

# DIGITAAL TOETSEN: KANSEN VOOR HET HOGER ONDERWIJS

INZICHT IN DE EFFECTEN VAN DIGITAAL TOETSEN;  
ERVARINGEN UIT HET PROGRAMMA TOETSING  
EN TOETSGESTUURD LEREN, 2010-2014

**SURF**

# INHOUDSOPGAVE

<b>Samenvatting</b>	<b>3</b>
<b>Samenwerken aan digitaal toetsen in het hoger onderwijs</b>	<b>4</b>
<b>Resultaten en bevindingen</b>	<b>5</b>
1. Verhoging studiesucces	5
2. Vermindering van de werkdruk van docenten	6
3. Kwaliteitsverbetering	7
4. Overige bevindingen	11
<b>Conclusies</b>	<b>12</b>
<b>Bijlage 1: De projecten</b>	<b>14</b>
<b>Bijlage 2: Literatuur</b>	<b>18</b>
<b>Colofon</b>	<b>19</b>

## SAMENVATTING

Samenwerking verbetert op verschillende manieren de kwaliteit van toetsen. Dat concludeert SURF op basis van de resultaten van 18 projecten die instellingen tussen 2010 en 2014 binnen het programma *Toetsing en Toetsgestuurd Leren* hebben uitgevoerd. In de projecten werkten 38 instellingen voor hoger onderwijs samen aan het ontwikkelen, testen en implementeren van diverse vormen van digitaal toetsen in het hoger onderwijs. De projecten richtten zich op het verminderen van studie-uitval, het verlagen van de werkdruk van docenten en het verbeteren van de kwaliteit van toetsen. Het programma is uitgevoerd in opdracht van het Ministerie van Onderwijs, Cultuur en Wetenschap. Na afronding van de projecten heeft SURF een evaluatie uitgevoerd naar de resultaten en effecten.

### **Digitaal toetsen: veelbelovend voor de kwaliteit**

Uit de projecten blijkt dat de kwaliteit van toetsen door samenwerking op verschillende manieren verbetert. Door toetsitems gezamenlijk te ontwikkelen en te gebruiken wordt meer aandacht besteed aan itemontwikkeling. Het gezamenlijk gebruik van itembanken leidt tot grotere hoeveelheden toetsresultaten, waardoor psychometrische analyses mogelijk zijn en de kwaliteit van itembanken toeneemt. Digitaal toetsen maakt het mogelijk andere, rijkere vraagvormen in te zetten waardoor toetsen beter aansluiten op de beroepspraktijk. Digitale afname zorgt er bovendien voor dat toetsen betrouwbaar worden uitgevoerd, toetsvragen objectief worden gecorrigeerd en de resultaten sneller beschikbaar zijn.

De meeste projecten die zich richtten op de verhoging van het studiesucces laten een positief effect zien. Op basis van de uitgevoerde effectmetingen en evaluaties is het aannemelijk dat de inzet van digitaal toetsen bijdraagt aan de leerprestaties. Gezien de beperkte looptijd en omvang van de projecten is nader onderzoek nodig voor onderbouwing hiervan.

In de meeste projecten is tijdens het uitvoeren van de projecten nog geen werkdrukvermindering voor docenten vastgesteld. Bij de invoering van digitale toetsen is veel tijd en aandacht nodig voor het ontwikkelen en reviewen van toetsitems (met als voordeel dat de kwaliteit ervan toeneemt) en docenten hebben tijd nodig om vertrouwd te raken met de software. Op basis van de ervaringen in de projecten verwachten betrokkenen dat op langere termijn, als docenten eenmaal kunnen beschikken over goed gevulde itembanken, de tijdsinvestering minder wordt en dan wel leidt tot werkdrukverlaging. Het nakijken van digitale toetsen kost aanzienlijk minder tijd en het verwerken en analyseren van de resultaten kan sneller en efficiënter plaatsvinden dan bij papieren toetsen.

“Op basis van toetsen nemen hogeronderwijsinstellingen beslissingen over hun studenten. De kwaliteit van toetsen mag daarom niet ter discussie staan. Het programma *Toetsing en Toetsgestuurd Leren* maakt duidelijk dat ICT op verschillende aspecten de kwaliteit van het toetsproces kan verbeteren. Ik vind dit een waardevolle aanvulling op de adviezen voor externe validatie van de Commissie Bruijn. De conclusies uit het programma geven aanleiding om de mogelijkheden van digitaal toetsen in relatie tot kwaliteitsverbetering en innovatie verder te exploreren.”

**Ron Minnée**, directeur Hoger Onderwijs en Studiefinanciering, Ministerie van OCW

# SAMENWERKEN AAN DIGITAAL TOETSEN IN HET HOGER ONDERWIJS

Toetsen raakt het hart van het onderwijsproces. Door middel van summatief toetsen wordt vastgesteld of een student aan de vereiste kwaliteitseisen van een opleiding voldoet. Formatieve toetsen, zoals oefentoetsen en diagnostische toetsen, ondersteunen het leerproces van de student. Het verhogen van de kwaliteit van toetsen en beoordelen staat in het hoger onderwijs dat laatste jaren hoog op de agenda. Enerzijds zijn de inzichten gegroeid hoe toetsen het leergedrag van studenten kan sturen, anderzijds stimuleren strengere accreditatie-eisen om toetsing meer instellingsoverstijgend te organiseren en waar mogelijk te werken met externe beoordelaars.

Het digitaliseren van toetsen biedt kansen voor nieuwe vormen van toetsen, kan een bijdrage leveren aan de kwaliteit van toetsing en kan het toetsproces efficiënter en gemakkelijker maken voor studenten en docenten. Mede daardoor groeit de vraag naar digitaal toetsen in het hoger onderwijs snel.

In 2010 is SURF in opdracht van het Ministerie van Onderwijs, Cultuur en Wetenschappen gestart met het programma *Toetsing en Toetsgestuurd Leren*. Doel van het programma was een bijdrage te leveren aan:

- Het verhogen van het studiesucces
- Verlagen van de werkdruk van docenten
- Verhogen van de kwaliteit van toetsen.

