

Logistieke organisatie van de computer-adaptieve interuniversitaire voortgangstoets geneeskunde (iVTG)

■ Jeroen Donkers

Dr. J. Donkers is Universitair docent bij de vakgroep Onderwijsontwikkeling en Onderzoek van de Faculty of Health Medicine and Life Sciences, Universiteit Maastricht, en Psychometrisch Coördinator binnen de iVTG. E-mail: jeroen.donkers@maastrichtuniversity.nl

■ Judith van den Brink

J. van den Brink is Toetscoördinator bij het Onderwijsinstituut van de Faculty of Health Medicine and Life Sciences, Universiteit Maastricht en Coördinator binnen de iVTG. E-mail: j.vandenbrink@maastrichtuniversity.nl

■ Theo Keesom

Th. Keesom is ICT-ontwikkelaar/beheerder bij het Onderwijsinstituut van de Faculty of Health Medicine and Life Sciences, Universiteit Maastricht en Toetsbeheerder binnen de iVTG. E-mail: theo.keesom@maastrichtuniversity.nl

■ André Bremers

Dr. A. Bremers is Chirurg en Universitair Hoofddocent Toetsing bij Radboud UMC en voorzitter van de iVTG. E-mail: Andre.Bremers@radboudumc.nl

Sinds september 2022 wordt de interuniversitaire voortgangstoets geneeskunde (iVTG) afgenomen als een computer-adaptieve toets in zeven geneeskundeopleidingen. Sinds het aansluiten van de laatste ontbrekende opleiding (Utrecht) in september 2023 is het een landelijke adaptieve voortgangstoets geneeskunde en nemen ongeveer 13.000 studenten deel. In dit artikel leggen we eerst uit wat deze toets inhoudt, en gaan daarna vooral in op hoe deze grootschalige toets logistiek is georganiseerd.

Wat is voortgangstoetsing?

Een voortgangstoets is een objectief instrument om de progressie van parate kennis van de student te meten. Deze toets is niet direct gekoppeld aan het curriculum, maar heeft betrekking op de volledige opleiding.

Een voortgangstoets wordt meerdere keren per jaar afgenomen onder alle studenten. De inhoud van de toets verschilt niet voor studenten in verschillende fasen van de opleiding; de verwachting is dat de individuele student op iedere toets een hogere score behaalt dan de

voorgaande toetsen als gevolg van opgedane kennis in het curriculum.

Deze toetsvorm is in de jaren 70 van de vorige eeuw geïntroduceerd in onder andere Maastricht en sindsdien is het gebruik van voortgangstoetsen verspreid over vele opleidingen in vele landen. In Nederland is de voortgangstoets voor geneeskunde uitgegroeid tot een landelijke interuniversitaire Voortgangstoets Geneeskunde, de iVTG.

Wat is de iVTG?

De (i)VTG is een samenwerkingsverband tussen alle acht Nederlandse geneeskundeopleidingen (UMC's) met als doel gezamenlijk voortgangstoetsing te doen. De toets valt onder verantwoordelijkheid van de lokale examencommissies en wordt in ieder UMC door eigen personeelsleden georganiseerd. De iVTG

wordt bestuurd door de werkgroep interuniversitaire Voortgangstoets (WiV) waarin alle UMC's zijn vertegenwoordigd. Daarnaast is er een centraal logistiek bureau dat bij Universiteit Maastricht is ondergebracht. We zullen hieronder de term iVTG soms ook gebruiken voor de toets zelf.

Wie doen aan de iVTG mee?

De iVTG wordt afgenomen bij de geneeskundeopleiding van Maastricht, Groningen, Nijmegen, Leiden, Amsterdam (VU en AMC), Rotterdam en Utrecht. Bij sommige opleidingen doen (nog) niet alle studiejaar mee. Er zijn ook afgeleide opleidingen die mee doen, zoals de 4-jarige master Arts-Klinisch Onderzoeker (A-KO) in Maastricht en Selective Utrecht Medical Master (SUMMA) in Utrecht, en de Engelstalige bachelors geneeskunde in Groningen en Maastricht. Daarnaast zijn er nog kleine groepen deelnemers, zoals buitenlandse artsen die in Nederland aan de slag willen en BIG-herregistranten. Al met al zijn er per toets zo'n 13.000 studenten die momenteel de toets afleggen.



