

Voortgangstoetsing

□ Arno Muijtjens

In dit nummer van *EXAMENS* wordt aandacht besteed aan het overlijden van Wynand Wijnen. Dit artikel gaat over een van zijn belangrijkste bijdragen aan het hoger onderwijs, namelijk de voortgangstoets. Het is een in de jaren zeventig van de vorige eeuw ontwikkeld instrument dat inmiddels breed is ingevoerd.

Metten van de studievoortgang

Een voortgangstoets is een veelomvattende toets die gericht is op de cognitieve doelstellingen van het onderwijsprogramma, zoals een eindexamen dat ook is. Het verschil met een eindexamen is dat een voortgangstoets niet alleen wordt afgenomen bij studenten in het laatste jaar van de studie, maar bij alle studenten van alle jaargroepen. Een verschil in kennisniveau komt tot uitdrukking in het niveau van de toetsscores: de scores van hogerejaars zullen gemiddeld hoger liggen dan die van lagerejaars.

Om voortgang in de studie te kunnen meten dient een voortgangstoets herhaald te worden afgenomen. Bij een volgende gelegenheid wordt weliswaar niet exact dezelfde toets gebruikt (vanwege het gevaar van herkenning van vragen), maar wel een toets die vergelijkbaar is qua inhoud en moeilijkheidsgraad. Op die manier zal voor de opeenvolgende toetsmomenten gelden dat latere scores van een student in het algemeen hoger zullen zijn dan vroegere in verband met de toegenomen kennis. Voortgangstoetsing biedt dus de mogelijkheid om vast te stellen hoe ver een student qua kennisniveau verwijderd is van het einddoel en welke kennisgroei (of gebrek aan kennisgroei) zich in een bepaalde periode bij een student heeft voorgedaan.

Hoewel in principe diverse vraagvormen in een

voortgangstoets gebruikt kunnen worden, is de toets vrijwel altijd samengesteld uit meerkeuzevragen. Dat is op zich niet zo verwonderlijk: bij een veelomvattende toets (groot aantal vragen) die frequent wordt afgenomen bij grote groepen studenten, is het om praktische redenen vrijwel onontkoombaar dat er gebruik wordt gemaakt van een efficiënt te verwerken vraagvorm.

Het primaire doel van een voortgangstoets is het regelmatig longitudinaal toetsen van studenten met betrekking tot het beoogde eindniveau van kennis. Hiermee wordt de kennisontwikkeling van studenten gevolgd en worden studenten aangezet tot regelmatige en effectieve kennisverwerving in het domein van functionele kennis zoals gedefinieerd in de eindtermen van het curriculum. Daarnaast levert een voortgangstoets ook informatie over de kennisontwikkeling in de subdomeinen door de tijd heen.

Herkomst

Toen in de jaren zeventig de Medische Faculteit Maastricht startte met haar probleemgestuurd onderwijs (PGO), bleken toetsing en onderwijsdoelen elkaar danig in de weg te zitten. Het PGO beoogde zelfgestuurd, ontdekkend leren in onderwijsgroepen op basis van geïntegreerd studeren in plaats van vakgericht studiedrag. Echter, de summatieve bloktoetsen die aan het eind van elk onderwijsblok van zes weken werden afgenomen, bevorderden toetsgestuurd studeren en specifiek toetsgerichte voorbereiding. Het volgen van individuele leerpaden werd erdoor ontmoedigd en van buiten leren werd bevorderd ten koste van het verwerven van inzicht, begrip en beklijvende, functionele kennis. De oplossing werd gezocht in het doorbreken van de directe relatie tussen het onderwijsprogramma en de toetsing, wat uiteindelijk leidde tot de ontwikkeling van de Maastrichtse voortgangstoets (Wijnen & Van der Vleuten, 1985; Van der Vleuten e.a., 1996). De gerichtheid op de doelstellingen van de hele opleiding en de daaruit voortvloeiende veelomvattendheid van de voortgangstoets maken het vrijwel



Veelomvattendheid is een belangrijk kenmerk van een voortgangstoets

ondoenlijk om de toets specifiek voor te bereiden. Daardoor wordt toetsgericht studeren tegengegaan en een continue wijze van studeren bevorderd (Wijnen, 1977; Verwijnen e.a., 1982).

