

Gewogen en...?

De landelijke kennistoets geschiedenis

Politiek en samenleving roepen om meer basiskennis bij docenten. Een mogelijk antwoord hierop is de kennisbasis voor lerarenopleidingen. Maar wat is een kennisbasistoets geschiedenis? En hoe komt deze toets tot stand?

Desirée Joosten-ten Brinke, Wim Lansu, en Arnoud van Leuven zijn werkzaam aan de Fontys Lerarenopleiding Tilburg en betrokken bij het project '10voordeleraar'. **Dominique Sluijsmans** is werkzaam aan de Hogeschool van Arnhem en Nijmegen en lid van de kwaliteitscommissie van '10voordeleraar'.

In het kader van de Kwaliteitsagenda Krachtig Meesterschap heeft de HBO-raad het project '10voordeleraar' (www.10voordeleraar.nl) opgezet met als doel de gezamenlijke kennisbases voor de lerarenopleidingen te beschrijven. Door het vaststellen van deze kennisbases wordt ernaar gestreefd de kennis en vaardigheden van de studenten in alle hbo-lerarenopleidingen te vergroten en de kwaliteit van de opleidingen te borgen, mede door de ontwikkeling van landelijke digitale toetsen. In 2010 is gestart met het ontwikkelen van de landelijke digitale kennistoetsen voor de tweedegraads opleidingen. In dit artikel worden de resultaten van en ervaringen met het ontwikkelen van de landelijke kennistoets voor geschiedenis gepresenteerd. Uit de resultaten blijkt dat voor een duurzame kwaliteit van de toets structurele ondersteuning bij de itemconstructie essentieel is. Tevens is een goede aansluiting met de kennisbasis in het curriculum voorwaarde om

opdracht van de overheid en onder regie van de HBO-raad redactieteams van hogeschool-docenten aan landelijke digitale kennistoetsen waarmee het vakinhoudelijke eindniveau verhoogd en geborgd wordt. Vanaf 2013 krijgen alle studenten voor het vak geschiedenis een twee uur durende toets over de kernconcepten van dit vak. Pas na het slagen voor deze toets kan de student deelnemen aan het eind-assessment van de eigen opleiding.

Een belangrijk onderdeel bij de ontwikkeling van deze landelijke toets is de kwaliteitsborging en externe legitimering. Bij de toetsontwikkeling worden redactieteams kritisch gevolgd door onafhankelijke vakcommissies. Deze vakcommissies worden voorgezeten door een vertegenwoordiger uit het Landelijk Expertisecentrum Mens en Maatschappijvakken (www.expertisecentrum-mm.v.nl). Toetsdeskundigen die lid zijn van de voor het project ingestelde kwaliteitscommissie beoordelen de items en de toets als geheel voor en na afname van de toets op toetstechnische aspecten. Deze werkwijze wordt tot 2013 op projectmatige basis voortgezet. De projectleiders zijn verantwoordelijk voor het aansturen van het project en er is een adviesraad, bestaande uit hogeschool-leraren op het gebied van toetsen en beoordelen, die het geheel monitort.

Van kennisbasis naar kennistoets

De kennisbasis geschiedenis (www.kennisbasis.nl) beschrijft de kennis die startende tweedegraads leraren geschiedenis minimaal moeten bezitten. Van de kennisbasis zijn twee domeinen aangewezen voor de landelijke toetsing, namelijk de tijdvakken en historisch redeneren. De toetsredactie heeft deze domeinen nader geconcretiseerd in gebeurtenissen, personen, ontwikkelingen en verschijnselen. Op basis van deze operationalisatie maken

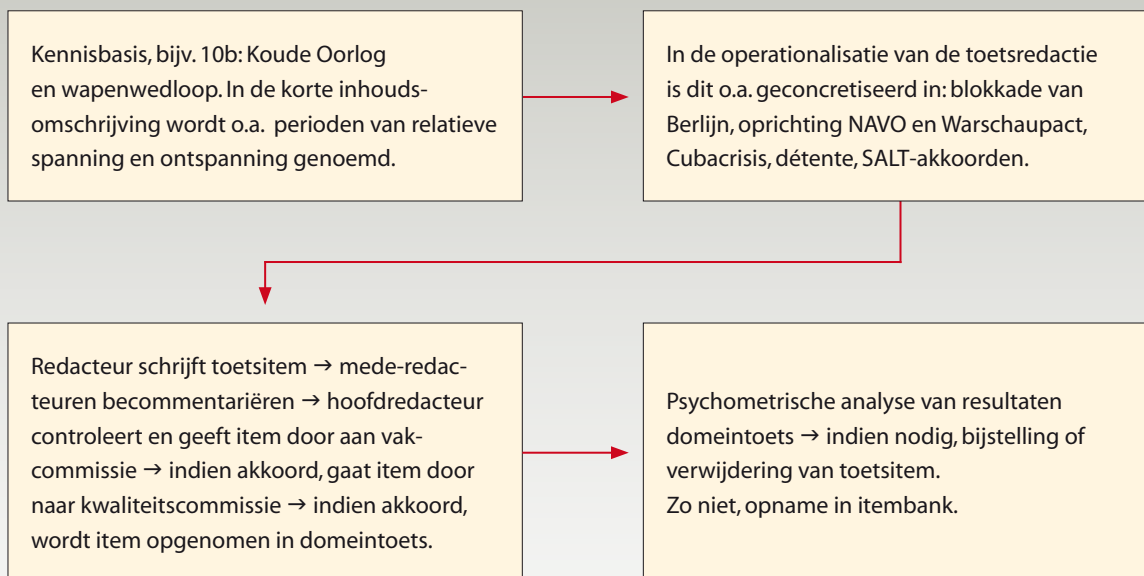
'Pas na het slagen voor deze toets kan de student deelnemen aan het eind-assessment van de eigen opleiding'