Longitudinale feedback op de voortgangstoets is een belangrijk studiemiddel in de geneeskundeopleidingen.

De vragen voor de toets worden door alle deelnemende UMC's aangeleverd en landelijk gereviewed (daarover hierna meer). Ieder UMC neemt dezelfde toets af en hanteert dezelfde cesuren. De voortgangstoets wordt viermaal per jaar afgenomen bij geneeskundestudenten van bachelor jaar 1 tot en met het laatste masterjaar. Studenten krijgen na iedere toets een score en een kwalificatie onvoldoende, voldoende, of goed. Deze kwalificatie hangt af van de cesuur die geldt voor de curriculumleeftijd van de student. Bij 4 toetsen per jaar en een 6-jarige opleiding zijn er 24 verschillende momenten waarop de student de toets maakt. We noemen dat meetmomenten. Op ieder volgend meetmoment geldt dus

een hogere cesuur waar de student aan moet voldoen.

De iVTG volgt een toetsmatrix die bestaat uit twee assen: disciplines (anatomie, huisartsgeneeskunde, interne geneeskunde, et cetera) en categorieën (Ademhalingsstelsel, Spier- en skeletstelsel, Hormonen en metabolisme, et cetera). Studenten ontvangen ook deelscores op deze disciplines en categorieën.

Studenten krijgen via het online dashboard, genaamd "ProF", de scores en deelscores van afgelopen voortgangstoetsen in grafische weergave te zien. Hiermee kunnen ze hun groei in kennis analyseren. Ook kunnen ze zien hoe andere studenten in dezelfde fase van de opleiding het hebben gedaan. Aan de hand hiervan kunnen ze analyseren aan welke disciplines en categorieën ze eventueel meer aandacht moeten besteden. Deze longitudinale feedback op de voortgangstoets is een belangrijk studiemiddel in de geneeskundeopleidingen.

De vragen in de voortgangstoets zijn meerkeuzevragen met één goed antwoord. De vragen gaan over het toepassen, analyseren en evalueren van medische kennis en hebben vaak de vorm van een patiënt-casus. De toets wordt in twee talen afgenomen: Nederlands en Engels. De Engelse vragen zijn vertalingen van de Nederlandse vragen.

Tot aan september 2022 werd de iVTG op papier afgenomen, daarna is de toets computer-adaptief geworden. De toetsaansameling, de scoring en de cesuurbepaling zijn daarbij gewijzigd.

Hoe werkte de papieren voorgangstoets?

Bij de papieren voortgangstoets kregen alle studenten per toets een identiek toetsboekje met 200 meerkeuzevragen en een antwoordvel uitgereikt. Studenten hadden vier uur de tijd om de vragen te beantwoorden. Vragen

die ze niet konden of wilden beantwoorden mochten ze onbeantwoord laten. Bij de scoring werden punten voor foute antwoorden afgetrokken; niet beantwoorde vragen leverden geen punten op. Omdat de moeilijkheid van de vragen per toets kon variëren, werd een relatieve cesuur gehanteerd, die afhing van hoe studenten als groep hadden gescoord. Na de toets mochten studenten het boekje met de toetsvragen mee naar huis nemen en later werden de goede antwoorden bekend gemaakt.

Omdat alle studenten dezelfde toets kregen waren de vragen in principe gericht op het vereiste niveau van studenten in het laatste masterjaar. Dit had als negatief effect dat veel van de vragen ver buiten het beheersingsniveau van bachelorstudenten lag. Hierdoor was voor deze studenten de feedback op hun kennisprogressie en sterktes en zwaktes minder nauwkeurig. De toets was daarnaast zeer omvangrijk en vermoeiend; de toetsboekjes bedroegen bijna 40 bladzijden in A5-formaat.

