

# Wiskunde valide getoetst?

De digitale landelijke kennistoets wiskunde van de tweedegraads lerarenopleiding vergeleken met de instituutstentamens

Paul Drijvers, Freudenthal Instituut, Universiteit Utrecht  
Hendrik Straat & Saskia Wools, Cito

Samenvatting

*De tweedegraads lerarenopleidingen in Nederland kennen sinds enkele jaren een digitale Landelijke Kennistoets (LKT), waarmee de vakinhoudelijke startbekwaamheid voor het leraarschap van studenten wordt vastgesteld. Nu hiermee enige ervaring is opgedaan, rijst de vraag hoe valide deze toets is. Om dit te onderzoeken voor het vak wiskunde zijn de resultaten van de LKT vergeleken met die van de instituutstentamens en is nagegaan in hoeverre de digitale toetsomgeving voldoet aan de huidige criteria. De conclusies zijn dat de LKT wiskunde ondanks een afwijkend karakter ten opzichte van de instituutstentamens in grote lijnen als valide kan worden beschouwd. De beperkingen van de digitale toetsomgeving zijn mede oorzaak van dit afwijkende karakter. Het advies is dan ook om het gebruik van een andere toetsomgeving te overwegen. Daarnaast is samenwerking tussen de verschillende opleidingen gewenst om een landelijk toetsbeleid verder vorm te geven.*

## Inleiding

Het niveau van de vakkennis van studenten van de tweedegraads lerarenopleidingen in Nederland is al enige tijd een punt van zorg. Om het gewenste minimumniveau vast te stellen zijn door de Vereniging van Hogescholen (voorheen HBO-raad) voor de verschillende schoolvakken kennisbases ontwikkeld. Hierin wordt beschreven over welke vakkennis studenten in Nederland aan het einde van de lerarenopleiding moeten beschikken om startbekwaam te zijn als docent in het tweedegraads onderwijsveld. Om vast te stellen of dit kennisniveau ook daadwerkelijk wordt beheerst, zijn onder regie van *10voordeleraar*, een samenwerkingsverband van alle lerarenopleidingen, sinds 2010 digitale landelijke kennistoetsen (LKT) ontwikkeld. De landelijke kennistoets bestaat voor 15 verschillende opleidingen van de tweedegraads lerarenopleiding. Deze landelijke ontwikkeling vraagt natuurlijk om een proces van afstemming met het onderwijs en de toetsing op de plaatselijke hogescholen.

In het geval van het vak wiskunde omvat de kennisbasis docent wiskunde bachelor (HBO-raad / VvH, 2009) zes domeinen, waarvan het eerste domeinoverstijgend is en de andere gerelateerd zijn aan specifieke deelgebieden van de wiskunde en de wiskundecurricula. Naast deze vakinhoudelijke kennisbasis is er ook een vakdidactische kennisbasis. De LKT wiskunde, waarvan er in de studie jaren 2013-2014 en 2014-2015 in totaal zes zijn afgenomen, wordt ontwikkeld en beoordeeld door opleiders van de verschillende lerarenopleidingen. Het betreft digitale toetsen die bestaan uit meerkeuze-items en numerieke antwoordvragen.<sup>1</sup>

Bij de tweedegraads lerarenopleidingen wiskunde wordt echter de vraag gesteld of deze vorm van toetsing valide is. Hoewel de tentamens van de opleiding en de LKT in principe dezelfde

<sup>1</sup>. Informatie over de werkwijze van *10voordeleraar* bij de ontwikkeling en afname van de kennistoets is te vinden op <https://www.10voordeleraar.nl/programma/zo-werkt-het>. De toetsgids wiskunde is beschikbaar via <https://www.10voordeleraar.nl/publicaties>.

