

Digitaal grootschalig summatief toetsen in het primair onderwijs

Onderzoeksrapportage

maart 2016,
Desirée Joosten-ten Brinke
Annemarijn Weber
Gregorij Ljubin Saveski

Open Universiteit



Dit rapport is geschreven in opdracht van Kennisnet.

VOORWOORD

Voor u ligt het onderzoeksrapport Digitaal grootschalig summatief toetsen in het primair onderwijs. Deze rapportage is het resultaat van een onderzoek dat in opdracht van Kennisnet is uitgevoerd.

In het laatste jaar van het primair onderwijs nemen alle leerlingen deel aan een eindtoets primair onderwijs. Hiervoor hebben scholen momenteel de keuze uit toetsen van drie aanbieders. Twee aanbieders werken met een schriftelijke toets en één aanbieder met een computer adaptieve toets. Bij een computer adaptieve toets krijgt elke leerling een unieke toets doordat toetsvragen geselecteerd worden op basis van gegeven antwoorden op voorgaande toetsvragen. In de toekomst zal steeds vaker adaptief getoetst worden. Maar wat betekent dat voor de scholen? Welke eisen stelt computer (adaptief) toetsen aan de afnamecondities en infrastructuur van de scholen in het primair onderwijs? Deze vraag staat aan de basis van dit onderzoeksrapport.

Aangezien computer (adaptief) toetsen niet volledig nieuw is, nam Kennisnet het initiatief om te kijken wat er te leren is van voorlopers op dit gebied. In een eerste desk research werden relevante aanbieders gevonden. Met veel interesse hebben we (een deel van) deze aanbieders geïnterviewd over de stappen die zij gezet hebben in het ontwikkelen van hun eindtoetsen. We danken de gesprekspartners voor hun bereidheid om deel te nemen aan de interviews en openheid te geven over de eisen die zij stellen aan de infrastructuur en organisatie van scholen.

Wij hopen dat de aanbevelingen in deze rapportage Kennisnet handvatten geeft voor verdere advisering over de implementatie van grootschalige digitale toetsen. Tevens bevat het rapport aanbevelingen voor toetsaanbieders, scholen en onderzoeks- en kenniscentra voor de verdere ontwikkeling en evaluatie van digitaal toetsen in het primair onderwijs.

Wij wensen u veel leesplezier!

Desirée Joosten-ten Brinke, Annemarijn Weber en Gregorij Ljubin

Heerlen, Open Universiteit, februari 2016

INHOUDSOPGAVE

Voorwoord.....	3
Samenvatting.....	7
1. Inleiding.....	9
1.1 Aanleiding en probleemstelling.....	9
1.2 Onderzoeksvraag.....	10
2. Methode.....	11
2.1 Participanten.....	11
2.2 Instrumenten.....	11
2.3 Opzet en procedure.....	12
2.4 Data-analyse.....	12
3. Resultaten.....	13
3.1 Beschrijvingen van de toetsen.....	13
3.2 Onderzoeksvraag 1. Afnamecondities en ict-infrastructuur.....	23
3.2.1 Welke eisen worden gesteld aan de netwerkinfrastructuur?.....	23
3.2.2 Welke eisen worden gesteld aan de devices?.....	26
3.2.3 Welke organisatorische eisen worden gesteld aan de scholen?.....	28
3.2.4 Welke eisen worden gesteld aan de privacy?.....	30
3.2.5 Geleerde lessen.....	31
3.3 Onderzoeksvraag 2. Impact van digitaal toetsen op toetsresultaten.....	36
3.3.1 Toetsresultaten bij digitale (adaptieve) toetsen vs. schriftelijke toetsen.....	36
3.3.2 Beïnvloedende variabelen.....	37
3.3.3 Wat zeggen de aanbieders over de impact?.....	38
4. Conclusie en discussie.....	39
4.1 Afnamecondities en ict-infrastructuur bij digitaal toetsen.....	39
4.2 Impact van digitaal adaptief toetsen.....	43
4.3 Aanbevelingen.....	45
5. Referenties.....	49
Bijlagen.....	52
A. Interviewprotocol.....	53
B. Urenverantwoording.....	56

SAMENVATTING

Computer adaptief toetsen is de toekomst van de Nederlandse eindtoetsen in het primair onderwijs. Maar welke eisen stelt dit aan de scholen voor wat betreft de schoolinrichting en de ict-infrastructuur en wat is de impact op toetsresultaten bij de verschuiving van de huidige schriftelijke toetsen naar computer adaptief toetsen? Deze vraag staat centraal in dit onderzoek.

Aan de hand van een gestructureerde vragenlijst hebben vier interviews met (inter)nationale aanbieders/ontwikkelaars van een computer adaptieve toets en drie interviews met aanbieders/ontwikkelaars/gebruikers met niet-adaptieve digitale summatieve toetsen plaatsgevonden. De vraag naar de impact op toetsresultaten is onderzocht met een korte literatuurstudie.

De vier aanbieders van een adaptieve toets zijn allen zeer enthousiast over het adaptieve karakter, doordat er in minder toetstijd een meer nauwkeurige uitspraak gedaan kan worden over de cognitieve vaardigheid van de leerling, de toets aansluit op het niveau van de individuele leerling en er meer mogelijkheden zijn met betrekking tot de organisatie van de toetsafname. Ondanks dat het een grote investering vraagt om te komen tot goede en voldoende grote itembanken, adviseren ze alle vier om de stap richting adaptief toetsen te zetten.

De netwerkinfrastructuur die nodig is voor het goed afnemen van de toetsen varieert. Sommige aanbieders geven daarvoor expliciet minimumeisen aan, maar veelal wordt aangegeven dat de eisen die landelijk ingesteld zijn voldoende zijn. Deze eisen staan los van de mate van adaptiviteit van de toets. Bij een adaptieve toets zijn aanvullende eisen ten aanzien van bijvoorbeeld een pretest en de omvang van de itembank. Om te voorkomen dat zich problemen voordoen met het downloaden van de toetsen (bij alle digitale toetsen), wordt de grootte van de te downloaden bestanden zo klein mogelijk gehouden. Dit beperkt de mogelijkheden van het gebruik van multimedia in items.

De toetsen worden vooral afgenomen op computers en laptops. Een enkele toets mag afgenomen worden op tablets. Geen van de aanbieders staat toe dat een eigen device wordt meegenomen. In een aantal landen worden de devices door de overheid geleverd. De benodigde randapparatuur en operating systemen variëren per aanbieder. De inrichting van de afname is een verantwoordelijkheid van de scholen.

Privacy van de leerlingen wordt door alle aanbieders belangrijk gevonden. Om dit te realiseren wordt gewerkt met persoonlijke inlogcodes en worden data geanonimiseerd indien ze gedeeld worden met derden. Voor de informatiebeveiliging zijn in een aantal landen protocollen beschikbaar. Hierin wordt expliciet vermeld welke data bewaard worden en op welke manier. In Amerika wordt daar expliciet aan toegevoegd wat niet opgeslagen wordt.

Het literatuuronderzoek naar de impact van digitaal (adaptief) toetsen laat zien dat de toetsresultaten niet altijd gelijk zijn voor een schriftelijke toets en een digitale (adaptieve) toets. Deze verschillen zijn in het algemeen klein en de verschillen lijken in de laatste jaren steeds minder vaak significant te zijn. Er tekent zich bij leerlingen echter wel een voorkeur af voor digitale toetsen. Ten aanzien van de perceptie van adaptief toetsen is vervolgonderzoek nodig.

De interviews en het literatuuronderzoek hebben geleid tot aanbevelingen voor ontwerpers en ontwikkelaars van computer (adaptieve) toetsen, voor aanbieders en scholen en voor onderzoeksinstituten en kenniscentra.

1. INLEIDING

1.1 AANLEIDING EN PROBLEEMSTELLING

In groep acht van het primair onderwijs wordt door de leerlingen een centrale eindtoets of andere eindtoets afgelegd. Welke eindtoets dit is, mogen scholen zelf besluiten (Ministerie van OCW, 2014). De eindtoets meet welk eindniveau de leerling heeft behaald ten opzichte van de referentieniveaus voor Nederlandse taal en Rekenen. Aan de hand van de resultaten levert de eindtoets vervolgens een advies op over het te volgen vervolgonderwijs.

De afnamevorm van elke toets, en daarmee ook van deze eindtoetsen moet passen bij de te meten vaardigheden. Om die reden wordt een aantal domeinen van Nederlandse taal en rekenen nu nog niet meegenomen in de eindtoets, omdat die niet goed meetbaar zijn met een schriftelijke afnamevorm. De verwachting is dat een digitale afname van de eindtoets ervoor kan zorgen dat taal en rekenen breder te toetsen zijn, waardoor de evaluatiedoelstelling van de eindtoets versterkt kan worden (Ministerie van OCW, 2014). Daarnaast zou een digitale toets ook beter kunnen aansluiten op de belevingswereld van de leerlingen en heeft een digitale toets organisatorische voordelen zoals een verkorte correctietijd en een grotere flexibiliteit rond de afname (CvTE, 2008). Bij deze digitale afnamevorm wordt het ook mogelijk om adaptief te toetsen.

Adaptief toetsen wil zeggen dat de items die een leerling moet maken gebaseerd worden op de vaardigheid van de leerling, zodat er nauwkeuriger gemeten kan worden. Nadat een leerling een toetsvraag (of een set toetsvragen) heeft beantwoord, selecteert de computer een volgende toetsvraag (of set van toetsvragen) op basis van het eerder gegeven antwoord. De selectie van een item uit een itembank gebeurt dus pas tijdens de afname. Maakt een leerling een item goed dan krijgt hij een moeilijker item en maakt de leerling een item fout dan krijgt hij een makkelijker item. Op die manier krijgt elke leerling een unieke toets en wordt de vaardigheid van de leerling efficiënter vastgesteld dan bij een niet-adaptieve toets. Eén van de andere eindtoetsen die momenteel beschikbaar is voor het primair onderwijs, Route-8 van A-Vision, is adaptief en de centrale eindtoets zal naar verwachting vanaf 2018 adaptief zijn.

Een belangrijke voorwaarde voor digitale afname van een eindtoets is een goede ict-infrastructuur. Kennisnet heeft geïnventariseerd wat de huidige stand van zaken is ten aanzien van de ict-infrastructuur in het primair onderwijs. Deze is op dit moment nog niet op alle scholen voldoende om binnen enkele jaren de eindtoets digitaal af te nemen (Ministerie van OCW, 2014). Bovendien heeft Kennisnet een advies opgesteld over de gevolgen van de beschikbare ict-infrastructuur, of beter gezegd het nog ontbreken van een volledig geschikte ict-infrastructuur op alle scholen voor digitalisering van de eindtoets (Kennisnet, 2015). In 2014 is een online verkenning uitgevoerd om buitenlandse initiatieven op het gebied van grootschalige digitale summatieve toetsing in kaart te brengen (Kennisnet, 2014). In dat onderzoek is gezocht naar toetsen die gebruikt worden in het primair of voortgezet onderwijs, op grote schaal afgenomen worden, een summatieve functie hebben en cognitieve vaardigheden toetsen op basis van nationale standaarden. De conclusie van dat onderzoek is dat de implementatie van (adaptieve) digitale toetsing verschillende eisen met zich mee brengt:

- Eisen aan deskundigheid van de toetsontwerpers en –ontwikkelaars met betrekking tot de ontwikkeling van toetsvragen. Het maken van een digitale toets is niet gelijk aan het vertalen van een schriftelijke toets naar beeldscherm.
- Eisen aan de docenten ten aanzien van het gebruik van de informatie die digitale toetsen opleveren (bv. formatief gebruik van de summatieve toetsen).
- Eisen aan de begeleiding van de leerlingen in het afleggen van de toets en hun ict-vaardigheden.
- Eisen aan de randvoorwaarden voor afname (voldoende devices, beschikbaarheid van handleidingen, ...).

De online verkenning leverde nog onvoldoende informatie op over de inrichting van de toetslocaties, de kenmerken van de ict-infrastructuur en de impact van digitaal toetsen op de toetsresultaten van leerlingen. Ook de vertaling van deze informatie naar de Nederlandse situatie ontbreekt nog.

Op basis daarvan en op basis van de behoefte vanuit het onderwijs naar meer informatie over de mogelijkheid om de eindtoetsen digitaal af te nemen heeft Kennisnet de Open universiteit de opdracht gegeven een verdiepend onderzoek uit te voeren om de expertise op dit gebied te verdiepen en te vertalen naar het primair onderwijs. Doel van dit verdiepende onderzoek is inzicht in succes- en faalfactoren bij de implementatie van digitaal toetsen in het primair onderwijs in Nederland. Met de uitkomsten van dit onderzoek kan Kennisnet een koers bepalen bij de advisering over implementatie van grootschalig summatief toetsen.

1.2 ONDERZOEKSVRAAG

De centrale vraag in dit onderzoek is: Aan welke afnamecondities en ict-infrastructuur-eisen moeten scholen voor primair onderwijs in Nederland voldoen, zodat het afnemen van (adaptieve) digitale toetsen geen negatieve gevolgen heeft voor de toetsresultaten van de leerlingen?

Deze onderzoeksvraag wordt uitgewerkt aan de hand van twee deelvragen:

1. Welke afnamecondities en eisen aan de ict-infrastructuur gelden voor het digitaal afnemen van een summatieve toets?
2. Welke impact kan (adaptief) digitaal toetsen hebben op toetsresultaten van leerlingen en hoe kan een eventuele negatieve impact voorkomen worden?

2. METHODE

2.1 PARTICIPANTEN

Interviews hebben plaatsgevonden met ontwerpers, ontwikkelaars of gebruikers van de volgende toetsen:

- CAT op de Folkeskole, Denemarken: ict-coördinator
- Smarter Balanced Assessments, VS: technology team SBAC
- Eindexamens, Noorwegen: teamleider van een school en daarvoor projectleider tijdens de toetsontwikkeling
- 10voordeleraar, Nederland; coördinator van het kernteam 10voordeleraar
- Route-8 van A-Vision, Nederland; operationeel manager, verantwoordelijk voor de interne aansturing
- Eindexamens, Macedonië; lerares primair onderwijs en gebruiker van de toets
- NILA en NINA, Noord-Ierland; programma en onderwijsmanager CCEA

Uit het vooronderzoek leken ook het Partnership for Assessment of Readiness for College and Careers (PARCC), het National Assessment of Educational Progress (NAEP) en Georgië interessant om te interviewen. PARCC en SBAC zijn echter gebaseerd op dezelfde standaard en subsidieregeling, waardoor de gestelde eisen vergelijkbaar zijn. Het enige verschil tussen beide consortia is dat SBAC een adaptieve variant aanbiedt. Helaas bleek de aanbieder uit Georgië geen Engelssprekende contactpersoon te hebben. Om deze reden is er gekozen om een lerares uit Macedonië te interviewen aangezien zij gekozen hebben voor een soortgelijke nationale toets. NAEP was in de periode van het onderzoek niet in de gelegenheid om geïnterviewd te worden.

2.2 INSTRUMENTEN

In dit onderzoek is gebruik gemaakt van semi-gestructureerde telefonische interviews en bronanalyse (voor de beantwoording van onderzoeksvraag 1) en van een gestructureerd literatuuronderzoek (voor het beantwoorden van onderzoeksvraag 2).

Semi-gestructureerde interviews

Er is gekozen voor interviews in plaats van vragenlijsten, omdat in dit onderzoek behoefte is aan verdieping. Vragenlijstonderzoek is geschikt voor het verkrijgen van veel informatie, maar de verkregen informatie is over het algemeen vrij algemeen (Verhoeven, 2014). De interviews zijn telefonisch afgenomen, aangezien de respondenten (inter)nationaal verspreid zitten, er weinig reistijd beschikbaar is, en de sociale context geen onderwerp van het onderzoek is (Opdenakker, 2006).

Voor de interviews is een interviewprotocol ontwikkeld (zie bijlage A), waarbij vooraf relevante bronnen over ict-infrastructuur zijn geraadpleegd (Kennisset, 2015; Open Universiteit, 2008; SURF, 2014).

Literatuurstudie

Om onderzoeksvraag 2 te beantwoorden is een korte literatuurstudie uitgevoerd (Jesson, Matheson, & Lacey, 2011). Via de onderzoeksdatabase Ebsco is gezocht op zoektermen 'computer based' OR 'computer adaptive' OR 'paper-pencil' AND 'assessment' OR 'testing' AND

'impact' OR 'perception' AND 'performance' or 'score'. Er is niet beperkt in publicatieperiode. In totaal zijn 24 publicaties meegenomen in de analyse.

2.3 OPZET EN PROCEDURE

Het opzetten van de studie en de interviews is in verschillende fases gebeurd. Hieronder een korte uitleg van deze opzet en procedure.

2.3.1 Benaderen van respondenten

De lijst van landen en desbetreffende toetsen is van te voren vastgelegd aan de hand van een eerder rapport van Kennisnet (Kennisnet, 2014). Macedonië, 10voordeleraar en A-Vision zijn toegevoegd aan deze lijst omdat zij een waardevolle toevoeging konden zijn voor dit onderzoek gezien de afnamevorm en/of context. Per land en toets is online gezocht naar contactpersonen. Voor Noord-Ierland, de Verenigde Staten, Noorwegen en Nederland (A-Vision) leidde deze zoekopdracht tot respondenten die nauw betrokken zijn geweest bij de ontwikkeling van de desbetreffende toetsen. Voor Denemarken, Macedonië en Nederland (10voordeleraar) is er gebruik gemaakt van contacten uit het eigen netwerk. De respondenten van Denemarken en Macedonië zijn gebruikers van de toets. De respondent van 10voordeleraar is zelf ook nauw betrokken bij de ontwikkeling van de toetsen.