Logistiek betekende de papieren toets dat vier keer per jaar een nieuwe set van 200 vragen moest worden samengesteld en vertaald, en dat hiervan toetsboekjes en antwoordvellen moesten worden gedrukt (voor alle studenten) en verzonden. De toets moest door alle studenten van alle UMC's op exact hetzelfde moment worden afgenomen onder toezicht. Ieder UMC moest faciliteiten organiseren om 1500 tot 2000 studenten simultaan te toetsen. Na de toets moesten de antwoordvellen worden verzameld en per koerier naar het centraal bureau in Maastricht worden gebracht, waar ze centraal werden verwerkt. Uiteindelijk werden vanuit Maastricht lijsten met resultaten en individuele feedback per student opgesteld en verstuurd naar de UMC's.

Waarom een adaptieve voorgangstoets?

De logistiek rondom de papieren toets werd met het uitbreiden van het aantal UMC's

steeds lastiger te coördineren en duurder. Er werd in de WiV besloten een toetsvorm te kiezen waarbij niet alle studenten op exact hetzelfde moment hoefden deel te nemen, zodat er met minder grote toetsruimtes en flexibelere planning kon worden gewerkt, en waarbij, mede door het vervallen van alle verplaatsingen van papieren materialen, de centrale verwerking sneller kon.

Door de toets digitaal af te nemen kon de logistiek rond de papieren afhandeling vervallen. Om de tijdstippen waarop studenten de toets maken te kunnen spreiden moeten de vragen in de afgelegde toetsen verschillend zijn, terwijl de inhoud en de moeilijkheid vergelijkbaar moeten blijven. Het steeds moeten samenstellen van meerdere vergelijkbare toetsen van 200 vragen was niet haalbaar. Bovendien, om te beslissen of toetsen vergelijkbaar zijn, moet objectief zijn vastgesteld hoe moeilijk iedere vraag is. Echter, als dit eenmaal bekend is, dan komt de mogelijkheid in beeld om de toets adaptief af te nemen. Bij een adaptieve toets wordt immers gebruikt gemaakt van de vooraf bekende moeilijkheid van elke vraag om daarmee een toets op maat samen te stellen voor elke individuele student. Iedere student krijgt dus een eigen unieke toets, getrokken uit een grote vragenbank, maar de scores van de studenten zijn wel vergelijkbaar.

De iVTG beschikt over een ruime databank van vragen die eerder in de papieren voortgangstoets zijn gebruikt. Op basis van antwoordgegevens is de moeilijkheid van al deze vragen statistisch bepaald (gekalibreerd). Daarnaast is een leverancier van toetssoftware gevonden (TestVision) die voor de iVTG het algoritme en de voorzieningen voor een adaptieve voortgangstoets heeft geïmplementeerd.

In de coronatijd is het proces versneld omdat massaal op papier toetsen tijdens lockdowns niet of beperkt mogelijk was. In die periode is de voortgangstoets deels online afgenomen, maar wel volgens de regels van de papieren

toets. Dit heeft ertoe geleid dat de logistiek en infrastructuur voor digitale afname van de voortgangstoets al in gang was gezet voordat de adaptieve voortgangstoets werd ingevoerd.

Hoe werkt de adaptieve voortgangstoets?