Vorm en lengte

Als voorbeeld van een voortgangstoets wordt in dit artikel de interuniversitaire voortgangstoets geneeskunde (iVTG) gebruikt, die momenteel door vier medische faculteiten in Nederland (Universiteiten van Groningen, Leiden, Maastricht en Nijmegen) gezamenlijk wordt geconstrueerd. De iVTG wordt afgenomen bij alle geneeskundestudenten van deze vier universiteiten en daarnaast bij alle geneeskundestudenten van de Vrije Universiteit Amsterdam vanaf het derde jaar (Van der Vleuten e.a., 2004; Schuwirth e.a., 2010). De toets bestaat uit 200 meerkeuzevragen met maximaal vijf alternatieven waarvan er één correct is. Om gokken te ontmoedigen en voor gokken te corrigeren is een 'weet-niet' optie toegevoegd en worden strafpunten in rekening gebracht bij een fout antwoord. Door te kiezen voor de weet-niet optie kan een student deze boete ontlopen. De score wordt berekend als een percentage van het totaal aantal vragen in de toets, waarbij een correct antwoord wordt beloond met +1, een fout antwoord wordt bestraft met -1/(aantal alternatieven minus 1) en een 'weet-niet' levert 0 punten op. Deze scoringsmethode wordt in de literatuur aangeduid als Formula scoring. Een belangrijke reden om deze methode bij voortgangstoetsing toe te passen is gelegen in het feit dat veel lagerejaars studenten een groot deel (40-80%) van deze vragen, die gericht zijn op het op eindniveau van de opleiding, niet kunnen beantwoorden. Zonder Formula scoring zou het antwoord op deze vragen blind gegokt worden met als gevolg de introductie van veel ruis en dus een onbetrouwbare meting.

Ter illustratie staan in het kader twee voorbeelden van vragen uit het medische domein.

Vragen kunnen verschillende subdomeinen betreffen, bijvoorbeeld het ademhalingsstelsel en het spijsverteringsstelsel. De veelomvattendheid is een belangrijk kenmerk van een voortgangstoets: alle subdomeinen die van belang zijn voor de eindtermen dienen immers bevraagd te worden. Om dat te waarborgen

Voorbeelden van toetsvragen uit de interuniversitaire voortgangstoets

Vraag 1.

De verwekkers die het meest gevonden worden bij patiënten die een pneumonie buiten het ziekenhuis oplopen verschillen in het algemeen van de verwekkers die gezien worden bij patiënten die tijdens een ziekenhuisopname een pneumonie oplopen. Dit verschil kenmerkt zich doordat een pneumonie die tijdens een ziekenhuisopname opgelopen wordt:

- A. minder vaak door anaerobe verwekkers veroorzaakt wordt;
- B. minder vaak door *Staphylococcus aureus* veroorzaakt wordt;
- C. vaker door *Chlamydia psittaci* veroorzaakt wordt;
- D. vaker door Gramnegatieve verwekkers veroorzaakt wordt.

Vraag 2.

Buikpijn ontstaat door prikkeling van sensibele zenuwen in de buik. Hierbij kan onderscheid worden gemaakt tussen viscerale en pariëtale buikpijn. Viscerale buikpijn is:

- A. scherp en nauwkeurig te lokaliseren;
- B. diffuus en moeilijk te lokaliseren.

wordt elke toets samengesteld conform een blauwdruk waarin aangegeven wordt hoeveel vragen per subdomein in de toets opgenomen dienen te worden.

Voordelen

Bij een beheersingstoets van een onderwijsmodule zijn de bijbehorende toetsitems gekoppeld aan de in de module behandelde stof. Items in een voortgangstoets daarentegen zijn gericht op alle functionele kennis die een afgestudeerde conform de eindtermen van het onderwijsprogramma dient te verwerven. Dat brengt een aantal voordelen met zich mee. Een dergelijke toets roept vanwege de omvangrijkheid van zijn inhoud veel minder op tot specifiek toetsgericht

Voortgangstoetsing zet studenten aan tot regelmatige en effectieve kennisverwerving



studeren. De breedte van de toets zorgt ervoor dat in principe elke leeractiviteit die relevant is voor de eindtermen gehonoreerd wordt. Essentieel is dat het gaat om functionele, beklijvende kennis, dat wil zeggen de kennis die een afgestudeerde student paraat moet hebben om in de praktijk adequaat te kunnen functioneren. In een voortgangstoets kan in elke fase van het programma een onderwerp (opnieuw) bevestigd worden. Daardoor wordt een continue en op het verwerven van inzicht gerichte methode van studeren gestimuleerd en onregelmatig en oppervlakkig studeren (incidenteel hard blokken voor een tentamen) ontmoedigd.