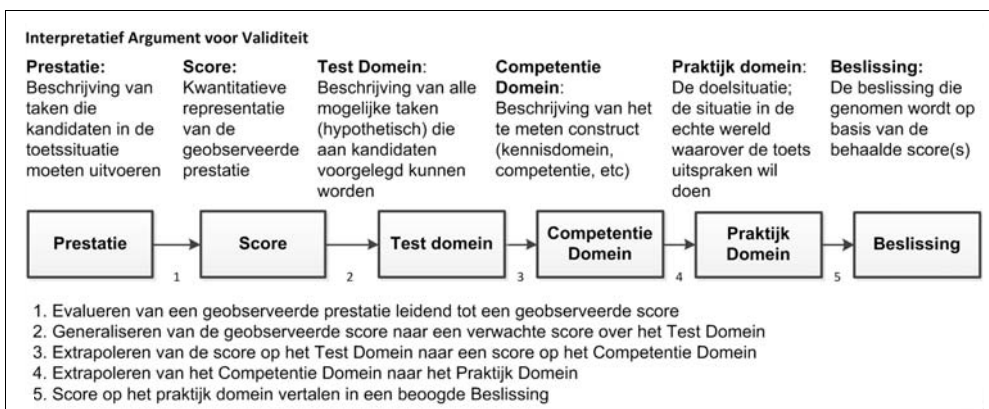
kennisbasis toetsen, vermoeden de opleiders een lage correlatie tussen resultaten op beide typen toetsen. Om een oordeel te vellen over de validiteit van de LKT en de resultaten te vergelijken met die van de instituuftentamens heeft *10voordeleraar* in het najaar van 2015 een onderzoek laten uitvoeren. In dit artikel doen we verslag van de belangrijkste bevindingen daarvan.

## Theoretisch kader

Om de geschetste problematiek nader te onderzoeken moeten we weten wat we onder validiteit verstaan. Validiteit wordt gezien als het belangrijkste kwaliteitsaspect van een toets (AERA, APA, & NCME, 2014). We definiëren het als de mate waarin een toetsscore geschikt is voor een bedoelde interpretatie en een bedoeld gebruik van de toets (Kane, 2013; Wools, 2015). De validiteit hangt daarmee samen met het doel en het gebruik van de toets en is dus geen kenmerk van de toets op zichzelf. In het geval van de LKT wiskunde betekent dit dat de toetsscore geschikt moet zijn om te beoordelen of de kandidaat beschikt over de kennis die de leraar aan het einde van zijn opleiding minimaal moet hebben om professioneel bekwaam en zelfstandig aan het werk te kunnen in het onderwijs.

In navolging van Kane (2006, 2013) en zijn *argument-based approach* to validation beschouwen we validiteit en daarmee ook validering als een keten van inferenties. In Figuur 1 wordt deze keten uitgebeeld. De rechthoeken geven de verschillende stadia aan om van een Prestatie, in dit geval het werk van de student op de toets, tot een Beslissing te komen, namelijk vakinhoudelijk startbekwaam zijn of niet. De genummerde pijlen geven aan hoe we van de ene fase naar de volgende gaan en welke aannames daaraan ten grondslag liggen. De geldigheid van deze aannames bepalen de validiteit van de toets als geheel. In het geval van de LKT hebben we weinig twijfels bij de correcte omzetting van studentprestaties in scores, of in de vertaling van score naar testdomein. De potentieel zwakke schakels van de keten lijken in dit geval met name de stappen 3 en 4, de extrapolaties van de resultaten naar het competentiedomein van de kennisbasis wiskunde en naar het praktijkdomein van het wiskundeonderwijs, en eventueel stap 5, het nemen van de beslissing.

Naast deze visie op validiteit betreft een tweede theoretische invalshoek de digitale toetsing van wiskunde. Digitale toetsing in het algemeen kent een aantal voordelen boven schriftelijke toetsing, zoals tijd- en plaatsafhankelijke afname, automatische, snelle en betrouwbare scoring,



Figuur 1. Interpretatieve validiteitsketen (Wools, Eggen & Sanders, 2010).

eenvoudige samenstelling van toetsen van vergelijkbare moeilijkheidsgraad, en mogelijkheden voor adaptieve toetsing. Wat is in het algemeen bekend over digitale toetsing van wiskunde in vergelijking tot schriftelijke toetsen en welke eisen kunnen vanuit het perspectief van validiteit aan een digitale toetsomgeving voor wiskunde worden gesteld?

**Omgaan met wiskundige objecten in een digitale omgeving verschilt van de praktijk met pen & papier.**

Uit een recente literatuurstudie blijkt dat de effecten van de afname-modus op de toetsresultaten in het algemeen klein zijn (Cito, 2015a). Als belangrijkste oorzaken van deze moduseffecten worden genoemd de noodzaak voor de kandidaat om te switchen tussen beeldscherm en kladpapier, de geringe vertrouwdheid met de digitale toetsomgeving, en het invoeren van antwoorden via het toetsenbord. Toch wordt er bij wiskunde ook wel zwaar getild aan de overgang van schriftelijke naar digitale toetsing, omdat het omgaan met wiskundige objecten in een digitale omgeving zo verschilt van de praktijken met pen en papier, waaraan leerlingen en studenten gewend zijn:

*"However, examinees overwhelmingly expressed a preference for taking a paper-based rather than computer-based test, because inputting mathematical objects was so cumbersome. The difficulties arising from the sharp contrast between hand written mathematics and keyboard-entered mathematics is a recurring theme in reports of computer-based assessment of all types and at all levels of education." (Stacey & Wiliam, 2013, p. 736-737)*

Om de digitale LKT te evalueren, identificeren we de volgende criteria waaraan een afspeelomgeving voor digitale toetsing van wiskunde zou moeten voldoen (Drijvers et al., 2016).

