebruik

Welke vier woorden passen er bij 'hospital'?

Geef bij elk woord aan of dit wel of niet bij 'hospital' past.

	wel	niet
court	☐	☐
emergency room	☐	☐
hostility	☐	☐
knitting	☐	☐
prescription	☐	☐
surgeon	☐	☐
wound	☐	☐

1 2 3 4 5 6 7 8 9 10 11 12 ...
Inleveren
Volgende >

DTT Nederlands schrijfvaardigheid vmbo voorbeeldopgaven
Vraag 3 van 12

Niveau: BB

Op het plein rondom jouw school staan bijna geen prullenbakken. Daarom gooien veel leerlingen hun afval gewoon op de grond. Je wilt de directeur overtuigen snel meer prullenbakken te plaatsen. Je besluit daarom een mail te sturen naar de directeur.

Geef van elke zin aan of deze wel of niet geschikt is om de directeur te overtuigen.

	wel	niet
Als er meer prullenbakken staan, gooien leerlingen hun rommel minder snel op de grond.	☐	☐
Een rommelig schoolplein is geen goede reclame voor de school.	☐	☐
Ik heb op internet een bedrijf gevonden waar je goedkoop prullenbakken kunt bestellen.	☐	☐
Ik vind dit echt niet normaal.	☐	☐
Zelf gooi ik mijn kauwgum en snoeppapiertjes nooit op de grond als ik geen prullenbak zie.	☐	☐

1 2 3 4 5 6 7 8 9 10 11 12
Inleveren
Volgende >

DTT wiskunde vmbo voorbeeldopgaven

Vraag 5 van 11



Bij de formules hoort een grafiek.

Geef aan of de bijbehorende grafiek daalt, stijgt of constant is.

	daalt	stijgt	constant
$temperatuur = -2 + 3 \times tijd$	☐	☐	☐
$temperatuur = 2 - 3 \times tijd$	☐	☐	☐
$temperatuur = 3 \times tijd$	☐	☐	☐

1234567891011

InleverenVolgende >

Meerkeuze-opgave

De leerling kiest het juiste antwoord uit een lijst van alternatieven (teksten, afbeeldingen, filmpjes, geluidsbestanden) of uit actieve gebieden in een afbeelding.

Soort interactie: Klikken (meerkeuze)

DTT Engels schrijfvaardigheid Voorbeeldopgaven

Vraag 16 van 18





Niveau: Havo/Vwo
Aspect: Woordenschat en woordgebruik

Why did Mr Ross have that astorished look on his face?

Welk woord betekent ongeveer hetzelfde als 'astorished' in de zin?

- ☐ arrogant
- ☐ furious
- ☐ scared
- ☐ surprised

... 7 8 9 10 11 12 13 14 15 16 17 18

Inleveren Volgende >

DTT Nederlands schrijfvaardigheid havo-vwo voorbeeldopgaven

Vraag 5 van 15





Niveau: havo/vwo
Nieuw: Expeditie Ribbelroute

Het muZIEum is een bijzonder museum: het is namelijk helemaal gericht op blindheid. Alle bezoekers ervaren het museum alsof ze blind zijn. Daarom organiseren we verschillende donkerbelevingen.

In het muZIEum zelf zijn twee verschillende belevingen. Bij de donkerbeleving *Op pad in de stad* word je rondgeleid in het dagelijks leven van iemand die blind is. Bij de donkerbeleving *Op vakantie* word je ondergedompeld in vakantiesferen en ervaar je typische vakantieactiviteiten zonder zicht.

Maar we hebben nu ook een buitenbeleving! Bij deze nieuwe *Expeditie Ribbelroute* word je begeleid door een gids met een visuele beperking. De expeditie is natuurlijk geen gewone stadswandeling, want met behulp van een virtueel reality-bril word je zelf tijdelijk slechtziend. Zo ervaar je zelf hoe het is om niet goed te zien.

De expeditie is spannend en is geschikt voor iedereen vanaf 13 jaar. De donkerbelevingen in het museum zijn toegankelijk voor iedereen vanaf 8 jaar. Als je jonger bent, kan de rondleiding nog best verwarrend zijn, daarom hebben we liever dat je nog even wacht tot je 8 jaar bent.

Een medewerker schrijft een krantenbericht over activiteiten die het museum organiseert. Het bericht is bedoeld om mensen aan te sporen naar het museum te komen. De tekst heeft nog geen inleiding.

Welke inleiding past het best bij de tekst?

- ☐ Het muZIEum waarin ik werk geeft jou een onvergetelijke ervaring. En ik kan dat weten, want al veel bezoekers hebben mij dat verteld.
- ☐ In Nederland zijn ruim 320.000 mensen slechtziend of blind. Je spreekt volgens onderzoekers van een visuele beperking als iemand niet of slechts gedeeltelijk kan zien.
- ☐ Stel je eens voor dat je geen auto's meer ziet of dat je in een groot gebouw de uitgang niet ziet. Welke invloed heeft dat op jou? In het muZIEum ga je op ontdekkingsstocht in het donker!
- ☐ Veel mensen zouden zich beter moeten inleven in de wereld van blinden en slechtzienden. Zij verdienen namelijk veel meer respect van ons!

1 2 3 4 5 6 7 8 9 10 11 12 ...

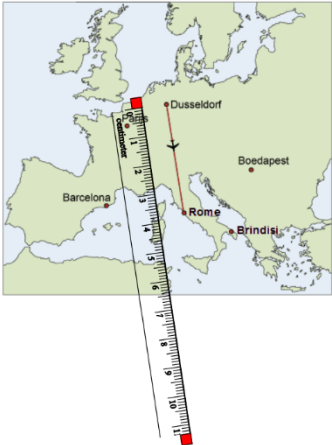
Inleveren Volgende >

DTT wiskunde vmbo voorbeeldopgaven

Vraag 8 van 11

Welke plaats heeft dezelfde vliegafstand tot Düsseldorf als Rome?
Gebruik de liniaal.

☐ Boedapest
☐ Parijs
☐ Barcelona
☐ Brindisi



1 2 3 4 5 6 7 8 9 10 11

Inleveren Volgende >

Multiplerespons opgave

De leerling kiest één of meer antwoorden uit een lijst van alternatieven (teksten, afbeeldingen, filmpjes, geluidsbestanden). Deze vorm komt veel voor met opgaven over internetformulieren en is in feite ook een vorm van een sorteervraag.

Soort interactie: Klikken, multiple respons

DTT Nederlands schrijfvaardigheid vmbo voorbeeldopgaven
Vraag 2 van 12

Niveau: KB
Aan: De Glasunie
Schoolstraat 12
3493 AC Schiebroek

Zeddam, 15 april 2016

Geachte heer, mevrouw,

.....

Met vriendelijke groet,
Feline Hendriksen

Feline heeft een mooie nieuwe vissenkomp gekocht. Er zit een sticker op de doos waarop staat: *geen garantie op glasbreuk*. Als Feline de kom thuis uitpakt, ziet ze dat er een grote barst in het glas zit. Feline besluit een brief naar de fabrikant te schrijven. Ze hoopt dat ze haar willen helpen.

Welke drie zinnen passen in de brief?

- ☐ Als u me geen nieuwe kom wilt geven, klaag ik u aan.
- ☐ Ik begrijp dat er geen garantie zit op glasbreuk, maar de barst zat er al in voordat ik de kom kon gebruiken.
- ☐ Ik hoop dat u me tegemoet kunt komen.
- ☐ Ik vind het echt schandalig dat uw bedrijf zulke vissenkomen verkoopt.
- ☐ Tot nu toe heb ik hele goede ervaringen met vissenkomen van uw merk.

1 2 3 4 5 6 7 8 9 10 11 12
Inleveren Volgende >

Sleep-opgave

De leerling sleept een antwoord in de vorm van tekst. Geschikt om te matchen of te sorteren/categoriseren.

Soort interactie: Slepen

DTT Engels schrijfvaardigheid Voorbeeldopgaven **Vraag 10 van 18**

Drag and drop

Je bestelt online treintickets voor een uitstapje naar London voor een gezin van vijf. Je vertrekt vanaf Liverpool.

Sleep de onderdelen van het formulier naar de juiste plek.

Let op! Er blijft één optie over.

03-14-17 16:00 26HJKie44 Liverpool London Passengers
Railcards Return Tickets

Niveau: VMBO
Aspect: Woordenschat en woordgebruik

Welcome. Please use the form below to plan your journey.

Leaving from _____
Destination _____
Date _____
Time _____
Journey type _____
_____ 2 Adults / 3 Children
_____ Family and Friends Pass
Promotional code _____

Check fares and availability >

... 4 5 6 7 8 9 10 11 12 13 14 ... Inleveren Volgende >

DTT Nederlands schrijfvaardigheid havo-vwo voorbeeldopgaven **Vraag 9 van 15**

Niveau: havo/vwo

Voor een schoolopdracht ga je een betoog schrijven tegen roken. Voordat je het betoog schrijft, maak je een schema waarin je alles op een rij zet.

Sleep de zinnen naar de juiste plek in het schema.

Let op! Er staan meer zinnen dan je moet gebruiken.

Standpunt: _____

Argument: _____

Argument: _____

Tegenargument: _____

Allerlei rookwaar, zoals sigaretten en shag, is tegenwoordig peperduur.
Echte rokers vinden stoppen met roken wel heel erg moeilijk.
Je kunt maar beter stoppen met roken.
Roken leidt tot vroegtijdig overlijden.
Veel kinderen hebben vroeger thuis geleerd om te roken.
Vroeger vond men het heel gewoon als je op je 14e begon met roken.

... 3 4 5 6 7 8 9 10 11 12 13 ... Inleveren Volgende >

DTT wiskunde vmbo voorbeeldopgaven

Vraag 11 van 11



Sleep de passende maat naar het plaatje.



2 deciliter

2 hectoliter

2 liter

2 ml

20 hectoliter

20 liter

20 ml

1

2

3

4

5

6

7

8

9

10

11

Inleveren

Volgende >

Sleep-opgave afbeelding

De leerling sleept een antwoord. Geschikt om plaatjes aan woorden te koppelen.

Soort interactie: Slepen

DTT Engels schrijfvaardigheid Voorbeeldopgaven
Vraag 5 van 18

Niveau: VMBO
Aspect: Woordenschat en woordgebruik

Je vakantieviend gaat een jaar geschiedenis studeren in Korea. Je stuurt hem een berichtje voor hij vertrekt.

Wat zet je in het berichtje?

Sleep de juiste zin naar het berichtje.



Congratulations on your new house

Enjoy your stay

Get well soon

Happy retirement

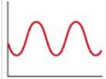
Wishing you joy on your special day

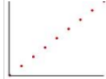
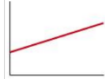
1 2 3 4 5 6 7 8 9 10 11 12 ...
Inleveren Volgende >

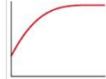
DTT wiskunde vmbo voorbeeldopgaven
Vraag 2 van 11

Sleep de juiste grafiek bij elk verhaal.

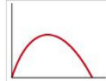
VERHAAL	GRAFIEK	VERHAAL	GRAFIEK
1. Een glazenwasser vraagt € 25,- voornijkosten en € 20,- voor elk gewerkt uur.		2. In de vijver van Bart zit kroos. Dit is een soort onkruid dat snel groeit. Elke dag verdubbelt de hoeveelheid zich.	
3. Het aquarium van Bart heeft een lek. Elke minuut stroomt er een halve liter water uit.		4. Bart heeft een maand lang elke dag de gemiddelde temperatuur bijgehouden.	









1 2 3 4 5 6 7 8 9 10 11
Inleveren Volgende >

Volgorde-opgave

De leerling sleept tekstdelen of getallen in een nieuwe volgorde.

Soort interactie: Slepen

DTT Engels schrijfvaardigheid Voorbeeldopgaven **Vraag 18 van 18**   

Niveau: Havo/Vwo
Aspect: Coherentie

Drag and drop
Je schrijft een recensie van de film 'Finding Nemo'.

Zet de alinea's in de juiste volgorde.

And they're right: 'Finding Nemo' is a kid's movie that doesn't stink.

'Finding Nemo' is the new treasure from Pixar Animation Studios. The expectations of kids have been stoked by Disney's marketing campaign, which has made sure that the little ones once again had no trouble finding the many pricey 'Finding Nemo' trinkets and edibles.

'Nemo' falls a bit short there. With more emphasis on action scenes - it does not stint on spectacle - it's not as funny, not as sharply written, not as heartfelt as the studio's other movies.

Parents, meanwhile? They're most probably expecting that this is going to be a children's movie they can actually sit through. After all, it's from Pixar, those creative and witty computer-animation geniuses who made the flawless 'Toy Story' and 'Monsters, Inc.'

The news isn't all great, though; this is Pixar on an off day. It looks great, of course: All the Pixar movies do. But the genius of those other movies lay in their actual stories.

... 7 8 9 10 11 12 13 14 15 16 17 **18** Inleveren Volgende >

DTT wiskunde havo-vwo voorbeeldopgaven **Vraag 12 van 12**   

Zet de breuken op volgorde van klein naar groot.

$\frac{1}{3}$

$\frac{1}{2}$

$\frac{1}{4}$

1 2 3 4 5 **6** 7 8 9 10 11 **12** Inleveren Volgende >

17.2 Bijlage accuratesse, grenswaardes en priors bij de opgeleverde blokindelingen

17.2.1 Resultaten blokindeling schrijfvaardigheid Nederlands

Tabel 17-1. Percentages correcte diagnoses schrijfvaardigheid Nederlands, aantal opgaven en responsen in de simulaties in havo/vwo (exclusief de opgaven in de doortoetsblokken)

School- soort/ Leerweg	Hoofdaspect						Deelaspecten				
		Onder niveau	Op niveau	Boven niveau	Aantal opgaven	Aantal responsen		Onder niveau	Op niveau	Boven niveau	Aantal responsen
havo	H1	0,816	0,673	0,797	42	126	D11	0,767	0,604	0,637	64
							D12	0,616	0,624	0,764	37
							D13	0,816	0,673	0,797	25
	H2	0,805	0,767	0,727	67	160	D21	0,595	0,605	0,593	29
							D22	0,855	0,738	0,705	35
							D23	0,660	0,656	0,751	32
							D24	0,805	0,767	0,727	64
	H3	0,921	0,928	0,933	43	174	D31	0,799	0,653	0,603	60
							D32	0,682	0,611	0,634	39
							D33	0,853	0,841	0,927	75
	H4	0,950	0,950	0,950	36	204	D41	0,799	0,837	0,883	82
							D42	0,834	0,731	0,795	77
							D43	0,777	0,713	0,774	45
vwo	H1	0,964	0,829	0,719	50	106	D11	0,590	0,571	0,510	42
							D12	0,621	0,623	0,691	25
							D13	0,964	0,829	0,719	39
	H2	0,766	0,744	0,714	67	164	D21	0,672	0,687	0,667	43
							D22	0,879	0,775	0,710	56
							D23	0,766	0,744	0,714	40
							D24	0,564	0,596	0,583	25
	H3	0,916	0,916	0,916	51	168	D31	0,772	0,742	0,726	79
							D32	0,728	0,757	0,817	40
							D33	0,720	0,714	0,707	49
	H4	0,978	0,976	0,976	34	176	D41	0,946	0,877	0,910	82
							D42	0,778	0,739	0,834	59
							D43	0,835	0,806	0,875	35

Tabel 17-2. Percentages correcte diagnoses schrijfvaardigheid Nederlands, aantal opgaven en responsen in de simulaties in vmbo (exclusief de opgaven in de doortoetsblokken)