Doordat in principe elke nieuwe voortgangstoets bestaat uit een bemonstering van hetzelfde kennisdomein is aan de reeks voortgangstoetsresultaten de ontwikkeling van het kennisniveau van een student af te lezen. Die ontwikkeling kan gevolgd worden voor wat betreft het geheel van kennis, maar ook per subdomein (per categorie of discipline in de blauwdruk). De monitoring van de individuele kennisontwikkeling en het daaruit af te leiden toekomstige verloop van de resultaten biedt mogelijkheden om onder- en overpresteren tijdig betrouwbaar vast te stellen en remediërende maatregelen te treffen of een aangepast versneld onderwijsprogramma aan te bieden.

Het verloop van de gemiddelde resultaten voor het totale kennisdomein of per subdomein biedt informatie over de algemene kennisontwikkeling. Dat levert voor docenten en onderwijsbeleidsmakers informatie over de effecten van het curriculum en biedt aanknopingspunten voor verbetering van het onderwijs.

Bij een voortgangstoets worden beslissingen over bevordering naar een volgend jaar of het met goed gevolg afsluiten van de propedeuse, bachelor- of masterfase gebaseerd op de combinatie van resultaten in een reeks van toetsen en niet op het resultaat van één allesomvattende eindtoets. Dit gebruik van herhaalde metingen om tot een beslissing te komen leidt tot minder invloed van toevallige fluctuaties in een toetsresultaat en daardoor tot betrouwbaardere en beter verdedigbare beslissingen. Additionele herkansingen zijn niet nodig omdat elke nieuwe voortgangstoets gebaseerd is op een steekproef uit hetzelfde kennisdomein en dus per definitie een herkansingstoets is.

Omdat een voortgangstoets gericht is op de eindtermen van het curriculum en niet op het curriculum zelf is het goed mogelijk om dezelfde voortgangstoets te gebruiken voor meerdere instituten met verschillende curricula maar wel dezelfde eindtermen. Dat biedt goede mogelijkheden om gezamenlijk eenzelfde voortgangstoets te construeren en af te nemen. De inspanningen voor het maken van toetsitems kunnen dan gedeeld worden en de validiteit van de toets is eveneens gediend met vragenmakers uit verschillende instituten. Een interinstitutionele voortgangstoets biedt ook mogelijkheden om te komen tot 'benchmarking' waarbij effecten van verschillende curricula met elkaar vergeleken kunnen worden en een sterkte/zwakke analyse uitgevoerd kan worden (Verhoeven e.a., 2005; Muijtjens, 2008).

Tenslotte bieden voortgangstoetsresultaten veel mogelijkheden voor het doen van onderwijskundig onderzoek. De data zijn zowel cross-sectioneel als longitudinaal van aard: ze bevatten metingen van verschillende (groepen van) individuen verkregen op eenzelfde moment zowel als herhaalde metingen binnen individuen.

Voortgangsresultaten bieden de mogelijkheid om feedback te geven aan studenten met betrekking tot het kennisniveau van het totale domein en voor alle subdomeinen (in het voorbeeld van de iVTG de 17 categorieën en 19 disciplines). In de regel worden daarbij de resultaten van een individuele student vergeleken met die van collega-studenten die zich in dezelfde fase van het curriculum bevinden (de referentiegroep). De feedback biedt de student de mogelijkheid om een sterkte-zwakke analyse te maken door de eigen resultaten te vergelijken met de gemiddelde prestatie van de referentiegroep. Dat kan door het resultaat van bijvoorbeeld de meest recente toets te inspecteren (momentane informatie), maar dat kan ook door het verloop van het kennisniveau te bekijken (longitudinale informatie), waarbij dat laatste uit oogpunt van betrouwbaarheid natuurlijk de voorkeur verdient.

Nadelen

Een voortgangstoets heeft natuurlijk ook nadelen. Omdat de toets gericht is op de eindtermen kunnen eerstejaars studenten niet veel vragen beantwoor-



Een voortgangstoets is gericht op de eindtermen van het curriculum en niet op het curriculum zelf

den waardoor de betrouwbaarheid voor deze groep relatief laag is.

De veelomvattendheid van de toets heeft als nadeel dat er per subdomein maar weinig vragen in de toets zitten waardoor de subdomeinscores (profielscores) onbetrouwbaar zijn.