- ▶ *Veiligheid en betrouwbaarheid*  
Vanzelfsprekend moet de toetsomgeving veilig zijn en betrouwbaar. Veilig houdt onder andere in dat vanuit de toets geen toegang tot andere bronnen mogelijk is (lock down). Betrouwbaar betekent onder meer dat de toetsomgeving robuust is en geen uitval vertoont.
- ▶ *Variatie aan vraag- en interactietypes*  
Een omgeving waarin alleen meerkeuze- of multiple response vragen kunnen worden gesteld is zeer beperkt voor het toetsen van wiskunde. De volgende criteria specificeren dit punt.
- ▶ *Formule-invoer door kandidaten*  
Een specifiek criterium voor het uitbreiden van vraag- en interactietypes is de beschikbaarheid van een gebruiksvriendelijke formule-editor waarmee de kandidaat expressies en vergelijkingen (inclusief breuken en wortels) kan invoeren.
- ▶ *Grafieken-invoer door kandidaten*  
Een tweede criterium voor het uitbreiden van vraag- en interactietypes is de beschikbaarheid van een gebruiksvriendelijke grafieken editor waarmee de kandidaat grafieken kan laten tekenen en wijzigen.
- ▶ *Meetkundige constructies door kandidaten*  
Een derde criterium voor het uitbreiden van vraag- en interactietypes is de beschikbaarheid van een gebruiksvriendelijke meetkunde module waarmee de kandidaat meetkundige objecten kan definiëren en meetkundige constructies kan uitvoeren.
- ▶ *Intelligente automatische beoordeling*  
Naarmate de vraag- en interactietypes rijker worden, is meer intelligente beoordeling nodig. In praktijk vraagt dit om de beschikbaarheid van computeralgebra software, die bijvoorbeeld

herkent dat  $x^2 + 3x + 2$  en  $(x + 1)(x + 2)$  equivalent zijn en dat beide antwoorden in een bepaald item als correct beoordeeld dienen te worden.

#### ► *Beoordeling deelstappen*

In het geval dat meerstaps oplossingsprocedures aan de orde zijn en dat is bij de LKT wiskunde veelal aan de orde - is het voor de aansluiting met de gangbare schriftelijke toetspraktijk bij wiskunde wenselijk dat deelstappen met behulp van deelscores kunnen worden gehonoreerd.

Op basis van deze overwegingen over validiteit en digitale toetsing formuleren we de volgende onderzoeksvragen.

- 1 In hoeverre is de LKT wiskunde valide in termen van de extrapolatie van de resultaten naar de beheersing van de kennisbasis en de praktijk van het wiskundeonderwijs?
- 2 In hoeverre voldoet de toetsomgeving van de LKT wiskunde aan de criteria die we aan een digitale wiskundetoets stellen?

## **Methode**

In de oriëntatiefase is gesproken met een aantal betrokkenen bij de LKT wiskunde. Daarnaast heeft een tweetal besprekingen plaatsgevonden met het Landelijk Vakoverleg Opleiders wiskunde (LVO).

Voor de beantwoording van de eerste onderzoeksvraag is bij *10voordeleraar* een aantal inhoudelijke, praktische en psychometrische gegevens van de LKT wiskunde opgevraagd en geanalyseerd. Het gaat met name om de toetsen zelf, de resultaten en de cesuurbepaling. Deze gegevens betreffen de zes LKT-afnames in de studie jaren 2013-2014 en 2014-2015. We hebben ons gericht op de vier domeinen analyse, meetkunde, algebra en statistiek uit de kennisbasis. De andere domeinen zijn hiermee buiten beschouwing gelaten, omdat ze een te algemeen karakter hebben en niet goed te koppelen zijn aan onderdelen uit het onderwijsaanbod van de hogescholen. Omdat in het veld het idee leefde dat de resultaten van de LKT slecht correleren met die van de instituuftentamens, terwijl beide typen toetsen gericht zijn op dezelfde kennisbasis, is bij de hogescholen een aantal studiegidsen, tentamens, correctievoorschriften, uitwerkingen en resultaten verzameld. Het gaat om tentamens die in de ogen van de opleiders het meest bijdragen aan de kennis van de vier bovengenoemde domeinen van de kennisbasis. Daarnaast is middels een vragenlijst geïnventariseerd op welke manier studenten op de LKT worden voorbereid. Voor de beantwoording van de tweede onderzoeksvraag is de implementatie van de huidige LKT wiskunde beoordeeld aan de hand van de hiervoor geschetste criteria. Tevens is verdere literatuur geraadpleegd, is met een aantal deskundigen gesproken.