School- soort/ Leerweg	Hoofdaspect					Deelaspecten					
		Onder niveau	Op niveau	Boven niveau	Aantal opgaven	Aantal responsen		Onder niveau	Op niveau	Boven niveau	Aantal responsen
vmbo- bb	H1	0,950	0,949	0,949	43	160	D11	0,809	0,782	0,824	60
							D12	0,818	0,721	0,721	42
							D13	0,767	0,826	0,866	58
	H2	0,933	0,873	0,848	51	129	D21	0,844	0,720	0,660	44
							D22	0,615	0,604	0,666	19
							D23	0,933	0,873	0,848	66
	H3	0,955	0,955	0,954	47	157	D31	0,785	0,775	0,847	29
							D32	0,736	0,834	0,855	84
							D33	0,865	0,769	0,759	44
	H4	0,949	0,944	0,944	40	213	D41	0,878	0,712	0,702	65
							D42	0,738	0,774	0,831	91
							D43	0,921	0,851	0,863	57
vmbo- kb	H1	0,936	0,935	0,935	44	190	D11	0,726	0,797	0,853	59
							D12	0,861	0,712	0,720	44
							D13	0,853	0,783	0,814	87
	H2	0,906	0,906	0,906	56	118	D21	0,651	0,662	0,732	28
							D22	0,607	0,676	0,627	21
							D23	0,884	0,804	0,851	69
	H3	0,970	0,968	0,968	50	192	D31	0,725	0,811	0,898	62
							D32	0,830	0,803	0,882	74
							D33	0,887	0,748	0,800	56
	H4	0,962	0,957	0,957	37	212	D41	0,766	0,722	0,734	65
							D42	0,856	0,836	0,867	101
							D43	0,905	0,822	0,869	46
vmbo-gt	H1	0,890	0,890	0,890	43	170	D11	0,807	0,737	0,772	56
							D12	0,716	0,699	0,656	59
							D13	0,818	0,761	0,759	55
	H2	0,952	0,945	0,947	64	127	D21	0,747	0,660	0,772	27
							D22	0,602	0,609	0,699	24
							D23	0,966	0,883	0,885	76
	H3	0,947	0,945	0,945	43	163	D31	0,827	0,781	0,815	57
							D32	0,738	0,808	0,857	61
							D33	0,846	0,705	0,786	45
	H4	0,859	0,789	0,736	41	212	D41	0,710	0,701	0,717	66
							D42	0,777	0,790	0,861	92
							D43	0,859	0,789	0,736	54

Tabel 17-3. Uitstapkans schrijfvaardigheid Nederlands in havo/vwo

School- soort/ Leerweg	Hoofdaspecten			Deelaspecten				
		Aantal responsen	Aantal responsen	Aantal responsen		Onder niveau	Op niveau	Boven niveau
havo	H1	0,010	0,947	0,766	D11	0,605	0,953	0,990
					D12	0,644	0,990	0,867
					D13	0,567	0,990	0,912
	H2	0,010	0,841	0,010	D21	0,796	0,877	0,920
					D22	0,640	0,953	0,902
					D23	0,727	0,990	0,802
	H3	0,620	0,864	0,010	D24	0,666	0,948	0,884
					D31	0,922	0,971	0,590
					D32	0,944	0,980	0,010
	H4	0,854	0,956	0,737	D33	0,860	0,922	0,010
					D41	0,318	0,405	0,277
					D42	0,892	0,694	0,990
vwo	H1	0,010	0,921	0,778	D43	0,878	0,990	0,889
					D11	0,010	0,947	0,990
					D12	0,847	0,947	0,918
	H2	0,010	0,882	0,691	D13	0,010	0,973	0,990
					D21	0,490	0,978	0,885
					D22	0,010	0,967	0,980
	H3	0,741	0,922	0,827	D23	0,010	0,968	0,902
					D24	0,620	0,942	0,959
					D31	0,279	0,419	0,302
	H4	0,010	0,955	0,694	D32	0,990	0,634	0,851
					D33	0,836	0,667	0,990
					D41	0,240	0,416	0,344
					D42	0,731	0,990	0,903
					D43	0,647	0,960	0,831

Tabel 17-4. Uitstapkansens schrijfvaardigheid Nederlands in vmbo

School- soort/ Leerweg	Hoofdaspecten				Deelaspecten			
		Onder niveau	Op niveau	Boven niveau		Onder niveau	Op niveau	Boven niveau
vmbo-bb	H1	0,010	0,978	0,872	D11	0,207	0,434	0,360
					D12	0,990	0,990	0,990
					D13	0,620	0,846	0,760
	H2	0,010	0,840	0,590	D21	0,010	0,971	0,990
					D22	0,990	0,913	0,903
					D23	0,010	0,968	0,787
	H3	0,010	0,960	0,657	D31	0,228	0,425	0,347
					D32	0,726	0,968	0,908
					D33	0,413	0,895	0,795
	H4	0,010	0,997	0,784	D41	0,170	0,430	0,400
					D42	0,990	0,990	0,917
					D43	0,010	0,994	0,864
vmbo-kb	H1	0,010	0,994	0,874	D11	0,176	0,442	0,382
					D12	0,010	0,990	0,844
					D13	0,010	0,971	0,990
	H2	0,010	0,982	0,909	D21	0,200	0,427	0,373
					D22	0,990	0,990	0,784
					D23	0,010	0,982	0,990
	H3	0,587	0,977	0,762	D31	0,226	0,402	0,372
					D32	0,746	0,905	0,878
					D33	0,990	0,618	0,571
	H4	0,010	0,984	0,809	D41	0,181	0,421	0,398
					D42	0,010	0,934	0,929
					D43	0,010	0,989	0,871
vmbo-gt	H1	0,010	0,991	0,910	D11	0,172	0,453	0,375
					D12	0,990	0,990	0,990
					D13	0,010	0,990	0,990
	H2	0,010	0,765	0,506	D21	0,010	0,953	0,990
					D22	0,010	0,893	0,851
					D23	0,010	0,913	0,801
	H3	0,010	0,983	0,869	D31	0,191	0,417	0,392
					D32	0,508	0,752	0,757
					D33	0,010	0,948	0,932
	H4	0,010	0,795	0,544	D41	0,552	0,986	0,914
					D42	0,010	0,990	0,809
					D43	0,010	0,935	0,985

Tabel 17-5. Prior modelkansen schrijfvaardigheid Nederlands

School- soort/ Leerweg	Hoofdaspecten				Deelaspecten		
		Onder niveau	Op niveau	Boven niveau	Onder niveau	Op niveau	Boven niveau
havo	H1	0,136	0,482	0,382	0,280	0,380	0,330
	H2	0,172	0,498	0,330	0,340	0,400	0,260
	H3	0,511	0,449	0,040	0,422	0,420	0,220
	H4	0,302	0,477	0,221	0,318	0,405	0,277
vwo	H1	0,046	0,552	0,402	0,400	0,443	0,230
	H2	0,069	0,512	0,419	0,290	0,430	0,270
	H3	0,224	0,505	0,271	0,279	0,419	0,302
	H4	0,147	0,498	0,355	0,240	0,416	0,344
vmbo-bb	H1	0,080	0,534	0,386	0,207	0,434	0,360
	H2	0,066	0,521	0,413	0,300	0,420	0,280
	H3	0,123	0,516	0,361	0,228	0,425	0,347
	H4	0,007	0,527	0,466	0,170	0,430	0,400
vmbo-kb	H1	0,018	0,551	0,430	0,176	0,442	0,382
	H2	0,067	0,520	0,414	0,200	0,427	0,373
	H3	0,119	0,471	0,410	0,226	0,402	0,372
	H4	0,029	0,508	0,463	0,181	0,421	0,398
vmbo-gt	H1	0,011	0,572	0,417	0,172	0,453	0,375
	H2	0,054	0,547	0,399	0,330	0,500	0,366
	H3	0,049	0,500	0,451	0,191	0,417	0,392
	H4	0,084	0,558	0,358	0,250	0,420	0,330

17.2.2 Resultaten blokindeling schrijfvaardigheid Engels

Tabel 17-6. Proporties correcte diagnoses schrijfvaardigheid Engels, aantal opgaven en responsen in de simulaties in havo/vwo (exclusief de opgaven in de doortoetsblokken)

School- soort/ Leerweg	Hoofdaspecten						Deelaspecten				
		Onder niveau	Op niveau	Boven niveau	Aantal opgaven	Aantal responsen		Onder niveau	Op niveau	Boven niveau	Aantal responsen
havo	H1	0,932	0,932	0,932	48	149	D11	0,822	0,753	0,787	55
							D12	0,841	0,852	0,907	94
	H2	0,913	0,913	0,913	54	84	D21	0,827	0,780	0,864	42
							D22	0,841	0,761	0,802	42
	H3	0,946	0,946	0,947	78	130	D31	0,896	0,863	0,883	75
							D32	0,821	0,811	0,851	55
	H4	0,933	0,933	0,933	73	109	D41	0,880	0,844	0,877	58
							D42	0,822	0,777	0,819	51
vwo	H1	0,882	0,882	0,882	39	100	D11	0,627	0,641	0,629	54
							D12	0,855	0,830	0,870	46
	H2	0,925	0,925	0,925	51	92	D21	0,812	0,728	0,745	37
							D22	0,847	0,849	0,909	55
	H3	0,917	0,917	0,917	72	114	D31	0,812	0,765	0,800	57
							D32	0,827	0,808	0,872	57
	H4	0,876	0,876	0,876	52	82	D41	0,835	0,783	0,836	44
							D42	0,738	0,657	0,707	38

Tabel 17-7. Proporties correcte diagnoses schrijfvaardigheid Engels, aantal opgaven en responsen in de simulaties in vmbo (exclusief de opgaven in de doortoetsblokken)

School- soort/ Leerweg	Hoofdaspecten						Deelaspecten				
		Onder niveau	Op niveau	Boven niveau	Aantal opgaven	Aantal responsen		Onder niveau	Op niveau	Boven niveau	Aantal responsen
vmbo- bb	H1	0,903	0,903	0,903	51	107	D11	0,744	0,710	0,725	54
							D12	0,863	0,820	0,861	53
	H2	0,941	0,941	0,941	61	80	D21	0,873	0,867	0,891	42
							D22	0,837	0,799	0,809	38
	H3	0,964	0,965	0,966	58	129	D31	0,923	0,906	0,945	80
							D32	0,813	0,838	0,864	49
	H4	0,929	0,928	0,928	65	70	D41	0,881	0,828	0,874	33
							D42	0,818	0,771	0,790	37
vmbo- kb	H1	0,890	0,890	0,890	52	124	D11	0,796	0,756	0,835	64
							D12	0,811	0,747	0,777	60
	H2	0,954	0,954	0,954	65	99	D21	0,842	0,848	0,891	45
							D22	0,898	0,852	0,883	54
	H3	0,971	0,971	0,971	68	147	D31	0,942	0,897	0,929	82
							D32	0,839	0,841	0,911	65
	H4	0,960	0,960	0,960	70	109	D41	0,828	0,863	0,944	60
							D42	0,914	0,843	0,824	49
vmbo-gt	H1	0,956	0,956	0,956	43	134	D11	0,906	0,861	0,868	88
							D12	0,840	0,829	0,904	46
	H2	0,961	0,961	0,962	73	96	D21	0,888	0,851	0,873	40
							D22	0,890	0,874	0,912	56
	H3	0,960	0,959	0,960	60	131	D31	0,927	0,890	0,922	81
							D32	0,806	0,807	0,879	50
	H4	0,936	0,936	0,936	70	99	D41	0,845	0,809	0,863	48
							D42	0,850	0,838	0,866	51

Tabel 17-8. Uitstapkansens schrijfvaardigheid Engels in havo/vwo

School- soort/ Leerweg	Hoofdaspecten				Deelaspecten			
		Onder niveau	Op niveau	Boven niveau		Onder niveau	Op niveau	Boven niveau
havo	H1	0,599	0,972	0,813	D11	0,691	0,802	0,811
					D12	0,792	0,920	0,911
	H2	0,714	0,957	0,863	D21	0,823	0,945	0,718
					D22	0,827	0,990	0,698
	H3	0,763	0,948	0,811	D31	0,855	0,855	0,860
					D32	0,875	0,770	0,615
	H4	0,813	0,961	0,591	D41	0,868	0,922	0,831
					D42	0,807	0,897	0,813
vwo	H1	0,748	0,990	0,919	D11	0,752	0,990	0,708
					D12	0,990	0,605	0,587
	H2	0,825	0,976	0,757	D21	0,867	0,753	0,649
					D22	0,880	0,881	0,783
	H3	0,767	0,961	0,874	D31	0,921	0,990	0,607
					D32	0,850	0,844	0,829
	H4	0,743	0,876	0,894	D41	0,750	0,781	0,906
					D42	0,792	0,990	0,850

Tabel 17-9. Uitstapkansen schrijfvaardigheid Engels in vmbo

School- soort/ Leerweg	Hoofdaspecten				Deelaspecten			
		Onder niveau	Op niveau	Boven niveau		Onder niveau	Op niveau	Boven niveau
vmbo-bb	H1	0,761	0,937	0,812	D11	0,620	0,990	0,831
					D12	0,685	0,803	0,848
	H2	0,736	0,970	0,796	D21	0,756	0,929	0,888
					D22	0,796	0,684	0,990
	H3	0,694	0,957	0,010	D31	0,770	0,967	0,564
					D32	0,872	0,990	0,481
	H4	0,774	0,894	0,742	D41	0,807	0,814	0,820
					D42	0,557	0,990	0,880
vmbo-kb	H1	0,806	0,882	0,864	D11	0,813	0,990	0,901
					D12	0,990	0,990	0,906
	H2	0,726	0,966	0,778	D21	0,850	0,854	0,831
					D22	0,803	0,990	0,764
	H3	0,010	0,922	0,651	D31	0,010	0,936	0,844
					D32	0,895	0,876	0,862
	H4	0,698	0,958	0,646	D41	0,896	0,990	0,519
					D42	0,753	0,954	0,692
vmbo-gt	H1	0,721	0,965	0,617	D11	0,878	0,990	0,990
					D12	0,795	0,922	0,659
	H2	0,792	0,965	0,666	D21	0,896	0,755	0,990
					D22	0,885	0,990	0,782
	H3	0,743	0,963	0,739	D31	0,794	0,922	0,753
					D32	0,854	0,907	0,839
	H4	0,784	0,966	0,717	D41	0,923	0,888	0,719
					D42	0,867	0,858	0,808

Tabel 17-10. Prior modelkansen schrijfvaardigheid Engels

School- soort/ Leerweg	Hoofdaspecten			Deelaspecten			
		Onder niveau	Op niveau	Boven niveau	Onder niveau	Op niveau	Boven niveau
havo	H1	0,171	0,479	0,350	0,330	0,340	0,330
	H2	0,232	0,507	0,262	0,330	0,340	0,330
	H3	0,286	0,504	0,209	0,330	0,340	0,330
	H4	0,288	0,502	0,210	0,330	0,340	0,330
vwo	H1	0,193	0,522	0,285	0,330	0,340	0,330
	H2	0,303	0,521	0,176	0,330	0,340	0,330
	H3	0,300	0,493	0,207	0,330	0,340	0,330
	H4	0,218	0,451	0,330	0,330	0,340	0,330
vmbo-bb	H1	0,263	0,509	0,228	0,330	0,340	0,330
	H2	0,219	0,525	0,256	0,330	0,340	0,330
	H3	0,379	0,506	0,115	0,330	0,340	0,330
	H4	0,234	0,495	0,271	0,330	0,340	0,330
vmbo-kb	H1	0,174	0,493	0,333	0,330	0,340	0,330
	H2	0,253	0,507	0,239	0,330	0,340	0,330
	H3	0,260	0,486	0,254	0,330	0,340	0,330
	H4	0,352	0,504	0,144	0,330	0,340	0,330
vmbo-gt	H1	0,326	0,479	0,195	0,330	0,340	0,330
	H2	0,351	0,479	0,170	0,330	0,340	0,330
	H3	0,273	0,499	0,228	0,330	0,340	0,330
	H4	0,324	0,508	0,168	0,330	0,340	0,330

17.2.4 Resultaten blokindeling wiskunde

Tabel 17-11. Proporties correcte diagnoses wiskunde, aantal opgaven en responsen in de simulaties in havo/vwo (exclusief de opgaven in de doortoetsblokken)