2.3.2 Afname interviews

De respondenten kregen bij het vastleggen van hun interview de interviewvragen toegestuurd (zie bijlage A). Hierdoor konden zij de antwoorden op de vragen voorbereiden en eventueel navragen bij collega's. Ook was het voor enkele respondenten noodzakelijk om de vragen van te voren te krijgen in verband met wettelijke verplichtingen vanuit hun bedrijf.

Er is gekozen voor semigestructureerde interviews in dit onderzoek. De vragen en thema's, en de volgorde hiervan, waren van te voren bepaald, maar er was tegelijkertijd ook nog ruimte binnen het interview om vragen toe te voegen naar aanleiding van eerdere antwoorden. Door deze vorm van interviews aan te houden is er meer diepgang ontstaan en kwamen de individuele bijzonderheden van elke toetsvorm naar voren. Elk interview duurde ongeveer 1,5 uur. Tijdens de interviews is er zoveel mogelijk gewerkt met Skype, zodat de interviewer ook kon letten op non-verbale communicatie. Dit zorgde ook voor een natuurlijker gesprek dan wanneer het interview telefonisch werd afgenomen. Bij het interview met A-Vision (Nederland) en met de CCEA (Noord-Ierland) was een Skypegesprek helaas niet mogelijk. Alle interviews zijn opgenomen met een voice recorder en naderhand uitgeschreven in een Word document. Het gebruik van de voice-recorder stelde de interviewer in staat zich enkel te richten op het gesprek met de geïnterviewde.

2.4 DATA-ANALYSE

De interviewdata zijn geclusterd op basis van de afbakening die Kennisnet hanteert voor de beschrijving van ict-infrastructuur: Connectiviteit, Devices, Software, en Privacy. Daarnaast is ingegaan op de toetsinhoud en aanbevelingen uit de praktijk. Op basis van de data is gekeken naar overeenkomsten en verschillen tussen de verschillende toetsen. Daarna is gekeken in hoeverre de resultaten te vertalen zijn naar de context van de eindtoetsen in het primair onderwijs in Nederland.

3. RESULTATEN

Achtereenvolgens volgt een algemene beschrijving van elke toets. Daarna wordt het doel, de inhoud, de doelgroep, de eigenaar van de toets en de wijze van het delen van de data beschreven.

3.1 BESCHRIJVINGEN VAN DE TOETSEN

CAT op de Folkeskole, Denemarken

Algemene beschrijving van de toets

De CAT is een computer adaptieve formatieve toets die leerlingen één keer per jaar maken in het 3^e tot het 8^e leerjaar van het basisonderwijs. De toetsmomenten worden verspreid over een periode van drie weken om het toetsysteem zoveel mogelijk te ontlasten. Leerkrachten moeten hun klas inschrijven voor een datum en tijdstip binnen deze periode. De overheid heeft op de inschrijvingen per tijdstip een limiet gezet. Leerlingen moeten daarom tijdig ingeschreven worden. Op de dag van het examen loggen de leerlingen allemaal in op de website van de toets met hun persoonlijke YouAndI inloggegevens. Binnen deze portal worden al hun cijfers en eerdere prestaties bewaard; YouAndI kan dus ook gebruikt worden om de voortgang te bekijken van elke leerling. De toets zelf wordt gemaakt op iPad Mini's die verstrekt worden door de overheid en bestaat uit meerkeuzevragen en andere simpele vraagvormen die één voor één worden gesteld en die geschikt zijn voor een touchscreen. De standaardtijd die gegeven wordt is 45 minuten. Er wordt echter bijna altijd gekozen om deze tijd met 15 minuten te verlengen. Een interessant aspect van deze toets is dat de leerkracht 'achter de schermen' mee kan kijken naar de voortgang van een leerling en op basis van een stoplichtsysteem kan besluiten wanneer een leerling mag stoppen met de toets. Als de status nog rood is, dan heeft de leerling nog niet genoeg vragen beantwoord om een goed beeld te scheppen van het niveau van de leerling. Wanneer de status vervolgens oranje kleurt is de leerling bijna klaar. Als de status vervolgens groen kleurt kan de leerkracht ervoor kiezen om de leerling te laten stoppen. Dit kan voor langzame leerlingen zeer prettig zijn omdat ze dan niet onnodig veel langer zitten dan hun klasgenoten. Leerlingen kunnen ook stoppen met de toets en vervolgens na de pauze of zelfs de dag erna weer verder gaan, mochten ze zich niet meer kunnen concentreren. Leerkrachten krijgen één dag na het examen de uitslagen en deze delen ze dan met de leerlingen en hun ouders. De cijfers worden gegeven op een schaal van 1 - 5. Deze schalen worden vergezeld door een advies om de prestaties te verbeteren.

Wat is het doel van de toets?

Het doel van deze toets is om de leerkracht een indruk te geven van de voortgang van een leerling. Door de cijfers van voorgaande jaren te vergelijken met huidige jaren, of individuele leerlingen met elkaar te vergelijken, kan een goed beeld worden verkregen van wie waar moeite mee heeft. Ook kan dit een leerkracht helpen om een lesplan op te stellen waar de hele klas iets aan heeft. De schoolleiding kan ook de scores van een klas bekijken en checken of een leerkracht ergens moeite mee heeft of niet goed presteert.

Welke domeinen worden getoetst?

Het verschilt per jaar wat verplicht is gesteld. Deens, leesvaardigheden, wetenschap worden verplicht getoetst in het 4e, 6e en 8e leerjaar. Wiskunde en Engels worden verplicht getoetst in het 3e en 6e leerjaar. Veel gemeentes kiezen er echter voor om in het jaar voor en na de

verplichte afname dezelfde toets ook af te nemen. Deze toetsen zijn dan alle drie op hetzelfde niveau, maar worden verspreid over drie jaar genomen. Een leerling zal dus in het 5e leerjaar voor het eerst de wiskundetoets van het 6e jaar maken, en dan in het 6e en in het 7e leerjaar nogmaals om zo een goed beeld te krijgen van de voortgang.

Welke doelgroep?

Deelname aan deze toetsen is verplicht gesteld voor elke school in Denemarken voor alle leerlingen in het 3e leerjaar tot en met het 8e leerjaar.

Wie is eigenaar van de toets?

De Ministerie van Onderwijs gaat over alles wat er op de openbare basisscholen (Folkeskole) gebeurt. Zij zijn verantwoordelijk voor het uitzetten en uitvoeren van de toetsen.

Hoe worden de landelijk verzamelde gegevens gedeeld?

De gegevens worden uitgebreid bekeken binnen een gemeente om te achterhalen of alles nog goed verloopt op de plaatselijke scholen. Vorig jaar was er bijvoorbeeld opgemerkt dat veel leerlingen voor sommige vakken slecht scoorden. Dit is doorgegeven aan de administratie van de school en waar nodig zijn er veranderingen gemaakt in het onderwijs. Of de gegevens ook landelijk gebruikt worden was niet bekend.

Smarter Balanced Assessments, SBAC, VS

Algemene beschrijving van de toets

De Smarter Balanced Assessments vormen een gebalanceerde set van formatieve, summatieve en tussentijdse digitale toetsen waarbij een deel adaptief is. De toetsresultaten worden summatief voor verantwoording gebruikt en formatief voor instructiedoeleinden. Daarnaast zijn er tussentijdse toetsen.

Deze assessments worden ontwikkeld en aangeboden door Smarter Balanced Assessments Consortium (SBAC). Zij hebben samen met Partnership for Assessment of Readiness for College and Careers (PARCC) in 2010 een overheidssubsidie ontvangen uit het Race to the Top assessments fonds om toetsen voor wiskunde en Engels te ontwikkelen en aan te bieden voor het basisonderwijs. De toetsen sluiten aan bij de Common Core State Standards, waarin vastgesteld is wat leerlingen aan het eind van elk jaar moeten beheersen. SBAC en PARCC zijn twee concurrerende organisaties. Het verschil tussen de toetsen van SBAC en PARCC is dat een deel van de toetsen van SBAC adaptief zijn en van PARCC niet. SBAC is de grootste aanbieder. Vijftien van de 52 staten gebruiken de SBAC toetsen. In totaal zijn er 6 miljoen deelnemers aan de toetsen. De interim assessments worden niet gezien als high stakes (bedoeld voor zakken/slagen), maar geven advies over waar de leerling staat in zijn ontwikkeling.

De summatieve toets wordt afgenomen tijdens de laatste 12 weken van een schooljaar. De toets bestaat uit twee delen die beide via de computer afgenomen worden: een adaptieve deel voor kennisdoelen en een niet adaptief deel voor het meten van hogere orde vaardigheden, zoals kritisch denken. De toets beschrijft het niveau van de student en de ontwikkeling van de student ten opzichte van anderen binnen de school, het district en de staat.

De formatieve toetsen en de tussentijdse toetsen worden gepland door de scholen zelf. Inhoudelijk zijn ze vergelijkbaar met de summatieve toets. Ze worden alleen niet in een beveiligde omgeving afgenomen.

De toetsen zijn gebaseerd op open source architectuur en specifiek voor dit doel ontwikkeld onder de naam SmarterApp.org. De toetsen worden afgenomen met een secure browser om te voorkomen dat leerlingen naar andere applicaties switchen. De toets bestaat uit verschillende vraagvormen.

Welke domeinen worden getoetst?

De toets sluit aan bij de inhoudelijke eisen die gesteld worden in de Common Core State Standards (CCSS) voor Engels en wiskunde voor grade 3-8 en grade 11. Voor alle toetsen zijn voorbeeldtoetsen beschikbaar.

Welke doelgroep?

Leerlingen in grade 3-8 en grade 11 leggen elk jaar de formatieve toets af.

Wie is eigenaar van de toets?

De University of California is de intellectuele eigenaar. De federal law bepaalt dat de staat eigenaar is van de student data.

Hoe worden de landelijk verzamelde gegevens gedeeld?

Elke staat is zelf verantwoordelijk voor hoe ze met de gegevens om gaan. SBAC heeft een duidelijk privacyrichtlijn waarin aangegeven staat welke data wel en welke data niet opgeslagen worden.

Eindexamens in Noorwegen

Algemene beschrijving van de toets

Alle examens die afgenomen worden in Noorwegen zijn sinds 2008 digitaal. Deze keuze is gemaakt omdat leerlingen meer gewend waren aan werken met computers dan met papier en pen. De non-adaptieve summatieve eindexamens worden vlak voor de zomervakantie door het hele land op hetzelfde tijdstip afgenomen. Tijdens de 5 uur durende toets zitten leerlingen in een grote hal samen, soms met wel 500 leerlingen tegelijkertijd. De examens worden aangeboden via een examen management systeem (genaamd PAS) dat draait op een server van het Ministerie van Onderwijs en Training. Toegangscontrole, het versturen van de toetsvragen en het verzamelen van de antwoorden voor verwerking gebeurt via dit systeem. De eindexamens bestaan alleen maar uit open vragen. Leerlingen loggen hier met een eenmalige inlogcode in zodat alle examens totaal geanonimiseerd nagekeken worden.

Gedurende de rest van het school jaar worden alle overige formatieve en summatieve toetsen en opdrachten gegeven via het leermanagementsysteem (LMS; bv. ItsLearning) van de school. Hiervoor worden de persoonlijke inlogcodes gebruikt. Het is gebruikelijk dat alle leerlingen in een klas dezelfde toets krijgen met daarin meerkeuzevragen en open vragen, beide met multimedia, die dan vervolgens door de docent wordt nagekeken. Wat dat betreft lijken deze eindexamens op traditionele toetsen. Alle vragen worden ook tegelijkertijd gegeven. De reden

dat er voor PAS is gekozen voor de eindexamens in plaats van de school eigen LMS is dat PAS beter beveiligd is en dat alle examens automatisch verstuurd worden naar diegene die moet nakijken. De eisen die door het Ministerie gesteld zijn aan het examensysteem zijn:

- voor de leerlingen een vertrouwde omgeving;
- beschikken over de mogelijkheid om het resultaat om het examen aan te leveren bij op de server van het ministerie;
- uitvoeren van het examen zonder storingen en met een back-up faciliteit;
- voorkomen dat niet toegestane (communicatie-)hulpmiddelen of applicaties worden gebruikt;
- een eenvoudig systeem van toezicht voor de surveillanten;
- een eenvoudig systeem voor te bereiden en het opzetten van examens.

Beide softwarepakketten bevatten geen save-as-you-go functie omdat de examens gemaakt worden in Word Office software. Dit probleem is deels opgelost door het gebruik van het 3amiMAS logging systeem. Op de ochtend van een toets of eindexamen worden de leerlingen op de hoogte gebracht dat al hun acties gemonitord worden door 3amiMAS. Alles wat leerlingen doen wordt opgeslagen in dit systeem. Dit functioneert niet alleen als een save-as-you-go, maar voorkomt ook dat leerlingen frauderen op de examens.

Nadat de scholen de cijfers weer terugkrijgen van de desbetreffende beoordelaars, worden deze op het rapport van de leerling gezet en geanonimiseerd op de schoolportal. De uitslagen van de examens zijn niet alleen voor de leerlingen, maar worden ook gebruikt om scholen met elkaar te vergelijken.

Wat is het doel van de toets?

Het doel van de eindexamens die voor de zomervakantie worden afgenomen is om leerlingen individuele cijfers te geven voor hun rapportkaart. De cijfers worden geanonimiseerd ook gebruikt om scholen met elkaar te vergelijken. Ouders zouden in principe de prestaties van scholen kunnen vergelijken op internet om dan hun kinderen naar de beter presterende scholen te sturen. Dit kan echter alleen in stedelijke regio's, omdat er in de rest van Noorwegen teveel reistijd tussen scholen zit.

Leraren worden niet openbaar met elkaar vergeleken. Deze data zijn alleen intern beschikbaar, zodat de schoolleiding wel kan ingrijpen mocht een leerkracht slecht presteren.

Welke domeinen worden getoetst?

Dat ligt aan het vakkenpakket van de leerlingen. Iedereen moet minstens vier geschreven toetsen maken waarvan één voor het vak Noors. Verder ligt het aan de vakken die een leerling heeft gekozen: wiskunde, scheikunde, biologie, natuurkunde, Engels, Spaans, geschiedenis, enz.

Welke doelgroep?

De eindtoets wordt enkel afgenomen in het laatste jaar van de middelbare school. Leerlingen krijgen in de andere twee jaren van middelbare school ook digitale eindtoetsen maar deze zijn formatief, terwijl de eindtoets summatief is. De summatieve toets moet gehaald worden willen leerlingen hun diploma krijgen.

Wie is eigenaar van de toets?

De overheid – Ministerie van Onderwijs

Hoe worden de landelijk verzamelde gegevens gedeeld?

Alle gegevens worden geanonimiseerd in een schoolportal gezet. Dit is een openbare website waar iedereen de resultaten kan bekijken per school en ze met elkaar kan vergelijken.

10voordeleraar, Nederland

Algemene beschrijving van de toets

Deze toetsen worden afgenomen in het Hoger beroepsonderwijs bij lerarenopleidingen voor primair onderwijs (PABO's) en voor voortgezet onderwijs. Onder druk van de politiek heeft de ministerie van OCW het project 10voordeleraar opgestart onder de vlag van de Vereniging Hogescholen. In dit project zijn kennisbases opgesteld voor 61 vakken en voor 17 vakken zijn landelijke kennistoetsen ontwikkeld. Deze kennistoetsen zijn summatieve, niet adaptieve eindtoetsen die meestal in het derde of het vierde jaar van de lerarenopleiding afgenomen worden. Alle 17 toetsen zijn gemaakt met de software Questionmark Perception (QMP) en opgebouwd in blokken van meerkeuzevragen over verschillende domeinen. Binnen één blok kunnen studenten heen en weer gaan van item naar item, maar de items worden wel gelijk opgeslagen. Wanneer ze doorgaan naar het volgende blok kunnen ze niet meer terug naar het vorige blok. Alleen wiskunde is opgebouwd uit één groot blok. Studenten loggen in op de website met hun studentID en hieraan wordt uiteindelijk ook hun uitslag gekoppeld. De toetsen nemen 2 – 3 uur in beslag en studenten met een studiebeperking krijgen standaard een half uur extra.

Wat is het doel van de toets?

Het doel van deze toetsen is voornamelijk gericht op het beoordelen of studenten voldoende kennis bezitten om zelf voor een klas te staan. Deze toetsen worden allemaal digitaal afgenomen in het derde of vierde studjaar. Studenten moeten deze toets met een voldoende afsluiten om hun diploma te behalen.

Hiernaast kunnen hogescholen ook zien hoe zij scoren ten opzichte van de andere hogescholen, in totaal maar ook per domein. De Colleges van bestuur en de directeurs krijgen deze cijfers te zien met het duidelijke doel om van elkaar te leren. Opleidingen die het beter dan andere doen, delen hun aanpak met opleidingen die minder goed scoren. De gemiddelde cijfers zijn er ook over alle hogescholen heen, maar deze worden niet openbaar bekend gemaakt.

Welke domeinen worden getoetst?

De twee domeinen die getoetst worden bij de PABO's zijn rekenen en wiskunde en de Nederlandse taal. Tweedegraads lerarenopleidingen zijn er per schoolvak. Voor 17 vakken is een toets beschikbaar, bijvoorbeeld aardrijkskunde, geschiedenis, natuurkunde, biologie, scheikunde, Frans, Engels, Duits, omgangskunde, gezondheidszorg en welzijn, maatschappijleer, algemene bedrijfskunde en economie.

Welke doelgroep?

De toets is bestemd voor studenten van de lerarenopleidingen in het hogere beroepsonderwijs.

Wie is eigenaar van de toets?

Dat zijn de deelnemende lerarenopleidingen. Dat concentreert zich bij de Vereniging hogescholen waar die hogescholen lid van zijn.

Hoe worden de landelijk verzamelde gegevens gedeeld?