Op uitnodiging van SURF hebben instellingen voor hoger onderwijs in 18 projecten onderzocht of zij met digitaal toetsen een bijdrage konden leveren aan de doelstellingen van het programma (zie bijlage 1 voor een korte beschrijving van de projecten). Op basis van de rapportages, effectmetingen en evaluaties van de projecten en interviews met projectleiders en toetsdeskundigen is onderzocht in hoeverre de projecten daarin geslaagd zijn.

“In verschillende domeinen werken hogescholen samen aan toetsbanken. Dat brengt een hele nieuwe dynamiek op gang: docenten discussiëren met elkaar over wat goede toetsitems zijn. Ik ben ervan overtuigd dat zo'n instellingsoverstijgende dialoog grote impact heeft op de kwaliteit van toetsen. Dankzij de praktische kennis uit het programma *Toetsing en Toetsgestuurd Leren*, kunnen die domeinen waar nog niet wordt samengewerkt een vliegende start maken. De gezamenlijke kennisopbouw en kennisuitwisseling binnen SURF op dit onderwerp zouden we moeten voortzetten.”

**Huug de Deugd**, lid College van Bestuur Hogeschool Inholland

# RESULTATEN EN BEVINDINGEN

8 projecten waren gericht op het verhogen van het studiesucces, 11 projecten op het verlagen van de werkdruk van docenten en 14 projecten op het verbeteren van de toetskwaliteit. De meeste projecten richtten zich op meer dan één programma-doelstelling. Om die doelstellingen te bereiken, zijn gezamenlijke itembanken en toetsvormen ontwikkeld voor zowel formatieve als summatieve toetsen, en is er geëxperimenteerd met specifieke vormen van toetsen, zoals het digitaal nakijken van werkstukken of het inzetten van een digitaal portfolio. De projecten waren primair gericht op het realiseren van onderwijsinnovaties, daarnaast is aandacht besteed aan het meten van effecten. In veel gevallen zijn evaluaties uitgevoerd op basis van kortlopende pilots, waarna docenten en studenten zijn ondervraagd. In sommige gevallen kon een voor- en nameting plaatsvinden. Desalniettemin levert dit een goede indicatie op van de kwaliteit van de innovatie en de te verwachten effecten op werkdruk en studiesucces.

## 1. Verhoging studiesucces

De inzet van formatieve toetsen en het geven van feedback heeft een gunstig effect op het leren van studenten. Het geeft studenten inzicht in hun eigen prestaties ten opzichte van hun leerdoelen, ze leren van hun fouten en krijgen feedback waar ze terecht kunnen voor aanvullende kennis. Daardoor kunnen ze tijdig aan de slag met hiaten in hun kennis en vaardigheden, wat vaak een positief effect heeft op hun studieresultaten. Bovendien stimuleert het maken van toetsen het ophalen van informatie uit het geheugen (het zogenaamde *testing effect*), waardoor het geleerde beter en langduriger beklijft. Binnen het programma Toetsing en Toetsgestuurd Leren is digitaal toetsen vooral ingezet voor formatieve doeleinden.

### Vakrendement stijgt door voortgangstoetsing

De faculteit Economie en Bedrijfskunde van de UvA heeft in het project 'Voortgangstoetsing in de propedeuse' op grote schaal ervaring opgedaan met digitale voortgangstoetsen bij 8 vakken in de gezamenlijke propedeuse van de bacheloropleidingen Economie en Bedrijfskunde en Fiscale Economie. De belangrijkste conclusie is dat er een grote correlatie is tussen de resultaten van de voortgangstoetsen en het eindcijfer. Het project toont aan dat de invoering van voortgangstoetsen de doelstelling van activerend onderwijs ondersteunt en het vakrendement verbetert.

Bij de meeste projecten zijn aanwijzingen gevonden dat de inzet van digitaal toetsen een positief effect heeft op de leerprestaties. Digitalisering van toetsen maakt het gemakkelijker om toetsen formatief in te zetten; ook complexere vaardigheden kunnen getoetst worden en digitaal toetsen maakt het mogelijk studenten direct van feedback te voorzien.

In de projecten *Bewust en bekwaam toetsen*, *E-Flow Nursing* en *ONBETWIST* is een voormeting, een interventie (in de vorm van een pilot met een digitale toets) en een nameting uitgevoerd. Bij andere projecten, waaronder *Toets- en vragenbank bedrijfsadministratie en bedrijfseconomie*, moesten studenten die de toetsen hadden gemaakt een vragenlijst invullen waarin gevraagd werd naar het effect dat zij verwachtten van de oefentoetsen op hun studieprestaties. In totaal werd bij zes projecten een positief verband gevonden tussen het gebruik van oefentoetsen en studiesucces.

De projecten *Voortgangstoetsing in de doorstroom mbo-hbo* en *Computer Adaptieve Monitoring van het Statistiekonderwijs* voerden metingen op meerdere tijdstippen uit. Daarbij is een vergelijking gemaakt van de prestaties van dezelfde studenten over een langere periode, respectievelijk een vergelijking van prestaties van verschillende jaargroepen binnen een bepaalde cursus. In de eerstgenoemde analyse is geen effect te zien van digitale voortgangstoetsing op het scoreverloop over langere tijd. Uit de andere analyse blijkt dat de prestaties van studenten na de introductie van digitale oefentoetsen waren toegenomen.

### Verbeteren aansluiting

Drie projecten hadden als doel de aansluiting tussen onderwijssectoren te verbeteren: *ONBETWIST* (aansluiting tussen vo en ho), *Digitaal Toetsen Master ANP* (aansluiting tussen hbo en wo) en *Voortgangstoetsing in de doorstroom mbo-hbo* (aansluiting tussen mbo en hbo). Vanwege de relatief korte doorlooptijd van de projecten en de kleine groepen studenten die aan de pilots meededen kon niet worden aangetoond dat de ontwikkelde toetsen daadwerkelijk de overgang van de ene naar de andere onderwijssector gemakkelijker hebben gemaakt. In de projecten *ONBETWIST* en *Voortgangstoetsing in de doorstroom mbo-hbo* is de relatie bepaald tussen de scores op digitale toetsen en de scores op de tentamens, waarbij met name de effectmetingen in *ONBETWIST* nieuwe inzichten opleveren.