► Inspectie van het Onderwijs (2004). *Onderwijsverslag 2002/2003*. Utrecht: Inspectie van het Onderwijs. NVAO (2004). *Notitie: meta-evaluatie pabo's, d.d. 3 maart 2004*. Den Haag: NVAO.
► Ministerie van OC&W (2008). *Krachtig meesterschap, Kwaliteitsagenda voor het opleiden van leraren 2008-2011*. Den Haag: Ministerie van OC&W.

de toets optimaal in te zetten. Op dit moment worden de curricula van de opleidingen aangepast, waarbij de eigenheid van de opleiding bewaard wordt, maar het kennisniveau van de studenten landelijk vergelijkbaar is.

Gebrek aan vakkennis

Het project '10voordeleraar' is een antwoord op de zorg in de samenleving en de politiek over het gebrek aan vakkennis van docenten in het basis- en voortgezet onderwijs. (1) Op basis van deze zorg heeft de minister aangegeven dat het eindniveau van hbo-studenten transparant en onderling vergelijkbaar moet worden vastgesteld. (2) Om dit te realiseren werken in



Figuur 1.
Schematische weergave van het itemconstructieproces.

de redacteurs toetsitems. Dit redactieproces is in figuur 1 schematisch weergegeven en geïllustreerd met een voorbeeld.

Op basis van de operationalisatie van de kennisbasis is een toetsmatrijs ontwikkeld waarin de inhoud en het gewenste beheersingsniveau tegen elkaar zijn afgezet. Kennisvragen vergen reproductie van kennis, terwijl toepassingsvragen van studenten verwachten dat ze op basis van nieuwe informatie vraagstukken kunnen oplossen. Hiervoor worden uitsluitend meerkeuzevragen gebruikt. Figuur 2 laat een deel van deze toetsmatrijs zien. Er is een vaste verdeling van de in de toets gestelde vragen over de verschillende domeinen. Deze verdeling komt overeen met de bandbreedte van de in de toetsmatrijs vermelde percentages. Iedere student krijgt dezelfde vragen, deze vragen worden per blok echter gerandomiseerd uitgeserveerd. Binnen een vraag worden ook de antwoordalternatieven *at random* gegeven, tenzij er een logische chronologische antwoordvolgorde is.

Er zijn tien blokken met vragen in de toets, 140 vragen in totaal. Per blok wordt een tijdvak getoetst. Vragen over historisch redeneren (verzamen, ordenen, verklaren, beeld vormen) worden gesteld in de context van een tijdvak. Alle items samen vormen een itembank waaruit vergelijkbare toetsen kunnen worden samengesteld. Bij elk item geven de constructeurs de verwachte moeilijkheid van de vragen aan en de tijd die de student nodig zal hebben om de vraag te beantwoorden. Deze informatie wordt door de hoofdredacteur gebruikt bij het samenstellen van de toets. De toets moet af te nemen zijn in twee uur.

De pretest en domeintoets

Om een eerste beeld te krijgen van de kwaliteit van de items, werd in 2010 een pretest en in 2011 een domeintoets uitgevoerd. Voor de pretest zijn uit de itembank van ongeveer 500 items, handmatig 140 items geselecteerd op basis van de toetsmatrijs. Bij de domeintoets is een toets samengesteld uit alle items die betrekking hadden op een specifiek domein, te weten de eerste drie tijdvakken en items over historisch redeneren waarbij de nadruk ligt op het op de juiste manier informatie verzamelen. In 2012 zullen nog drie domeintoetsen en een pretest worden uitgevoerd.