Elke hogeschool krijgt alleen de eigen gegevens. De data zijn niet openbaar voor de buitenwereld. Op termijn zal meer openbaar zijn, maar aangezien het project nog in een aanloopfase zit waarin de aansluiting van de toets op het curriculum nog aandacht nodig heeft, worden data alleen nog binnen de hogescholen gedeeld. De gegevens kunnen niet gebruikt worden om opleidingen mee te beoordelen.

Route-8 van A-Vision, Nederland

Algemene introductie op de eindtoets primair onderwijs in Nederland

Scholen in het primair onderwijs zijn vanaf schooljaar 2014-2015 verplicht om bij alle leerlingen een onafhankelijke, objectieve eindtoets af te nemen in het laatste schooljaar. Dit is vastgelegd in de Wet op het primair onderwijs. Scholen kunnen kiezen voor de centrale eindtoets van het College voor Toetsen en Examen (CvTE) of voor een andere eindtoets die door de minister van OCW is toegelaten, zoals de Route-8 van A-Vision. Om de kwaliteit van de eindtoetsen te waarborgen, moet de eindtoets aan een aantal andere voorwaarden voldoen. Deze voorwaarden zijn beschreven in het Toetsbesluit PO (Ministerie van OCW, 2014). Voor de eindtoets gelden de volgende voorwaarden:

- wordt jaarlijks aangeboden voor afname in het tijdvak tussen 15 april en 15 mei;
- meet ten minste de kennis en vaardigheden van de leerling op het terrein van Nederlandse taal en rekenen en wiskunde;
- heeft een mogelijkheid voor een tweede afnamemoment voor het geval een leerling verhinderd is bij de eerste afname;
- leidt tot een eenduidig advies aan de leerling omtrent het te volgen vervolgonderwijs;
- bestaat jaarlijks uit nieuwe opgaven voor taal en rekenen;
- kan ook gemaakt worden door leerlingen met een specifieke ondersteuningsbehoefte;
- biedt inzicht in de beheersing van de referentieniveaus taal en rekenen, en
- beantwoordt aan de kerneisen van psychometrische kwaliteit: inhoudelijke validiteit, betrouwbaarheid en deugdelijke normering.

De inhoudelijke kwaliteitseisen zijn nader uitgewerkt in de Toetswijzer eindtoets PO algemeen deel (CvTE, 2014). Er worden geen eisen gesteld aan de afnamevorm, bijvoorbeeld digitaal of schriftelijk.

Algemene beschrijving van de toets

De Route-8 toets van A-Vision is een eindtoets primair onderwijs. Het is een formatieve adaptieve toets die een ondersteunende functie heeft voor het advies wat eerder door de leerkracht is gegeven wat betreft het niveau van de middelbare school.

De toets moet worden afgenomen tussen 15 april en 15 mei, maar scholen mogen zelf kiezen welke dag en welk tijdstip hen het beste uitkomt. Leerlingen hebben 2 tot 3 uur nodig voor het maken van de toets. Leerlingen mogen zelf kiezen met welk onderdeel ze beginnen en in welke volgorde ze de onderdelen maken. Tussen de onderdelen door kan een pauze worden ingelast mocht dit nodig zijn. De save-as-you-go functie garandeert dat er geen gegevens verloren gaan en dat leerlingen weer kunnen doorgaan waar ze waren gebleven. Daarnaast is er een onderdeel optioneel, te weten Functioneren dat bestaat uit werkhouding en zelfconcept. Scholen mogen kiezen of ze deze willen afnemen of niet. Dit onderdeel wordt niet meegenomen in het advies, maar scholen kunnen het wel gebruiken om hun advies te onderbouwen of juist aan te passen.

Route-8 is gemaakt binnen de software RouteWijs en bevat alleen meerkeuzevragen. Deze toets kan gemaakt worden op laptops, desktops en tablets, zolang het scherm minstens 10" is zodat de vragen goed leesbaar zijn. Wanneer leerlingen gaan beginnen aan de toets loggen ze in met een eenmalige persoonlijke inlogcode. Hierna komen ze in de toetsomgeving waar ze al eerder kennis mee hebben gemaakt in een instructiefilmpje. Leerlingen selecteren vervolgens een specifiek onderdeel. De vragen komen één voor één op en de software berekend aan de hand van eerder gegeven antwoorden welk volgend item een leerling krijgt. Er is geen vast aantal vragen. Het programma gaat zo lang door totdat het weet op welk niveau de leerling zit. Ondertussen kan de leerkracht kijken op een apart scherm om te zien hoe ver een leerling is. Na afname krijgt de school automatisch de scores doorgestuurd. De Route-8 toets geeft scores tussen de 100 en 300. Het leerlingrapport is beschikbaar voor de leerkracht en die bespreekt het met ouders en leerlingen.

Wat is het doel van de toets?

Het doel is om aan leerlingen in het primair onderwijs een advies voor het voortgezet onderwijs te geven. Leerkrachten hebben al voor 1 maart een schooladvies gegeven. En dat betekent dat deze toets als het ware een check is van het advies dat de leerkracht al gegeven heeft. Als leerlingen op deze toets hoger of lager uitkomen, en dat geldt voor alle deelttoetsen, dan kan de leerkracht het advies heroverwegen. Dat betekent niet dat het advies wordt bijgesteld, maar de leerkracht of de school moet er nog wel een keer naar kijken. Leerkrachten worden niet beoordeeld aan de hand van deze toets. Het resultaat betreft puur alleen de leerling. Op schoolniveau kunnen scholen nagaan of zij het gemiddeld goed doen. De inspectie kijkt voor de verantwoording ook naar de resultaten van de eindtoets, dus deze informatie kan worden meegenomen in de evaluatie van een school.

Welke domeinen worden getoetst?

De domeinen vallen uiteen in taal- en rekenonderdelen waarvan er een aantal delen verplicht zijn gesteld. Er is een beoordelingskader opgesteld door Stichting Leerplan Ontwikkeling (SLO) in opdracht van het ministerie en ze hebben met elkaar bepaald wat er minimaal in de eindtoets naar voren moet komen. Dit resulteerde in verschillende taal- en rekenonderdelen die verplicht getoetst moeten worden. Dat is bij rekenen: meetkunde, verbanden, verhoudingen, getallen. En bij taal is dat taalverzorging en leesvaardigheid. Deze toets bevat ook leesvaardigheid,

taalverzorging, kijken-luistervaardigheid. Dat laatste is een onderdeel dat andere eindtoetsen niet hebben. A-Vision heeft daar bewust voor gekozen omdat zij dat een meerwaarde vinden.

Welke doelgroep?

Groep acht leerlingen. Deelname aan een eindtoets is verplicht voor alle basisscholen in Nederland. Hierbij mogen scholen kiezen uit één van de drie eindtoetsen die momenteel beschikbaar zijn.

Leerlingen krijgen extra advies boven op het advies wat hun leraren al hadden gegeven. Dit advies kan het eerder gegeven advies versterken, maar het kan ook aanleiding zijn om het advies te herzien. Ook kunnen leerlingen een beeld krijgen van hun werkeethos en instelling.

Wie is eigenaar van de toets?

A-Vision is eigenaar van Route-8. De bekostiging van de toets loopt via het ministerie, zodat scholen zelf niks hoeven te betalen voor de toets.

Hoe worden de landelijk verzamelde gegevens gedeeld?

De uitslagen van de toets gaan via een standaardrapportage naar de toetsleider van de school toe. De school zet deze dan weer door naar ouders en kinderen. Scholen zijn verplicht om het advies voor het voortgezet onderwijs op te nemen het leerlingenadministratiesysteem en te delen via een landelijke databank met alle gegevens van alle leerlingen in het onderwijs.

De data die worden verzameld tijdens de toetsconstructie zijn voor intern gebruik. Op basis van deze data wordt een verantwoording van de kwaliteit van de toets geschreven. Deze verantwoording wordt verstrekt aan de expertgroep toetsen PO, die het ministerie adviseert over het al dan niet toelaten van een nieuwe eindtoets basisonderwijs.

Eindexamens, Macedonië

Algemene beschrijving van de toets

De eindexamens in het basisonderwijs in Macedonië zijn digitaal, maar niet adaptief. Ze hebben een summatieve functie. De toetsen bestaan uit één groot blok van meerkeuzevragen die leerlingen in elke volgorde mogen beantwoorden. Ook al zitten leerlingen allemaal naast elkaar en is de toets niet adaptief, wordt er wel voorkomen dat leerlingen bij elkaar kunnen afkijken doordat in elk klaslokaal leerlingen van verschillende jaren door elkaar zitten. Ook krijgt elke leerling een willekeurige code voor het examen waarmee ze kunnen inloggen op de website. Deze code is verbonden aan een specifiek examen met een set van vragen die ad random is geselecteerd. Dit betekent dat leerlingen die in dezelfde klas zitten, andere examens krijgen. Tien tot 15 dagen na het examen krijgen de scholen vervolgens de uitslagen van de toets. Deze cijfers komen op het rapport te staan, maar worden ook openbaar bekend gemaakt.

Wat is het doel van de toets?

Het doel van deze toets is om een beter algemeen beeld te krijgen van schoolniveau en om leerkrachtprestaties te bekijken en te vergelijken. Leraren moeten leerlingen op basis van hun

inzicht in de leerlingen een cijfer geven tussen een één en een vijf; deze cijfers worden dan vergeleken met de cijfers die leerlingen krijgen voor deze eindtoets.

Daarnaast wordt het ook gebruikt om te meten welke kennis leerlingen gedurende het schooljaar hebben opgedaan. Het wordt niet zo zeer ingezet om leerlingen advies te geven op hun voortgang, maar zij krijgen een cijfer die naast hun andere cijfers op het rapport komt te staan.

Welke domeinen worden getoetst?

In principe wordt elk vak dat leerlingen krijgen op school ook getoetst met uitzondering van muziek of gymnastiek. Op de lagere klassen van de basisschool wordt wiskunde, Macedonisch, wetenschap, natuur en de samenleving en Engels getoetst. In de hogere klassen van de basisschool worden wiskunde, Macedonisch, Engels, biologie, natuurkunde, scheikunde, aardrijkskunde en geschiedenis getoetst.

Welke doelgroep?

Deze toetsen worden aan het einde van het schooljaar afgenomen. Deelname hieraan is verplicht voor alle leerlingen in het hele land. De toetsen worden afgenomen vanaf het 5e jaar tot het 9e jaar op de basisschool en daarna tot het 12e jaar op de middelbare school. Kinderen worden op deze wijze getoetst van hun 10e tot hun 18e.

Wie is eigenaar van de toets?

De toetsen zijn eigendom van het Ministerie van Onderwijs. Zij zijn ook diegene die deze levert en ervoor zorgt dat alles goed wordt uitgevoerd.

Hoe worden de landelijk verzamelde gegevens gedeeld?

De verzamelde gegevens worden gepresenteerd op de algemene website van de Ministerie van Onderwijs. Op deze site worden ook de prestaties van leerkrachten weergegeven en vergeleken. Scholen krijgen ook nog de uitslagen toegestuurd om te delen met hun studenten. Dit wordt meestal gedaan door de cijfers op te hangen in de gang.

NILA en NINA, Noord-Ierland

Algemene beschrijving van de toets

De Northern Ireland Literacy Assessment (NILA) en de Northern Ireland Numeracy Assessment (NINA) zijn allebei diagnostische toetsen om de prestatie en voortgang van leerlingen te toetsen in het 4^e, 5^e, 6^e en 7^e lesjaar. Dit betekent dat de toetsen als doel hebben de ontwikkeling en de leerbehoefte van leerlingen vast te stellen, zelfbeoordeling en het stellen van doelen mogelijk te maken, leerkrachten te ondersteunen bij het aanpassen van hun onderwijs aan de leerbehoeften van leerlingen en ouders voorzien van informatie over de sterke en zwakkere punten van hun kinderen ten aanzien van taal en rekenen. De toetsen zijn adaptief; de toetsvragen die de leerling aangeboden krijgt is afhankelijk van de beantwoording van de voorgaande toetsvraag. Deze toetsen zijn eigendom van het Ministerie van Onderwijs maar worden geïmplementeerd met behulp van Council for the Curriculum, Examinations and Assessment (CCEA). C2k, een project gefinancierd door datzelfde Ministerie, is verantwoordelijk voor de infrastructuur en middelen van scholen om te voldoen aan de eisen van de onderwijskundige technologiestrategie van het

Ministerie. C2k implementeerde bijvoorbeeld een breedband LMS waarbij 50.000 desktops zijn geplaatst op 1.200 scholen. Het Ministerie van Onderwijs, CCEA en C2k werken dus alle drie samen om scholen te voorzien van de NILA en NINA. Deelname aan beiden toetsen was eerst verplicht vanuit de overheid, maar na een grote hoeveelheid technische problemen in de eerste jaren is deelname nu volledig vrijwillig. Scholen kunnen zelf een tijdstip kiezen tussen september en december dat voor hen het beste uitkomt. Voordat leerlingen de toets daadwerkelijk maken krijgen zij eerst nog een instructiefilmpje te zien en mogen ze een aantal vragen oefenen. Hierna loggen ze in op een door C2k beveiligd internet platform met hun C2k gebruikersnaam en wachtwoord. De toetsen zelf nemen ongeveer twintig tot vijftwintig minuten elk. Wanneer een leerling heel veel langzamer is, dan stopt de toets automatisch en berekend hij de uitslag op basis van de beantwoorde vragen. Dit voorkomt dat een leerling zich heel erg slecht voelt vanwege de tijd die hij of zij nodig had. Omdat de toetsen van twee verschillende aanbieders zijn, Tribal en Rising Stars, zijn er ook twee verschillende soorten software. Niet alleen komen de lay-outs van de toetsen niet overeen met elkaar, maar binnen NILA zijn de vragen opgebouwd in blokken van 8 vragen terwijl dit in blokken van 5 vragen gaat bij NINA. Wel bevatten beiden toetsen uitsluitend meerkeuzevragen die één voor één worden gevraagd. Deze keuze voor meerkeuzevragen is gemaakt aan de hand van eerder onderzoek die uitwijst dat dit de beste en meest praktische manier is om leesvaardigheid en wiskunde te testen. Aangezien de tentamens adaptief zijn kunnen leerlingen niet meer terug naar eerder gestelde vragen. Het is belangrijk om dit vooraf, bij de start en tijdens de toetsafname duidelijk aan te geven. Leerlingen stellen het ook op prijs als ze weten hoe ver ze zijn met de toets. In de adaptieve toetsen wordt dit weergegeven per inhoudelijk blok (bv. drie blokken afgerond, nog twee te gaan), aangezien het aantal vragen dat een leerling krijgt niet op voorhand bekend is. Over het algemeen beginnen leerlingen bij niveau twee en gaan ze dan omhoog of omlaag aan de hand van de resultaten per blok van vragen. Kinderen met een leerbeperking beginnen bij een lager startpunt. 48 uur na de toets krijgen docenten de feedback en mogen zij bepalen of deze informatie wordt gedeeld met ouders en kinderen. Er zit geen cijfer aan deze tentamens vast, alleen een leeftijdgerelateerde uitkomst en een gestandaardiseerde score met enige diagnostische feedback.

Wat is het doel van de toets?

De NILA en NINA assessments zijn diagnostische toetsen die eenmalig in de herfst worden afgenomen om een beeld te krijgen van het niveau van leerlingen. Scholen kunnen zelf een tijdstip kiezen tussen september en december dat voor hen het beste uitkomt.

Welke domeinen worden getoetst?

De domeinen die hiermee worden getoetst zijn geletterdheid (NILA: Northern Ireland Literacy Assessment) en rekenvaardigheden (NINA: Northern Ireland Numeracy Assessment).

Welke doelgroep?

Alle basisscholen in Noord-Ierland mogen gebruik maken van de NILA en NINA en tot een paar jaar geleden was deelname aan deze toetsen ook verplicht. In 2015 namen 272 scholen van de in totaal 3900 basisscholen deel aan de NILA en NINA. De doelgroep bestaat uit leerlingen in jaren 4, 5, 6 en 7. Deze laatste vier jaren van de basisschool noemen zij ook wel 'key stage 1 & 2'.

Wie is eigenaar van de toets?

Het Ministerie van Onderwijs is de eigenaar van alle centraal gefinancierde examens in Noord-Ierland. Council for the Curriculum, Examinations and Assessment (CCEA) werkt voor het Ministerie en is verantwoordelijk voor het laten ontwikkelen en beschikbaar stellen van de toetsen, maar het Ministerie heeft eigendom over de toetsen. NINA en NILA worden door twee verschillende organisaties ontwikkeld.

Hoe worden de landelijk verzamelde gegevens gedeeld?

De gegevens van de individuele studenten worden binnen 48 uur geleverd aan de scholen. De scholen mogen dan zelf bepalen of ze de resultaten delen met ouders of niet. Toen deelname aan de assessments nog verplicht was, hoorde daar een gesprek tussen leerkracht en ouders bij over het niveau van hun kinderen. Dit is in Noord-Ierland redelijk uitzonderlijk aangezien dit het enige examen was waarbij dit verplicht was. De resultaten kunnen ook gebruikt worden door leerkrachten om te kijken hoe het algemene niveau is binnen hun klas en waar er nog ruimte voor verbetering is. Aan de hand van een beschrijving van de vragen waarop minder is gescoord, kan een leerkracht makkelijk komen te weten welke leerling waarmee moeite heeft. De schoolleiding kan ook kijken naar de prestatie en functioneren van individuele leerkrachten aan de hand van de gegevens op klasniveau vergeleken met andere klassen binnen de school. Scholen worden echter nooit met elkaar vergeleken en de gegevens van scholen worden nooit gedeeld door de overheid. De enige manier waarop het NILA/NINA team en de overheid zicht op gegevens krijgen is door rapporten die geschreven worden door de aanbieders. Binnen deze rapporten worden trends beschreven waarbij alle data zijn geanonimiseerd. Deze rapporten zijn niet openbaar.