School- soort/ Leerweg	Hoofdaspecten					Deelaspecten					
		Onder niveau	Op niveau	Boven niveau	Aantal opgaven	Aantal responsen		Onder niveau	Op niveau	Boven niveau	Aantal responsen
havo	B	0,790	0,790	0,790	34	69	B1	0,667	0,526	0,712	28
							B2	0,513	0,603	0,735	20
							B3	0,670	0,612	0,878	21
	C	0,800	0,800	0,800	38	47	C1	0,647	0,612	0,720	11
							C2	0,667	0,371	0,749	15
							C3	0,641	0,582	0,739	21
	D	0,774	0,773	0,774	37	49	D1	0,730	0,399	0,689	18
							D2	0,751	0,344	0,635	14
							D3	0,751	0,462	0,726	17
	E	0,820	0,820	0,820	36	56	E1	0,828	0,571	0,858	15
							E2	0,692	0,397	0,745	13
							E3	0,674	0,436	0,795	28
	F	0,767	0,767	0,767	41	56	F1	0,778	0,525	0,736	19
							F2	0,594	0,533	0,629	23
							F3	0,580	0,360	0,599	14
vwo	B	0,779	0,779	0,779	38	67	B1	0,637	0,362	0,676	20
							B2	0,396	0,572	0,667	18
							B3	0,728	0,571	0,867	29
	C	0,774	0,774	0,774	34	46	C1	0,558	0,485	0,722	15
							C2	0,688	0,429	0,769	11
							C3	0,692	0,423	0,777	20
	D	0,847	0,847	0,848	41	55	D1	0,777	0,514	0,840	18
							D2	0,733	0,477	0,701	17
							D3	0,749	0,541	0,810	20
	E	0,853	0,852	0,853	42	48	E1	0,803	0,664	0,839	18
							E2	0,770	0,461	0,715	17
							E3	0,722	0,459	0,690	13
	F	0,754	0,754	0,754	36	48	F1	0,649	0,417	0,612	16
							F2	0,626	0,657	0,791	23
							F3	0,445	0,317	0,667	9

Tabel 17-12. Proporties correcte diagnoses wiskunde, aantal opgaven en responsen in de simulaties in vmbo (exclusief de opgaven in de doortoetsblokken)

School- soort/ Leerweg	Hoofdaspecten					Deelaspecten					
		Onder niveau	Op niveau	Boven niveau	Aantal opgaven	Aantal responsen		Onder niveau	Op niveau	Boven niveau	Aantal responsen
vmbo- bb	B	0,878	0,878	0,878	70	90	B1	0,750	0,576	0,716	37
							B2	0,742	0,566	0,794	31
							B3	0,814	0,605	0,852	22
	C	0,862	0,862	0,862	56	76	C1	0,859	0,721	0,833	32
							C2	0,606	0,523	0,619	25
							C3	0,658	0,380	0,678	19
	D	0,835	0,835	0,836	49	77	D1	0,733	0,482	0,761	23
							D2	0,791	0,618	0,846	42
							D3	0,735	0,355	0,553	12
	E	0,795	0,795	0,795	43	49	E1	0,736	0,447	0,703	20
							E2	0,648	0,384	0,585	10
							E3	0,746	0,567	0,800	19
vmbo- kb	B	0,805	0,805	0,805	57	84	B1	0,716	0,508	0,691	28
							B2	0,673	0,564	0,715	36
							B3	0,657	0,502	0,714	20
	C	0,814	0,814	0,814	46	70	C1	0,849	0,643	0,755	27
							C2	0,672	0,368	0,647	24
							C3	0,656	0,337	0,595	19
	D	0,836	0,836	0,836	55	70	D1	0,674	0,446	0,763	26
							D2	0,770	0,582	0,804	25
							D3	0,693	0,516	0,725	19
	E	0,886	0,886	0,887	41	63	E1	0,817	0,589	0,711	26
							E2	0,741	0,380	0,752	17
							E3	0,795	0,635	0,887	20
vmbo-gt	B	0,826	0,826	0,826	53	70	B1	0,719	0,585	0,801	33
							B2	0,679	0,603	0,673	18
							B3	0,745	0,477	0,814	19
	C	0,833	0,833	0,833	67	77	C1	0,807	0,645	0,855	29
							C2	0,604	0,406	0,726	25
							C3	0,699	0,482	0,774	23
	D	0,838	0,838	0,838	52	63	D1	0,767	0,561	0,819	23
							D2	0,756	0,457	0,789	21
							D3	0,743	0,493	0,764	19
	E	0,872	0,872	0,872	46	61	E1	0,799	0,657	0,846	28
							E2	0,775	0,470	0,705	15
							E3	0,797	0,457	0,772	18

Tabel 17-13. Uitstapkansen wiskunde in havo/vwo

School- soort/ Leerweg	Hoofdaspecten				Deelaspecten			
		Onder niveau	Op niveau	Boven niveau		Onder niveau	Op niveau	Boven niveau
havo	B	0,939	0,877	0,010	B1	0,580	0,990	0,990
					B2	0,963	0,608	0,990
					B3	0,847	0,682	0,010
	C	0,872	0,734	0,846	C1	0,792	0,990	0,817
					C2	0,438	0,990	0,990
					C3	0,461	0,470	0,990
	D	0,674	0,990	0,873	D1	0,463	0,990	0,990
					D2	0,990	0,990	0,912
					D3	0,990	0,990	0,576
	E	0,990	0,857	0,697	E1	0,657	0,736	0,788
					E2	0,990	0,672	0,677
					E3	0,990	0,990	0,990
	F	0,894	0,990	0,894	F1	0,816	0,799	0,541
					F2	0,990	0,990	0,990
					F3	0,990	0,990	0,684
vwo	B	0,927	0,924	0,783	B1	0,603	0,990	0,462
					B2	0,474	0,990	0,990
					B3	0,921	0,990	0,579
	C	0,866	0,773	0,803	C1	0,474	0,990	0,398
					C2	0,576	0,990	0,445
					C3	0,731	0,990	0,990
	D	0,889	0,788	0,830	D1	0,884	0,671	0,546
					D2	0,990	0,990	0,406
					D3	0,771	0,990	0,990
	E	0,865	0,905	0,695	E1	0,833	0,990	0,828
					E2	0,471	0,990	0,610
					E3	0,990	0,990	0,990
	F	0,925	0,719	0,562	F1	0,643	0,990	0,990
					F2	0,923	0,990	0,990
					F3	0,792	0,990	0,990

Tabel 17-14. Uitstapkansen wiskunde in vmbo

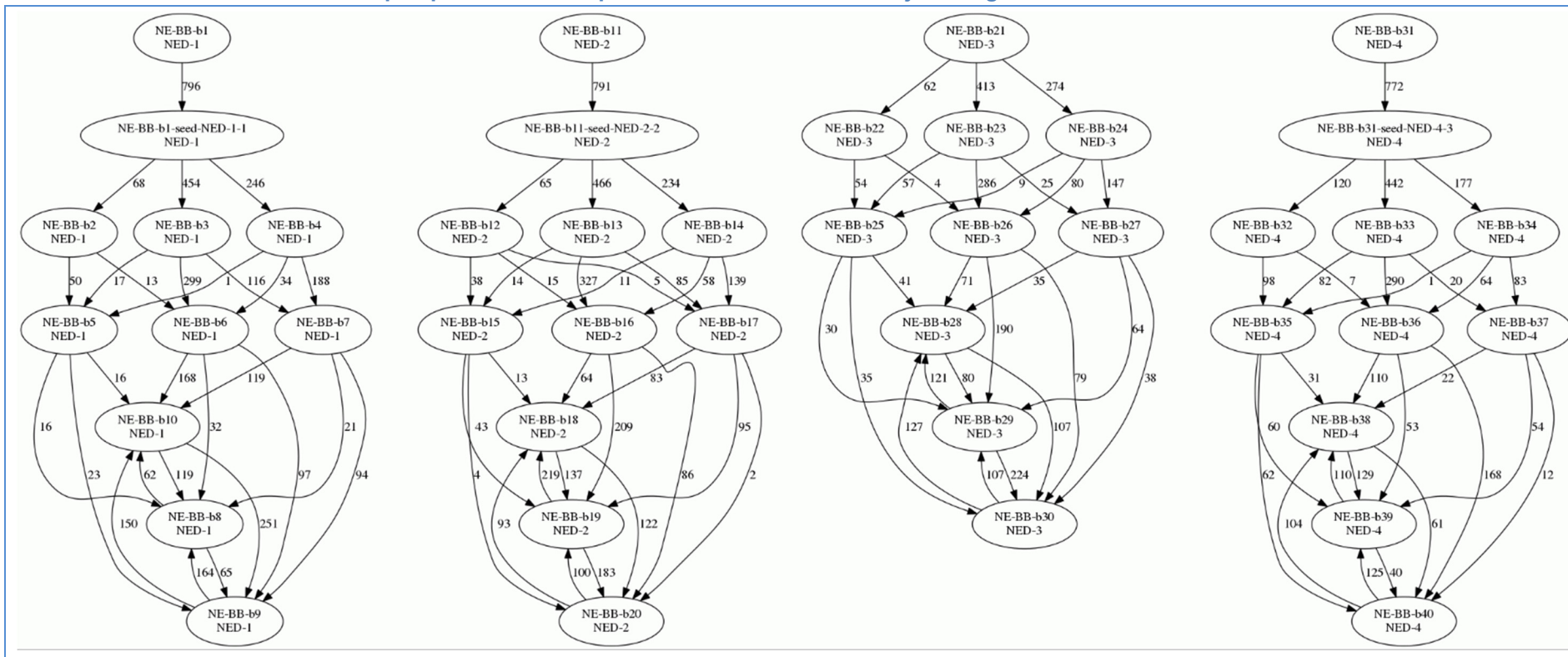
School- soort/ Leerweg	Hoofdaspecten				Deelaspecten			
		Onder niveau	Op niveau	Boven niveau		Onder niveau	Op niveau	Boven niveau
vmbo-bb	B	0,893	0,919	0,740	B1	0,990	0,608	0,990
					B2	0,990	0,717	0,578
					B3	0,857	0,875	0,824
	C	0,898	0,923	0,858	C1	0,931	0,990	0,869
					C2	0,990	0,990	0,431
					C3	0,990	0,990	0,990
	D	0,898	0,949	0,800	D1	0,870	0,648	0,990
					D2	0,990	0,911	0,798
					D3	0,990	0,990	0,421
	E	0,765	0,797	0,869	E1	0,463	0,570	0,601
					E2	0,470	0,990	0,477
					E3	0,518	0,608	0,643
vmbo-kb	B	0,908	0,895	0,866	B1	0,821	0,990	0,838
					B2	0,659	0,990	0,849
					B3	0,990	0,990	0,990
	C	0,861	0,790	0,909	C1	0,762	0,547	0,564
					C2	0,990	0,990	0,990
					C3	0,990	0,639	0,907
	D	0,912	0,814	0,809	D1	0,522	0,990	0,990
					D2	0,748	0,990	0,826
					D3	0,757	0,990	0,493
	E	0,905	0,925	0,847	E1	0,764	0,596	0,586
					E2	0,990	0,990	0,990
					E3	0,661	0,787	0,653
vmbo-gt	B	0,943	0,858	0,682	B1	0,630	0,990	0,822
					B2	0,990	0,990	0,481
					B3	0,990	0,621	0,687
	C	0,913	0,938	0,509	C1	0,752	0,857	0,523
					C2	0,990	0,990	0,990
					C3	0,737	0,674	0,726
	D	0,880	0,990	0,700	D1	0,701	0,990	0,747
					D2	0,584	0,649	0,990
					D3	0,634	0,990	0,990
	E	0,906	0,873	0,871	E1	0,472	0,990	0,826
					E2	0,713	0,990	0,990
					E3	0,530	0,990	0,990

Tabel 17-15. Prior modelkansen wiskunde

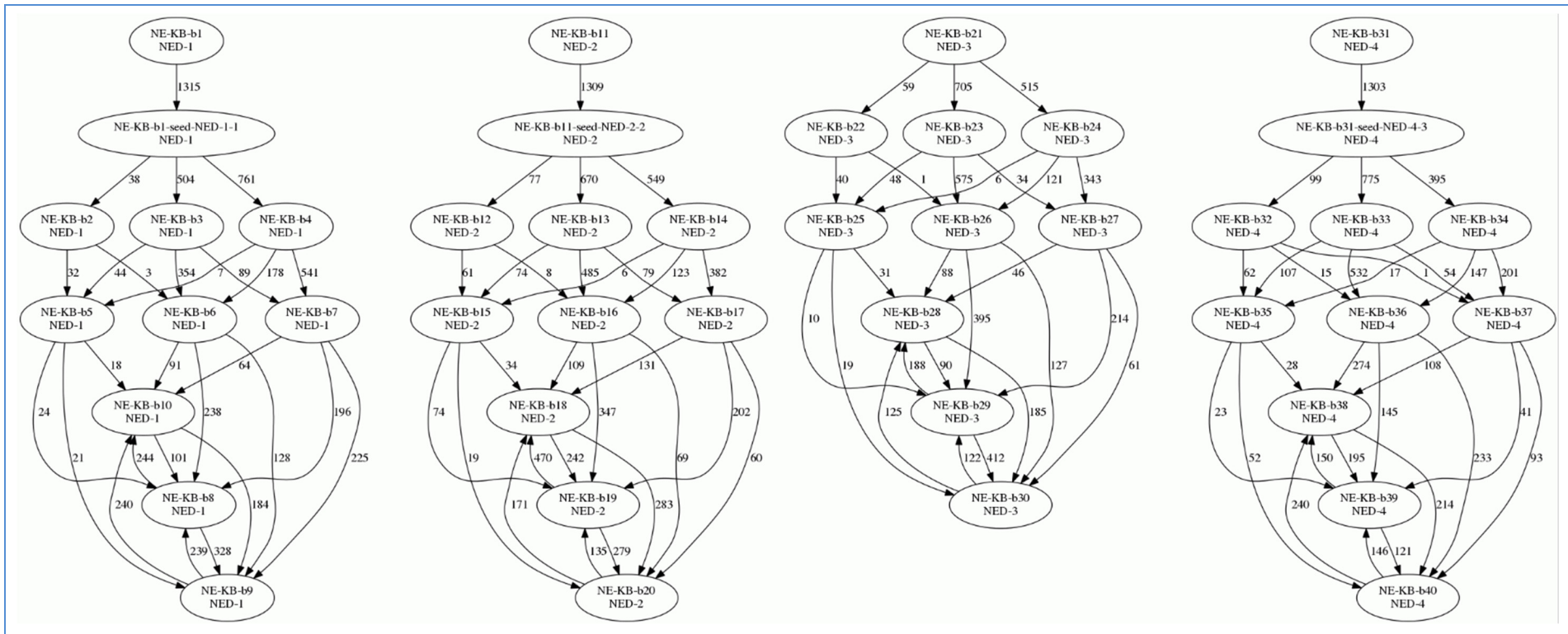
School- soort/ Leerweg	Hoofdaspecten				Deelaspecten		
		Onder niveau	Op niveau	Boven niveau	Onder niveau	Op niveau	Boven niveau
havo	B	0,495	0,472	0,034	0,330	0,340	0,330
	C	0,353	0,436	0,211	0,330	0,340	0,330
	D	0,306	0,435	0,259	0,330	0,340	0,330
	E	0,394	0,486	0,120	0,330	0,340	0,330
	F	0,336	0,444	0,220	0,330	0,340	0,330
vwo	B	0,454	0,426	0,120	0,330	0,340	0,330
	C	0,363	0,418	0,219	0,330	0,340	0,330
	D	0,351	0,463	0,186	0,330	0,340	0,330
	E	0,370	0,468	0,163	0,330	0,340	0,330
	F	0,430	0,405	0,165	0,330	0,340	0,330
vmbo-bb	B	0,370	0,471	0,159	0,330	0,340	0,330
	C	0,304	0,444	0,252	0,330	0,340	0,330
	D	0,321	0,461	0,218	0,330	0,340	0,330
	E	0,288	0,454	0,258	0,330	0,340	0,330
vmbo-kb	B	0,305	0,429	0,266	0,330	0,340	0,330
	C	0,170	0,457	0,373	0,330	0,340	0,330
	D	0,340	0,447	0,213	0,330	0,340	0,330
	E	0,318	0,472	0,210	0,330	0,340	0,330
vmbo-gt	B	0,382	0,477	0,141	0,330	0,340	0,330
	C	0,413	0,480	0,107	0,330	0,340	0,330
	D	0,342	0,504	0,154	0,330	0,340	0,330
	E	0,290	0,463	0,246	0,330	0,340	0,330