#### Digitale toetsen als opstap naar zelfsturend leren

Binnen het project *ONBETWIST* is een database met wiskunde-opgaven ontwikkeld om de wiskundekennis van instromers en eerstejaarsstudenten op voldoende niveau te krijgen en te houden. Wiskundeleraars in het hoger onderwijs konden daardoor beschikken over meer dan 1000 interactieve wiskunde-opgaven.

Uit effectmetingen bij verschillende instellingen blijkt dat niet alle typen studenten van digitale oefentoetsen lijken te profiteren. Studenten die stapsgewijs leren en bij wie het leergedrag extern gestuurd moet worden, oefenden meer met behulp van de toetsen, en presteerden beter ten opzichte van de studenten die diep en zelfsturend leren. Op basis van deze uitkomsten concluderen de onderzoekers dat digitale toetsen de transformatie van voortgezet onderwijs - gekenmerkt door strak, extern gestuurd leren - naar hoger onderwijs - gekenmerkt door diep, zelfsturend leren - kan vergemakkelijken.

## 2. Vermindering van de werkdruk van docenten

Om inzicht te krijgen of digitaal toetsen kan bijdragen aan het verlagen van werkdruk is docenten gevraagd een vragenlijst in te vullen. Daarin moesten zij een schatting geven van het aantal uren dat zij hadden bespaard ten opzichte van de oude situatie en aangeven of ze verwachtten tijd te besparen in de toekomst. Uit de projecten bleek dat het ontwikkelen van items voor digitale toetsen in de beginfase vaak extra tijd kostte. Ook kostte het docenten tijd om te wennen aan het werken met nieuwe toetssoftware. Daarnaast werd er veel tijd gestoken in het afstemmen van leerdoelen en bijvoorbeeld het ontwikkelen van rubrics<sup>1</sup>.

De verwachting is dat na deze aanloopfase de tijdsinvestering zal afnemen. Digitalisering van toetsing biedt docenten in potentie aanzienlijke tijdswinst omdat het inleveren, beheren, nakijken en analyseren van toetsen en het geven van feedback geheel of gedeeltelijk automatisch kan worden uitgevoerd. Ook de logistiek en archivering van toetsen kan veel efficiënter verlopen. Bij centraal opgezette itembanken kunnen docenten bovendien items met elkaar delen, waardoor zij minder tijd kwijt zijn aan het ontwikkelen van toetsen.

<sup>1</sup> Analytische beoordelingsschaal voor het beoordelen van vaardigheden.

**Tijdswinst door digitaal nakijken van werkstukken**

In het project SCALA werd de verwerking van schriftelijke werkstukken vereenvoudigd en versneld met tools om teksten digitaal na te kijken en van commentaar te voorzien. Uit interviews met de betrokken docenten blijkt dat ze met digitale tools zoals *Turnitin* niet alleen minder tijd kwijt waren met het beoordelen zelf, maar dat ook de logistiek (verslagen laten inleveren, feedback teruggeven, cijfers communiceren en archiveren) minder tijd kostte. De digitaal ingeleverde verslagen worden automatisch bij elkaar opgeslagen, hoeven niet meer geprint en docenten kunnen hun commentaar rechtstreeks in de digitale versie verwerken. Bijna 40% van de docenten in het SCALA-project ervoer al tijdswinst bij de eerste keer dat ze digitaal feedback gaven en essays beoordeelden.

**3. Kwaliteitsverbetering**

In de *Kwaliteitspiramide voor Eigentijds Toetsen en Beoordelen* (Sluijsmans et al., 2012) worden criteria benoemd die van invloed zijn op de kwaliteit van toetsen. Digitaal toetsen kan vooral van invloed zijn op de volgende criteria:

- validiteit en relevantie van de toets en toetsitems
- efficiëntie: haalbaarheid in termen van menskracht en middelen
- objectiviteit: onafhankelijkheid van storende invloeden van beoordelaars
- zin van het toetsprogramma voor studenten: is de toets een leermoment (naast een beoordelingsmoment)
- cognitieve complexiteit van de toetsen: is het mogelijk denkprocessen en cognitieve vaardigheden te beoordelen
- authenticiteit: overeenkomst tussen de beoordeling en de toekomstige beroepssituatie
- ontwikkeling van zelfsturend leren: leren de studenten zelfstandig hun leerproces te sturen, zichzelf en elkaar te beoordelen, te reflecteren en hun eigen leerdoelen formuleren
- gevolgen voor het onderwijs: effect op instructie en begeleiding van docent en leergedrag student.

**Validiteit en relevantie**

Het gezamenlijk ontwikkelen van digitale toetsen vereist intensieve afstemming vooraf. Zo is bij samenwerken aan toetsbanken vooraf afstemming nodig tussen deelnemende instellingen over de onderwijsdoelen en de eisen aan items en toetsen die daarbij horen. Bijvoorbeeld door een gezamenlijk reviewsysteem te ontwikkelen om de kwaliteit en validiteit van ontwikkelde items te borgen. Binnen de projecten *Toetsing in beeld*, *Digitaal toetsen Master ANP* en *Digitaal toetsen van klinisch redeneren* is onderzoek gedaan om te bepalen welke competenties en cognitieve processen er betrokken zijn bij de te meten vaardigheden en met welke typen vragen dit het beste kan worden getoetst.

Uit het project *Nieuwe scoreregel voor digitaal toetsen* blijkt dat bij de verwerking van toetsresultaten de validiteit van toetsing nog verder kan worden vergroot.

### Nieuwe regel verhoogt validiteit

In het project Nieuwe scoreregel voor digitaal toetsen is de validiteit van de High Speed High Stakes -regel onderzocht. Bij deze scoreregel (HSHS) geven snelle, goede antwoorden hoge scores en snelle, foute gokken sterke negatieve scores. Alles wijst erop dat deze scoreregel een valide weergave geeft van vaardigheden zoals statistiek en rekenen. De werking van de regel is onderzocht op basis van gegevens uit de systemen van Oefenweb, die online applicaties biedt voor het oefenen en toetsen van vaardigheden. Hieruit blijkt een hoge samenhang tussen de HSHS-scores en scores op toetsen die dezelfde vaardigheid meten met een andere scoreregel. In veel gevallen presteert de HSHS-regel zelfs beter.