3.2 ONDERZOEKSVRAAG 1. AFNAMECONDITIES EN ICT-INFRASTRUCTUUR

De resultaten met betrekking tot de vergelijking tussen aanbieders over het netwerk, de gebruikte devices, organisatorisch kenmerken en privacy worden in tabellen gepresenteerd.

3.2.1 WELKE EISEN WORDEN GESTELD AAN DE NETWERKINFRASTRUCTUUR?

De connectiviteit is een belangrijke eigenschap van de netwerkinfrastructuur om goed gebruik te kunnen maken van toetsystemen die draaien via Internet of externe servers. In de interviews is gevraagd of er gewerkt wordt via breedbandinternet (ADSL, VDSL, glasvezel en kabel), wat de benodigde downloadsnelheid is, wat het betrouwbaarheidslevel is van de toetsapplicatie, of de toetsapplicatie is ingesteld op piekbelasting, of de afname bekabeld of via wifi uitgevoerd wordt, wat de kwaliteit is van de wifi en of er voor ingebruikname een stresstest is uitgevoerd. Tabel 1 geeft een overzicht van de resultaten.

Bevindingen

De aanbieders leggen de verantwoordelijkheid voor de kwaliteit van de infrastructuur in het algemeen bij de school neer. Bij het ontwerp houden ze zoveel mogelijk rekening met de gangbare situatie op scholen. De downloadsnelheid die nodig is om de toetsen te downloaden is toetsafhankelijk. Bij SBAC zorgen ze er door een richtlijn bij de itemconstructie voor dat de omvang van de items zeer klein blijft, waardoor er geen problemen ontstaan met het downloaden. De aanbieders stellen hier wel minimumeisen aan. Het betrouwbaarheidslevel varieert per land en uit de interviews bleek dat dit geen groot punt van aandacht was. Met piekbelasting wordt soms technisch rekening gehouden, maar meestal wordt aangeraden om de toetsafnames te spreiden. Bij adaptieve toetsen is spreiding van toetsafnames geen probleem

omdat elke toets voor elke leerling uniek is samengesteld. De wifi moet goed zijn, maar indien dit niet het geval is het belangrijk dat bekabeling mogelijk is.

Tabel 1. De netwerkinfrastructuur

	CAT, Dene- marken	Smarter Balanced Assessme nts, VS	Eindexa mens in Noor- wegen	10voordel eraar, Nederland	Route-8 A-Vision, Neder- land	Eindex amens , Mace- donië	NILA en NINA, Noord- lerland
Breedbandinte rnet (ADSL, VDSL, glasvezel en kabel)	Geen breedbandi nternet voor leerlingen – Wifi	ADSL/VDSL zou moeten werken. Is verantwoorde lijkheid van de Staat. Bandbreedte kan de school zelf meten via SchoolSpeedT est en SpeedTest.net	Glasvezel	Breedbandinter net, maar soort verschilt per school	Breedbandin ternet, maar soort verschilt per school	?	Eis breedband is afhankelijk van grootte school
Downloadsnelh eid	Niet bekend – maar examen staat op een website	Elk item is < 5kb. 15.000 toetsen kunnen gelijktijd gedownload worden. minimum van 10–20 Kbps	Meer dan 1000MBit	Niet bekend, het verschilt per hogeschool. Alle hogescholen hebben hoog netwerk	0.6 MBit per leerling of 2.0 MBit per klas	?	Afhankelijk van de school: 4 – 200 Mbit
betrouwbaarhe idslevel	100% - nooit problemen mee gehad	Verschilt per school	99% - nog nooit problemen mee gehad	Verschilt per school	Verschilt per school	Veel proble men mee, valt geregeld uit tijdens examens	Niet 100%, geen precies percentage
Ingesteld op piekbelasting	Toetsmome nten worden gereserveer d per klas met een maximum per test moment.	Dat is afhankelijk van de school.	Ja, iedereen wordt tegelijktijd getoetst	Ja, tijden voor afname worden gebaseerd op de resultaten van de stresstest	Ja	?	Ja, systeem getest op maximale hoeveelheid gelijktijdige afnames
Bekabeling vereist?	Nee	Nee	Nee	Verschilt per school	Nee	Ja	Nee
Wifi kwaliteit	Erg goed, elk lokaal heeft een access point voor 25-30 apparaten	Verschilt per school	Erg goed – Wifi netwerk alleen toegankelijk vanaf gecertificeer de apparaten en met studenten inloggegeven s	Verschilt per school – moet wel voldoen aan de eisen van het examen	0.6 MBit per leerling of 2.0 MBit per klas	Geen Wifi	Goed, verzorgd door C2k
Stresstest uitgevoerd?	Nee	Ja, door scholen zelf. Smarter Balanced Technology Readiness Calculator	Netflix en andere programma's via internet worden dagelijks gebruikt – stresstest niet nodig	Ja, aan de kant van 10voordeleraar en bij de scholen zelf	Ja, scholen kunnen van te voren een 'ben ik geschikt' stappenplan volgen	Ja, door ict coördinat oren en door de overheid	Ja, door CkK, NILA & NINA en door scholen zelf (pre- flight check)

De meeste organisaties bieden de mogelijkheid voor het meten van de geschiktheid van de eigen toetslocatie voor de toetsafname. Voorbeelden zijn de Pre-Flight check en de, Ben ik geschikt test. Met deze testen kan een school nagaan of de computers up-to-date zijn en of de infrastructuur voldoet aan de eisen van de toets. Met een test als de Smarter Balanced Technology Readiness Calculator wordt op basis van het aantal leerlingen, aantal beschikbare computers en aantal uren dat de computers beschikbaar zijn bepaald hoeveel dagen en bijbehorende bandbreedte nodig is om de toetsen af te nemen.

3.2.2 WELKE EISEN WORDEN GESTELD AAN DE DEVICES?

Toetsen kunnen afgenomen worden op verschillende devices. In Tabel 2 is weergegeven welke devices door de scholen gebruikt worden voor het afleggen van de verschillende toetsen en wat de eisen aan die devices zijn. Daarbij is tevens geïnventariseerd hoeveel devices er beschikbaar zijn, hoe ze gebruikt worden in de school en wat de benodigde randapparatuur is.

Tabel 2. Toegestane en gebruikte devices en randapparatuur

	CAT Dene-mark en	SBAC, VS	Eindexame ns, Noor-wegen	10voordeler aar, Nederland	Route-8 van A-Vision, Nederla nd	Eindexame ns, Macedonië	NILA/NI NA, Noord-Ierland
Type devices	iPad Mini	geen beperking en	HP ProBook 430Intel Core i3 2,10 GHz3 MB cache256 GB SSD HD4 GB DDR3 SDRAM	Computers en laptops - Verschilt per school	Computers, laptops, tablets	Computers	Computers en laptops
Aantal devices op school	1 per leerling	Verschilt per school	1 per leerling	Verschilt per school	Verschilt per school	1 per leerling	Ligt aan de school
Plaats devices	Bij de leerling	Verschilt per school	Bij de leerling	Verschilt per school - sommige scholen hebben computer lokalen bijv.	Verschilt per school	Verschilt per school - sommige scholen hebben computer lokalen bijv.	In school
Opstelling devices	Flexibel	Verschilt per school	Flexibel	Verschilt per school	Flexibel		Ligt aan de school
Ander gebruik van de devices dan toetsen?	Ja, gebruikt voor alles. School en privé.	Ja, worden voor andere doeleinde n gebruikt.	Ja, gebruikt voor alles. School en privé.	Ja	Ja	Ja	Ja, wel alleen op school. Geen huiswerk.
Server	Geen eigen server, scholen maken gebruik van de cloud	Portal servers, web server, applicatio n server , data storage hosting by SBAC	Elke school heeft een fysieke server voor zichzelf. Een virtuele server dient als backup en wordt gedeeld door meerdere scholen.	Scholen maken gebruik van een externe virtuele server waar de toets op staat. Ze kunnen wel hun eigen server hebben, maar deze haalt de toets binnen via de ander.	Scholen maken gebruik van een externe server waar de toets op staat. Deze staan op meerdere locaties	?	Ligt aan de school, grotere scholen wel en kleinere niet

Vervolg Tabel 2. Toegestane en gebruikte devices en randapparatuur

BYOD (bring your own device)	Nee	Nee	Nee	Nee	Nee	Nee	Ja, maar niet voor toetsen
Eisen voor kwaliteit beeldscherm	Diagonaal:7.9 inch	Minimum 9.5 inches diagonal ("10-inch class"). Resolutie minimum 1024x768.	13" LED scherm (1366 x 768)	Verschilt per school – geen minimum eis bekend	Minstens 10" voor de leesbaarheid	?	Minimum eisen worden opgesteld door C2K
Eisen voor operating system	Apple iOS7	Windows, Mac OS X, Linux, iOS, Android, Chrome OS Minimum en aanbevolen versies worden gegeven	Windows	Windows	Microsoft Windows Vista of hoger, Apple OS X Lion of hoger, Apple iOS 5.1 of hoger, Google Android 4.4.4 of hoger	Linux	Windows
Eisen voor internet browser	Safari	Alle	Alle	Internet Explorer en Firefox	Google Chrome 24 of hoger, Mozilla Firefox 24 of hoger, Microsoft Internet Explorer 10 of hoger, Apple Safari 5 of hoger	Google Chrome	Internet Explorer
Beeldschermresolutie	2048 x 1536	10" class of groter met 1024 x 768 display resolution	Verschillend per jaar		Geen specificatie		C2k
Operating system	Apple A7	Windows, Mac OS X,	Windows	Windows	Kan allemaal		Windows
Processor		1GHz processor					
Intern geheugen	16GB	1 GB Ram	Verschillend per jaar		NVT		NVT
Harde schijf	NVT	80 GB hard drive of minstens 1GB vrije ruimte	SSD – kan tegen een stootje		Geen specificatie		NVT
CD/DVD station	Nee	Nee	NVT	NVT	NVT		Nee
Randapparatuur	Nee	muis, touch screen, touchpad, ... En fysiek toetsenbord	Nee	Nee	NVT		Raden het gebruik van een muis aan
Netwerkadapter	NVT	Nvt	NVT	NVT	NVT		Nee
Streaming audio/video	Aanwezig – NVT	Nvt	NVT	Nee	OGG met Vorbis codec of MPEG met MP3 Codec (audio); VP8 of VP9 Codec (video)		MP3 voor de audio
Audio output	Ja	Ja	NVT	NVT	Geen specificatie		Ja

Vervolg Tabel 2. Toegestane en gebruikte devices en randapparatuur

Video RAM	NVT	Nvt	NVT	NVT	Geen specificatie		NVT
Printer	NVT	Voor toets coördinator voor sessie informatie en leespassages	NVT	NVT	NVT	NVT	NVT
Pop-up blokkeringen	Nee	Nee, secure omgeving	Nee	Ja -	Nee	Nee	Beveiligde internet verbinding – C2k
Headset	Ja	Ja, voor taaltoets en voor leerlingen die de functie tekst-to-speech gebruiken bij wiskunde	Alleen bij dyslexie		Ja, wordt geleverd door A-Vision	Nee	Ja

Bevindingen

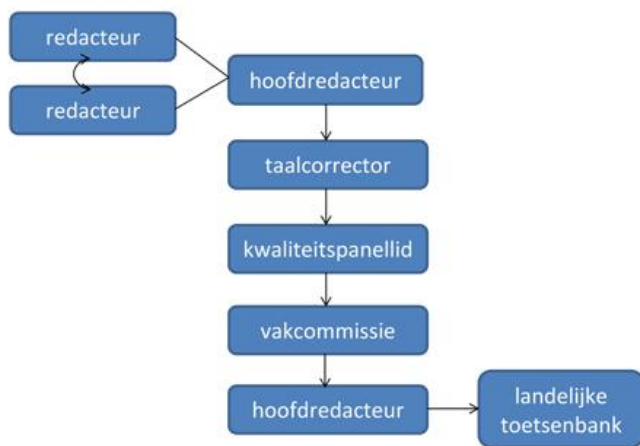
De toetsen worden vooral afgenomen op computers en laptops. Een enkele toets mag afgenomen worden op tablets. Geen van de aanbieders staat Bring your own device toe. De reden hiervoor is dat de aanbieders dan niet kunnen garanderen dat de toets goed draait of dat het device up-to-date is. Devices worden bij Denemarken, Noorwegen en Macedonië door de overheid geleverd. Bij NINA/NILA mag alleen apparatuur gebruikt worden die ondersteund wordt door C2k. Het aantal beschikbare devices per school verschilt, behalve als ze aan de school geleverd worden door de overheid. In dat geval is er voor elke leerlingen een device beschikbaar. Aan de inrichting van de afdelingslocatie worden geen specifieke eisen gesteld. De inrichting van de toetsruimte is een verantwoordelijkheid van de scholen. De toetsen kunnen in de eigen klas of in specifieke computerruimtes worden gedaan. Alle devices worden ook voor andere onderwijsactiviteiten gebruikt. De minimumeisen aan de randapparatuur variëren per aanbieder. Bijna elke aanbieder geeft wel een aanbeveling voor een device met randapparatuur.

3.2.3 WELKE ORGANISATORISCHE EISEN WORDEN GESTELD AAN DE SCHOLEN?

Het hele toetsproces bestaat uit ontwerp, ontwikkeling, uitvoering, analyse en bijstelling. Organisatorische eisen hebben vooral betrekking op de personen die betrokken zijn bij de ontwikkeling en de afname. Deze staan opgenomen in Tabel 3.

Bevindingen

Bij alle toetsen zijn in het toetsproces zeer veel personen betrokken en speelt de betrokkenheid zich af op meerdere niveaus. Zo heeft 10voordeleraar op het niveau van toetsontwikkeling taken voor een redacteur, taalcorrector, kwaliteitspanellid, en een vakcommissie beschreven (zie Figuur 1).



Figuur 1. Betrokkenen bij toetsconstructie bij 10voordeleraar. (Bron: <https://www.10voordeleraar.nl/programma/zo-werkt-het>)

Bij de afname wordt vooral benadrukt dat er expertise nodig is op ict-gebied. Tevens dient er een surveillant aanwezig te zijn. Dit kan de leerkracht zijn, maar in Noorwegen worden hiervoor bijvoorbeeld gepensioneerden ingehuurd.

Tabel 3. Organisatorische eisen

	CAT, Dene - mark en	SBAC, VS	Eindexam ens, Noor- wegen	10voordeler aar, Nederland	Route-8 van A- Vision, Nederla nd	Eindexam en, Macedoni ë	NILA/NI NA, Noord- Ierland
Wie zijn betrokken bij ontwikkeling ?	?	Brede groep betrokken en. Eerst een pilot toets en een veldtoets	ict coördinatoren en didactisch personeel	Brede groep, werken met workflow controle	ict personeel, statisticus, methodoloog , contactperso nen, pr, collega's die vragen maken	?	NILA/NINA, aanbieders, leerkrachten, C2k
Wie zijn betrokken bij uitvoering?	Leraren en ict coördinatoren	Met name de leerkrachten	Schoolleiding heeft een systeem voor elke leerling. Surveillanten en logging controle Oefententamen s: leerkrachten zelf	Instituutsbeheerder, surveillanten en roosterbureau	Directeur van de school, intern begeleider of leerkrachten van groep 8, ict coördinator als die aanwezig is	Ministerie van onderwijs, ict coördinatoren	Leerkrachten , ict coördinatoren, assessment coördinatoren
Welke expertise is nodig in de school?	ict coördinator	ict coördinator	ict coordinator & externe ict coordinator voor back-up	Niet gespecificeerd	Geen specifieke expertise nodig	ict coördinator	Leerkrachten , assessment coördinatoren
Wordt de toets groepsgewijs of individueel afgenomen?	Individueel	Individueel	Individueel – behalve Engels: van te voren groepsoverleg	Individueel	Individueel	Individueel	Individueel

3.2.4 WELKE EISEN WORDEN GESTELD AAN DE PRIVACY?

Bij digitale toetsafnames is een aantal risico's te onderscheiden (Surf, 2014):

- Het niet kunnen aantonen dat de toets rechtmatig verlopen is;
- Onduidelijkheid over rollen en verantwoordelijkheden, waardoor bijvoorbeeld op ad-hoc-basis met incidenten en calamiteiten wordt omgegaan;
- Technisch is het mogelijk om te frauderen, bijvoorbeeld door ongeoorloofd samen te werken of af te kijken.

Deze risico's betreffen de authenticiteit van de toetsresultaten. In hoeverre zijn de verkregen toetsresultaten toe te schrijven aan de specifieke leerling? Aan de andere kant is er een risico ten aanzien van de privacy van leerlingen. Bij digitaal toetsen kan zeer veel informatie opgeslagen worden. In Tabel 4 is weergegeven op welke manier informatie beveiligd is.

Tabel 4. Eisen aan de informatiebeveiliging

	CAT, Dene-marken	SBAC, VS	Eindexamens, Noorwegen	10voordeleraar, Nederland	Route-8 A-Vision, Nederland	Eindexamens, Macedonië	NILA/NINA, Noord-Ierland
Welke privacy eisen?	Log in system: YouAndI	Leerlingen hebben persoonlijke inlogcode alle communicatie is encrypted via https	Leerlingen krijgen een eenmalige inlogcode: nakijken is volledig geanonimiseerd.	StudentID en uitslag wordt vastgelegd. En vooropleiding en herkansing.	Leerlingen krijgen een persoonlijke inlogcode, VDOD bepalingen	Leerlingen loggen in met hun naam en geboortedatum.	Log in system: personal C2k codes
Informatiebeveiliging?	Log in system: YouAndI	Informatie die verzameld wordt en niet verzameld wordt is expliciet beschreven	Examens worden minstens 3 maanden bewaard door externe partij, examens kunnen worden opgeslagen door leerlingen zelf	De boel is stevig beveiligd. Binnenkort hebben we een audit. De data worden bewaard	Proberen zo weinig mogelijk persoonlijke gegevens te verzamelen, beperkt aantal collega's die bij data kunnen	?	Suppliers meet security standards & log their data into escrow

Bevindingen

Leerlingen hebben allemaal een persoonlijke inlogcode waarmee ze op de website inloggen. Bijna alle data worden geanonimiseerd gepubliceerd. Alleen in Macedonië worden de resultaten met naam erbij bekend gemaakt via lijsten in de gangen van de school. De toets van Noorwegen bevat als enige open vragen en voor het nakijken worden de antwoorden van de leerlingen anoniem verstuurd naar beoordelaars. De andere tentamens bestaan uit meerkeuzevragen die binnen de software wordt nagekeken en opgeslagen.