In de adaptieve voortgangstoets worden dezelfde type meerkeuzevragen gebruikt als voorheen. De basis is een vragenbank van nu rond de 6000 vragen die actief meedoen in de toets. Van iedere vraag in deze actieve bank is de moeilijkheid bekend. Als een student de toets start worden eerst een aantal (6) random-geselecteerde vragen getoond. De antwoorden daarop geven het adaptief algoritme een eerste schatting van het niveau van de student. Daarna volgt het adaptieve deel van 114 vragen. Die worden zodanig gekozen dat de moeilijkheid van de vraag steeds dicht bij het geschatte niveau van de student ligt. Het algoritme houdt daarbij ook rekening met de toetsmatrijs en met vragen die de student al in eerdere afnames heeft gezien. Ten slotte worden zogenaamde “enemy-vragen”, vragen die inhoudelijk overlappen, vermeden. Na ieder antwoord van de student stelt het algoritme het geschatte niveau van de student bij op basis van alle al gegeven antwoorden en de moeilijkheid van de vragen. Deze schatting wordt dus steeds beter. Studenten kunnen vanwege deze schatting niet terugbladeren in de toets om eerdere antwoorden aan te passen. Om dezelfde reden kunnen ze ook geen vragen overslaan of openlaten. De schatting aan het eind van de toets is de score van de student op de toets.

Willekeurig verspreid tussen de 120 vragen die meetellen voor de score, krijgen studenten ook 15 pretest-vragen te zien. Dit zijn nieuw ontwikkelde vragen waarvan de moeilijkheid bepaald moet worden zodat ze opgenomen kunnen worden in de actieve vragenbank. Uiteraard kunnen deze vragen nog niet meetellen in de score.

De hele toets omvat dus 135 vragen die aansluiten bij het kennisniveau van de student, tegenover de 200 vragen van diverse moeilijkheidsniveaus in de papieren toets. Studenten krijgen drie uur de tijd om alle vragen te beantwoorden, een uur minder dan nodig was bij de papieren toets. De toets wordt niet meer door alle studenten tegelijkertijd afgenomen, maar er is steeds een toetsperiode van anderhalve week waarin alle afnamesessies door de UMC's moeten worden georganiseerd. Het aantal sessies hangt af van de capaciteit van de beschikbare computerzalen en de onderwijsroosters. De afnames vinden nog steeds altijd onder toezicht plaats.

De cesuren zijn ook in de adaptieve toets afhankelijk van het meetmoment waarop de student aan de toets meedoet. In tegenstelling tot relatieve cesuren in de papieren toets, zijn deze cesuren nu vast omdat de moeilijkheid van de vragen bekend is. Hiervoor is het wel nodig de afnamesessies binnen een beperkte tijd te groeperen omdat studenten met hetzelfde meetmoment anders te ver uit elkaar gaan lopen in ontwikkeling.

In mei 2022 is een pilotstudie gedaan naar deze nieuwe toetsvorm waarin een grote groep studenten zowel de papieren als de adaptieve toets heeft afgelegd. Hierbij bleek betrouwbaarheid van de adaptieve toets voldoende te zijn (de test-retest betrouwbaarheid was 0.83) en de beleving door studenten positief.

Hoe is het toetsproces ingericht?

De toets wordt afgenomen in een eigen iVTG-toetsomgeving van TestVision. Hierin is een speciaal type toets en afname ontwikkeld voor een adaptieve voortgangstoets. Het toetsproces bestaat uit twee delen: vraagproductie en toetsafname. We beginnen met de laatste.

Toetsafname

De basis van de toets is de actieve vragenbank. Bij iedere vraag in deze bank is opgeslagen: de



Door de overgang op adaptief toetsen konden de logistieke problemen rondom planning en organisatie van de afnames fors worden beperkt.

discipline en categorie van de vraag, bij welke toets de vraag in de actieve bank is gekomen en het volgnummer daarbij, de moeilijkheid van de vraag, de taal van de vraag (Nederlands of Engels), en eventuele enemy-vragen. Daarnaast bevat iedere actieve vraag een korte omschrijving van het onderwerp, een leerstofverwijzing en tekstuele feedback. Alle actieve vragen hebben een identieke Nederlandse en Engelse variant. De actieve vragenbank is dynamisch: tussen de toetsperiodes door worden sommige vragen verwijderd en nieuwe vragen toegevoegd.