### Efficiëntie

Samenwerken aan itembanken blijkt te kunnen leiden tot een hogere efficiëntie: door samen toetsvragen te maken en te reviewen wordt het werk gedeeld. Daarnaast kan het nakijken van gesloten toetsvragen vaak grotendeels of volledig worden geautomatiseerd. Binnen de looptijd van de projecten werd de vergroting van efficiëntie over het algemeen nog niet goed zichtbaar omdat docenten nog moesten wennen aan een nieuw toetssysteem en de itembanken nog moesten worden opgebouwd.

### Efficiënt samenwerken in de cloud

Vijf hogescholen zetten in het project Toets- en vragenbank BABE een gezamenlijke itembank op in de cloud en waren zo in staat in relatief korte tijd veel vragen te genereren. Docenten kunnen in de itembankomgeving precies zien hoeveel vragen er gemaakt zijn en wat de status van de vragen is. Ze kunnen filteren op onderwerp, toetsterm, auteur of status van een vraag. Zo is bijvoorbeeld in één oogopslag duidelijk welke vragen nog 'gerepareerd' moeten worden. Naast het maken van de vragen kunnen docenten ook de oefeningen en toetsen in de cloudomgeving samenstellen.

### Objectiviteit

In bijna alle gevallen werd toetsing objectiever. Met software als *Turnitin* (getest in het SCALA-project) is tussen beoordelaars meer overeenstemming in de beoordeling van verslagen en open vragen bereikt. Het werken met deze tool dwong docenten samen rubrics te ontwikkelen, waarin zij leerdoelen en de verschillende beheersingsniveaus die hiermee verbonden zijn opnemen. Zo ontstond een bruikbaar format voor de online beoordeling van teksten van studenten. Verschillende docenten gaven aan voorheen niet zo snel een zware onvoldoende of juist een extreem hoog cijfer te geven, maar dankzij de rubrics waren ze beter in staat extreme cijfers zoals een 3 of een 9 te beargumenteren. De rubrics komen dus de objectiviteit en transparantie van de beoordeling ten goede. Tegelijkertijd kan met de criteria beter onderscheid worden gemaakt tussen de prestaties van studenten, zelfs als zij hetzelfde eindcijfer hebben. Zo kunnen twee studenten allebei overall een 7 krijgen, maar binnen de criteria van de rubrics een heel andere scoreverdeling hebben.

### Nut en zin van het toetsprogramma voor studenten

Digitale toetsvormen maken het mogelijk de student direct te voorzien van feedback, bijvoorbeeld met een uitwerking van het goede antwoord en een verwijzing naar relevante bronnen. Ook is het mogelijk de student een terugkoppeling te geven van scores op verschillende onderwerpen of vakgebieden binnen de toets en een vergelijking te maken met de norm of met de scores van andere studenten uit de groep.



### Oefenen in een spelomgeving

In het project *Computeradaptieve monitoring van het statistiekonderwijs* is de Statistiekfabriek ontwikkeld: een oefenvolgsysteem in de vorm van een spel. Studenten kunnen de fabriek onderhouden en 'pimpen' als ze beter presteren. Het spel laat studenten op hun eigen niveau oefenen, motiveert doordat ze de meeste opgaven correct zullen beantwoorden en geeft feedback op grond van fouten die ze vaak maken. Binnen de Statistiekfabriek kunnen studenten (de ontwikkeling van) hun eigen niveau en hun relatieve score binnen de groep zien. Ook zien ze hun sterke en zwakke punten. Docenten kunnen de ontwikkeling van studenten binnen de Statistiekfabriek volgen en zo een beeld krijgen van individuele en groepsontwikkeling. Met deze informatie kan de docent instructies beter afstemmen op individueel en groepsniveau. Uit effectmetingen blijkt dat speelfrequentie en toetscijfers duidelijk samenhangen; ook het niveau in de Statistiekfabriek en de toetscijfers hangen duidelijk samen. Bij de grootste groep gebruikers (circa 480 studenten) is het slagingspercentage op het vak van 68% naar 84% gestegen.

### Cognitieve complexiteit

Digitale toetsing biedt mogelijkheden voor verschillende vraagtypen. Er is in de projecten veel variatie aangetroffen binnen de ontwikkelde vragen (zoals multiple choice, meer-uit-meervragen, aanwijsvragen, vragen met stellingen). Vaak werd context toegevoegd aan de items in de vorm van casussen: om een vraag goed te beantwoorden moet de student niet alleen kennis reproduceren en begrijpen, maar die kennis ook kunnen toepassen en analyses kunnen maken. Een voorbeeld van een toets op hoog cognitief niveau is de Script Concordance Test voor het klinisch redeneren, die is ontwikkeld in het project *Master ANP*.

### Digitaal toetsen van klinisch redeneren

In het project *Master ANP* werkten drie hogescholen aan het toetsen van klinisch redeneren. Het gaat hierbij om complexe cognitieve vaardigheden: de student koppelt de eigen observaties en interpretaties aan de medische kennis om zodoende vast te stellen welke volgende stappen nodig zijn in het verpleegkundig handelen. Om deze vaardigheden digitaal te toetsen gebruikten ze de Script Concordance Test (SCT). Dit is een innovatieve toets met korte klinische casussen. Bij het valideren van de toetsvragen worden experts uit de klinische praktijk betrokken. Docenten en studenten zijn enthousiast over het resultaat en de mogelijkheden.

### Authenticiteit

Met digitale middelen is het mogelijk toetsen te ontwikkelen die veel beter aansluiten op de latere werkpraktijk van de student dan papieren toetsen. De duidelijkste voorbeelden hiervan zijn de projecten *Toetsing in beeld* en *Digitaal toetsen met beelden*, waarbij beeldinterpretatievaardigheden binnen de opleiding Radiologie voor het eerst werden getoetst met behulp van authentieke 3D-beelden in plaats van met 2D-beelden. Met deze vorm van toetsing kunnen artsen in opleiding beter worden voorbereid op de praktijk, omdat daar met 3D-beelden wordt gewerkt.

In het project *Van wie is dit portfolio?* werd een nieuwe beoordelingssystematiek gecreëerd voor de beoordeling van digitale portfolio's van pabo-studenten, waarin hen meer ruimte werd geboden om hun eigen, unieke aanpak als leraar en hun handelen in de praktijk te laten zien. Zo konden studenten naast de traditionele talige reflecties ook video-opnamen van hun lessen aan het portfolio toevoegen om hun competenties aan te tonen.