Informatiebeveiliging varieert enorm. Bij Noorwegen en Ierland zorgen externe partijen hiervoor. In Nederland bij zowel A-Vision als 10voordeleraar worden zo weinig mogelijk

persoonlijke gegevens verzameld. Ook mogen er bij A-Vision maar een beperkt aantal mensen bij de data.

3.2.5 GELEERDE LESSEN

Bij alle toetsen is gevraagd wat goed bevalt aan de huidige manier van toetsen en wat aandachtspunten zijn. Deze resultaten worden hieronder per toets gepresenteerd.

CAT op de Folkeskole, Denemarken

Wat bevalt goed?

Er is een aantal factoren dat erg goed bevalt bij deze toets:

- Het feit dat elke leerling moet inloggen met hun persoonlijke inloggegevens betekent dat het examen niet alleen beveiligd is, maar ook dat resultaten van verschillende jaren met elkaar vergeleken kunnen worden.
- Doordat de items telkens per stuk worden getoond kunnen leerlingen zich goed concentreren.
- Het adaptieve karakter van het examen. Samen met het voorgaande punt wordt voorkomen dat leerlingen bij elkaar kunnen afkijken.
- De simpele en begrijpbare opmaak van de toets bevalt ook goed.
- De toets kan 'gepauzeerd' worden. Mocht een leerling zich niet meer kunnen concentreren of zich niet lekker voelen, dan kan de leerkracht ervoor kiezen om het examen even stil te leggen en op een later tijdstip van hetzelfde punt door te gaan.

Wat zijn aandachtspunten?

- Ondanks het feit dat het ook een kracht is, levert het adaptieve karakter van de toets ook een aandachtspunt op. Leerlingen krijgen gedurende de toets geen feedback, en het komt geregeld voor dat ze gefrustreerd worden door de steeds moeilijker wordende vragen. Vooral in de eerste jaren hebben de leerlingen nog niet door dat dit betekent dat ze het juist goed doen.
- Leerkrachten krijgen geen advies of richtlijn over wat zij het beste kunnen doen als een leerling slecht gescoord heeft. Het zou wenselijk zijn om advies te krijgen over hoe en met welke oefeningen je dit het beste kan aanpakken. Nu krijgen leerkrachten alleen maar data hoe hun klas het heeft gedaan, maar verder wordt er weinig met de data gedaan. Leerlingen krijgen nu nog alleen een advies over hoe zij presteren, maar nog geen advies over wat ze nog kunnen veranderen of waar ze aan kunnen werken.

Wat zijn aanbevelingen?

Laat een soortgelijke toets niet het enige instrument zijn waarmee leerlingen getoetst worden. Je kan niet alles te weten komen van een leerling en hun denkwijze door één specifieke toets. De toets is formatief om vast te stellen wat de leerlingen al weten en hoe ze er over het algemeen voor staan. In Denemarken beginnen ze nu echter te ondervinden dat leraren hun lesinhoud aanpassen aan de toetsinhoud zodat hun klas beter scoort. Dit zou voorkomen kunnen worden als ook de dagelijkse bevindingen van leerkrachten meegenomen worden in de beoordeling.

Smarter Balanced Assessments, VS

Wat bevalt goed?

- Er is een breed spectrum van toetsen beschikbaar die 90% dekken van alle talen die in Californië gesproken worden.
- Het adaptieve karakter is zeer positief.
- Het complete uitgebalanceerde systeem van zowel de formatieve toetsen die gericht zijn op instructie en leren en de summatieve toetsen.

Wat zijn aandachtspunten?

Er is meer onderzoek nodig naar het gebruik van plaatjes in plaats van woorden. Een toets voor leerlingen met dyslexie en een verbetering van de brailletoets voor blinden.

Technisch willen we van 15.000 gelijktijdige toetsafnames naar een ongelimiteerd aantal. Hiervoor worden aanpassingen gedaan in de architectuur.

Wat zijn aanbevelingen?

- Start met een vaste toetsvorm voordat je adaptief gaat toetsen. Het ontwikkelen van een adaptieve toets is erg ingewikkeld. De ontwikkeling van 20000 items in 2 jaar was een haast onmogelijke klus.
- Stel een realistisch tijdpad op.
- Zorg voor een oefentoets

Wat zou je anders doen als je opnieuw begon?

De communicatie met de ouders zouden we nog beter en eerder oppakken. Dat is cruciaal. Daar zijn we nu te laat mee gestart. Als het opnieuw zou kunnen is het beter om op een hoger niveau te beginnen en dan naar beneden toe door te ontwikkelen.

Eindexamens in Noorwegen

Wat bevalt goed?

Het fijnste aan digitale toetsen is dat de distributie van examens en de organisatie redelijk makkelijk verloopt. Eerder moesten alle schriftelijke uitwerkingen worden verzameld en dan worden verstuurd naar de desbetreffende leerkracht die was toegewezen om het examen na te kijken. Hij /zij stuurde het dan vervolgens naar de volgende leerkracht voor een tweede check, en uiteindelijk werd het nagekeken examen weer terug gestuurd. Dit hield veel administratief en organisatorisch werk in en er was altijd een kans dat het examen kwijt zou raken in de post. Sinds het digitaliseren is dit proces veel efficiënter geworden.

Voor leerlingen zelf is het ook een stuk prettiger om op computers te werken aangezien zij dit in het dagelijkse leven ook doen voor andere toetsen, huiswerk en in hun vrije tijd.

Het loggingsysteem dat gebruikt wordt in de Nord Trondelag regio bevalt ook erg goed. Deze software, 3ami MAS, zit op elke laptop die leerlingen van de school krijgen en wordt gebruikt bij elke toets die ze maken. (zie <http://www.3ami.com/latest-case-studies.htm>). Het systeem logt vanaf het moment dat je het activeert alles wat de leerling doet. Elke knop die hij of zij indrukt,

elke website die ze openen, elke correctie die ze maken, enz. Op de dag van een examen krijgen ze nog een waarschuwing te zien: 'remember, you are being logged'. Hier moeten ze dan aanvinken dat ze dit hebben gelezen, en dan kunnen ze beginnen met het examen. Deze logdata worden dan rechtstreeks naar de schoolserver gestuurd en twee weken bewaard. Als een leerkracht een student verdenkt dat ze een programma hebben gebruikt of iets hebben gekopieerd op een examen dan kunnen ze dit makkelijk nakijken. Tijdens de eindtoetsen wordt dit gedurende het examen zelf gedaan door een team van gepensioneerden. Zij gaan gedurende elke 5 uur durende toets telkens door elke computer heen om te kijken of alles eerlijk verloopt.

Wat zijn aandachtspunten?

Het feit dat alle vragen nog steeds tegelijkertijd getoond worden is een aandachtspunt. Een tweede aandachtspunt betreft de inhoud van de toetsen. Leraren van taalvakken vinden dat hun leerlingen niet genoeg de kans krijgen om te laten zien dat ze een vreemde taal goed beheersen. Leerlingen hebben nu geen mogelijkheid om te laten zien dat ze ook daadwerkelijk een vreemde taal kunnen spreken of schrijven tijdens de examens. In wiskunde krijgen leerlingen bijvoorbeeld twee verschillende examens waarvan één op papier is en één op de computer. Taalleerkrachten willen zoiets graag terug zien bij hun vakken omdat leerlingen dan moeten laten zien dat ze ook zonder alle hulpjes kunnen die aanwezig zijn op een computer. Zij vinden dat leerlingen hier te afhankelijk van zijn en dat hierdoor de essentie van een vreemde taal niet goed wordt getoetst.

Wat zijn aanbevelingen?

Probeer zoveel mogelijk software te gebruiken die de leerlingen al kennen vanuit hun huiswerk en andere toetsen. Het is namelijk niet de bedoeling dat bijvoorbeeld flexibiliteit in het omgaan met software getoetst wordt.

Wat zou je anders doen als je opnieuw begon?

De leerkrachten veel meer betrekken bij het hele proces. Op dit moment zijn de toetsen allemaal top-down en de leerkrachten voelen zich niet genoeg betrokken bij het examen. Dit wordt wel al beter, maar soms is het ook gewoon onmogelijk om iedereen tevreden te houden. In een school met 250 werknemers is het niet te doen om iedereen naar hun mening te vragen en moeten er gewoon beslissingen genomen worden. Het vinden van een goede balans hierin is het streven.

10voordeleraar, Nederland

Wat bevalt goed?

Als het systeem opgestart is, dan werkt het ook goed. Het is heel stabiel en heeft nog geen problemen opgeleverd.

Wat zijn aandachtspunten?

Het opbouwen van de toetsenbanken zouden we op een andere manier willen organiseren. We werken met constructieteams die een afvaardiging zijn van de hogescholen. We zouden dat wat meer in eigen hand willen hebben, meer zoals het Cito-model. Hiernaast kan ook de analysefunctionaliteit van de software beter. Ook is de beperkte afnamecapaciteit op de

hogescholen iets waar we in de toekomst aan willen werken. Het invoegen van meer multimedia zoals filmpjes en geluidsfragmenten is ook een aandachtspunt.

Wat zijn aanbevelingen?

Het is erg belangrijk om goed contact te onderhouden met de doelgroep. 10voordeleraar houdt om de 1 à 2 jaar tevredenheidsonderzoeken. Hiernaast kunnen mensen ook altijd vragen stellen op onze website. Ook zijn studenten betrokken geweest bij de vragenontwikkeling om te voorkomen dat het een top-down constructie werd.

Wat zou je anders doen als je opnieuw begon?

Niets.

Route-8 van A-Vision, Nederland

Wat bevalt goed?

Dat de toets adaptief is. Daar zijn wij zelf erg groot voorstander van, maar dat scholen daar ook heel erg blij mee zijn, merk je aan hun reacties. Door die adaptiviteit is de toets korter en dat bevalt ook. De centrale eindtoets neemt drie dagdelen in beslag en de Route-8 neemt één dagdeel in beslag. Dat is voor leerlingen prettig. We zijn ook tevreden over de toevoeging van het onderdeel Functioneren. Je vraagt in een toets heel erg naar de harde gegevens maar het kan ook wel eens goed zijn om naar meer te kijken dan naar alleen dat gedeelte. Scholen hoeven hier ook niks extra's voor te betalen. En wij hebben bepaalde onderdelen van taal opgenomen in de toets die wij belangrijk vinden voor de bepaling van het advies voor het voortgezet onderwijs die niet in alle eindtoetsen voorkomen. Dat zijn een beetje de speerpunten waarvan wij denken dat onze toets anders of beter is dan de andere eindtoetsen.

Wat zijn aandachtspunten?

Scholen nog beter voorbereiden op het feit dat het een adaptieve toets is, zodat leerlingen van tevoren weten dat ze niet terug kunnen naar een vorig item. En het digitale, dat blijft gewoon altijd een aandachtspunt. Daarmee zitten we toch aan de gesloten vragen vast. Daarnaast is het ontwikkelen van een app voor ons een aandachtspunt.

Wat zijn aanbevelingen?

Wij hebben ook al andere adaptieve toetsen ontwikkeld en wij zijn daar ontzettend tevreden over. Maar het is wel een erg ingewikkeld proces en het neemt veel meer tijd en geld in beslag om een adaptieve toets te maken dan een standaardtoets. Het is dus ook een kostenoverweging om het eventueel niet te doen. De keuze voor adaptief toetsen moet je doen op basis van het argument dat een adaptieve toets een leerling meer recht doet dan een reguliere toets, niet omdat je denkt dat het de makkelijkste weg is.

Wat zou je anders doen als je opnieuw begon?

Het was prettig geweest om meer tijd te hebben. Het is enorm intensief geweest.

Eindexamens, Macedonië

Wat bevalt goed?

Sommige vragen zijn erg goed en deze toets biedt een mooie manier om het hele jaar door te nemen. Ook voor het toetsen van algemene kennis is deze toets erg goed.

Wat zijn aandachtspunten?

Sommige vragen zijn niet relevant, te moeilijk of juist te makkelijk voor leerlingen. Het Ministerie zou dit kunnen verbeteren door leerkrachten te raadplegen bij het ontwerpen van de toetsen. Een ander probleem is dat alle vragen zijn gebaseerd op een specifieke set van boeken. Scholen mogen echter zelf besluiten welk boek zij gebruiken in hun curriculum. Dit betekent echter dat leerlingen belangrijk informatie kunnen missen en vragen krijgen waar ze niks van af weten. Leerkrachten proberen hierbij te helpen door eerdere tentamens te oefenen.

Een punt waar zware kritiek op gekomen is, is dat de overheid de inkomens van leerkrachten naar beneden of juist naar boven wilde aanpassen op basis van het verschil tussen de twee verschillende cijfers die een leerling krijgt.

Wat zijn aanbevelingen?

Het hele examen in één keer geven aan de leerlingen zodat ze zelf de volgorde van de vragen kunnen bepalen. Meerkeuzevragen zijn ook fijn.

Wat zou je anders doen als je opnieuw begon?

Meer contact met leerkrachten zodat het geen top-down toets zou worden zoals hij nu wel is.

NILA en NINA, Noord-Ierland

Wat bevalt goed?

De adaptiviteit van de toetsen is één van de kernaspecten die deze toetsen zo goed maken. Het NINA/NILA team zou niemand aanraden om terug te gaan naar een lineair toetsingssysteem. Een adaptieve toets geeft elke leerling de kans om zijn of haar volle potentie te bereiken en voorkomt dat leerlingen gefrustreerd raken door het niveau van de toets. Twee andere aspecten die zij als zeer belangrijk beschouwen zijn de hoge betrouwbaarheid en validiteit.

Ook het lagere instappunt voor bepaalde leerlingen die Engels als een tweede taal hebben, een leerachterstand hebben of andere speciale behoefte hebben, bevalt heel goed. Dit bestond in het begin niet en dat leidde er toe dat bepaalde leerlingen grote moeite hadden om de toets te maken en dat hierdoor veel frustraties werden waargenomen. Leerkrachten kunnen er sinds twee jaar voor kiezen om deze leerlingen te laten beginnen bij een lager startniveau. Deze leerlingen komen meestal niveau 1 niet uit. Het toevoegen van dit lagere niveau heeft hun ervaring met de toets verbeterd.

Naast deze punten is het automatische opslaan van de voortgang en dat soort standaard faciliteiten ook als zeer prettig ervaren door de deelnemende scholen.

Wat zijn aanbevelingen?

Een toets met deze kenmerken moet je niet overhaast ontwikkelen. Dat is in het verleden wel gebeurd waardoor er veel slordigheden in zaten. Op basis daarvan zijn een aantal scholen

gestopt met het afnemen van deze toetsen. Nu de kwaliteit weer op orde is, neemt het aantal gebruikers weer toe.

Wat zou je anders doen als je opnieuw begon?

De eerste trial was tien jaar geleden en sinds dien zijn veel dingen veranderd.

3.3 ONDERZOEKSVRAAG 2. IMPACT VAN DIGITAAL TOETSEN OP TOETSRESULTATEN

Indien schriftelijke toetsen vervangen worden door computer (adaptieve) toetsen, dan is vereist dat de toetsresultaten niet door deze veranderde toetsvorm mogen worden beïnvloed. Via een literatuurstudie is onderzocht welke impact digitaal toetsen kan hebben op toetsresultaten van leerlingen en hoe deze eventuele negatieve impact voorkomen kan worden.

Bij de beantwoording van deze vraag is de zoektocht beperkt tot de vraag naar het verschil tussen de impact van digitaal toetsen op toetsresultaten ten opzichte van schriftelijke toetsen en is niet in het algemeen gekeken naar de impact die een toets heeft op het presteren van een student. Een verschil in toetsresultaat kan veroorzaakt worden door de toetsvorm (schriftelijk vs. digitaal) of door het karakter van de toets (lineair vs. adaptief). Het is echter bekend dat het toetsresultaat ook door andere aspecten beïnvloedt kan worden. Zo is de perceptie die een student heeft over toetsen van invloed op de prestatie. Als een toets niet overeenkomt met de verwachting, kunnen studenten gedemotiveerd raken of delen van een toets verkeerd interpreteren, waardoor het toetsresultaat niet het werkelijke prestatieniveau van de student weergeeft (Birenbaum, 2007; Downing & Haladyna, 1997). Maar ook de kwaliteit van een toets heeft consequenties voor het uiteindelijke resultaat van de leerling. In dit onderzoek is ervanuit gegaan dat de gebruikte toetsen in de onderzoeken kwalitatief goed zijn. Dat betekent dat zij inhoudsvalide en betrouwbaar zijn. Voor computer adaptieve toetsen betekent dit dat de itembank waar de items uit geselecteerd worden voldoende groot is en beschikt over alle informatie per item die nodig is om nauwkeurig te meten.

In eerste instantie is gekeken wat de verschillen zijn ten aanzien van toetsresultaten tussen schriftelijke toetsen en digitale (adaptieve) toetsen (paragraaf 3.3.1). Deze kunnen vergeleken worden met betrekking tot de validiteit, psychometrische gegevens en wijze van het vaststellen van de score. In de meeste studies wordt de vergelijking gedaan tussen een schriftelijke toets en een gedigitaliseerde versie van dezelfde toets. Bij de vergelijking tussen toetsresultaten van een digitale toets met een digitale adaptieve toetsen is de toets niet meer identiek, aangezien er bij de adaptieve toets een selectie van items gemaakt wordt op basis van eerder gegeven antwoorden. In paragraaf 3.3.2. wordt beschreven welke andere factoren een rol spelen bij de totstandkoming van toetsresultaten bij digitale (adaptieve) toetsen.