17.3 Bijlage Visualisaties van de doorlopen paden per vak en schoolsoort/leerweg (per package)

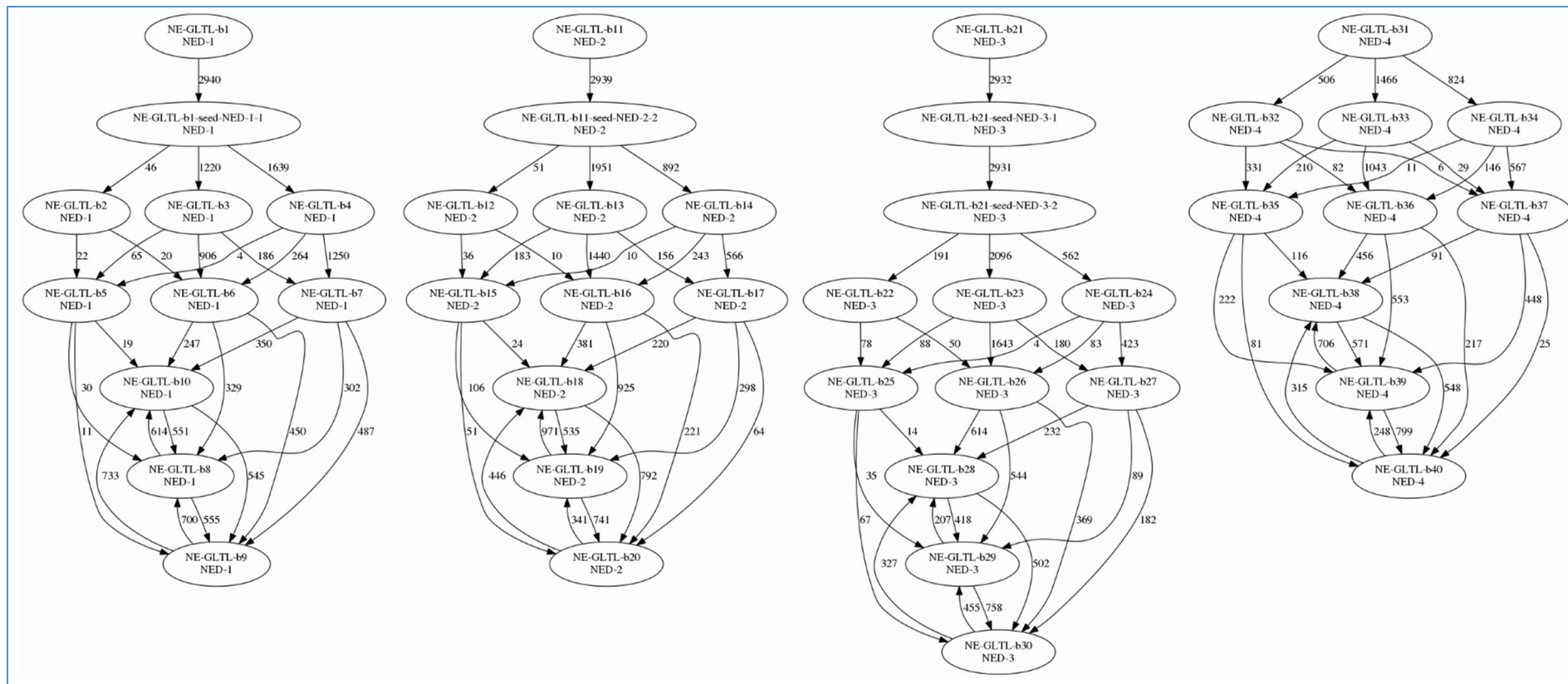
17.3.1 Visualisaties van de doorlopen paden in de adaptieve afname 2017 van schrijfvaardigheid Nederlands



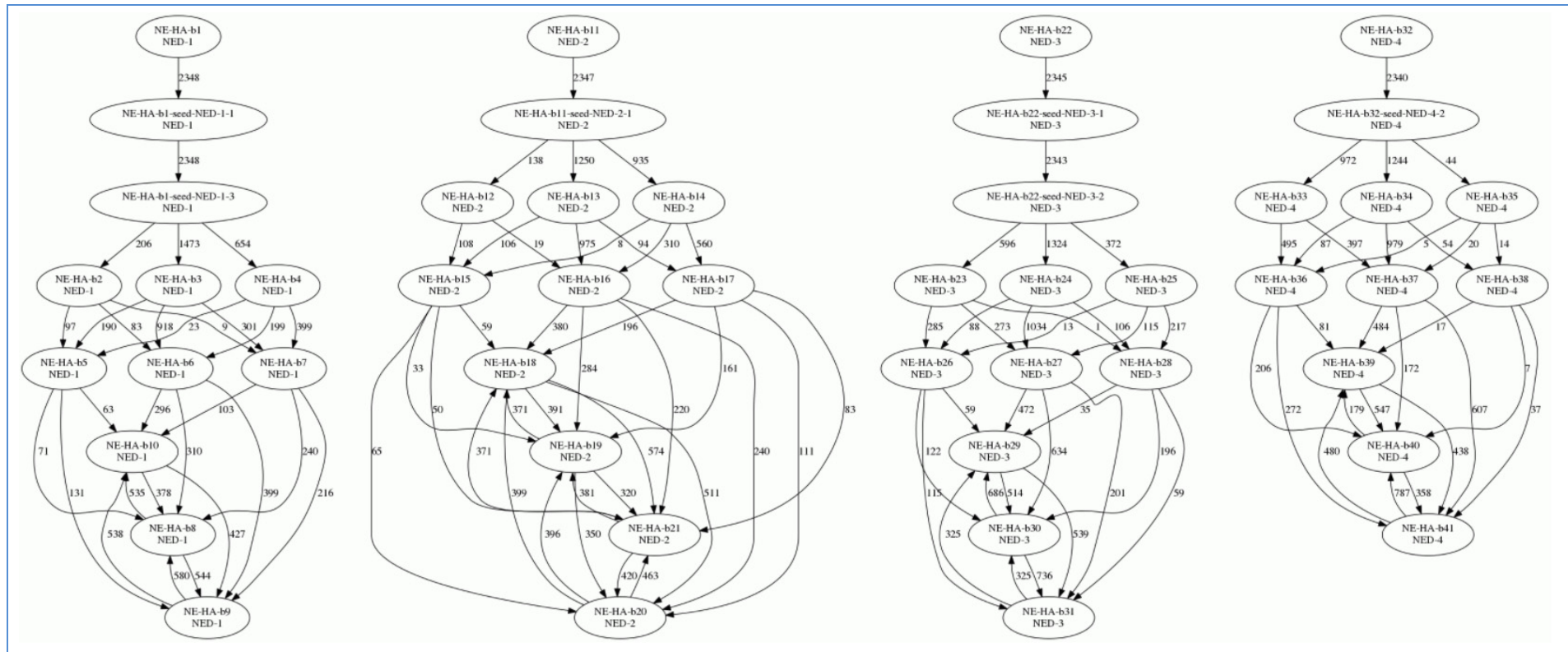
Figuur 17-1. Visualisatie van de doorlopen paden in de adaptieve afname 2017 van schrijfvaardigheid Nederlands in vmbo-bb



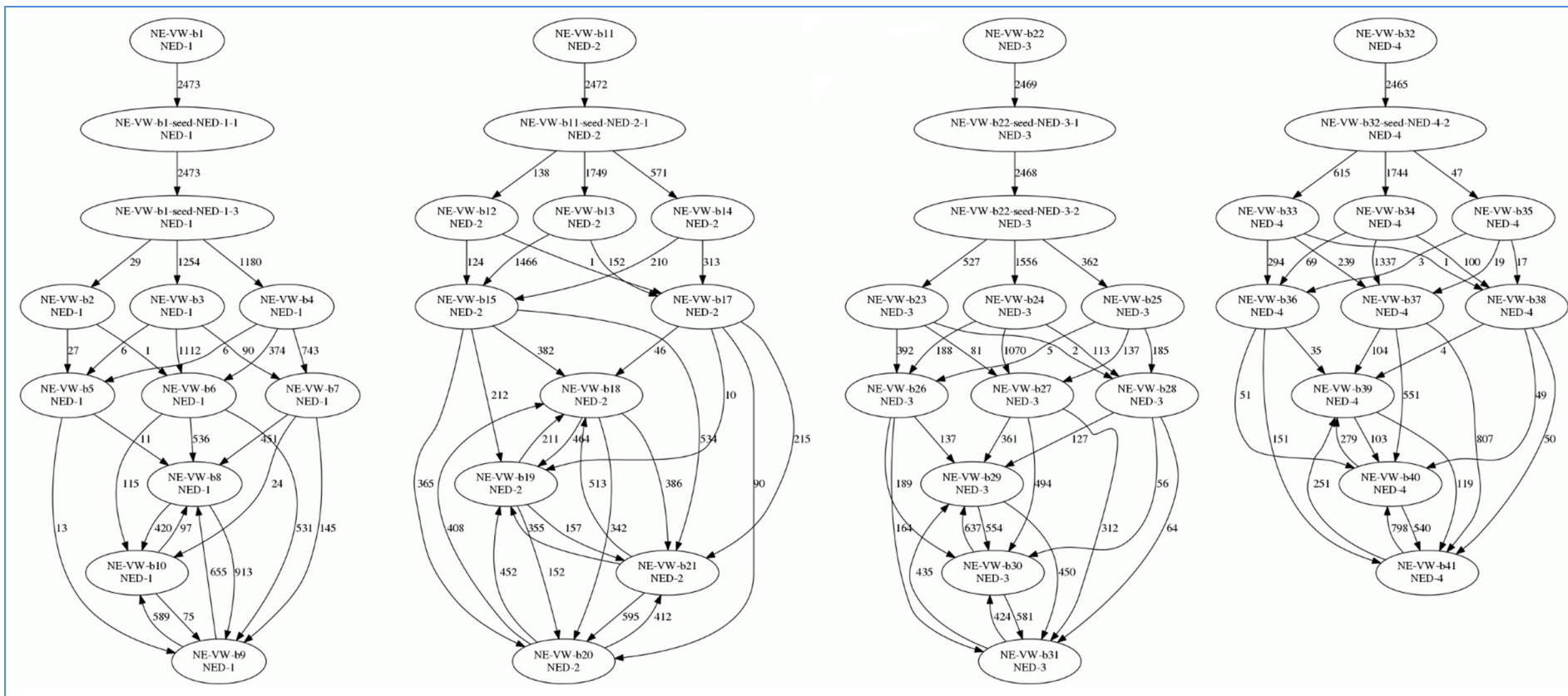
Figuur 17-2. Visualisatie van de doorlopen paden in de adaptieve afname 2017 van schrijfvaardigheid Nederlands in vmbo-kb



Figuur 17-3. Visualisatie van de doorlopen paden in de adaptieve afname 2017 van schrijfvaardigheid Nederlands in vmbo-gt

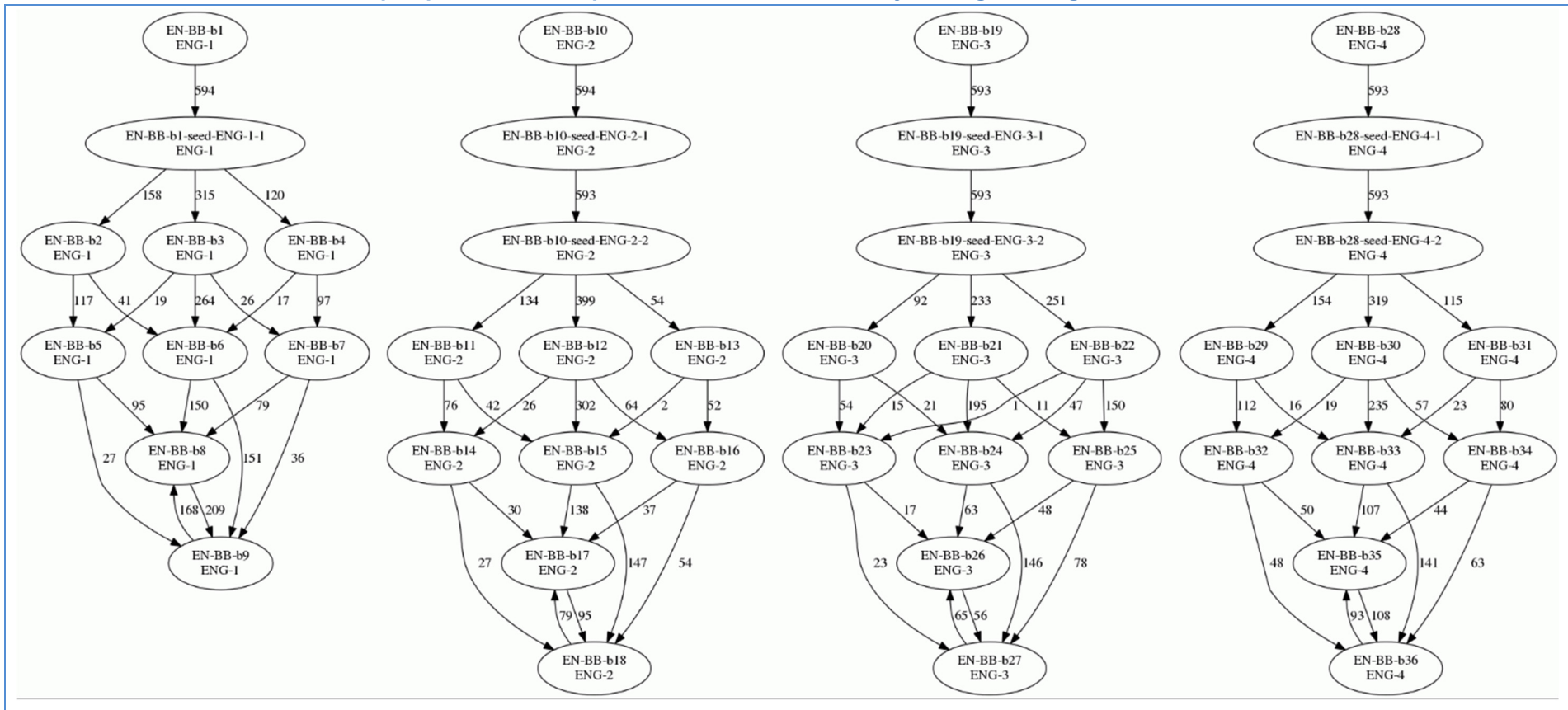


Figuur 17-4. Visualisatie van de doorlopen paden in de adaptieve afname 2017 van schrijfvaardigheid Nederlands in havo

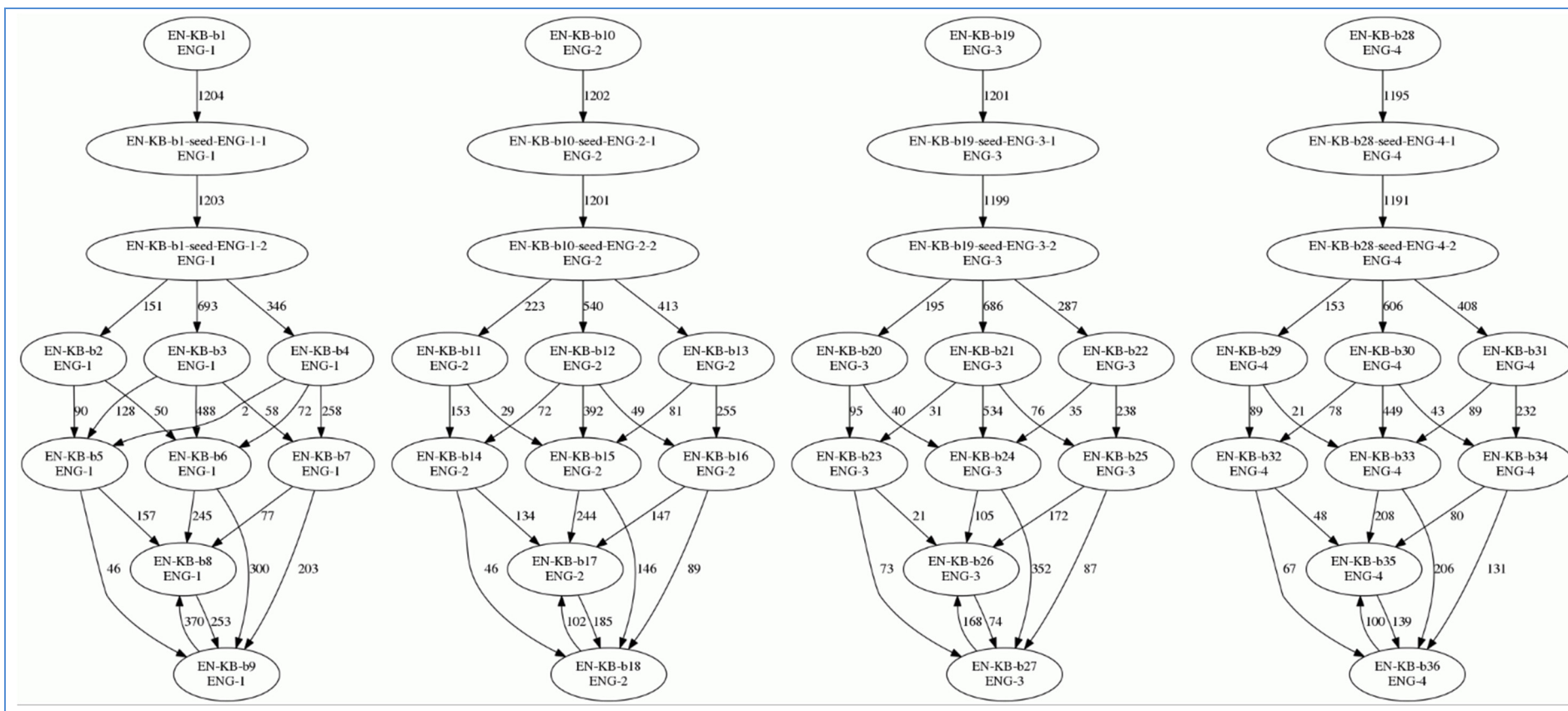


Figuur 17-5. Visualisatie van de doorlopen paden in de adaptieve afname 2017 van schrijfvaardigheid Nederlands in vwo