### Radiologen beter voorbereid op praktijk

In het project Digitaal toetsen met beelden is gebruik gemaakt van authentieke 2D- en 3D-beeldvragen in het digitale toetsprogramma VQuest. De kwaliteit van digitaal toetsen met radiologische beelden met VQuest blijkt op meerdere aspecten beter dan schriftelijke toetsing. De digitale voortgangstoets radiologie is inmiddels opgenomen als vast onderdeel van het reguliere toetsprogramma van de landelijke medisch-specialistische vervolgopleiding radiologie.

### Ontwikkeling van zelfsturend leren

Digitaal toetsen kan eraan bijdragen dat studenten hun leerproces meer in eigen hand nemen. In het geval van formatieve oefentoetsen kunnen studenten vaststellen op welk niveau zij zich bevinden ten opzichte van een leerdoel, en welke vaardigheden en/of kennis extra aandacht behoeven. Veel toetsen bevatten ook feedback (meestal bestaande uit een toelichting of uitwerking bij het juiste antwoord of een overzicht van scores per toets of toetsonderwerp) of een verwijzing naar bronnen voor verdere studie, wat een stimulans is voor studenten om de regie te nemen over hun eigen leeractiviteiten.

### Meer zelfinzicht studenten

In het project *Voortgangstoetsing in de doorstroom mbo-hbo* is een doorgaande leerlijn van mbo naar hbo ontwikkeld voor de kernvakken Nederlands, rekenen en Engels. Dankzij digitale formatieve toetsen krijgen de studenten inzicht in hoeverre ze de basiskennis al beheersen en op welke onderdelen zij nog moeten investeren om hun doel te bereiken. Aan de voortgangstoetsen is feedback per vakinhoudelijk domein gekoppeld met een indicatie van hun niveau op dat domein. Deze feedback wordt zichtbaar voor de studenten als zij de toets gemaakt hebben.

### Leergedrag

Bij enkele projecten had het inzetten van digitale oefentoetsen het effect dat studenten eerder aan de slag gingen met de stof die ze voor de tentamens of eindtoets moesten bestuderen; in andere projecten bleek dat studenten zich door de digitale toetsen beter bewust werden van hun hiaten in kennis. Studenten maakten gebruik van de bronnen die in de feedback van de toets is opgenomen en begrepen hun beoordeling beter. Ook peer-feedback levert nuttige tips op van medestudenten.

### Efficiënt, flexibel en authentiek toetsen

Bij adaptief toetsen krijgt een student een reeks vragen voorgelegd die naarmate de reeks vordert steeds beter passen bij het vaardigheidsniveau van de student. Zo'n geïndividualiseerde, automatisch samengestelde toets kan op ieder gewenst moment worden afgenomen en is door de afstemming op het niveau van de kandidaat zeer efficiënt. In het project *AdAPT* is onderzocht of het haalbaar is om van de interuniversitaire Voortgangstoets Geneeskunde een adaptieve toets te maken. De conclusie is dat dit zeker mogelijk is. Men weet nu aan welke eisen toetsvragen moeten voldoen willen ze geschikt zijn voor deze vorm van toetsen. Toepassing van de adaptieve toets is in de masterfase op korte termijn mogelijk. Volledige vervanging van de papieren toets zal nog meer tijd in beslag nemen, omdat er nog te weinig geschikte toetsvragen beschikbaar zijn.

#### 4. Overige bevindingen

Naast de resultaten op de drie hoofddoelen van het programma hebben de projecten geleid tot:

##### Meer kennis bij en afstemming tussen instellingen

Door de samenwerking in projecten tussen docenten van verschillende instellingen is de gezamenlijke kennis over wat goede toetsvragen zijn toegenomen. In het project *VG Together* kregen docenten bijvoorbeeld met behulp van een nieuw digitaal systeem voor de review van items voor de interuniversitaire VoortgangsToets Geneeskunde meer zicht op de items van collega-auteurs en hoe die items door studenten waren gemaakt.

##### Bewustwording bij docenten

De samenwerking heeft bij veel projectteams bewustwording over toetsing in gang gezet. Voor de gezamenlijke ontwikkeling van een toetssysteem of een itembank moeten vooraf doelen worden vastgelegd, evenals de criteria waaraan toetsitems moeten voldoen.

Het *SCALA*-project laat bij uitstek zien hoe docenten bewuster gaan nadenken over wat zij hun studenten willen leren, onder meer door *rubrics* te ontwikkelen waarin toetscriteria zijn verwerkt.

Bij het project *Voortgangstoetsing in de propedeuse* zijn voortgangstoetsen afgenomen over stof die al in de colleges aan bod was geweest. Het ontwikkelen van nieuwe voortgangstoetsen bracht een soortgelijk bewustzijnsproces tot stand: docenten gingen uitgebreider dan voorheen na welke stof zij de week ervoor hadden behandeld en welk deel daarvan zij wilden toetsen.

##### Zelfbeoordelingsinstrument toetsbekwaamheid

Binnen het project *Bewust en bekwaam toetsen* is een digitaal zelfbeoordelingsinstrument ontwikkeld waarmee docenten kunnen vaststellen hoe het met hun toetsvaardigheden- en kennis is gesteld. Pilots met docenten geven aan dat het instrument een positieve invloed kan hebben op de toetsbekwaamheid van docenten.

##### Aandacht voor toetsbekwaamheid

Toetskwaliteit, dus ook de kwaliteit van digitale toetsen, wordt in hoge mate bepaald door de toetsbekwaamheid van opleiders en hun bewustzijn van goed toetsen. In de context van het programma Toetsing en Toetsgestuurd Leren is in het project *Bewust en bekwaam toetsen* een onderzoek uitgevoerd gericht op de toetsbekwaamheid van lerarenopleiders. Op basis van analyse van bestaande instrumenten is een ontwikkelingsgericht zelfbeoordelingsinstrument ontwikkeld in samenwerking met leraren en assessmentexperts. Het instrument is te vinden op [www.bewustenbekwaamtoetsen.nl](http://www.bewustenbekwaamtoetsen.nl). Het gebruik van het instrument is onderzocht: docenten scoorden na gebruik beter op kennis, inzicht en toepassing van toetsen. Het instrument werd ervaren als zinvol, authentiek, betekenisvol is en stimulerend voor de zelfbeoordeling. Docenten gaven aan dat het instrument hen bewust had gemaakt van de omvang van het domein van toetsen en beoordelen.