3.3.1 TOETSRESULTATEN BIJ DIGITALE (ADAPTIEVE) TOETSEN VS. SCHRIFTELIJKE TOETSEN

In een reviewstudie van Bunderson, Inouye en Olsen uit 1988 werden resultaten op verschillende schriftelijke studietoetsen en psychologische tests vergeleken met de resultaten op een identieke digitale versie. 52% van deze studies (n=23) lieten een verschil in resultaat zien; 13% een hogere score voor de digitale versie en 39% een hogere score voor de schriftelijke versie. De verschillen waren echter wel vrij klein en de drie studietoetsen in deze review (Olsen, Maynes, Slawson, & Ho, 1986; Wise, Boettcher, Harvey, & Plake, 1987; Wise & Wise, 1986) lieten alle drie geen significante verschillen zien. De studies hadden betrekking op een adaptieve

wiskundetoets voor basisschoolleerlingen. De leerlingen in de studie van Olsen et al. hadden al ervaring met het werken met een computer in de klas.

Begin jaren negentig is een aantal studies uitgevoerd waarbij wel verschillen gevonden zijn in studietoetsen. Schriftelijke toetsen werden beter gemaakt dan de digitale versies in de onderzoeken van Mazzeo, Druesne, Raffeld, Checketts, & Muhlstein, (1991) en van Mead en Drasgow (1993). Het betrof in deze studies toetsen voor wiskunde en Engels. De verschillen waren wel zeer klein. Het verschil bij de wiskundetoets werd niet teruggevonden in het onderzoek van Higgins, Patterson, Bozman, & Katz (2010). Ook als er gecorrigeerd werd voor computergebruik en voorkeur voor computertoetsen was er geen verschil.

Dat vakinhoud wel een rol kan spelen bleek ook uit een meta-analyse van Kingston (2009). Deze studie richtte zich op leerlingen in het Amerikaanse primair onderwijs (k-12 populatie). Er werden 81 studies onderzocht die uitgevoerd waren tussen 1997 en 2007. De conclusie was dat er een klein verschil is in leerprestaties als gevolg van de twee toetsvormen, maar dat de effectgroottes over de 81 studies zeer klein waren ($-.01$). Dit betekent dat het effect van de toetsvorm heel erg klein is. Er waren echter wel verschillen per vak. Bij Engels en Social studies was adaptief toetsen gunstiger en bij wiskunde was de schriftelijke versie positiever.

In het onderzoek van Clariana en Wallace (2002), Watson (2001) en Butters en Walstad (2011) presteerden studenten beter op de digitale toets. Geslacht, mate van gedrevenheid, leeftijd en ict-vaardigheid waren niet gerelateerd aan het verschil in prestatie. De beter presterende studenten hadden in beide studies het meeste profijt van de digitale toetsvorm.

Onderzoek uit de laatste jaren toont geen significante verschillen tussen de toetsresultaten op de verschillende toetsvormen (Anakwe, 2008; Boo & Vispoel, 2012; Bowman, Seo, & Taherbhai, 2012; Mojarrad, Hemmati, JafariGohar, & Sadeghi, 2013; Srivastava, Gray, Nippold, & Schneider, 2012; Tsai & Shin, 2013). Uit het onderzoek van Anakwe (2008) bleek dat het ook niet uit maakte of de studenten mannelijk of vrouwelijk waren.

Onderzoek naar de vergelijking tussen adaptief toetsen en lineaire digitale toetsen is beperkt. De vergelijking is ook lastig, aangezien elke adaptieve toets voor elke leerling uniek is doordat de samenstelling van de toets bepaald wordt door de antwoorden van de leerling. Schaeffer et al. (1995) bestudeerde de vergelijkbaarheid van toetsresultaten bij een digitale adaptieve toets (computer adaptieve test; CAT) met de toetsresultaten van leerlingen bij een lineaire digitale toets. De toetsresultaten waren bij de verbale en kwantitatieve toetsen vergelijkbaar, maar bij de analytische toets werd hoger gescoord op de adaptieve toets. In 2002 onderzochten Goldberg en Pedulla digitale toetsing waarbij de leerling de mogelijkheid heeft om heen en weer te bladeren vergeleken met digitale toetsing waarin de leerling deze mogelijkheid niet had (en daarmee lijkt op eigenschappen van een adaptieve toets); De leerlingen scoorden hoger op de test waar ze heen en weer konden bladeren. De vergelijking tussen de impact van digitaal adaptief toetsen en de impact van lineaire digitale toetsen op presteren vereist vervolgonderzoek (Rezaie & Golshan, 2015). Daarbij is het van belang om de mogelijkheden en eigenschappen van de verschillende vormen van toetsing mee te nemen, zoals het al dan niet kunnen bladeren tussen de items, inzicht in de lengte van de toets of mogelijkheid van het gebruik van multimedia.

3.3.2 BEÏNVLOEDENDE VARIABLEN

De basisvraag in dit onderzoek was wat de impact is van digitaal toetsen op de toetsresultaten van de leerlingen. De literatuurstudie laat echter zien dat de toetsvorm ook effect kan hebben op

andere aspecten, zoals motivatie en toetsangst. Daarmee kan de toetsvorm eventueel indirect effect hebben op de toetsresultaten. Het onderzoek van Wise et al. (1987) laat geen significant effect zien in toetsangst of in ervaring met computers. Een onderzoek naar toetsangst bij leerlingen van middelbare scholen laat daarentegen een hogere score voor toetsangst zien bij de schriftelijke toetsvorm dan de digitale adaptieve toetsvorm. Geslacht speelde daarbij geen rol (Fritts & Marszalek, 2010).

Meerdere onderzoeken tonen aan dat de studenten een voorkeur hebben voor een digitale variant en dat ze de werking van digitale toetsen waarderen (bijvoorbeeld Boo & Vispoel, 2012; Higgins, et al., 2010; Mojarrad et al., 2013). In een studie waarin studenten hun ervaringen met digitaal toetsen in vergelijking met schriftelijke toetsen beschrijven, geven ze aan dat ze de digitale toetsen efficiënt, onderwijsgericht, beter en helpend vinden (Ogilvie, Trusk, & Blue, 1999). Wat minder positieve reacties waren er bij taaltoetsen waarbij veel gescrold moest worden. Bij de overgang van schriftelijke toetsen naar digitale toetsen met open vragen bij de Open universiteit geven studenten aan dat deze toetsvorm beter aansluit bij de dagelijkse praktijk (Joosten-ten Brinke, et al., 2010). Daarbij vonden ze het typen (rammelen op het toetsenbord) wel storend voor de medestudenten.

Digitaal toetsen betreft niet alleen de toetsafname, maar ook de manier waarop de leerlingen feedback krijgen. Onderzoek dat zich richt op het schriftelijk dan wel digitaal geven van feedback bij een toets laat ook gemengde resultaten zien (anonieme auteur, 2016). Ten aanzien van houding en gedrag ten opzichte van digitale feedback toont Mumm & Mutlu (2011) een hoge motivatie en vinden Lin, Atkinson, Christopherson, Joseph, & Harrison (2013) en Sträfling, Fleischer, Polzer, Leutner, & Krämer (2010) geen effect. Voor de motivatie van studenten maakt het niet uit of de computer of een docent de toetsresultaten bekend maakt. Studenten accepteren digitale toetsing snel. Om dit verder te bevorderen is het belangrijk om duidelijk te zijn over het doel van de toets en ervoor te zorgen dat er sprake is van een gebruikersvriendelijk systeem (Terzis & Exonomides, 2011).

3.3.3 WAT ZEGGEN DE AANBIEDERS OVER DE IMPACT?

In de interviews is ook gevraagd naar de impact van digitaal toetsen. Ten aanzien van de toetsresultaten werd geen impact gemeld of werd expliciet gemeld dat daar nog geen onderzoek naar gedaan is (SBAC). De impact van digitaal toetsen betrof vooral de leerkrachten.

In Denemarken stellen leerkrachten op basis van de toetsresultaten hun lesinhoud bij, zodat hun klas beter scoort en zij zelf daardoor ook een betere beoordeling krijgen. Aangezien het een formatieve toets is, is dat niet wenselijk.

Aangezien de toets in Noorwegen, bij 10voordeleraar en in Macedonië summatief is, kan de impact van de toets groot zijn voor de leerlingen en studenten. Als leerlingen in Noorwegen hun toetsen niet halen in het derde jaar dan krijgen ze geen diploma en kunnen ze niet verder studeren. Alle vier de toetsen en de twee mondelingen moeten gehaald worden en ze kunnen niet compenseren met eerdere cijfers van eerder in het schooljaar. Bij de lerarenopleidingen in Nederland mogen studenten die de toets niet halen niet voor de klas staan. Als leerlingen in Macedonië het examen niet halen dan krijgen ze nog een kans om deze te nemen augustus. In principe zou iemand door slechte cijfers een jaar kunnen blijven zitten. In de praktijk komt dit echter niet voor vanwege het lage niveau van de vragen. Deze impact is echter niet toe te schrijven aan de toetsvorm.

4. CONCLUSIE EN DISCUSSIE

In dit onderzoek stonden twee deelvragen centraal:

1. Welke afnamecondities en eisen aan de ict-infrastructuur gelden voor het digitaal afnemen van een summatieve toets?
2. Welke impact kan digitaal toetsen hebben op toetsresultaten van leerlingen en hoe kan eventuele negatieve impact voorkomen worden?

4.1 AFNAMECONDITIES EN ICT-INFRASTRUCTUUR BIJ DIGITAAL TOETSEN

Voor het beschrijven van de afnamecondities en de eisen aan de ict-infrastructuur zijn interviews uitgevoerd met verschillende aanbieders. Geen van de aanbieders (uiteeraard m.u.v. Route-8 van A-vision) komt volledig overeen met de Nederlandse situatie, maar ze hebben wel kenmerken die geschikt zijn om te vertalen naar de Nederlandse situatie. Het interview in Macedonië was met een lerares. Zij was minder op de hoogte van gestelde eisen en beschreef de afnamecondities binnen de eigen school. Vier aanbieders werken met een computer adaptieve toets en deze toetsen hebben allemaal een formatieve functie, drie aanbieders hebben wel een eindtoets die summatief is, maar deze zijn niet adaptief. In Tabel 5 is een overzicht opgenomen van de algemene kenmerken van de aanbieders.

Voor het afnemen van de toets kan gebruik gemaakt worden van generieke toetssoftware of software die specifiek ontworpen en ontwikkeld is voor de toetsafname. In dit onderzoek was de toetssoftware meestal specifiek ontworpen en ontwikkeld voor de specifieke toetsafname. Alleen 10voordeleraar maakt gebruik van een commercieel beschikbaar toetservicesysteem QuestionMark Perception. In Noord-Ierland zijn er als gevolg van de aanbestedingsprocedure zelfs twee verschillende aanbieders voor de twee toetsen.

Tabel 5. Overzicht algemene toetskenmerken

	CAT, Dene- marken	SBAC, VS	Eindexa mens, Noor- wegen	10voord eleraar, Nederlan d	Route- 8 van A- Vision, Nederl and	Eindexa mens, Macedoni ë	NILA/ NINA, Noord- Ierland
Functie	formatief	Formatief/ summatief	Summatief	Summatief	formatief	summatief	Formatief
Adaptief	ja	combinatie	Nee	Nee	Ja	Nee	Ja
Onderwijstype	Leerjaar 3-8 PO	Grade 3-8 PO en 11 VO	Laatste jaar PO	Hbo	Laatste jaar PO	Laatste jaar PO	Leerjaar 3 - 8
Afname	Per klas	Per klas of individueel	Grote groepen	Per klas	Per klas of individueel	Klassen door elkaar	Per klas of individueel
Vraagvorm	mc	meerdere	Open vragen	Mc	Mc	Open	Mc
Vakken	Deens, leesvaardigh eid, wetenschap, wis, Engels	Wiskunde en Engels		17 hoofdvakken	Taal en rekenen	Alle PO- vakken	taal en rekenen
Volgsysteem	ja	ja	Nee	Nee	Nee	Nee	Nee

Adaptiviteit

Er zijn vier organisaties die werken met een adaptieve toets. Alle vier zijn zij hier zeer positief over. Voordelen van een adaptieve toets zijn dat ten eerste dat het adaptieve karakter ervoor zorgt ervoor dat de toetsafname veel korter wordt. Ten tweede sluit elke toets beter aan bij het niveau van de leerling. De vaardigheid van de leerling wordt met een hogere nauwkeurigheid gemeten, niet alleen bij de leerlingen die rond het gemiddelde scoren, maar ook voor zeer goed en zeer zwak scorende leerlingen. Hierdoor kan bij de zwakke leerlingen de frustratie van toetsen afnemen en is de toets ook uitdagend voor de sterkere leerlingen. Ten derde biedt een adaptieve toets meer organisatorische vrijheid. Aangezien elke leerling een andere toets voor zich krijgt, heeft afkijken geen zin, of is het mogelijk om leerlingen de toets na elkaar te laten maken. Dit betekent dat met een adaptieve toets een grotere spreiding mogelijk is en er niet direct meer devices aangeschaft moeten worden. Bij sommige aanbieders hebben leerkrachten een meekijkfunctie. Door deze meekijkfunctie kan een leerkracht een leerling eerder laat stoppen met de toets of kan een pauze worden ingelast.

Het adaptieve karakter van een toets moet wel goed met de leerlingen en ouders besproken worden om te voorkomen dat leerlingen denken dat ze ongelijk worden behandeld. Elke leerling krijgt namelijk een unieke toets, waardoor het voor leerlingen kan lijken of er oneerlijk getoetst wordt. Daarnaast waarschuwen de aanbieders wel voor het intensieve en tijdrovende traject dat voorafgaat aan het aanbieden van een adaptieve toets. Er moet een zeer groot aantal items ontwikkeld en getest worden. Ondanks dat, zou geen van de aanbieders terug willen naar het lineaire formaat.

Adaptief toetsen heeft ook enkele nadelen. Het is niet mogelijk om terug te bladeren in de toets, aangezien een volgende vraag geselecteerd wordt op basis van het antwoord op de vorige vraag. Een optie die hier iets aan tegemoet komt is om de toets niet item voor item op te bouwen, maar per blok. Binnen een blok van items kan de leerling dan wel terugbladeren. Een tweede nadeel is de grootte van de itembank. Deze moet bestaan uit een groot aantal items op alle niveaus en voor alle te meten vaardigheden. Om te voorkomen dat items bekend raken, is het belangrijk dat ze na een bepaalde tijd of na een bepaald aantal keren gebruikt te zijn uit de itembank gehaald worden.

Naast de genoemde voor- en nadelen vereist het maken van een adaptieve toets ook specifieke kennis bij de ontwerpers over de onderliggende theoretische modellen waarbij de vaardigheidsniveaus van leerlingen worden geschat op basis van gegeven antwoorden en de eigenschappen van de items. Het werken met adaptieve toetsen vereist echter dat er bij het ontwerp keuzes worden gemaakt met betrekking tot:

- het aantal te pretesten items. Er is een groot aantal gepreteste items nodig, waardoor de moeilijkheid en onderscheidend vermogen van de items op voorhand bekend is;
- de toetslengte. De toets mag niet te kort zijn, waardoor er geen accurate meting is, en die niet te lang waardoor items onnodig gebruikt worden en de leerling zich langer dan noodzakelijk moet concentreren;
- stopregels, bijvoorbeeld als er voldoende nauwkeurig gemeten is, bij een maximale toetslengte, bij het uitputten van de itembank of bij ongewenst gedrag van een leerling (bv. sneller antwoorden dan dat de vraag te lezen is);
- regels om juist niet te stoppen, bijvoorbeeld als de ondergrens van de toetslengte nog niet is bereikt, als bepaalde thema's nog niet zijn getoetst of dat er een vergelijkbare lengte van de toets is voor de hele groep.

Netwerkinfrastructuur

De aanbieders stellen minimumeisen aan de netwerkinfrastructuur, maar leggen de verantwoordelijkheid voor de werking van de systemen in het algemeen bij de school of overheid neer. De downloadsnelheid die nodig is om de toetsen te downloaden is toetsafhankelijk, maar de aanbieders houden rekening met de downloadsnelheid door beperkingen op te leggen aan de omvang van de toetsen. Niet alle aanbieders en gebruikers zijn goed op de hoogte van het betrouwbaarheidslevel van de software en de mate waarin rekening gehouden wordt met piekbelasting. Om problemen met piekbelasting te voorkomen wordt scholen geadviseerd om de afname te spreiden.

Voor de kwaliteit van de toetsen is het belangrijk dat de netwerkinfrastructuur geen belemmering mag opleveren. De aanbieders houden daar rekening mee door de omvang van de afzonderlijke items/toetsen te beperken, daarnaast wordt vooral door overheidsinstanties en ondersteunende bedrijven gezorgd voor de juiste voorzieningen op de scholen. Bij vragen van de scholen dient er altijd een technische vraagbaak beschikbaar te zijn.