Voor de afname van de toets wordt gebruik gemaakt van Question Mark Perception (QMP), een toetsservicesysteem waarbij de toetsen in een beveiligde omgeving direct op de computer afgenomen kunnen worden. Na ►

Figuur 2.
Toetsmatrijs geschiedenis.

	Kennis	Inzicht	Toepassing	Totaal
Tijdvak 1: Jagers en boeren	40-70%	10-30%	20-50%	5
Tijdvak 2: Grieken en Romeinen	40-70%	10-30%	20-50%	7
Historisch redeneren tijdvak 2			100%	
Tijdvak 3: Monniken en ridders	40-70%	10-30%	20-50%	8
Historisch redeneren tijdvak 3			100%	
Tijdvak 4: Steden en staten	40-70%	10-30%	20-50%	8
Historisch redeneren tijdvak 4			100%	
...	...	...	...	...
Totaal				140

de toets ontvangen de studenten direct hun percentuele score.

De pretest werd in november 2010 bij 624 studenten (eerstejaars en derdejaars) van de lerarenopleidingen geschiedenis afgenomen en de domeintoets in november 2011 bij 46 studenten. De context voor deze eerste pretest was vrijblijvend: de pretest was niet verplicht voor studenten en het resultaat van de pretest had geen gevolgen voor de student. Daarnaast konden de studenten niet specifiek voor de pretest leren en waren de curricula van de opleidingen nog niet volledig gebaseerd op de kennisbasis. Als de studenten in 2013 aan de toets deelnemen moet dit natuurlijk wel gerealiseerd zijn.

'De toetsscore was voor een aantal studenten hoger dan verwacht en voor een even groot aantal studenten lager dan verwacht'

De resultaten van de toetsen zijn na de afname door toetsdeskundigen geanalyseerd. Om een oordeel te geven over de kwaliteit van de pretest zijn de volgende vragen beantwoord:

- 1. Meet de toets wat hij beoogt te meten? (validiteit)
- 2. Meet de toets consistent? (betrouwbaarheid)
- 3. Hoe hebben de studenten op de toets gescoord?
- 4. Wat is de kwaliteit van de afzonderlijke items?

Bij de domeintoets is alleen gekeken naar de kwaliteit van de afzonderlijke items.

Op itemniveau is gekeken naar (a) de moeilijkheid van het item (p-waarde), (b) de afleiders en (c) het onderscheidend vermogen van het item. In figuur 3 staat een voorbeeld van een itemanalyse waarbij het juiste alternatief een p-waarde heeft van 0,61 en een hoog

onderscheid gemaakt wordt tussen de betere studenten (de bovenste groep) en de zwakkere studenten (de onderste groep).

Items met een lage p-waarde zijn items die door veel studenten foutief zijn beantwoord. Deze items en de items met een laag of negatief onderscheidend vermogen tussen goede en zwakke studenten worden door de toetsdeskundige besproken met de hoofdredacteur. Op basis van deze analyse worden items bijgesteld en soms zelfs weggegooid.

De ervaringen van studenten

Na afloop van het maken van de toetsen evalueerden de studenten vervolgens het gemaakte werk. Bij de *pretest* in 2010 wisten studenten vooraf nog niet goed wat het doel van de toets was, bij de domeintoets is dit verbeterd. De studenten vinden de toets wel moeilijk en geven aan dat nog niet alle stof van de toets in het curriculum behandeld is. 95 Procent van de studenten heeft zich niet specifiek voor de domeintoets voorbereid. De studenten geven aan dat deze kennistoets helpt om de vakinhoudelijke kennis beter te beheersen. Ze vinden de toets wel lang en de teksten bij het historisch redeneren te uitgebreid. De toetsscore was voor eenderde van de studenten hoger dan verwacht en voor een even groot aantal studenten lager dan verwacht.

Het tussentijds aanscherpen van het ontwikkelproces en het uitvoeren van de *pretest* en domeintoets hebben een bijdrage geleverd aan het tot stand komen van kwalitatief goede items. De domeintoetsen die dit jaar nog volgen zullen een compleet beeld geven van de relevantie, moeilijkheid en het onderscheidend vermogen van de items. Deze informatie zal er voor zorgen dat er goed onderbouwde keuzes gemaakt kunnen worden voor het vaststellen van een kwalitatief goede landelijke kennistoets. ■

Figuur 3.
Een itemanalyse
voor geschiedenis.

Antwoordmogelijkheden	Aantal keren gekozen	p-waarde	Bovenste groep	Onderste groep	Onderscheidend vermogen	Correlatie itemtoets
a*	28	0,61	0,92	0,17	0,75	0,53
b	4	0,09	0,08	0,00	0,08	0,08
c	4	0,09	0,12	0,08	0,04	-0,11
d	10	0,21	0	0,75	-0,75	-0,49
* Juiste antwoord						