# CONCLUSIES

## Verhogen van het studiesucces

Bij het grootste deel van de projecten die zich richtten op de verhoging van het studiesucces, is een positief effect waargenomen. Op basis van de uitgevoerde effectmetingen en evaluaties is het aannemelijk dat de inzet van digitaal toetsen een positief effect heeft op de leerprestaties, al is dat in dit stadium nog lastig te onderbouwen.

Digitalisering van toetsen ondersteunt formatieve toetsing. Digitalisering maakt het mogelijk dat een student 24/7 toegang heeft tot oefentoetsen en dat hij automatisch feedback ontvangt. De inzet van multimedia zorgt ervoor dat toetsen veel realistischer worden en digitaal toetsen biedt mogelijkheden ook complexere vaardigheden te toetsen. In verschillende projecten is aangetoond dat de inzet van formatieve digitale toetsing bijdraagt aan het resultaat op de eindtoets.

Van de digitale toetsen die zijn ontwikkeld om de aansluiting tussen onderwijssectoren te verbeteren kan nog niet worden vastgesteld of dit ook daadwerkelijk gelukt is. Daarvoor was de tijd te beperkt en waren de groepen studenten die werden gevolgd te klein.

## Vermindering van werkdruk

Digitalisering van toetsing biedt docenten in potentie aanzienlijke tijdswinst omdat het inleveren, beheren, nakijken, analyseren van toetsen en het geven van feedback geheel of gedeeltelijk automatisch kan worden uitgevoerd. Bij centraal opgezette itembanken kunnen docenten bovendien items met elkaar delen, waardoor zij minder tijd kwijt zijn aan het ontwikkelen van toetsen.

In de meeste projecten is tijdens het uitvoeren van de projecten nog geen werkdrukvermindering voor docenten vastgesteld. In de beginfase van de invoering van digitale toetsing is meer tijd en aandacht nodig voor het ontwikkelen en reviewen van toetsitems (met als voordeel dat de kwaliteit ervan toeneemt) en docenten hebben tijd nodig om vertrouwd te raken met de software. Vanuit de meeste projecten is de verwachting uitgesproken dat op langere termijn werkdrukvermindering wel mogelijk is, als docenten eenmaal kunnen beschikken over goed gevulde itembanken. Wel blijft het nodig om de items periodiek opnieuw te reviewen en te actualiseren. Men is het erover eens dat het nakijken van de toetsen en het verwerken en analyseren van de resultaten minder tijd kost bij digitale toetsen dan bij de papieren variant.

## Verbetering van toetskwaliteit

Digitaal toetsen kan een bijdrage leveren aan de kwaliteit van toetsing. Uit de projecten blijkt dat de kwaliteit van toetsen door samenwerking op verschillende manieren verbetert; de samenwerking wordt mogelijk door digitale ondersteuning, zoals een gezamenlijke auteursomgeving en vooral een gezamenlijke itembank. Door toetsitems gezamenlijk te ontwikkelen en te gebruiken wordt meer aandacht besteed aan itemontwikkeling, wat betere toetsvragen oplevert. Daarnaast levert het gezamenlijk gebruik van de itembank grotere hoeveelheden toetsresultaten, waardoor het beter mogelijk wordt psychometrische analyses uit te voeren. Die geven docenten inzicht in de kwaliteit van de toetsvragen en de toets als geheel, waardoor het gemakkelijker wordt te werken aan de kwaliteit van de itembanken.

Digitaal toetsen maakt het mogelijk andere, rijkere vraagvormen in te zetten waardoor toetsen beter aansluiten op de beroepspraktijk. Een sprekend voorbeeld hiervan is het toetsen van interpretatievaardigheden van 3D-beelden bij radiologiestudenten, wat op papier onmogelijk is. De adaptieve spelomgeving van de Statistiekfabriek en het project *Digitaal toetsen van klinisch redeneren binnen de medische opleidingen* zijn voorbeelden die laten zien dat een andere vraagvorm grote meerwaarde kan hebben.

Digitale afname zorgt er bovendien voor dat toetsen betrouwbaar worden uitgevoerd, toetsvragen objectief worden gecorrigeerd en de resultaten snel beschikbaar zijn. Daarnaast blijkt dat het samenwerken (binnen en tussen instellingen) aan itembanken gunstig is voor de kwaliteit. Docenten zijn bewuster gaan nadenken over de eisen waaraan toetsen en toetsvragen moeten voldoen, en hoe het systeem van vraagontwikkeling- en review eruit zou moeten zien.

“Binnen onze opleiding radiologie maken we dankzij een project van SURF gebruik van 3D-beelden, zodat we zorgvuldige en adequate beeldinterpretatie kunnen toetsen. Daarmee hebben we een belangrijke kwaliteitsslag gemaakt in de opleiding. Ik zie nog veel mogelijkheden voor digitaal toetsen. De beschikbaarheid van toetsbanken met grote hoeveelheden toetsitems maakt regelmatig tussentijds toetsen mogelijk. En dat heeft een positief effect op het leren van studenten. Het lokt ook een discussie uit over de rol van summatief en formatief toetsen in het onderwijs. Ik zou die discussie graag met andere instellingen blijven voeren.”

**Bert van der Zwaan**, Rector Magnificus Universiteit Utrecht

# BIJLAGE 1: DE PROJECTEN

Van de 18 projecten uit het programma Toetsing en Toetsgestuurd Leren is de opgedane kennis en ervaring via [www.surf.nl/digitaaltoetsen](http://www.surf.nl/digitaaltoetsen) beschikbaar voor alle hogeronderwijsinstellingen.