Privacy en informatiebeveiliging

Alle aanbieders gaan bewust met de beschikbare informatie om. Over het algemeen proberen alle landen zo weinig mogelijk informatie te verzamelen over hun respondenten voordat zij aan de toetsen beginnen. In sommige gevallen, zoals Macedonië, Noorwegen en A-Vision, krijgen leerlingen een eenmalige inlogcode om de data anoniem te houden. In Noord-Ierland en Denemarken loggen leerlingen in met hun gebruikelijke inlogcodes voor een beveiligd systeem waarin ook hun eerdere gegevens staan. Bij SBAC wordt specifiek beschreven welke informatie verzameld wordt en welke informatie niet. Bij A-Vision bewaren ze zelf de gegevens die ze ontvangen, slechts enkele werknemers mogen bij deze gevoelige data. In Noord-Ierland worden de NILA en NINA afgenomen binnen een beveiligde internetomgeving die alleen toegankelijk is met een C2k inlogcode. De verdere informatiebeveiliging is de verantwoordelijkheid van de twee verschillende providers. Hiervoor zijn wel strenge regels opgesteld. Beide providers moeten ook de data opgeven bij Escrow, een apart bedrijf die gespecialiseerd is in informatiebeveiliging. In Noorwegen krijgen leerlingen een eenmalige inlogcode zodat alle informatie compleet geanonimiseerd is. Hiernaast wordt de data ook minstens 3 maanden bewaard, zodat meerdere mensen het examen na kunnen kijken mocht dit nodig zijn. De aanbieders zijn allen terughoudend ten aanzien van het delen van de informatie die opgeslagen wordt in de systemen. In Macedonië en Denemarken konden de geïnterviewden geen informatie geven over de manier van data-opslag, maar in geen van de landen zijn er problemen mee geweest.

Privacy van de gebruikers wordt gegarandeerd door de inlogcodes, verwijderen van namen voor het nakijken van de open vragen, en anonimiseren van data voorafgaand aan het evalueren van de kwaliteit van de toets op schoolniveau.

Hardware and Software Requirements Overview

Operating System	Minimum Smarter Balanced Requirements for Current Computers ¹²³	Recommended Smarter Balanced Minimum for New Purchases
Windows	Windows XP (service pack 3) Pentium 233 MHz processor 128 MB RAM 52 MB hard drive free space	Windows 7 to Windows 8.1 1GHz processor 1GB RAM 80 GB hard drive or at least 1GB of hard drive space available
Mac OS X	Mac OS X 10.4.4 Macintosh computer with Intel x86 or PowerPC G3 (300 MHz) processor, 256 MB RAM, 200 MB hard drive free space	Mac OS X 10.7 to OS X 10.9 1 GHz processor 1GB RAM 80 GB hard drive or at least 1GB of hard drive space available
Linux	Linux (Ubuntu 9-10, Fedora 6) Pentium II or AMD K6-III 233 MHz processor 64 MB RAM 52 MB hard drive free space	Linux (Ubuntu 11.10, Fedora 16) 1 GHz processor 1GB RAM 80 GB hard drive or at least 1GB of hard drive space available
iOS	iPads 2 running iOS6	iPads 3+ running iOS6
Android	Please see http://www.smarterbalanced.org for the current set of approved secure browsers and the devices for which they are approved.	Please see http://www.smarterbalanced.org for the current set of approved secure browsers and the devices for which they are approved.
Chrome OS	Chromebooks running Chrome OS (v31–v34)	Chromebooks running Chrome OS (v31–v34)

Figuur 2. De minimum en aanbevolen eisen aan devices van SBAC

In een aantal landen worden de toetsresultaten gebruikt voor het vormen van een oordeel over de kwaliteit van de scholen. Bij het eindexamen van Macedonië wordt het ook gebruikt als een beoordelingsinstrument voor het bepalen van de kwaliteit van de docent. De eindtoets is hierbij een criteriummaat voor de eigen inschatting van de docent.

Devices

De toetsen worden vooral afgenomen op computers en laptops. Een enkele toets mag afgenomen worden op een tablet. Geen van de aanbieders staat Bring your own device toe. De reden hiervoor is dat de aanbieders dan niet kunnen garanderen dat de toets goed draait of dat het device up-to-date is. Devices worden bij Denemarken, Noorwegen, Macedonië door de overheid geleverd. Bij NINA/NILA mag alleen apparatuur gebruikt worden die ondersteund worden door C2k, anders vervalt de ondersteuning.

Er zijn verschillende aspecten die de keuze voor een device bepalen:

- In Denemarken zou de toets op elk device kunnen worden gemaakt, maar de staat schenkt elke leerling een iPad mini. Dit maakt de keuze (geen keuze) voor scholen eenvoudig. Ook in Noord-Ierland wordt de hardware beschikbaar gesteld. In dit geval door C2k. Scholen mogen ook eigen devices gebruiken, maar als het dan niet werkt, kan C2k niet helpen. In Macedonië zijn een aantal jaar geleden ook desktop computers geschonken door de staat.
- Bij SBAC vinden ze dat scholen in de gelegenheid moeten worden gesteld om de devices te gebruiken die ze hebben, zodat er geen extra investeringen nodig zijn.
- In Noorwegen wordt elk jaar onderzocht welke laptop het beste is voor het aanstaande jaar. De laptops worden door de leraren getest op geschiktheid voor de software. Op basis van de bevindingen van de leraren wordt een keuze gemaakt.

Het aantal beschikbare devices per school verschilt, behalve als ze aan de school geleverd worden door de overheid. Aan de inrichting van de afnamelocatie worden geen specifieke eisen gesteld. De inrichting van de toetsruimte is een verantwoordelijkheid van de scholen. Alle devices worden ook voor andere onderwijsactiviteiten gebruikt. De minimumeisen aan de randapparatuur variëren per aanbieder. Bijna elke aanbieder geeft wel een aanbeveling voor een device met randapparatuur.

Het te gebruiken operating system wordt bij de meeste aanbieders ook aanbevolen. SBAC is ten aanzien van alle eisen het meest flexibel. Voor alle devicetypen geven zij aan welk operating system wenselijk is (zie Figuur 2). Het is niet mogelijk om een algemene richtlijn af te geven over de te gebruiken devices. Dit is namelijk afhankelijk van de functionaliteiten van het toetsprogramma (bv. wel of niet gebruik moeten maken van een toetsenbord).

4.2 IMPACT VAN DIGITAAL ADAPTIEF TOETSEN

In 2005 heeft een groep experts een internationale richtlijn opgesteld voor het digitaal toetsen (International Test Commission, 2005). In deze richtlijn staat dat verschillende toetsvormen die ingezet worden voor een vergelijkbare meting, een vergelijkbare betrouwbaarheid moeten hebben, dat ze met elkaar correleren en dat ze vergelijkbare gemiddelde scores en standaard deviaties moeten opleveren of dat er een goed gekalibreerd design onder zit. De digitale vormen bieden echter ook mogelijkheid om meer valide te meten. Afhankelijk van de leerdoelen zal bepaald moeten worden of het gaat om een volledig vergelijkbare meting of om een aanpassing van de meting ten behoeve van hogere inhoudsvaliditeit. Het is dan niet wenselijk om de toets in meerdere modi aan te bieden.

Het literatuuronderzoek naar de impact van digitaal (adaptief) toetsen laat echter zien dat de toetsresultaten niet altijd gelijk zijn voor een schriftelijke toets en een digitale (adaptieve) toets. De onderzoeken die echter verschillen laten zien, tonen zeer kleine verschillen. Dat lijkt op zich niet bezwaarlijk, maar kan bij een summatieve toets net het verschil zijn tussen zakken en slagen. Voor de eindtoets basisonderwijs zal een zeer klein verschil minder vergaande consequenties hebben, aangezien de eindtoetsen gebruikt worden ter ondersteuning van het advies van de leerkracht. Als de scores echter gebruikt zouden worden als toelatingseis voor het voortgezet onderwijs, dan kan een verschil van 1 punt leiden tot een verschil tussen een havo toelating of een vwo toelating. De eindtoets wordt echter niet ingezet als toelatingstoets. Onderzoek uit de laatste jaren lijkt er echter steeds meer op te wijzen dat er geen verschil is in toetsresultaten gerelateerd aan de afname vorm (digitaal vs. schriftelijk).

De Amerikaanse consortia die de Common Core standaarden toetsen (PARCC en SBAC) geven allebei aan dat het belangrijk is om de vergelijkbaarheid van de scores tussen verschillende uitleverformaten en tussen de staten en de consortia te onderzoeken. Ze hebben dat echter tot nu toe nog niet gedaan en stellen, zonder onderzoek, dat de resultaten vergelijkbaar zijn (Colwell, 2013).

De vraag is echter of het wel interessant is om de digitale versies te vergelijken met schriftelijke versies. De reden dat er gekeken wordt naar een mogelijke overstap naar digitaal toetsen wordt, naast efficiëntie, ingegeven door de mogelijkheid om de inhoudsvaliditeit van de toets te verhogen. In de meeste studies waarin de toetsresultaten bij verschillende toetsvormen met elkaar vergeleken worden, wordt in beide situaties gewerkt met identieke items. Voor de vergelijkbaarheid van de resultaten van deze onderzoeken en voor de interpretatie van de resultaten is dat een gewenste situatie. Echter de reden om over te stappen naar een digitale vorm van toetsen wordt ingegeven door de behoefte om ook aspecten te meten die niet met een schriftelijke toets te meten zijn. Bij de keuze voor een schriftelijke dan wel een digitale toets dient validiteit van de meting dan ook voorop te staan. Bij de huidige eindtoets basisonderwijs worden bepaalde vaardigheden niet gemeten omdat dat niet mogelijk is met een schriftelijke toetsvorm. Het meten van bepaalde vaardigheden, zoals luistervaardigheden, is digitaal beter te realiseren dan in een schriftelijke vorm. Ook kan het digitaal weergeven van bijvoorbeeld foto's van kunstwerken van een hogere kwaliteit zijn dan het weergeven van diezelfde foto's op geprinte schriftelijke toetsen. Het gebruik van dit soort multimedia in het onderwijs kan bijdragen aan een diepere of andere verwerking van de leerstof (Mayer, 2005). Indien deze multimedia ook bij de toetsing wordt gebruikt kan dat bijdragen aan de inhoudsvaliditeit van de toets, maar of het impact (positief dan wel negatief) heeft op de toetsresultaten van leerlingen is nog nauwelijks onderzocht. Jarodzka, Janssen, Kirschner en Erkens (2014) hebben aangetoond dat de regels voor gebruik van multimedia in instructie niet een-op-een over te zetten is naar het gebruik van multimedia in toetsing (zie ook Jarodzka & Kirschner, 2014). In een pilotstudy van Kraaijeveld (2016) lijkt er geen verschil te zijn in toetsresultaat als gevolg van het gebruik van multimedia, maar dit onderzoek is op zeer kleine schaal uitgevoerd.

Leerlingen en studenten geven wel steeds vaker een voorkeur voor de digitale toetsvorm aan. Dit is mede het gevolg van het dagelijks gebruik van computers in school en thuis. De adaptieve toetsvorm is echter in Nederland nog vrij nieuw en de onbekendheid zorgt voor de vraag of deze vorm van toetsen wel eerlijk is. Goede voorlichting over de werking van adaptieve toetsing is noodzakelijk. Het ontwikkelen van een digitale toets vereist de nodige inspanning. Alle items moeten gepretest worden bij een grote steekproef die representatief is voor de groep die uiteindelijk de toets moet afnemen. Items moeten op hun psychometrische en onderwijskundige kwaliteit beoordeeld worden en dat moet leiden tot een itembank die groot genoeg is om een nauwkeurige uitspraak te kunnen doen over alle leerlingen. Door deze uitgebreide ontwerp en ontwikkelfase wordt de kwaliteit van de toets geborgd. Daarnaast is het belangrijk dat de kwaliteit geborgd blijft bij de toetsafname.

Leerkrachten in het basisonderwijs voelen zich met de huidige positie van de eindtoets wel meer verantwoordelijk voor het op te stellen advies. Leerkrachten die hier minder vertrouwen in hebben, zoeken andere manieren om meer zekerheid te krijgen over hun eigen advies, bijvoorbeeld door het inzetten van intelligentietesten.

4.3 AANBEVELINGEN

Uit de interviews werd duidelijk dat het proces om te komen tot de uiteindelijke toets voor alle aanbieders zeer leerzaam is. De geleerde lessen leiden tot aanbevelingen voor ontwerpers en ontwikkelaars van digitale (adaptieve) toetsen en voor de aanbieders en afnemende scholen van deze toetsen. Enkele algemene aanbevelingen zijn:

- Houdt rekening met aannames en overtuigingen van de verschillende betrokkenen, leerkrachten, schoolleiders, ouders en kinderen over digitaal toetsen.
- Beschouw digitaal toetsen als een geheel, van ict-infrastructuur, de schoolkenmerken en de ervaringen, bij het nemen van beslissingen over digitaal toetsen.
- Neem digitaal toetsen op in het toetsbeleid van de school.
- Ontwikkel trainingen voor docenten en leerlingen om ervoor te zorgen dat leerlingen optimaal kunnen presteren.

Voor de ontwerpers en ontwikkelaars gelden de volgende aanbevelingen:

- Betrek leerkrachten zo vroeg mogelijk bij het hele proces vanaf ontwerp en itemconstructie.
- Werk met persoonlijke inloggegevens. Dat is ten eerste belangrijk in het kader van beveiliging, maar ten tweede ook om resultaten over jaren heen met elkaar te vergelijken.
- Vormgeving
 - Toon de items per item op het scherm. Doordat de vragen telkens per stuk worden getoond kunnen leerlingen zich goed concentreren.
 - Zorg voor een eenvoudige opmaak die niet afleidt van de inhoud.
- Het is prettig als de toets 'gepauzeerd' kan worden, zodat leerkrachten ervoor kunnen kiezen om een toets tijdelijk stil te leggen bij onvoorziene zaken.
- Start met een vaste toetsvorm voordat je adaptief gaat toetsen. Het ontwikkelen van een adaptieve toets is erg ingewikkeld. Overweeg in welke mate het belangrijk of noodzakelijk is voor de leerling om antwoorden op gemaakte vragen te kunnen aanpassen.
- Stel een ruim en realistisch tijdpad op voor de ontwikkeling van grote en goede itembanken voor adaptief toetsen.
- Geef expliciet de minimumeisen voor de infrastructuur, devices en randapparatuur aan om de toetsen probleemloos af te kunnen nemen.
- Zorg voor een oefentoets bij elke toetsvorm.
- Gebruik effectief en efficiënt multimedia, niet omdat het leuk is.
- Digitaliseer niet alleen het afnameproces, maar ook het nakijkproces.
- Zorg voor goede analysefunctionaliteit.
- Kies software die de leerlingen al kennen vanuit hun huiswerk en andere toetsen.
- Maak de toetsen leerstofonafhankelijk, zodat een brede inzet mogelijk is.
- Zorg dat voor de zwakkere leerlingen een lager instappunt beschikbaar is, zodat de toets niet vanaf het begin frustrerend is.
- Zorg dat de voortgang automatisch wordt opgeslagen.
- Zorg dat er een test beschikbaar is voor het meten van de geschiktheid van de eigen toetslocatie voor de toetsafname. Voorbeelden zijn de Preflight test, Ben ik geschikt test, en de Smarter Balanced Technology Readiness Calculator. Deze testen schatten

op basis van het aantal leerlingen, aantal beschikbare computers en aantal uren dat de computers beschikbaar zijn het aantal dagen en bijbehorende bandbreedte in die nodig is om de toetsen af te nemen.

Voor de aanbieders en afnemende scholen gelden de volgende aanbevelingen:

- Begin vroegtijdig met de communicatie met de ouders.
- Geef goede voorlichting over het adaptieve karakter van de toets. Leg uit dat elke toets dezelfde vergelijkbare inhoud meet, maar dat elke leerling een unieke toets krijgt, die afgestemd is op het niveau van de leerling. Hele goede leerlingen krijgen dan ook veel moeilijke items. Dat zijn ze bij reguliere toetsen niet gewend en dat kan ze bij onwetendheid onzeker maken.
- Zorg dat leerkrachten weten op welke manier ze de beschikbare data kunnen interpreteren en gebruiken voor het verbeteren van de eigen lespraktijk.
- Zorg voor ict ondersteuning op school of via de aanbieder of een tussenpersoon.
- Zorg voor een goede balans van verschillende (formatieve en summatieve) toetsen, zodat niet alles afhangt van deze eindtoets.
- Kies een toetsvorm die past bij het te meten construct.
- Informeer de leerlingen over het doel van de toets en laat ze kennismaken met het afnamesysteem voor de daadwerkelijke afname.
- Verhoog de afnamecapaciteit op scholen.
- Evalueer de verschillende aspecten van de toetsing, zoals kwaliteit randapparatuur, afnamecondities, inhoud toets, met de leerlingen.

Voor overheid en andere ketenpartners (OCW/CvTE):

- Begin vroegtijdig met de communicatie met de ouders.
- Geef goede voorlichting over het adaptieve karakter van de toets. Leg uit dat bij alle leerlingen dezelfde vergelijkbare inhoud getoetst wordt, maar dat elke leerling andere vragen krijgt in een unieke toets die afgestemd is op het niveau van de leerling. Dit betekent dat de leerlingen die hoog scoren meer moeilijke items krijgen om nauwkeuriger hun hoge score vast te stellen en leerlingen die laag scoren juist meer makkelijke items krijgen om hun lage score meer nauwkeurig vast te stellen. Dat zijn leerlingen bij niet-adaptieve toetsen niet gewend en dat kan ze bij onzeker maken als ze daarover niet geïnformeerd zijn.
- Verstrek scholen eenduidige informatie over de kwaliteit van de toetsen
- Biedt eenduidige ondersteuning bij het selecteren van toetsen.