Voorname nadeln kunnen behoorlijk ingedamd worden door gebruik te maken van aggregatie van de herhaalde metingen zoals dat bijvoorbeeld in het feedbacksysteem ProF gebeurt door gebruik te maken van de gecumuleerde scores (Muijtjens e.a., 2010). Om voldoende betrouwbaar te zijn is een relatief lange toetsduur nodig. Om voor zesdejaars een betrouwbaarheid van 0.80 te bereiken zijn 200 meerkeuzevragen nodig waarvoor in de regel vier uur toetstijd volstaat.

Het gebruik van Formula scoring kent ook nadelen. De bereidheid om te antwoorden als je als student niet helemaal zeker bent dat je genoeg van het onderwerp weet, is van invloed op de score. Studenten die terughoudend zijn, kunnen daardoor benadeeld worden (Muijtjens e.a., 1999; McHarg e.a., 2005). Bij toepassing van Number-right scoring, waarbij de score op de test de optelsom is van het aantal correct beantwoorde vragen, gerelateerd aan het totaal aantal vragen, bleek echter de betrouwbaarheid in de lagere jaren dramatisch te dalen. Daarom wordt bij de iVTG toch de voorkeur aan Formula scoring gegeven.

Het is praktisch moeilijk om een voortgangstoets te baseren op minder efficiënte vraagvormen omdat deze onacceptabele testduur en verwerkingsinspanningen met zich mee zouden brengen. Aangewezen zijn op juist-onjuist of meerkeuzevragen bij het construeren van een voortgangstoets lijkt overigens niet zo'n ernstige beperking, omdat in het algemeen de inhoud en niet de vorm van de items bepalend is voor wat gemeten wordt door een toets (Norman e.a., 1987, 1996).

Literatuur

- McHarg, J., Bradley, P., Chamberlain, S., et al. (2005). Assessment of progress tests. *Medical Education*, 39(2), 221-227.
- Muijtjens, A.M.M., van Mameren, H., Hoogenboom, R.J.I., et al. (1999). The effect of a 'don't know' option on test scores: number-right and formula scoring

compared. *Medical Education*, 33(4), 267-275.

Muijtjens, A.M.M., Schuwirth, L.W.T., Cohen-Schotanus, J., et al. (2008). Benchmarking by cross-institutional comparison of student achievement in a progress test. *Medical Education*, 42(1), 82-88.

Muijtjens, A.M., Timmermans, I., Donkers, J., et al. (2010). Flexible electronic feedback using the virtues of progress testing. *Medical Teacher*, 32(6), 491-495.

Norman, G.R., Smith, E.K.M., Powles, A.C., et al. (1987). Factors underlying performance on written tests of knowledge. *Medical Education*, 21, 297-304.

Norman, G., Swanson, D., Case, S. (1996). Conceptual and methodology issues in studies comparing assessment formats: issues in comparing item formats. *Teaching and Learning in Medicine*, 8(4), 208-216.

Schuwirth, L., Bosman, G., Henning, R.H., et al. (2010). Collaboration on progress testing in medical schools in the Netherlands. *Medical Teacher*, 32(6), 476-479.

Van der Vleuten, C.P.M., Verwijnen, G.M., Wijnen, W.H.F.W. (1996). Fifteen years of experience with progress testing in a problem-based learning curriculum. *Medical Teacher*, 18(2), 103-109.

Van der Vleuten, C.P.M., Schuwirth, L.W.T., Muijtjens, A.M.M., et al. (2004). Cross institutional collaboration in assessment: a case on progress testing. *Medical Teacher*, 26(8), 719-725.

Verhoeven, B.H., Snellen-Balendong, H.A.M., Hay, I.T., et al. (2005). The versatility of progress testing assessed in an international context: a start for benchmarking global standardization? *Medical Teacher*, 27(6), 514-520.

Verwijnen, G.M., Imbos, T.J., Snellen, H., et al. (1982). The evaluation system at the medical school of Maastricht. *Assessment and Evaluation in Higher Education*, 3, 225-244.

Wijnen, W.H.F.W. (1977). Einddoeltoetsen: Waarom en hoe? *Onderzoek van Onderwijs*, 6, 16-19.

Wijnen, W.H.F.W. & Van der Vleuten, C.P.M. (1985). Toetsing: Hordenloop of voortgangscontrole? *Universiteit en Hogeschool*, 31, 270-279.

□ De heer dr. ir. A.M.M. Muijtjens is werkzaam aan de Universiteit Maastricht, FHML, vakgroep Onderwijsontwikkeling & Onderwijsresearch. E-mail: a.muijtjens@maastrichtuniversity.nl.