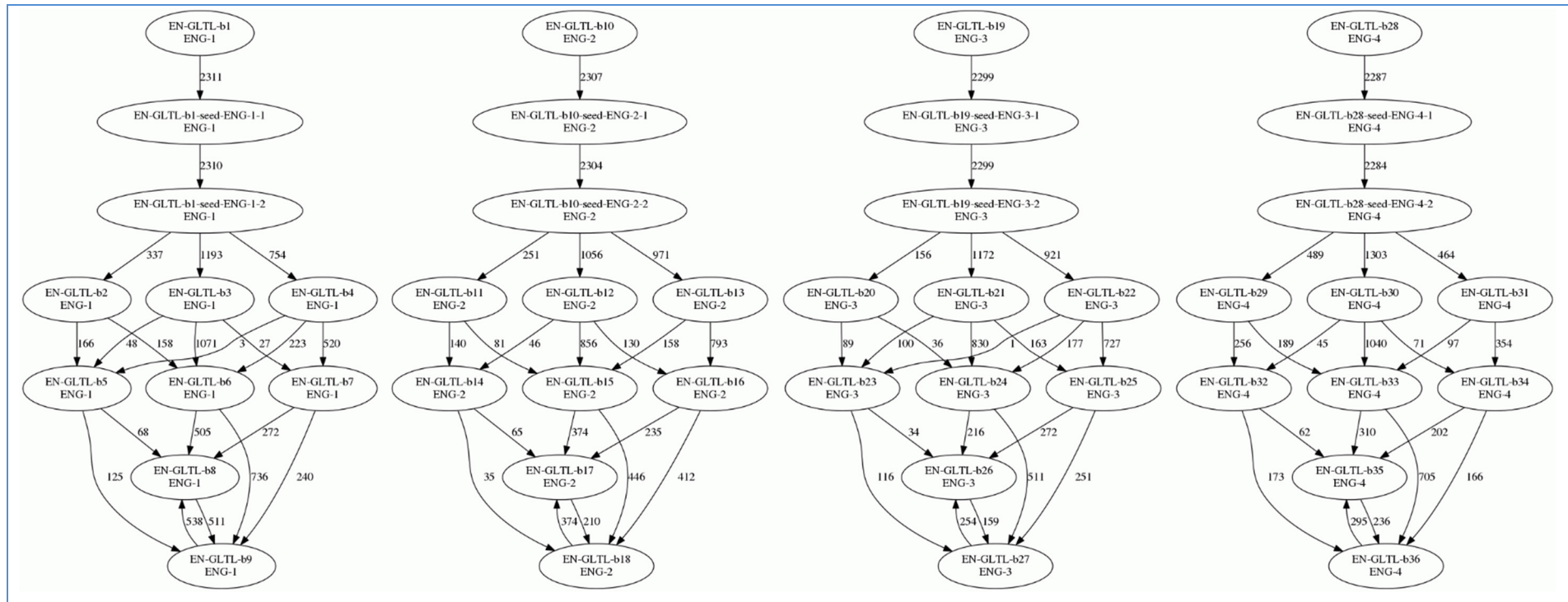
17.3.2 Visualisaties van de doorlopen paden in de adaptieve afname 2017 van schrijfvaardigheid Engels



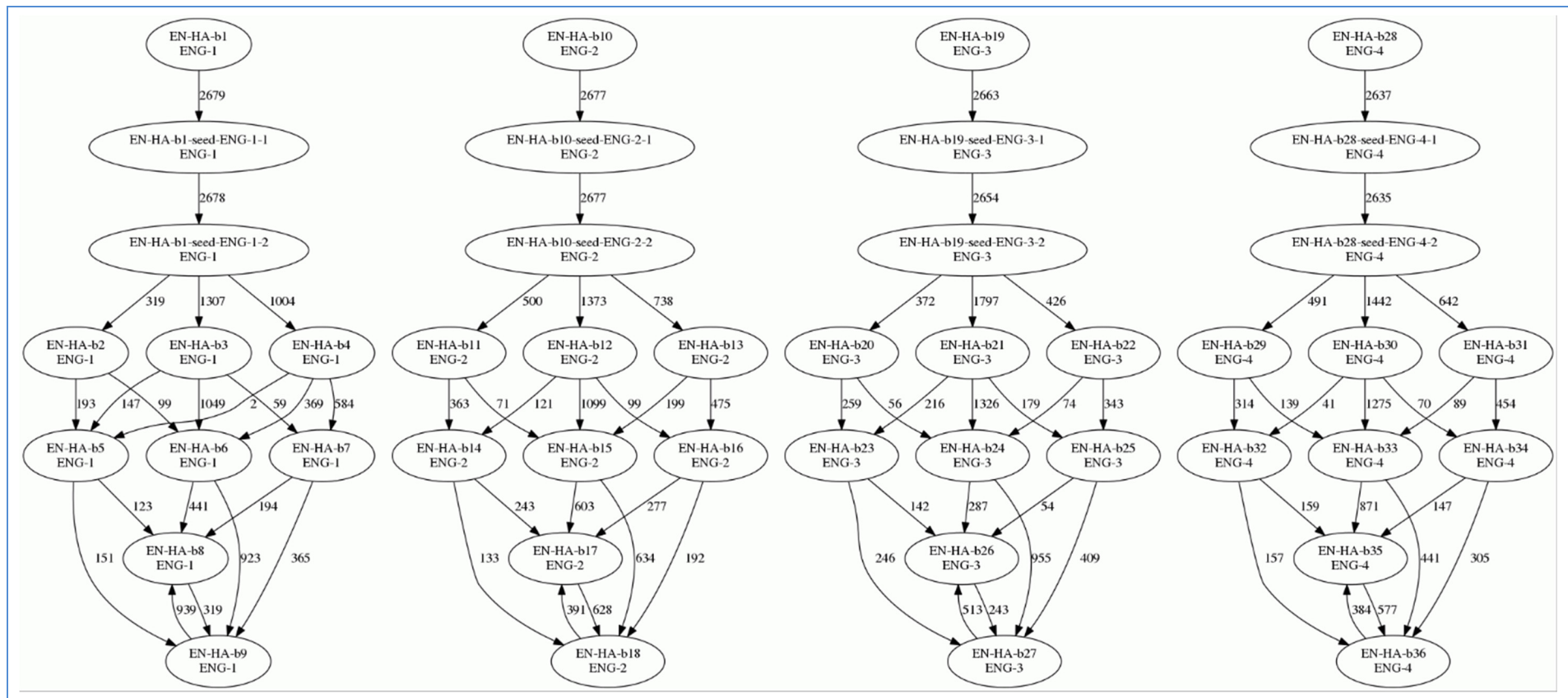
Figuur 17-6. Visualisatie van de doorlopen paden in de adaptieve afname 2017 van schrijfvaardigheid Engels in vmbo-bb



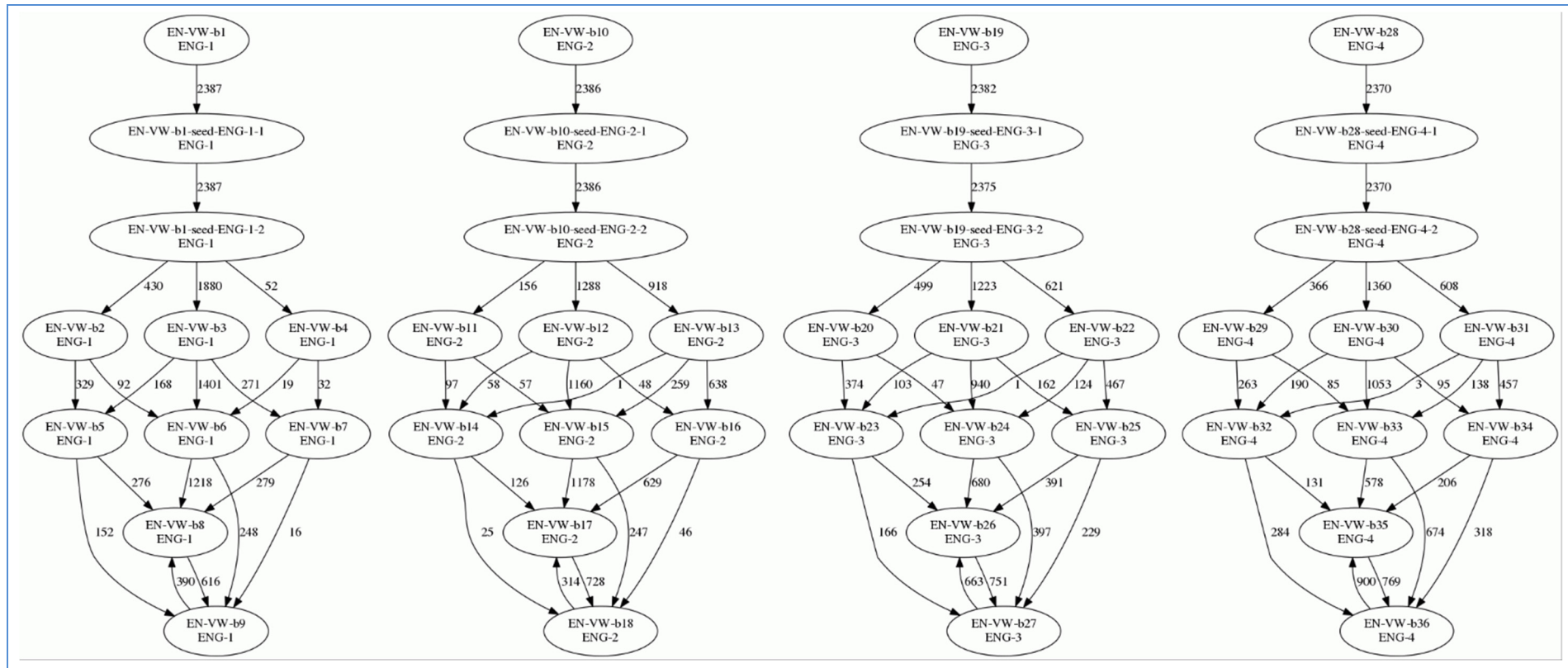
Figuur 17-7. Visualisatie van de doorlopen paden in de adaptieve afname 2017 van schrijfvaardigheid Engels in vmbo-kb



Figuur 17-8. Visualisatie van de doorlopen paden in de adaptieve afname 2017 van schrijfvaardigheid Engels in vmbo-gt

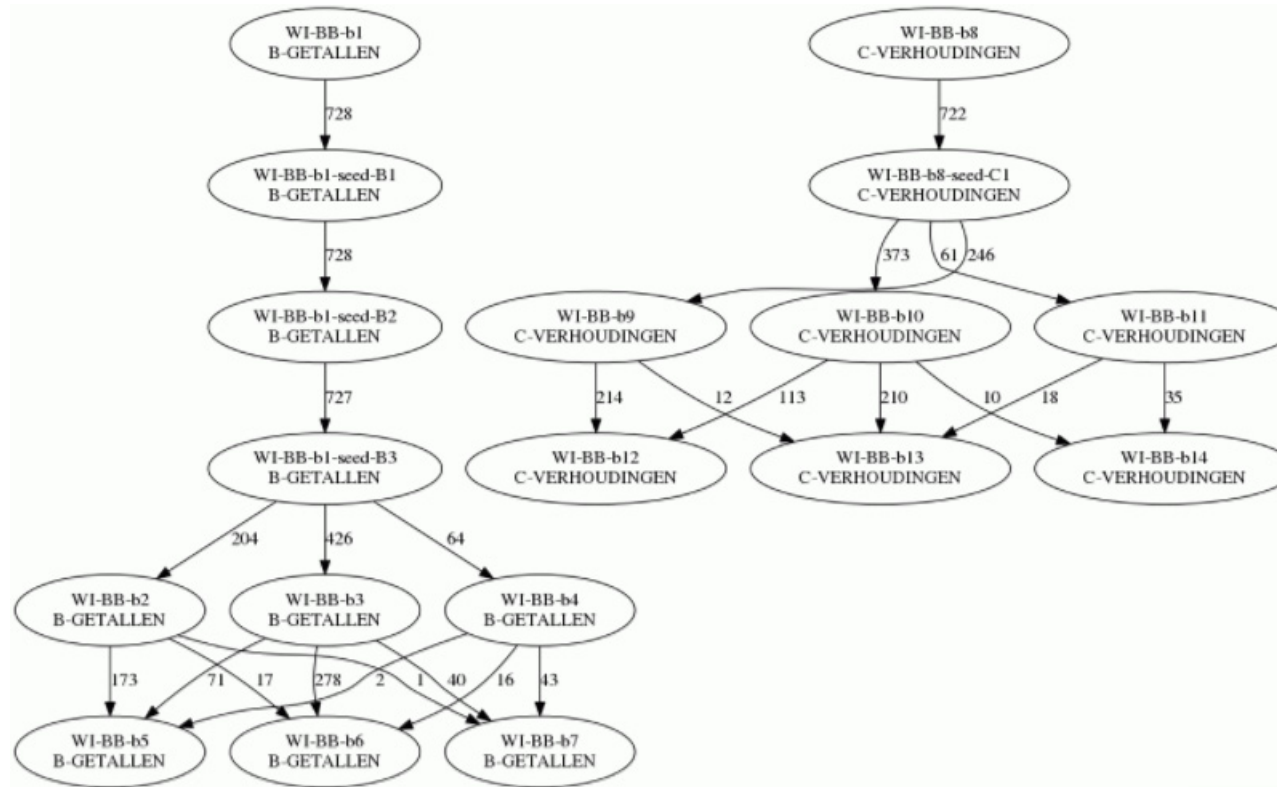


Figuur 17-9. Visualisatie van de doorlopen paden in de adaptieve afname 2017 van schrijfvaardigheid Engels in havo