Voor onderzoeksinstituten en kenniscentra:

- Uit het literatuuronderzoek bleek dat er weinig literatuur is die zich specifiek richt op computer adaptief toetsen in het primair onderwijs. Vooral de betekenis van het adaptieve karakter voor de doelgroep van kinderen tot 12 jaar blijft onderbelicht.
- Onderzoek de huidige praktijk en stel een onderzoeksprogramma op ten aanzien van het gebruik van digitaal (adaptief) toetsen bij deze doelgroep.
- Splits dit onderzoek uit naar evaluatieonderzoek en effectonderzoek
- Belangrijk subthema bij het onderzoek naar digitaal (adaptief) toetsen is onderzoek naar
 - het gebruik en impact van multimedia in toetsing;

- het gebruik van meer innovatieve toetsvormen.
- Geef docenten de mogelijkheid om hun eigen onderwijspraktijk te onderzoeken (praktijkonderzoek). Interessante thema's naar aanleiding van dit onderzoek zijn:
 - het gebruik van andere toetsvormen (digitaal in plaats van schriftelijk).
 - toetsvoorbereiding (werken met voorbeeldvragen)
 - de informatie die een docent nodig heeft om inzicht te krijgen in het niveau van de leerling (bv. via leerlingrapporten).

5. REFERENTIES

- Anakwe, B. (2008). Comparison of student performance in paper-based versus computer based testing. Dover: Delaware State University.
- Birenbaum, M. (2007). Evaluating the assessment: Sources of evidence for quality assurance. *Studies in Educational Evaluation*, 33, 29-49. doi:10.1016/j.stueduc.2007.01.004
- Boo, J. & Vispoel, W. (2012). Computer versus paper-and-pencil assessment of educational development: A comparison of psychometric features and examinee preferences. *Psychological Reports*, 111, 443-460. doi: 10.2466/10.03.11.PR0.111.5.443-460.
- Bowman, T., Seo, D., & Taherbhai, H. (2012). Comparison of paper-pencil and online performances of students with learning disabilities. *British Educational Research Journal*, 38(1), 61-74. doi:10.1080/01411926.2010.526193
- Bunderson, C. V., Inouye, D. K., & Olsen, J. B. (1988). The four generations of computerized educational assessment. Princeton, New Jersey: ETS Research Report. Verkregen op <http://onlinelibrary.wiley.com/doi/10.1002/j.2330-8516.1988.tb00291.x/epdf>
- Butters, R. B. & Walstad, W. B. (2011). Computer Versus Paper Testing in Precollege Economics. *The Journal of Economic Education*, 42, 366-374. DOI: 10.1080/00220485.2011.606087
- Clarian, R. & Wallace, P. (2002). Paper-based versus computer-based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology*, 33, 593-602.
- Chua, Y. P., & Don, Z. M. (2013). Effects of computer-based educational achievement test on test performance and test takers' motivation. *Computers in Human Behavior*, 29, 1889-1895. doi:10.1016/j.chb.2013.03.008
- Colwell, N. M. (2013). Test anxiety, computer-adaptive testing and the Common Core. *Journal of Education and Training Studies*, 1(2), 50-60. doi:10.11114/jets.v1i2.101
- CvTE. (2008). De computer bij de centrale examens 2009-2013. Duidelijk digitaal, 1 (oktober 2008).
- CvTE. (2014). Toetswijzer eindtoets PO algemeen deel. Verkregen op https://www.hetcvte.nl/nieuws/20141020/toetswijzer_eindtoets_po_algemeen.
- CvTE. (2015). De computer bij centrale examens en rekentoets VO 2015 - 2018. Duidelijk digitaal, 8 (september 2015).
- Downing, S. M., & Haladyna, T. M. (1997). Test item development: Validity evidence from quality assurance procedures. *Applied Measurement in Education*, 10, 61-82. doi:10.1207/s15324818ame1001_4
- Fritts, B. E. & Marszalek, J. M. (2010). Computerized adaptive testing, anxiety levels, and gender differences. *Social Psychological Education*, 13, 441-458. doi 10.1007/s11218-010-9113-3.
- Goldberg, A. L., & Pedulla, J. J. (2002). Performance differences according to test mode and computer familiarity on a practice graduate record exam. *Educational and Psychological Measurement*, 62, 1053-1067.

Higgins, J., Patterson, M.B., Bozman, M., & Katz, M. (2010). Examining the Feasibility and Effect of Transitioning GED Tests to Computer. *Journal of Technology, Learning, and Assessment*, 10(2). Verkregen via <http://www.jtla.org>.

International Test Commission (2005). International guidelines on computer-based and Internet delivered testing. www.intestcom.org

Jesson, J. K., Matheson, L., & Lacey, F. M. (2011). *Doing Your Literature Review. Traditional and Systematic Techniques*. Los Angeles: Sage.

Jarodzka, H., Janssen, N., Kirschner, P. A., & Erkens, G. (2014). Avoiding split attention in computer-based testing: is neglecting additional information facilitative? *British journal of educational technology*, 46(4), 803-817.

Jarodzka, H., & Kirschner, P. A. (2014). Digitale toetsen: waar moet je op letten? *4W*, 3(1), 24-31.

Joosten-ten Brinke, D. (2011). Eigentijds toetsen en beoordelen. Lectorale rede. Tilburg: Fontys Lerarenopleiding Tilburg. www.fontys.nl/generiek/bronnenbank/sendfile.aspx?id=251356.

Jordan, S. (2012). Student engagement with assessment and feedback: Some lessons from short-answer free-text e-assessment questions. *Computers & Education*, 58, 818-834. doi:10.1016/j.compedu.2011.10.007

Kennisnet. (2014). Digitaal toetsen in het (basis)onderwijs: de internationale ontwikkelingen. Een online verkenning. Zoetermeer: Kennisnet.

Kennisnet. (2015). Vier in balans-monitor 2015. Zoetermeer: Kennisnet.

Kingston, N. M. (2009). Comparability of computer- and paper-administered multiple-choice tests for K-12 populations: A synthesis. *Applied Measurement in Education*, 22, 22-37. Doi: 10.1080/08957340802558326.

Lin, L., Atkinson, R. K., Christopherson, R. M., Joseph, S. S., & Harrison, C. J. (2013). Animated agents and learning: Does the type of verbal feedback they provide matter?. *Computers & Education*, 67, 239-249. doi:10.1016/j.compedu.2013.04.017

Marriott, P. (2009). Students' evaluation of the use of online summative assessment on an undergraduate financial accounting module. *British Journal of Educational Technology*, 40, 237-254. doi:10.1111/j.1467-8535.2008.00924.x

Mayer, R. E. (2005). *The Cambridge handbook of multimedia learning*. New York: Cambridge University Press.

Mazzeo, J., Druesne, B., Raffeld, P. C., Checketts, K. T., & Muhlstein, A. (1991). Comparability of computer and paper-and-pencil scores for two CLEP general examinations. College Board report No. 91-5. (Available from ERIC: ED 344 902)

Mead, A.D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114, 449-458.

Ministerie van OCW. (2014). Toetsbesluit PO. Den Haag: Ministerie van OCW. Te raadplegen op <https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/besluiten/2014/01/20/toetsbesluit-po/toetsbesluit.pdf>

Mojarrad, H., Hemmati, F., JafariGohar, M., & Sadeghi, A. (2013). Computer-based assessment (CBA) vs. Paper/pencil-based assessment (PPBA): An investigation into the performance and attitude of Iranian EFL learners' reading comprehension. *International Journal of Language Learning and Applied Linguistics World*, 4, 418-428.

Mumm, J., & Mutlu, B. (2011). Designing motivational agents: The role of praise, social comparison, and embodiment in computer feedback. *Computers in Human Behavior*, 27, 1643-1650. doi:10.1016/j.chb.2011.02.002

Ogilvie, R. W., Trusk, T. C., & Blue, A. V. (1999). Students attitudes towards computer testing in a basic science course. *Med Educ*, 33, 828-831.

Olsen, J. B., Maynes, D. M., Slawson, D. A. & Ho, K. (1986). Comparison and equating of paper-administered, computer-administered and computerized adaptive tests of achievement. Paper presented at the meeting of the AERA, San Francisco.

Opdenakker, R. (2006, August). Advantages and Disadvantages of Four Interview Techniques in Qualitative Research. *Forum: Qualitative Social Research*, 7(4), Art. 11, <http://nbn-resolving.de/urn:nbn:de:0114-fqs0604118>.

Rezaie, M. & Golshan, M. (2015). Computer Adaptive Test (CAT): Advantages and Limitations *International Journal of Educational Investigations*, 2, 128-137.

Schaeffer, G. A., Steffen, M., Golub-Smith, M.L., Mills, C.N., & Durso, R. (1995). The introduction and comparability of the computer adaptive GRE general test (GRE Board Professional Report No. 88-08aP). Princeton, NJ: Educational Testing Service.

Srivastava, P., Gray, S., Nippold, M., & Schneider, P. (2012). Computer-Based and Paper-Based Reading Comprehension in Adolescents With Typical Language Development and Language-Learning Disabilities. *Language, Speech & Hearing Services in Schools*, 43, 424-437. doi:10.1044/0161-1461(2012/10-0108

Sträfling, N., Fleischer, I., Polzer, C., Leutner, D., & Krämer, N. C. (2010). Teaching learning strategies with a pedagogical agent: The effects of a virtual tutor and its appearance on learning and motivation. *Journal of Media Psychology: Theories, Methods, and Applications*, 22(2), 73-83. doi:10.1027/1864-1105/a000010.

Surf. (2014). Richtlijn veilige digitale toetsafname. Utrecht: Surf. <https://www.surf.nl/binaries/content/assets/surf/nl/kennisbank/2013/richtsnoer-veilige-digitale-toetsafname-versie1.05-ttl.pdf>

Terzis, V. & Economides, A. A. (2011). The acceptance and use of computer based assessment. *Computer & Education*, 56, 1032-1044.

Tsai, T. H., & Shin, C. D. (2013). A Score Comparability Study for the NBDHE Paper-Pencil Versus Computer Versions. *Evaluation & the health professions*, 36, 228-239.

Verhoeven, N. (2014). *Wat is onderzoek? Praktijkboek voor methoden en technieken* (vijfde druk). Amsterdam: Boom Lemma uitgevers.

Watson, B. (2001). Key factors affecting conceptual gains from CAL. *British Journal of Educational Technology*, 32, 587-593.

Wise, L. A., & Wise, S. L. (1986). Comparison of computer-administered and paper-administered achievement tests with elementary school children. In

Wise, S. L., Boettcher, L. L. Harvey, A. L., & Plake, B. S. (1987). Computer-based testing versus paper-pencil testing: Effects of computer anxiety and computer experience. Paper presented at the meeting of the AERA, Washington, DC.

BIJLAGEN

A. INTERVIEWPROTOCOL

Interview protocol: ICT - infrastructure of adaptive digital assessment

- 1. What is your name and function within this assessment?**
- 2. How broadly are these assessments used?**
- 3. What is the exact goal of the assessment?**
 - a. Is the exam used to provide a clearer image of general school level, teacher level, individual students? Or for the entire country or state?
 - b. Is the function of the assessment formative and/or summative?
 - c. Are the results from the exam used to advise students on their progress?
 - d. Which subjects are tested?
 - e. Who is the target population?
 - f. Who owns the exam?
 - g. How is the collected data shared with the outside world?

The following questions will be asked to give us some insight into the basic requirements of ICT-infrastructure and testing conditions for existing (adaptive) large-scale digital summative assessments. The following variables will be addressed within these questions:

Connectivity, Devices, Software, User privacy, Lessons learned.

- 1. Which requirements are imposed on the school in terms of connectivity (ex: Internet connection)?**
 - a. Broadband
 - i. Broadband via internet cable?
 - ii. Broadband via fiber optic connection?
 - iii. No broadband but instead via the phone line (ADSL or VDSL)?
 - iv. No broadband internet, but something else?
 - b. Download speed
 - i. Up to 100 Mbit
 - ii. 100 – 500 Mbit
 - iii. 500 -1000 Mbit
 - iv. > 1000 Mbit. How quick?
 - c. Reliability level of this download speed?
 - i. ...%
 - d. Set to a peak load?
 - i. Yes, testing moments are spread out
 - ii. Yes, computers and internet are strong enough for simultaneous testing
 - iii. No
 - e. Cabling?
 - i. Adequate cables
 - ii. WIFI
 - f. Wifi quality
 - i. Advanced and controllable access points?
 - ii. Based on industry standards?
 - g. Is a stress-test carried out before testing?
 - i. De minimal amount of time for the rendering of the exam?
 - ii. De maximum amount of time for the rendering of the exam?
 - iii. The average rendering time of the exam?
- 2. Which requirements are imposed on the school in terms of devices?**
 - a. Type of devices
 - i. Computers
 - ii. Laptops

- iii. Tablets
 - iv. Chromebooks
 - v. Laptop with specific software for the assessment
- b. Number of devices per school
 - i. 1 per student
 - ii. 1 per 5 students
 - iii. 1 per 10 students
 - iv. 1 per classroom
- c. Location of devices
 - i. Classrooms
 - ii. In a special computer room
 - iii. With the students throughout the day
 - iv. Outside the school
- d. Placing of the devices
 - i. Fixed vs. flexible
- e. Are the devices used for anything outside of testing?
- f. Do the schools have their own physical server or do they use a virtual server in a datacenter? Or do the schools use the cloud?
- g. Bring your own device?
 - i. If yes, how was the security of the exam ensured and how was cheating avoided?
- h. What should the quality be of the computer screens in terms of resolution?
- i. What operating system is needed for the assessment?
 - i. Windows (Microsoft), OSx (Apple), Chrome OS (Google) or Android?
- j. Internet browser?
- k. Are there any other requirements in terms of devices:
 - i. Monitor?
 - ii. Processor?
 - iii. Internal memory?
 - iv. Hard disk?
 - v. CD/DVD station?
 - vi. Extra equipment: headset? microphone?
 - vii. Network adapter?
 - viii. Audio/video streaming?
 - ix. Audio output?
 - x. Video RAM?
 - xi. Printer?
 - xii. Pop-up block?
- l. What are these requirements based on? Why were these specific choices made?
- m. Were there any other requirements which have not been covered by these questions?
- n. Did you carry out any research into which devices were best for your specific assessment and what impact they had?

3. The next questions will be about the software used for the assessments

- a. What software is/was used for the summative testing?
- b. What software is/was used for the formative testing?
 - i. Was this software designed especially for this assessment?
- c. What are the characteristics of the software?
 - i. Do questions come up one at a time?
 - ii. Are the questions build up in blocks?
 - iii. Do students receive feedback after each question?
 - iv. Do students receive feedback at the end of the exam? Do they receive a grade?

- v. Is the test/software adaptive? So does each student receive a personalized exam based on earlier answers? Or does each student receive the same standardized exam?
- vi. Is multimedia used in the exam?
- d. Does the software run in the cloud, in the server at school, or in each separate device?
- e. Does the exam save as you go?
- f. What are the functionalities of the software?
 - i. Item construction and item management?
 - ii. Test construct and profile: multiple choice, open questions, listening, writing assignments
 - iii. Import and export facilities
 - iv. Reporting and analysis of results
 - v. Process monitoring and workflow support.
 - vi. Is the assessment adaptive?

4. The following questions are about the organization of the assessment

- a. Who are involved in the development of the assessment?
- b. Who are involved in the implementation?
 - i. teachers, ict- coordinator, etc.
- c. Which expertise is necessary within the schools to successfully implement the assessment?
- d. Is the test taken in groups or individually?
 - i. How big is the group?
 - ii. Are people ever allowed to work together?

5. Which privacy requirements and guidelines are followed? How is data security ensured?

6. Where is the test taken?

- a. Which requirements are there of the physical location?

7. Lessons learned

- a. What aspects of the exam work well?
- b. Which would you recommend for use in other exams?
- c. Which aspects need to be worked out more? Or which aspects did you discover needed more attention once the first trial had been run?
- d. What would you do differently if you could start over?
- e. How was the contact with the target group from the beginning?
- f. How was the contact with the schools from the beginning?
- g. What impact do the results from this exam have on students?

B. URENVERANTWOORDING

De uitgevoerde activiteiten, geplande uren en gerealiseerde uren zijn opgenomen in Tabel 6.

Tabel 6. Overzicht van activiteiten, planning en inzet.

Activiteit	Datum/ periode	Geplande uren	Gerealiseerde uren	Onderbouwing
Startbijeenkomst in Zoetermeer (2 personen)	13/11/15	8	2	1 onderzoeker naar Zoetermeer
Ontwikkelen onderzoeksmaterialen (opstellen interviewprotocol en analyseprotocol voor 'andere bronnen')	16/11/15 - 25/11/15	40	30	2 onderzoekers à 20 uur pp
Vaststellen en benaderen te onderzoeken populatie en 'andere' bronnen	16/11/15 - 20/11/15	10	15	1 onderzoeker (gezamenlijk netwerk van collega's in het Welten-instituut)
Literatuuronderzoek naar 'impact'	16/11/15 - 16/12/15	20	30	1 onderzoeker
Afnemen telefonische interviews (n = 10)	14/12/15 - 15/12/15	15	10	10 interviews van 1,5 uur (7 interviews)
Verwerken informatie analyse 'andere bronnen'	15/12/15 - 21/12/15	20	10	2 onderzoekers
Verwerken telefonische interviews	15/12/15 - 16/12/15	40	50	4 uur per interview (bleek meer te zijn)
Verwerking en samenvoegen gegevens, analyse en bespreken Opstellen voorlopige resultaten	16/12/15 - 31/1/2016	60	54	3 onderzoekers à 20 uur
Schrijven eerste rapportage (voorlopige resultaten en conclusies)	21/12/15 - 4/2/2016	30	30	20 uur schrijven; 10 uur feedback en herschrijven
Eerste rapportage bij Kennisnet	31/12/15	PM		
Feedback geven op eerste rapportage	15/2/2016 - 7/3/2016	PM	Februari	Te reserveren tijd bij Kennisnet voor het geven van feedback
Bespreking rapport bij Kennisnet	7/3/16	8	8	2 personen
Feedback van Kennisnet verwerken	8/3/2016 - 31/3/2016	20	18	Completeren rapportage en verwerken feedback Kennisnet
Indienen eindrapport bij Kennisnet	31/3/16	PM		
Totaal		263 uur	257	