Voor iedere toetsperiode wordt in de toetsomgeving een nieuwe adaptieve voortgangstoets aangemaakt (of eigenlijk twee, voor Nederlands en Engels) waarbij de actuele inhoud van

de actieve bank aan de toets wordt gekoppeld, samen met pretest-vragen. De pretest-vragen worden daarbij in setjes van 15 vragen verdeeld, zodat de toetsomgeving per student één setje kan uitkiezen om aan die student voor te leggen.

Iedere UMC heeft een lokale afnameorganisatie (personeelsleden die bij de lokale afname zijn betrokken) die toegang heeft tot een apart deel van de toetsomgeving waarin de eigen afnames en studenten zijn ondergebracht. De afnames en studenten van andere UMC's zijn niet toegankelijk. De UMC's hebben ook geen toegang tot de toets zelf of tot de actieve vragenbank. Als de nieuwe adaptieve voortgangstoets (centraal) is aangemaakt kunnen de lokale afnameorganisaties hun afnamesessies gaan inrichten. Per afnamesessie wordt onder andere door hen ingesteld wanneer en waar die plaatsvindt, en welke beveiliging wordt gebruikt in de toetszaal.

Nadat de afnamesessies zijn aangemaakt, kunnen de lokale afnameorganisaties er zelf de studenten aan toewijzen. Hiervoor is een procedure ontwikkeld waarmee studenten van een externe lijst kunnen worden ingelezen, inclusief hun juiste actuele meetmoment. Studentaccounts worden tijdens het inlezen gekoppeld aan de afname en voorzien van het juiste meetmoment, nog niet bestaande studentaccounts worden dan meteen aangemaakt. Studentaccounts worden hierbij in een zodanige map-structuur in de toetsomgeving opgeslagen dat duidelijk is welke studenten met welk meetmoment aan welke afnamesessie meedoen. Hiermee is het mogelijk om, indien nodig, snel handmatige aanpassingen te doen. Studenten loggen via Surfconex met hun eigen instellingsaccount in, er worden dus geen wachtwoorden opgeslagen.

Tijdens de toetsafname houden de lokale afnameorganisaties zelf de eigen afnamesessies in de gaten via de afnamemonitor van de toetsomgeving.

Na de afnameperiode volgt eerst een (psychometrische) controle van de resultaten. Daarna worden de vaste cesuren toegepast en kunnen studenten in de toetsomgeving hun uitslag zien. Lokale afnameorganisaties kunnen dan ook de uitslagenlijsten downloaden voor verdere verwerking. Er worden inzagesessies gehouden waarin studenten de inhoud van de vragen kunnen terugzien en commentaar op de vragen kunnen indienen. Buiten de inzagesessies om kunnen studenten, naast hun uitslag ook van iedere door hun beantwoorde vraag de korte omschrijving, de score en de bijbehorende leerstofverwijzingen inzien.

Na de afnameperiode volgt de centrale psychometrische analyse. Hiervoor worden vanuit de toetsomgeving databestanden gedownload met alle benodigde detailgegevens om het gedrag van vragen en studenten te kunnen analyseren. In deze fase wordt met name de moeilijkheid (kalibratie) van zojuist gepreteste vragen bepaald, en een rapport gemaakt over de actieve vragenbank die aangeeft in welke deelgebieden behoefte is aan vragen van welke moeilijkheid.

Vraagproductie

Aan de andere kant van het toetsproces ligt de vraagproductie. In de papieren toets was er reeds sprake van een uitgebreide organisatie rond productie, review en recycling van toetsvragen, ondersteund door een centrale, offline, vragenbank. De bestaande procedures en kwaliteitschecks zijn zoveel mogelijk in stand gehouden in de adaptieve toets. De medische kennis die wordt getoetst is aan continue veranderingen onderhevig dus moeten er iedere toetsperiode nieuwe vragen worden gepretest. In de oude situatie waren er iedere toets weer 200 nieuwe vragen nodig voor het toetsboekje, in de nieuwe situatie steeds 210 pretest-vragen om de bank in stand te houden (uitgaande van een doelgrootte van 8000 actieve vragen, een gemiddelde levensduur van 10 jaar en 5% tussentijds verlies). Dat betekent een vergelijkbare vraagproductie en dus een

vergelijkbare organisatie.