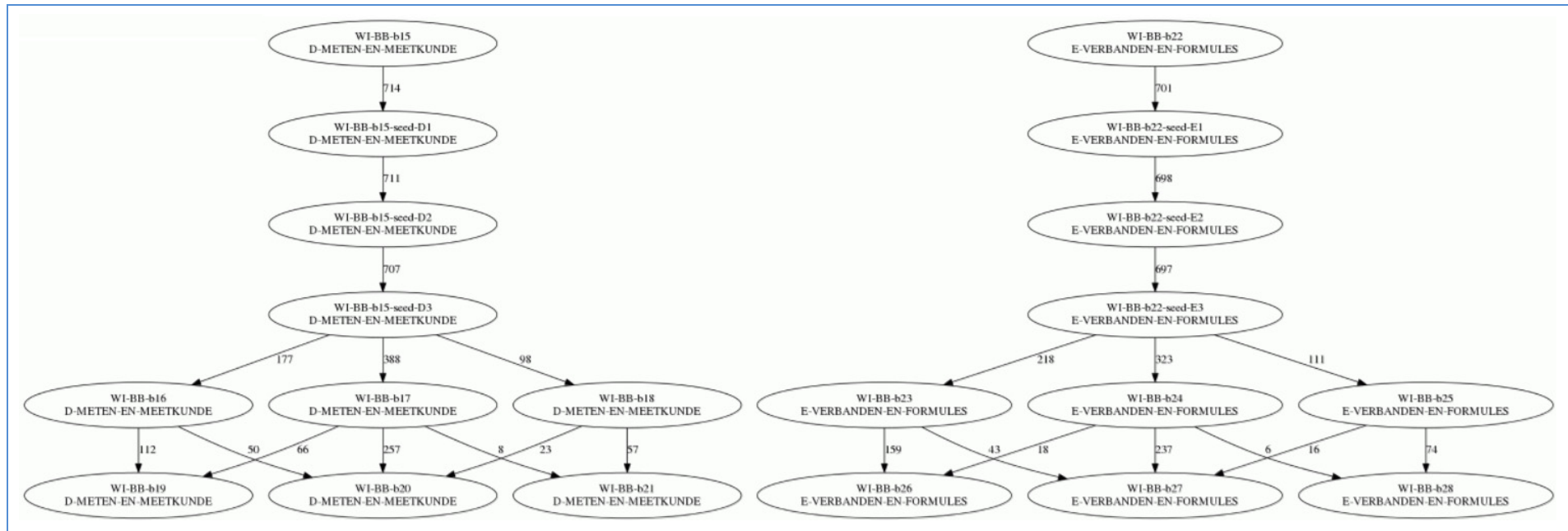


Figuur 17-10. Visualisatie van de doorlopen paden in de adaptieve afname 2017 van schrijfvaardigheid Engels in vwo

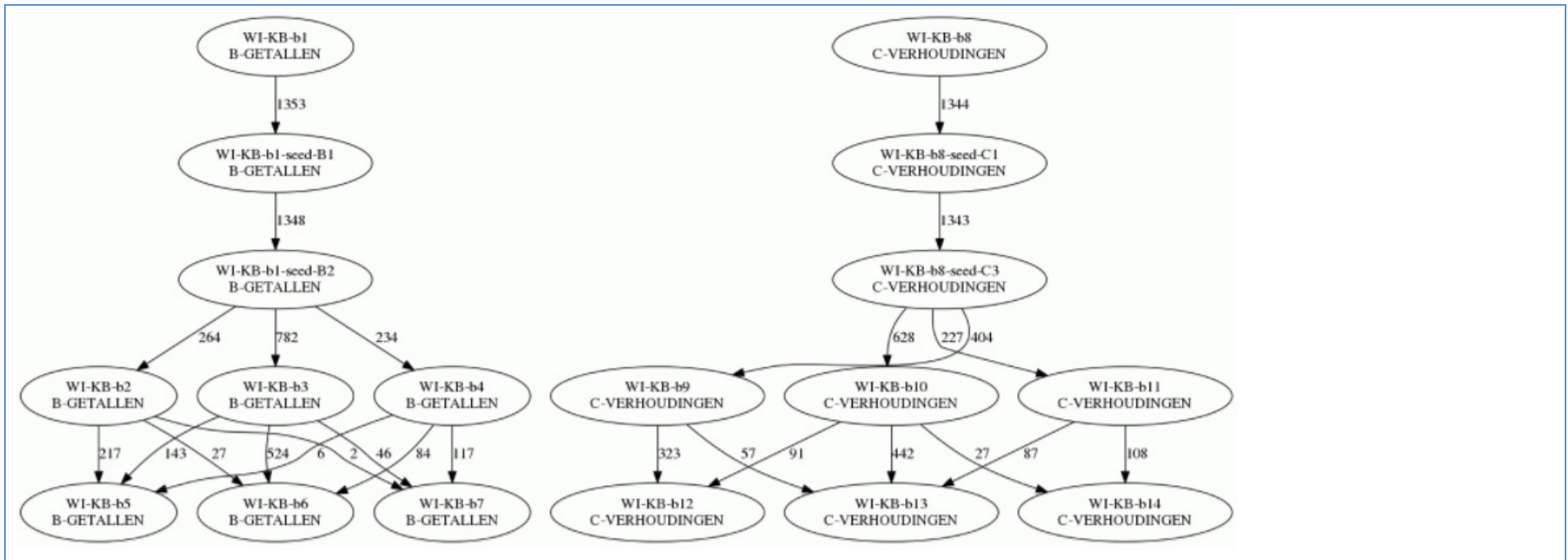
17.3.3 Visualisaties van de doorlopen paden in de adaptieve afname 2017 van wiskunde



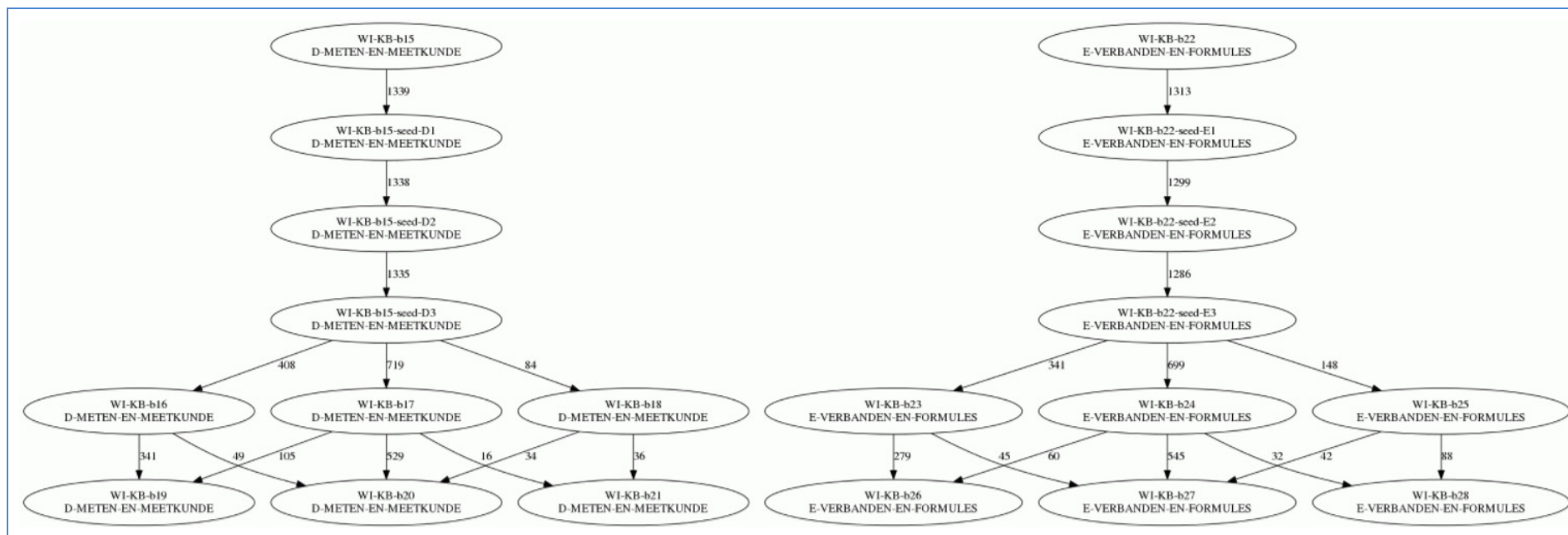
Figuur 17-11. Visualisatie van de doorlopen paden in de adaptieve afname 2017 van wiskunde in vmbo-bb domeinen B en C



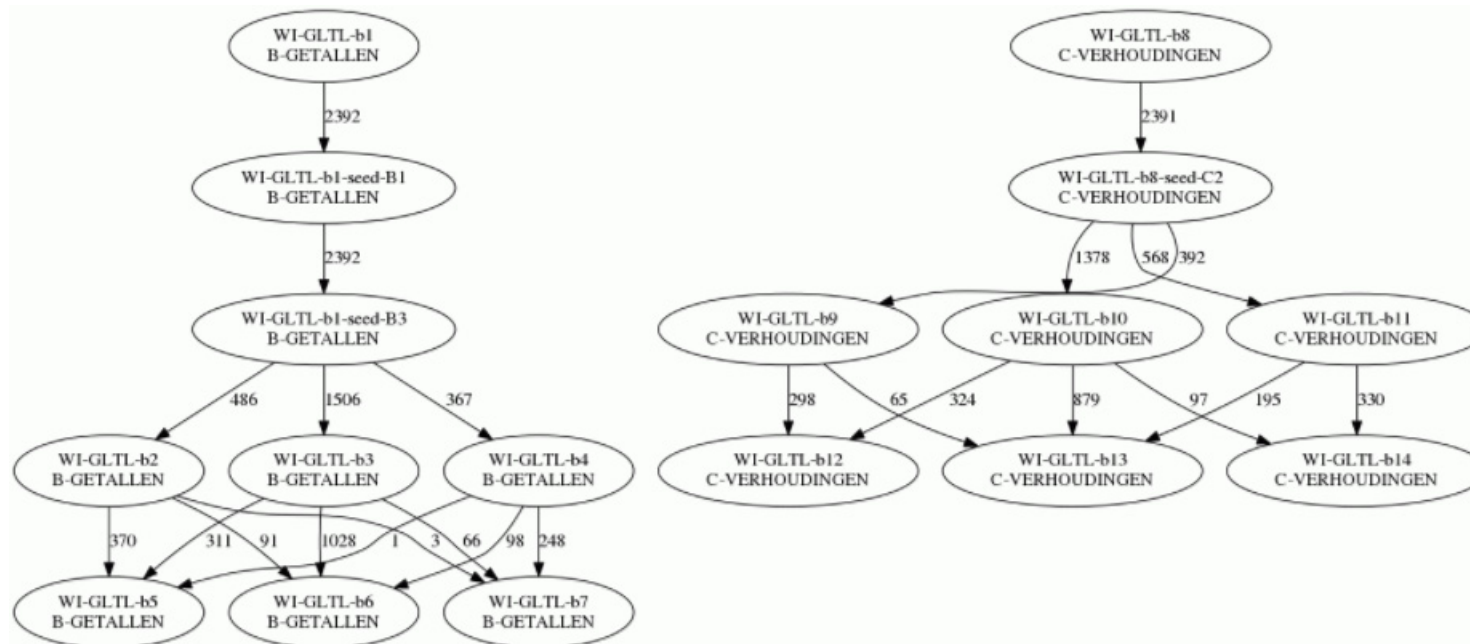
Figuur 17-12. Visualisatie van de doorlopen paden in de adaptieve afname 2017 van wiskunde in vmbo-bb domeinen D en E



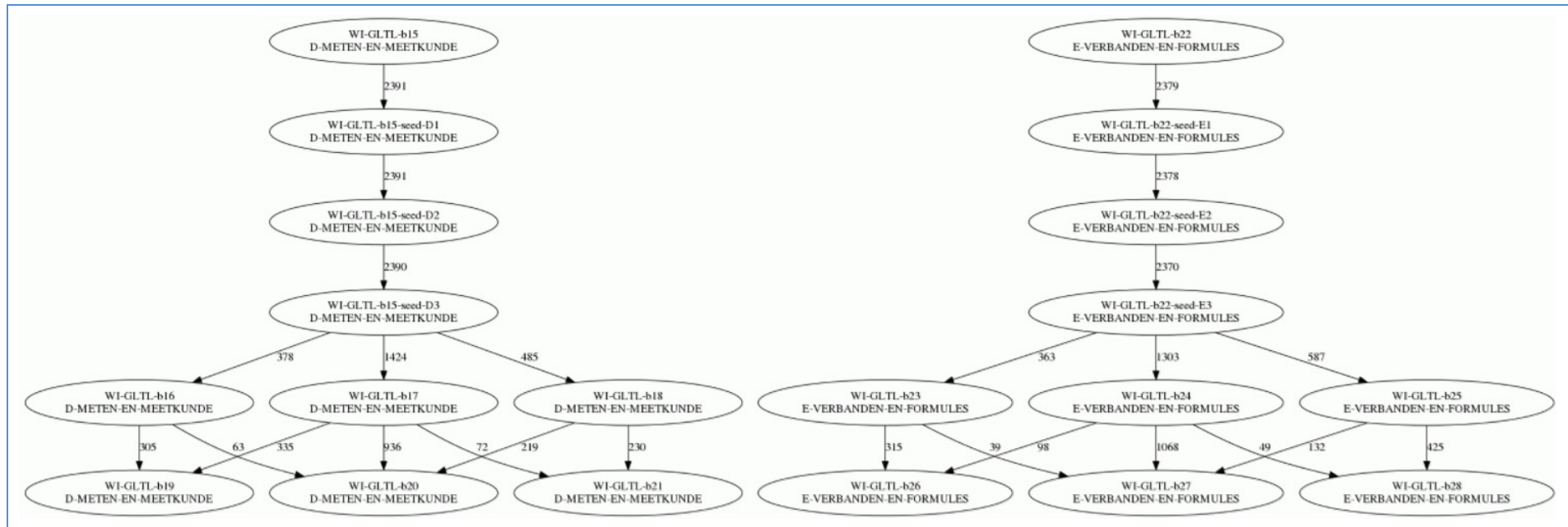
Figuur 17-13. Visualisatie van de doorlopen paden in de adaptieve afname 2017 van wiskunde in vmbo-kb domeinen B en C



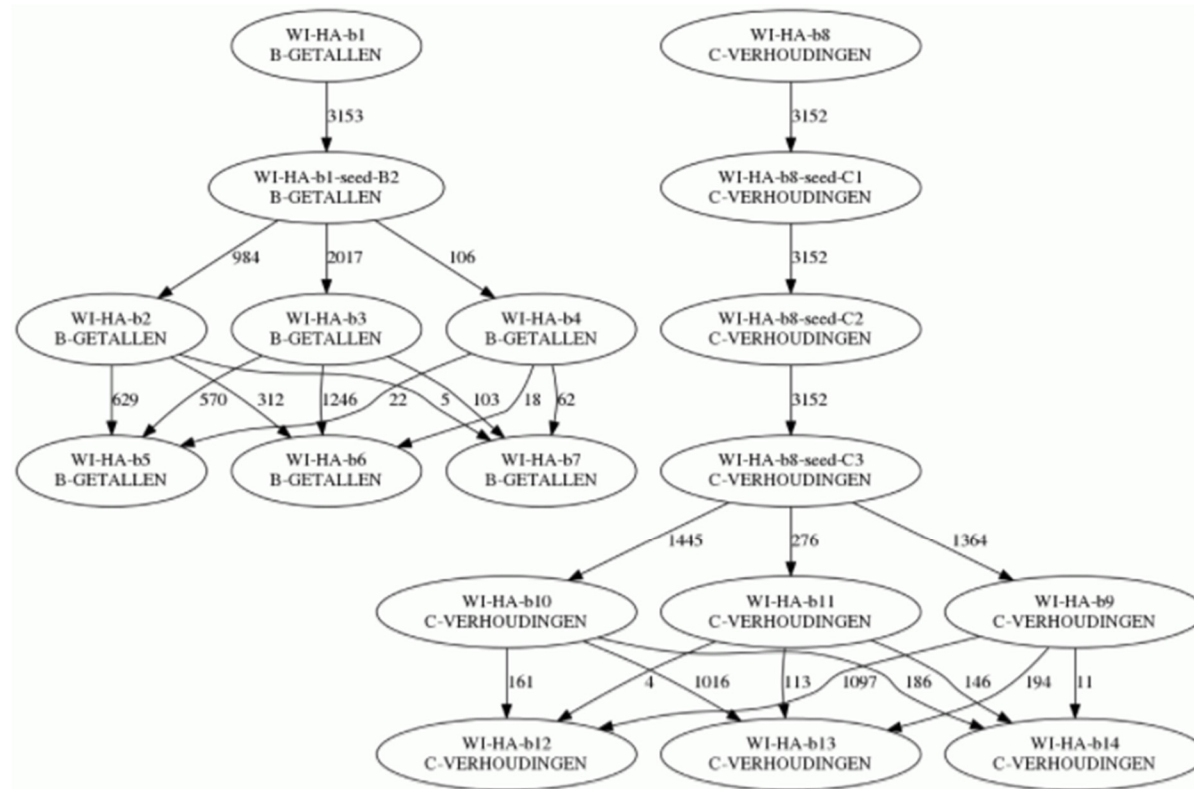
Figuur 17-14. Visualisatie van de doorlopen paden in de adaptieve afname 2017 van wiskunde in vmbo-kb domeinen D en E