Project	Omschrijving	Deelnemende instellingen
<a href="#"><u>AdaPT, adaptieve voortgangstoetsing</u></a>	ADAPT onderzocht hoe computergestuurd adaptief toetsen toepasbaar is bij voortgangstoetsing en kan bijdragen aan kwaliteitsverbetering. Resultaat: In het project is gebleken dat het haalbaar is om van de interuniversitaire Voortgangstoets een adaptieve toets te maken. Toepassing van de adaptieve toets is in de masterfase op korte termijn mogelijk. Voor volledige vervanging van de toets is meer tijd nodig voor de ontwikkeling van geschikte toetsvragen.	UM, Cito
<a href="#"><u>Bewust en bekwaam toetsen</u></a>	Dit project onderzocht toetspercepties, -kennis en -kunde en verdere professionalisering bij lerarenopleiders, met als concreet resultaat een zelfbeoordelingsinstrument voor toetskennis. Uit het onderzoek blijkt dat het zelfbeoordelingsinstrument een bijdrage kan leveren aan het verhogen van de kennis over toetsen bij docenten.	Fontys, HAN, HZ, VU
<a href="#"><u>Body of Knowledge and Skills (BoKS), verbreden en verbinden</u></a>	Dit project richtte zich op werkdrukvermindering van docenten hbo-verpleegkunde en kwaliteitsverbetering van toetsen en toetsitems, die aansluiten bij het curriculum van hbo-v. De kennistoetsbank van Leerstation Zorg is uitgebreid met items voor summatief toetsen van medische hbo-v-vakken. Bestaande toetsitems van hogescholen zijn verzameld, geredigeerd, gereviseerd en van meer metadata voorzien.	HAN, HR, NHL Hogeschool, Saxion, Stichting Leerstation Zorg
<a href="#"><u>Computer Adaptieve Monitoring in het Statistiekonderwijs ('Statistiekfabriek')</u></a>	'Statistiekfabriek' zette een universiteitsoverstijgend computeradaptief oefenvolgsysteem op voor statistiek. In de adaptieve spelomgeving met zo'n 2000 statistiekopgaven oefenen studenten op hun eigen niveau en monitoren docenten én studenten hun ontwikkeling.	UvA, UT, UU
<a href="#"><u>Digitaal Toetsen master Advanced Nursing Practice</u></a>	Drie hogescholen hebben voor de master Advanced Nursing Practice (MANP) een gezamenlijke digitale, diagnostische pre-mastertoets ingevoerd. Ook is er een digitale toetsbank farmacologie opgezet en een digitaal vraagtype voor klinisch redeneren ontwikkeld. Het samenwerken aan digitale toetsen heeft tot kwaliteitsverbetering geleid.	HR, Hogeschool Zuyd, Fontys, Stichting Leerstation Zorg

Project	Omschrijving	Deelnemende instellingen
<u>Digitaal toetsen met beelden, een extra dimensie</u>	Dit project, dat voortbouwde op 'Toetsing in beeld', implementeerde digitaal toetsen met 2D- en 3D-beelden op landelijke schaal in de medische specialistenopleidingen. Resultaat: digitaal toetsen met 3D-beelden kan veilig en betrouwbaar op grote schaal worden uitgevoerd en verbetert de kwaliteit van toetsen. De ontwikkelde digitale voortgangstoets van beeldinterpretatie-vaardigheden krijgt een vaste plek in de vervolgopleiding radiologie.	UMC Utrecht, AMC Amsterdam, UU
<u>Digitaal toetsen van klinisch redeneren binnen de medische opleidingen</u>	Dit project ontwikkelde een digitaal format om het vermogen tot klinisch redeneren te toetsen. De betrokken opleidingen ontwikkelden en deelden vragen, die resulteerden in een gezamenlijke vragendatabank. Resultaat: de keuze voor gesloten vraagtypes in plaats van open vragen resulteert in tijdwinst voor docenten/artsen tijdens het nakijken en voor ondersteunende medewerkers bij het analyseren van de toetsresultaten. De cijfers van de studenten kunnen sneller bekend zijn.	UvA, UMC Utrecht, Erasmus MC
<u>E-flow Nursing</u>	Dit project ontwikkelde een doorlopende en werkplekonafhankelijke digitale toets voor derde- en vierdejaars studenten mbo-verpleegkundige en eerste- en tweedejaars studenten hbo-v. Resultaat: het maken van de toetsen heeft een positief effect op het kennisniveau van studenten en de toetsen leiden tot meer kennisuitwisseling tussen begeleiders en studenten.	UMC Groningen, Hanzehogeschool Groningen, Stichting Leerstation Zorg
<u>Nieuwe scoreregel voor digitaal toetsen</u>	Dit project ontwikkelde en testte een nieuwe scoreregel (High Speed High Stakes) voor digitaal toetsen, die zowel accuratesse en snelheid als de gokfactor op meerkeuzevragen meeneemt. Hiermee wordt tegemoet gekomen aan de standaardbezwaren tegen scoreregels in klassieke toetsen. Resultaat: HSHS vergroot de validiteit van digitale toetsing van vaardigheden zoals statistiek.	UvA
<u>Non Satis Scire</u>	In dit project ontwikkelden dertien universiteiten een digitale itembank voor de toetsing en beoordeling in zowel de minor- als masterprogramma's voor de universitaire eerstegraads lerenopleidingen. Resultaat: door toetsen en procedures met elkaar te delen en te bespreken is de toets- en beoordelingspraktijk kritisch bekeken. Dat komt de kwaliteit ten goede. De waardering van opleiders en studenten voor de digitale beoordelingsprocedures is hoog.	UvA, UL, EUR, RUN, RUG, TUD, TU/e, UM, TU, UT, UU, VU, WU