Iedere UMC produceert in principe een afgesproken hoeveelheid nieuwe vragen (aangestuurd door de centrale productiecoördinator). Daarvoor zijn er binnen ieder UMC enkele auteurs aangewezen die voor het eigen vakgebied vragen schrijven. Per UMC is er een reviewcommissie (de Voortgangstoets Beoordelingscommissie, VBC) die auteurs aansturen en vragen reviewen. In de toetsomgeving is er voor ieder UMC een eigen gebied waar vragen kunnen worden geschreven en gerevied. Het is aan VBC's om te zien in hoeverre ze dit in de toetsomgeving doen of daarbuiten, zolang de vragen aan het eind van het proces maar in de toetsomgeving staan.

VBC's bieden vragen daarna ter review aan bij de VBC van een van de andere UMC's. Bij deze review wordt gelet op inhoud, relevantie en formulering. Bij akkoord voert de centrale productiecoördinator onder andere kwaliteitscontroles en bureauredactie op de vragen uit. Is de vraag dan ook akkoord, dan verhuist de vraag naar een afgesloten deel van de toetsomgeving waarin alle vragen staan die als pretest-vraag kunnen worden gekozen bij een volgende toets.

Voorafgaand aan iedere toetsperiode selecteren de voorzitter van de WiV en de productiecoördinator op basis van zowel de toetsmatrijs als de actuele behoefte van de actieve bank de 210 nieuwe pretest-vragen. Deze pretest-vragen worden in het Engels vertaald.

Voordat de pretest-vragen aan de toets worden toegevoegd wordt gecontroleerd of er doublures van deze vragen in de actieve bank zitten en of er enemy-relaties moeten worden aangemaakt. De pretest-vragen krijgen dan ook een toets-datumveld en een volgnummer toegewezen.

Na de psychometrische analyse van de pretest-vragen, worden de resultaten aan de VBC's

ter beoordeling voorgelegd en wordt in een centraal overleg besloten welke pretest-vragen in de actieve bank worden toegelaten. De afgekeurde vragen (meestal zo'n 15-20 stuks) gaan voor een periode van 3 jaar op slot en gaan dan terug naar de oorspronkelijke auteur om te worden gereviseerd. De overige pretest-vragen worden met hun gemeten moeilijkheidswaarde in de actieve bank opgenomen.

Eénmaal per jaar vindt er "groot onderhoud" op de actieve bank plaats: alle actieve vragen worden geïnspecteerd en sommige verwijderd. Denk hierbij aan vragen waarvan de moeilijkheid is veranderd (parameter drift), vragen die niet meer in het psychometrisch model passen, of vragen die om inhoudelijke redenen niet meer in de toets mogen komen. Deze vragen worden in een archief geplaatst en niet meer in roulatie genomen. Hierdoor blijft de psychometrische informatie van deze vragen bewaard voor toekomstige analyses.

Tot slot

Bij het schrijven van dit artikel zijn er inmiddels zes reguliere adaptieve toetsen geweest die zonder noemenswaardige problemen zijn verlopen. De toetsomgeving is betrouwbaar gebleken en de performance van het adaptief algoritme is bevredigend. Door de overgang op adaptief toetsen konden de logistieke problemen rondom planning en organisatie van de afnames fors worden beperkt. Zo is het niet meer nodig op één moment alle onderwijs stil te leggen en alle coassistenten van hun leer-werkplekken terug te halen. De toets is bij gehandhaafde betrouwbaarheid een stuk korter en minder zwaar geworden voor de student. Tegelijkertijd is de feedback, zeker voor de jongerejaars studenten een stuk betrouwbaarder geworden. De komende jaren zullen procedures en inrichting zeker aanpassingen ondergaan: veranderingen in de toetsomgeving, veranderingen in regelgeving, voortschrijdend inzicht, bevindingen uit onderwijskundig onderzoek, enzovoort. ■