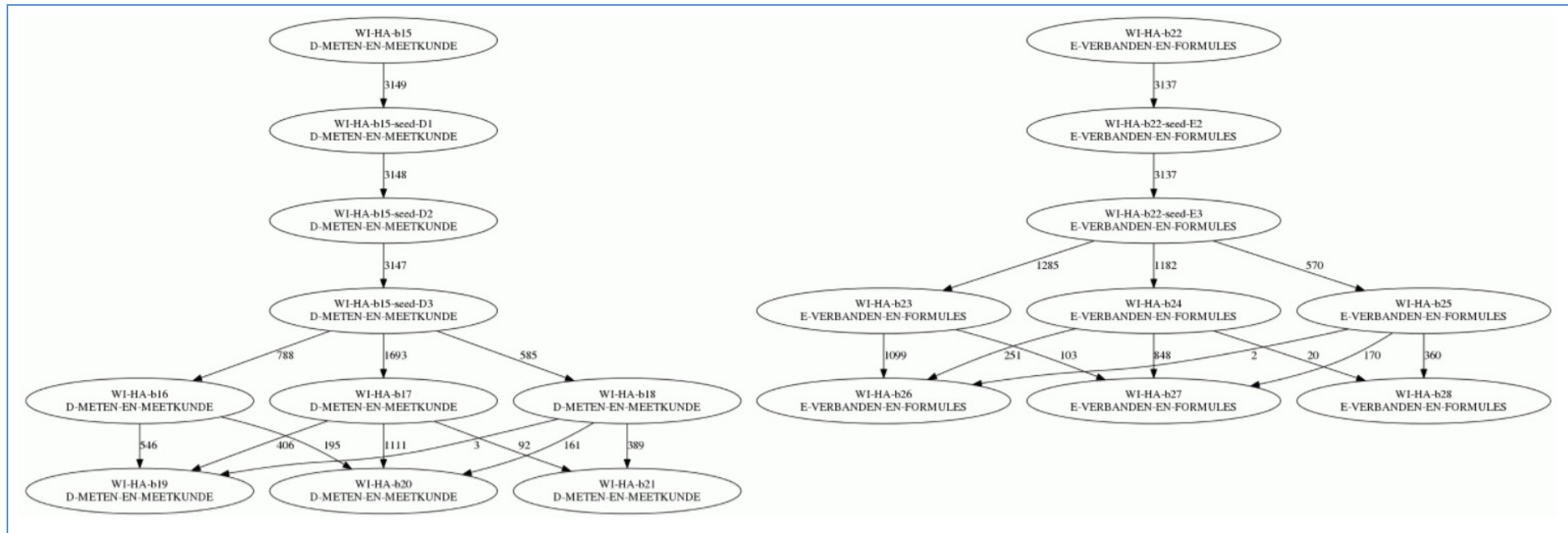
Figuur 17-15. Visualisatie van de doorlopen paden in de adaptieve afname 2017 van wiskunde in vmbo-gt domeinen B en C



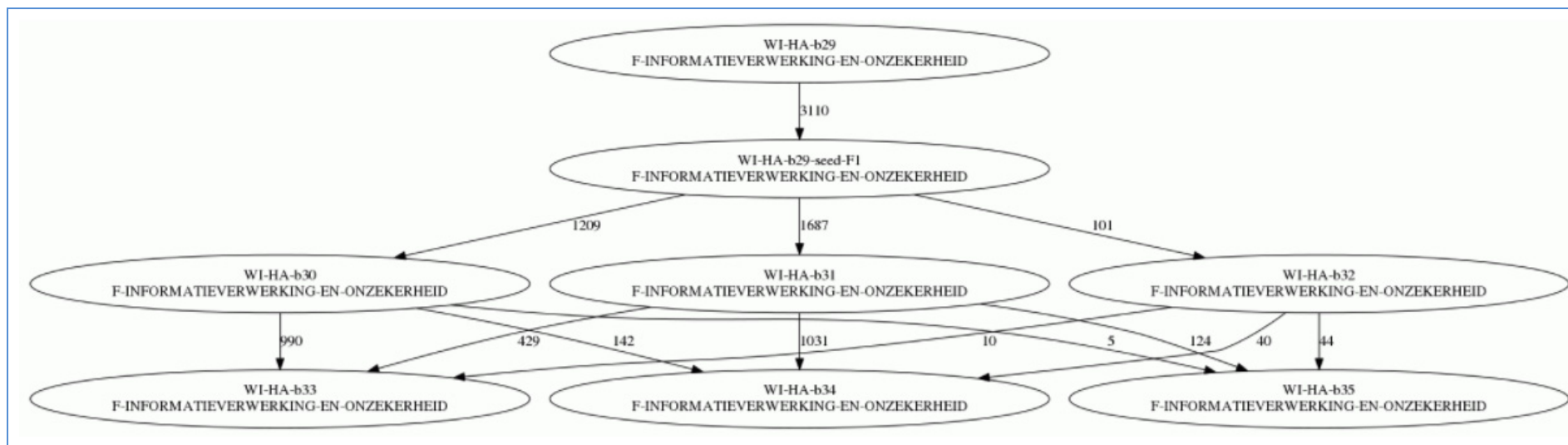
Figuur 17-16. Visualisatie van de doorlopen paden in de adaptieve afname 2017 van wiskunde in vmbo-gt domeinen D en E



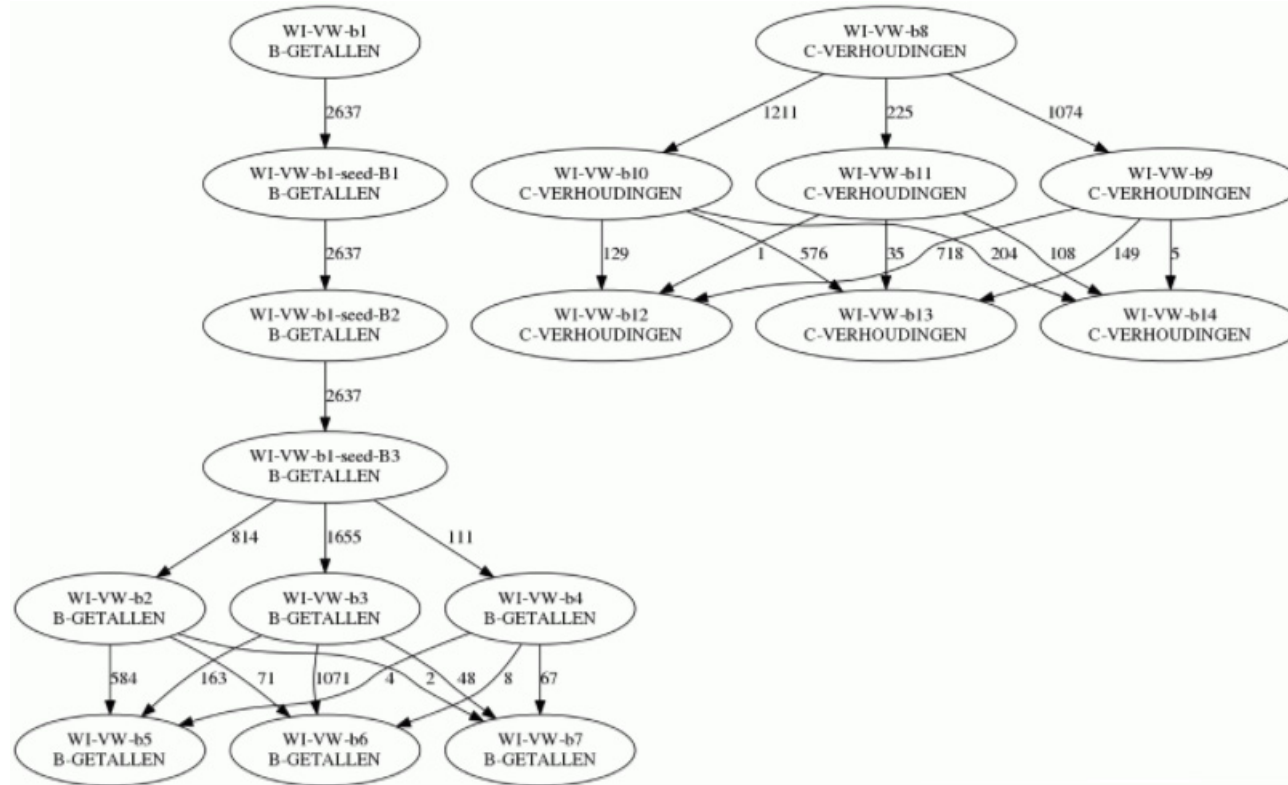
Figuur 17-17. Visualisatie van de doorlopen paden in de adaptieve afname 2017 van wiskunde in havo domeinen B en C



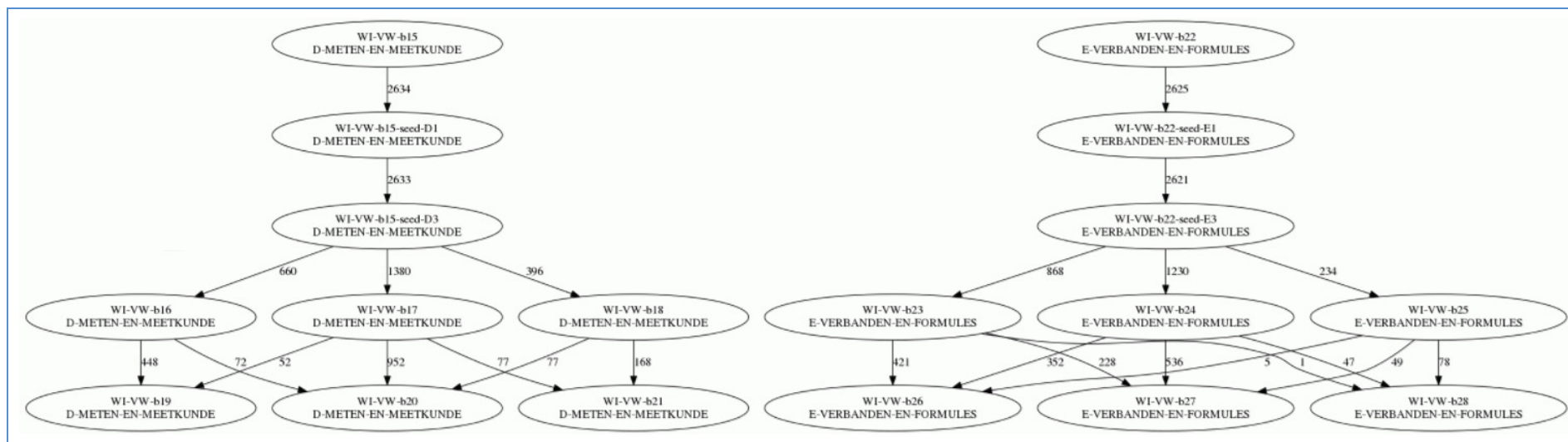
Figuur 17-18. Visualisatie van de doorlopen paden in de adaptieve afname 2017 van wiskunde in havo domeinen D en E



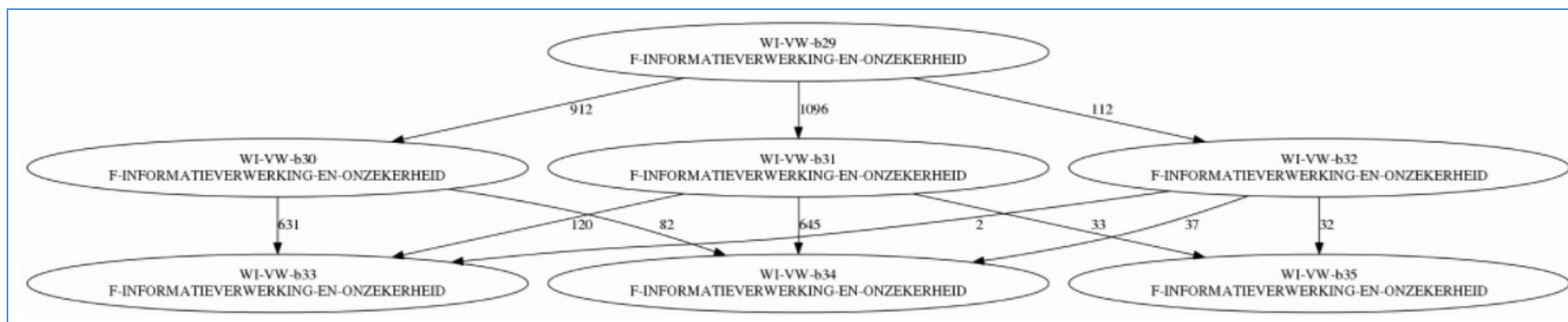
Figuur 17-19. Visualisatie van de doorlopen paden in de adaptieve afname 2017 van wiskunde in havo domein F



Figuur 17-20. Visualisatie van de doorlopen paden in de adaptieve afname 2017 van wiskunde in vwo domeinen B en C



Figuur 17-21. Visualisatie van de doorlopen paden in de adaptieve afname 2017 van wiskunde in vwo domeinen D en E



Figuur 17-22. Visualisatie van de doorlopen paden in de adaptieve afname 2017 van wiskunde in vwo domein F

17.4 Bijlage Resultaten van de enquête over het rapportagemodel

Het CvTE heeft via een online enquête feedback gevraagd over de rapportages uit de mock-up en de bijbehorende uitleg. De enquête bevatte 41 vragen of stellingen: zes over de onderwijsachtergrond van de respondent, twee over het startscherm van de rapportage, vijftien over de visualisatie en keuzemogelijkheden, acht over de leerlingrapportage, zeven over de groepsrapportage en tenslotte konden de respondenten indien gewenst nog de naam van de school, hun eigen naam en opmerkingen achterlaten. Er waren drie versies van de enquête. In een versie stonden voorbeeldrapportages voor schrijfvaardigheid Engels, in de tweede voorbeeldrapportages voor schrijfvaardigheid Nederlands en in de derde versie voorbeeldrapportages voor wiskunde.

17.4.1 De respondenten

In totaal zijn 69 vragenlijsten ingeleverd, maar 12 vragenlijsten waar leeg of vrijwel leeg (geheel zonder feedback op de rapportages). Van de 57 respondenten met bruikbare antwoorden hadden 23 de versie met voorbeelden over schrijfvaardigheid Engels beantwoord, 18 de versie met voorbeelden over schrijfvaardigheid Nederlands en 16 met voorbeelden over wiskunde.

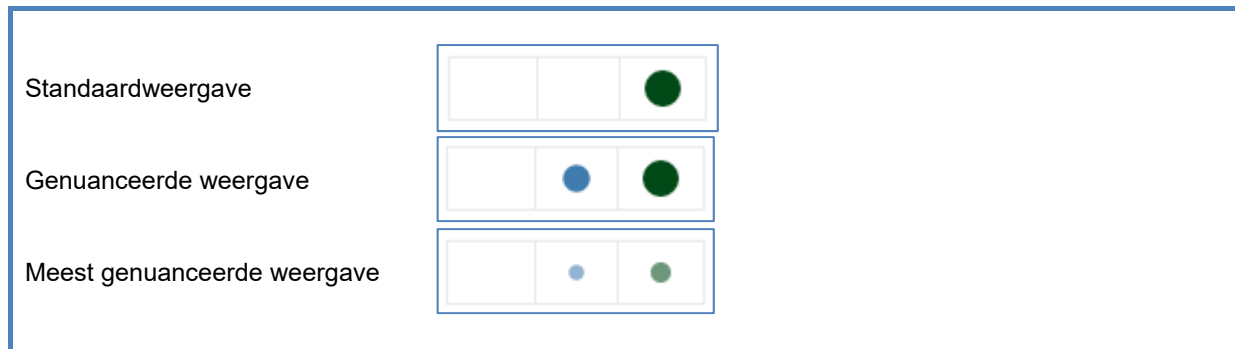
Van de 57 respondenten was 75% werkzaam als docent ($N = 43$). De andere respondenten waren onder andere directeur, conrector, afdelingsleider, domeinleider of teamleider. Docenten van de drie vakken, Engels, Nederlands en wiskunde kwamen vrijwel even vaak voor (zie Tabel 17-16). De docenten gaven vaker les in vmbo dan in havo/vwo en vaker in de eerste drie leerjaar. De helft van de docenten was betrokken bij bijspijkerlessen en ongeveer een derde bij verrijkingslessen.