Project	Omschrijving	Deelnemende instellingen
<u>ONBETWIST</u>	ONBETWIST ontwikkelde een database met wiskunde-opgaven om de wiskundekennis van instromers en eerstejaars op efficiënte wijze op voldoende hoog niveau te krijgen en te houden. Resultaat: via de database ONBETWIST hebben wiskundedocenten in het hoger onderwijs de beschikking over meer dan 1000 interactieve wiskunde-opgaven.	TU/e, UvA, UU, OU, UM
<u>SCALA - Scaffolding Assessment for Learning</u>	SCALA bevorderde aan de hand van pilots het gebruik van digitale middelen bij het beoordelen van en feedback geven op schriftelijke producten van studenten (zoals essays en onderzoeksverslagen). Resultaat: Digitaal feedback geven en beoordelen helpt om de werkdruk van docenten beheersbaar te houden, zonder dat dit ten koste gaat van de kwaliteit van de beoordeling	UU, HvA, HU, VU
<u>Toets- en vragenbank bedrijfsadministratie en bedrijfseconomie</u>	Dit project ontwikkelde in een gedeelde online toetsomgeving een landelijke toets- en vragenbank waarmee studenten kunnen oefenen en waaruit docenten diverse soorten toetsen kunnen samenstellen voor digitale of papieren afname. Resultaat: 2000 (oefen) toetsvragen met feedback.	HR, HU, ECABO, Saxion, Haagse Hogeschool, HvA
<u>Toetsing in beeld</u>	Dit project was gericht op het verhogen van de psychometrische kwaliteit en de validiteit van toetsing van beeldvaardigheden en het verkrijgen van inzicht in de benodigde cognitieve processen bij het interpreteren van 3D-beelden. Resultaat: toetsing van radiologische beeldinterpretatievaardigheden met 3D-beeldvragen scoort op diverse kwaliteitsaspecten beter dan met 2D-beelden. Het project geeft daarnaast meer inzicht in de kennis en vaardigheden die een rol spelen bij beeldinterpretatie in de radiologie.	UMC Utrecht, UU
<u>Van wie is dit portfolio? De beoordeling beoordeeld.</u>	Dit project formuleerde richtlijnen om elektronische portfolio's in het lerarenonderwijs te beoordelen en daarbij voldoende ruimte te bieden voor de authenticiteit van de student (eigenheid) en zijn praktijk (echtheid). Resultaat: digitale portfolio's worden authentiek door adaptieve begeleiding met (peer)feedback, ondersteuning van multimedia en ruimte voor studenten om competenties op eigen wijze aan te tonen.	Hogeschool de Kempel, Hogeschool Marnix Academie
<u>VG Together - samenwerken in voortgangstoetsing</u>	VG Together ontwikkelde een centrale, generieke te gebruiken infrastructuur voor voortgangstoetsing: een nieuw systeem voor itemconstructie, itembeoordeling en een omgeving voor kennisuitwisseling. Resultaat: met deze infrastructuur in combinatie met het itemmanagementsysteem IMS zijn de mogelijkheden voor toekomstige digitale afname van de voortgangstoets sterk verbeterd.	UM, RUN, RUG, UL, UM



Project	Omschrijving	Deelnemende instellingen
<u>Voortgangstoetsing doorstroom mbo-hbo</u>	Dit project ontwikkelde een doorgaande leerlijn van mbo naar hbo voor de kernvakken Rekenen, Nederlands en Engels, inclusief formatieve toetsing. Resultaat: de toetsen zijn een goed hulpmiddel voor studenten om vast te stellen welke basiskennis zij al beheersen en op welke onderdelen ze nog moeten investeren om op het gewenste niveau van de pabo te komen.	Driestar Hogeschool, Christelijke Hogeschool Ede, Hoornbeeck College
<u>Voortgangstoetsen in de propedeuse</u>	Dit project richtte zich op invoering en effectmeting van een digitale, wekelijkse voortgangstoets op het studierendement in de propedeuse Economie & Bedrijfskunde en Fiscale Economie. Resultaat: er bestaat een grote correlatie tussen de resultaten van de voortgangstoetsen en het cijfer. Het project toont aan dat de invoering van voortgangstoetsen de doelstelling van activerend onderwijs ondersteunt en vakrendementen verbetert.	UvA

## BIJLAGE 2: LITERATUUR

Baartman, L. & Dijkstra, A. (2011). *Zelfevaluatie van de kwaliteit van assessment*. Onderwijsinnovatie, maart 2011.

Bloom, B.S., Engelhart, M. D., Furst, E. J., Hill, W.H. & D.R. Krathwohl (1956). *Taxonomy of educational objectives: The classification of educational goals*. Handbook I: Cognitive domain. New York: David McKay Company.

Camp, G. & De Bruin, A. (2013). *Leerstrategieën als brug tussen cognitieve psychologie en onderwijspraktijk*. Onderwijsinnovatie, september 2013.

Denyer, Tranfield, & Van Aken (2008). *Developing Design Propositions through Research Synthesis*. Organization Studies, 29, 249-269.

*Effective Practice with e-Assessment* (2007). An overview of technologies, policies and practice in further and higher education. JISC-rapport.

Gulikers, J. (2006). *Authenticity is in the eye of the beholder*. Academisch proefschrift. Heerlen: Open Universiteit.

Hattie, J. (2009). *Visible Learning: A Synthesis of over 800 Meta-Analyses relating to Achievement*. London: Routledge.

Hattie, J. & Timperley, H. (2007). *The Power of Feedback*. Review of Educational Research, 77, 81-112.

Joosten-ten Brinke, D. (2011). *Eigentijds toetsen en beoordelen*. Lectorale rede. Tilburg: Fontys Lerarenopleiding Tilburg.

Koopmans, C., Tewes, F. en D. De Graaf (2009). *Economische projectbeoordeling Programma Toetsing en Toetsgestuurd Leren*. SEO Economisch Onderzoek.

Segers, M., Dochy, F. & Dierick, S. (2001). *Een ander onderwijsmodel...een ander toetsmodel?* In: A. Wald & J. van der Linden (red.) *Leren in perspectief*. Garant: Leuven.

Schuwirth, L. (2011). *Personal Communication*, July 2011.

Sluismans, D., Peeters, A., Jakobs, L. & S. Weijzen (2012). *De kwaliteit van toetsing onder de loep*. Onderwijsinnovatie, december 2012.

Tempelaar, D., Cuypers, H., Van de Vrie, E., Van der Kooij, H. & A. Heck (2012). *Toetsgestuurd leren en learning analytics*. Onderwijsinnovatie, september 2012.

# COLOFON

**Research**

Judith Evertse (Rosetta Lingua)

**Redactie**

Daphne Riksen (Ediction) en SURF

**Ontwerp**

Vrije stijl, Utrecht

**Datum**

Oktober 2014

**SURF**

Graadt van Roggenweg 340

Postbus 2290  
3500 GG Utrecht

T +31 (0)30 234 66 00

F +31 (0)30 233 29 60

info@surf.nl

www.surf.nl



2014

Beschikbaar onder de licentie Creative Commons Naamsvermelding 3.0 Nederland.  
[www.creativecommons.org/licenses/by/3.0/nl](http://www.creativecommons.org/licenses/by/3.0/nl)

**SURF**