Tabel 17-16. Onderwijsachtergrond van de 43 docenten

Achtergrond docenten	Ja		Nee	
	N	%	N	%
Docent Engels	15	34,9%	28	65,1%
Docent Nederlands	14	32,6%	29	67,4%
Docent wiskunde	13	30,2%	30	69,8%
docent ander vak	3	7,0%	40	93,0%
Geeft les in vmbo-bb	14	32,6%	29	67,4%
Geeft les in vmbo-kb	15	34,9%	28	65,1%
Geeft les in vmbo-gt	15	34,9%	28	65,1%
Geeft les in havo	20	46,5%	23	53,5%
Geeft les in vwo	22	51,2%	21	48,8%
Geeft les in andere leerweg	3	7,0%	40	93,0%
Geeft les in leerjaar 1	33	76,7%	10	23,3%
Geeft les in leerjaar 2	32	74,4%	11	25,6%
Geeft les in leerjaar 3	30	69,8%	13	30,2%
Geeft les in leerjaar 4	22	51,2%	21	48,8%
Geeft les in leerjaar 5	9	20,9%	34	79,1%
Geeft les in leerjaar 6	10	23,3%	33	76,7%
Bent u betrokken bij verrijkingslessen?	19	33,3%	38	66,7%
Bent u betrokken bij bijspijkerlessen?	29	50,9%	28	49,1%

17.4.2 Grafische weergave van de diagnoses (bolletjes)

In de leerling- en groepsrapportages kunnen de diagnoses op drie wijzen gevisualiseerd worden. In de standaardweergave wordt per aspect alleen de meest waarschijnlijke diagnose getoond met een bolletje. In de genuanceerde weergave is de relatieve waarschijnlijkheid van de andere diagnoses ook gevisualiseerd. Naast een bolletjes bij de meest waarschijnlijke diagnose, zijn de andere twee diagnoses ook getoond. Hoe waarschijnlijker de diagnose hoe groter en zichtbaarder het bolletje is. In de meest gedetailleerde weergave variëren alle bolletjes in grootte en zichtbaarheid (ook de meest waarschijnlijke diagnose), afhankelijk van de waarschijnlijkheid.



Figuur 17-23. De drie visualisaties van de diagnoses

Ondanks dat de meerderheid van de docenten de standaardweergave als eerste wilden zien (respectievelijk 49% en 63% bij de leerling- en groepsrapportage, zie Tabel 17-17), wilden de meeste docenten vooral de genuanceerde weergave aangeboden krijgen zowel in de leerlingrapportage (65%) als de groepsrapportage (74%, zie Tabel 17-18). Maar ook de standaardrapportage was vaak gewenst: respectievelijk 48% (bij de leerlingrapportage) en 62% van de respondenten (bij de groepsrapportage) wilden de standaardweergave aangeboden krijgen. De meest genuanceerde weergave was beduidend minder in trek (27% bij de leerlingrapportage en 22% bij de groepsrapportage, zie Tabel 17-18).

Er werd ook gevraagd of de docenten een voorkeur hadden voor een grafische (met bolletjes), getalsmatige of tekstuele weergave en of ze die wel eens zouden gebruiken. De docenten hadden duidelijk een voorkeur voor de grafische weergave: respectievelijk 60% bij de leerlingrapportage en 70% bij de groepsrapportage had een voorkeur voor de weergave met de bolletjes boven een getalsmatige of tekstuele weergave. De grote meerderheid zou zowel bij de leerlingrapportage (73%) als bij de groepsrapportage (84%) de grafische weergave wel eens gebruiken. Net iets minder dan de helft van de docenten (respectievelijk 49% bij de leerlingrapportage en 48% bij de groepsrapportage) gaf aan dat ze de getalsmatige rapportage wel eens zouden gebruiken. Bij de leerlingrapportage zou 37% een tekstuele weergave van de diagnose wel eens gebruiken en bij de groepsrapportage slechts 4%.

Tabel 17-17. Voorkeursweergave van de docenten

		De standaardweergave	De genuanceerde weergave	De meest genuanceerde weergave	Anders, namelijk	Totaal
Welke weergave zou u als eerste willen zien bij het openen van de leerlingrapportage?	F %	28 49,1%	20 35,1%	3 5,3%	0 0,0%	23
Welke weergave zou u als eerste willen zien bij het openen van de klasrapportage?	F %	36 63,2%	9 15,8%	2 3,5%	2 3,5%	11
		Weergave van de diagnose door bolletjes	Weergave van de diagnose door getallen	Weergave van de diagnose door een uitgeschreven tekst	Een andere weergave, namelijk	
Welke weergave van de diagnose heeft uw voorkeur [in de leerlingrapportage]?	F %	34 59,6%	10 17,5%	4 7,0%	2 3,5%	48
Welke weergave van de diagnose heeft uw voorkeur voor de klasrapportage?	F %	40 70,2%	7 12,3%	0 0,0%	2 3,5%	47

Tabel 17-18. Gewenste weergaves van de docenten

		Ja		Nee		Totaal
		F	%	F	%	F
Leerlingrapportage						
Welke weergaves zou u aangeboden willen krijgen in de leerlingrapportage?	De standaardweergave	25	48,1%	27	51,9%	52
	De genuanceerde weergave	34	65,4%	18	34,6%	52
	De meest genuanceerde weergave	14	26,9%	38	73,1%	52
	Anders	2	3,8%	50	96,2%	52
Welke weergave zou u wel eens gebruiken?	Weergave van de diagnose door bolletjes	37	72,5%	14	27,5%	51
	Weergave van de diagnose door getallen	25	49,0%	26	51,0%	51
	Weergave van de diagnose door een uitgeschreven tekst	19	37,3%	32	62,7%	51
	Een andere weergave	2	3,9%	49	96,1%	51
Groepsrapportage						
Welke weergaves zou u aangeboden willen krijgen in de klasrapportage?	De standaardweergave	31	62,0%	19	38,0%	50
	De genuanceerde weergave	37	74,0%	13	26,0%	50
	De meest genuanceerde weergave	11	22,0%	39	78,0%	50
	Anders	2	4,0%	48	96,0%	50
Welke weergave van de klasrapportage zou u wel eens gebruiken?	Weergave van de diagnose door bolletjes	42	84,0%	8	16,0%	50
	Weergave van de diagnose door getallen	24	48,0%	26	52,0%	50
	Weergave van de diagnose door een uitgeschreven tekst	2	4,0%	48	96,0%	50

De respondenten is ook gevraagd naar hun mening over de betekenis van de verschillende grafische weergaves van de diagnoses (de bolletjes). De overgrote meerderheid was het eens of helemaal eens dat de betekenis van de bolletjes in de standaardweergave (96%) en in de genuanceerde weergave (80%) duidelijk was. De betekenis van de bolletjes in de meest genuanceerde weergave was beduidend minder vaak duidelijk (52% was het eens of helemaal eens, zie Tabel 17-19). Meer dan zeventig procent van de respondenten vond de betekenis van de grootte en zichtbaarheid van de bolletjes duidelijk (71% was het eens of helemaal eens, zie Tabel 17-19) en 70% vond de additionele informatie waardevol. Wederom was men het minder positief over de meest genuanceerde weergave: 55% was het eens of helemaal eens dat de betekenis van de grootte en zichtbaarheid van de bolletjes duidelijk is en 38% vond de additionele informatie waardevol.

Tabel 17-19. Mening over de grafische weergave (bolletjes) van de diagnoses

		Helemaal mee eens	Mee eens	Enigszins mee eens	Enigszins mee oneens	Mee oneens	Helemaal mee oneens	Totaal
De betekenis van de kleurenbolletjes is mij duidelijk.	<i>F</i> %	28 53,8%	22 42,3%	2 3,8%	0 0,0%	0 0,0%	0 0,0%	52
De betekenis van de extra kleurenbolletjes in de genuanceerde weergave is duidelijk voor mij.	<i>F</i> %	11 22,0%	29 58,0%	6 12,0%	3 6,0%	1 2,0%	0 0,0%	50
De betekenis van de extra kleurenbolletjes in de meest genuanceerde weergave is duidelijk voor mij.	<i>F</i> %	6 12,0%	20 40,0%	14 28,0%	6 12,0%	4 8,0%	0 0,0%	50
De betekenis van de grootte en zichtbaarheid van de extra kleurenbolletjes in de genuanceerde weergave is duidelijk voor mij.	<i>F</i> %	8 16,3%	27 55,1%	7 14,3%	5 10,2%	1 2,0%	1 2,0%	49
De betekenis van de grootte en zichtbaarheid van de kleurenbolletjes in de genuanceerde weergave is duidelijk voor mij.	<i>F</i> %	5 10,0%	22 44,0%	11 22,0%	7 14,0%	4 8,0%	1 2,0%	50
De genuanceerde weergave biedt waardevolle informatie over de diagnose ten opzichte van de standaardweergave.	<i>F</i> %	6 12,0%	29 58,0%	12 24,0%	2 4,0%	1 2,0%	0 0,0%	50
De meest genuanceerde weergave biedt waardevolle informatie over de diagnose ten opzichte van de standaardweergave.	<i>F</i> %	5 10,0%	14 28,0%	18 36,0%	7 14,0%	6 12,0%	0 0,0%	50

17.4.3 Startscherm

Ten aanzien van het startscherm gaf 30% van de respondenten aan ook andere groepen te willen kunnen selecteren, zoals mentorgroepen, een heel leerjaar, clustergroepen (per vakdocent), jongens/meisjes of leerlingen uit verschillende klassen, die bijvoorbeeld een taalklas volgen. Bijna 93% van de respondenten miste niks anders op het startscherm (zie Tabel 17-20).

Tabel 17-20. Evaluatie startscherm

	Ja		Nee		Totaal
	<i>F</i>	%	<i>F</i>	%	<i>N</i>
Zou u in het startscherm van het rapportagemodel, naast de selectie van de klas, andere groepen willen kunnen selecteren? U kunt daarbij bijvoorbeeld denken aan het selecteren van mentorgroepen.	17	30,4%	39	69,6%	56
Zijn er overige zaken die u mist in het startscherm van het rapportagemodel?	4	7,3%	51	92,7%	55

17.4.4 Leerlingrapportages

Bij de leerlingrapportages vond 89% van de respondenten de rapportage duidelijk en vond 96% de diagnoses van de deelaspecten (verdiepende diagnoses) duidelijk. De grote meerderheid van de docenten zouden de leerlingrapportages laten zien aan de leerling (86%) en aan de ouders (73%), zoals in Tabel 17-21 te zien is. Indien ze de rapportage niet wilden tonen was tijdgebrek (en de tijd die de uitleg van de rapportage zou kosten) de belangrijkste reden.

Tabel 17-21. Duidelijkheid en mogelijk gebruik leerlingrapportages

	F	Ja %	F	Nee %	Totaal N
Is het diagnostisch profiel duidelijk voor u?	47	88,7%	6	11,3%	53
Is het diagnostisch profiel in combinatie met de verdiepende diagnose duidelijk voor u?	48	96,0%	2	4,0%	50
Zou u het diagnostisch profiel aan de ouders van een leerling laten zien, bijvoorbeeld tijdens een tienminutengesprek?	37	72,5%	14	27,5%	51
Zou u het diagnostisch profiel aan de leerling laten zien?	44	86,3%	7	13,7%	51

De meningen over de leerlingrapportages waren zeer positief, slechts een enkeling was in enige mate negatief of de informatie waardevol was ($F = 2$) en voldoende was ($F = 8$), zie Tabel 17-22. Van de respondenten was 72% het minstens enigszins eens dat de leerlingrapportage waardevol was, 83% dat de diagnoses op deelaspecten waardevol waren en 69% dat de deelaspect diagnoses essentieel waren. Zestig procent vond het diagnostisch profiel voldoende informatie geven.

Tabel 17-22. Mening over de leerlingrapportage

		Helemaal mee eens	Mee eens	Enigszins mee eens	Enigszins mee oneens	Mee oneens	Helemaal mee oneens	Totaal
Het diagnostisch profiel geeft waardevolle informatie over de sterke en minder sterke punten van de leerling.	F %	4 7.5%	34 64.2%	13 24.5%	1 1.9%	1 1.9%	0 0.0%	53
De verdiepende diagnose levert waardevolle informatie over de leerling, bovenop de informatie uit het diagnostisch profiel.	F %	7 13.5%	36 69.2%	7 13.5%	1 1.9%	1 1.9%	0 0.0%	52
Voor een waardevolle diagnose is de verdiepende diagnose essentieel.	F %	12 23.5%	23 45.1%	14 27.5%	1 2.0%	1 2.0%	0 0.0%	51
Het diagnostisch profiel geeft voldoende informatie over de sterke en minder sterke punten van de leerling.	F %	4 7,5%	28 52,8%	13 24,5%	4 7,5%	4 7,5%	0 0,0%	53

17.4.5 Groepsrapportages

Bijna alle respondenten waren het eens of helemaal eens (92%) dat de interpretatie van overall diagnose (eerste kolom) van de vakken duidelijk was (92%) en dat de diagnoses van de aspecten (diagnostisch profiel) duidelijk was (86%). Tachtig procent was het (helemaal) eens dat de weergave zorgde voor een goede interpretatie. In de groepsrapportage vond men de diagnoses van de deelaspecten (verdiepende diagnoses) wat minder belangrijk. Iets meer dan de helft (55%) was het eens of helemaal eens dat de verdiepende diagnoses essentieel zijn en 74% dat de verdiepende diagnoses van toegevoegde waarde zou zijn.

Over de aangebrachte arcering in de groepsrapportage was men minder uitgesproken positief: 62% was het (helemaal) eens dat de arcering functioneel was en 55% was het (helemaal) eens dat de arcering functioneel van toegevoegd waarde was (zie Tabel 17-23).

Tabel 17-23. Mening over de groepsrapportage

Groeps-rapportage		Helemaal mee eens	Mee eens	Enigszins mee eens	Enigszins mee oneens	Mee oneens	Helemaal mee oneens	Totaal
De interpretatie van de kolom 'Wiskunde' is duidelijk.	<i>F</i> %	16 32,7%	29 59,2%	3 6,1%	1 2,0%	0 0,0%	0 0,0%	49
De interpretatie van het diagnostisch profiel in de klasrapportage is mij duidelijk.	<i>F</i> %	7 14,3%	35 71,4%	2 4,1%	4 8,2%	1 2,0%	0 0,0%	49
De arceringen in het diagnostisch profiel zijn van toegevoegde waarde voor de interpretatie van de klasrapportage.	<i>F</i> %	4 8,2%	23 46,9%	12 24,5%	2 4,1%	8 16,3%	0 0,0%	49
Het tonen van een verdiepende diagnose zou van toegevoegde waarde zijn voor de interpretatie van de klasrapportage	<i>F</i> %	7 14,3%	29 59,2%	7 14,3%	4 8,2%	2 4,1%	0 0,0%	49
Het tonen van de verdiepende diagnose is essentieel voor de interpretatie van de diagnose op klasniveau.	<i>F</i> %	5 10,2%	22 44,9%	13 26,5%	3 6,1%	6 12,2%	0 0,0%	49
De manier van weergave van het diagnostisch profiel zorgt voor een goede interpretatie van de diagnose op klasniveau.	<i>F</i> %	5 10,2%	34 69,4%	7 14,3%	1 2,0%	2 4,1%	0 0,0%	49
De arceringen bij de groene en oranje bolletjes vind ik functioneel.	<i>F</i> %	8 16,0%	23 46,0%	8 16,0%	3 6,0%	8 16,0%	0 0,0%	50

Cito

Amsterdamseweg 13
6814 CM Arnhem
Postbus 1034
6801 MG Arnhem
T (026) 352 11 11

Fotografie: Gijs Versteeg