## **Resultaten**

In de volgende paragrafen beschrijven we eerst de psychometrische kenmerken van de LKT. Vervolgens vergelijken we de resultaten van de LKT met die van de instituuftentamens. Ten slotte gaan we in op de digitale toetsvorm van de huidige LKT.

### **Psychometrische kenmerken van de LKT wiskunde**

Tabel 1 bevat de belangrijkste kengetallen van de zes afnames van de LKT wiskunde die zijn onderzocht. In de eerste kolom staan het studiejaar en rangnummer van de toets in dat jaar.

In de tweede kolom staat het aantal studenten dat heeft deelgenomen. Dit kan een eerste deelname zijn of een herkansing. Kolom 3 bevat gemiddelde kennispercentages: het gemiddelde percentage van de maximale ruwe score die studenten hebben behaald. Kolom 4 geeft de betrouwbaarheden (Cronbach's alfa). In de laatste kolom staan de slagingspercentages.

In Tabel 1 valt op dat het aantal deelnemers aan 1314T01 zo laag is dat we deze afname in onze analyse buiten beschouwing laten. Verder zien we in de derde kolom dat de gemiddelde kennispercentages variëren van 53.7 tot 62.7 %.

De slagingspercentages in de laatste kolom vertonen meer spreiding. Als we deze slagingspercentages vergelijken met die van de andere vakken die een LKT kennen, dan verschilt dit niet spectaculair. Het (naar het aantal deelnemers gewogen) gemiddelde slagingspercentage van de kennistoetsen 1415T03 over de vijftien verschillende vakken heen is bijvoorbeeld 67%. Dit is wel wat hoger dan de 61% voor wiskunde, maar niet spectaculair.

*Tabel 1* Overzicht van de zes kennistoetsen in 2013-2014 en 2014-2015

<b>Kennistoets</b>	<b>N</b>	<b>Kennispercentage</b>	<b>Betrouwbaarheid</b>	<b>Perc. geslaagd</b>
1314T01	38	62.7 %	0.81	84 %
1314T02	120	61.1 %	0.78	73 %
1314T03	149	55.2 %	0.78	68 %
1415T01	80	56.6 %	0.72	25 %
1415T02	152	53.7 %	0.73	25 %
1415T03	278	59.1 %	0.76	61 %

Stap 2 in de valideringsketen van Figuur 1 betreft de vraag of de score op een specifieke kennistoets kan worden gegeneraliseerd naar het testdomein, in dit geval de verzameling items in de LKT itembank. Criteria zijn betrouwbaarheid, representativiteit en stabiliteit. Tabel 1 laat zien dat de betrouwbaarheid van de verschillende toetsen voldoende<sup>2</sup> is. Wat betreft de representativiteit is gebleken dat de LKT wiskunde een toetsmatrijs kent, waarin de spreiding van de items over de verschillende domeinen wordt gegarandeerd. Dat leidt tot een representatieve en vergelijkbare inhoudelijke samenstelling van de toetsen. De matrijs bevat behalve domeinindicaties ook een indeling in de categorieën Kennis, Inzicht en Toepassing. Ter bevordering van de stabiliteit bestaat de LKT wiskunde voor ongeveer 20% uit zogeheten ankeritems, items die ook bij eerdere afnames zijn gebruikt. Deze ankeritems maken het mogelijk om de verschillende cohorten deelnemers met elkaar te vergelijken.

Stap 3 in de validiteitsketen van Figuur 1 betreft de vraag of de scores kunnen worden geëxtrapoléerd naar het competentiedomein, de kennisbasis wiskunde. De hierboven genoemde toetsmatrijs garandeert dat de verschillende domeinen van de kennisbasis (met uitzondering van domein 0) in een goede verhouding aan de orde komen. Verder zijn de lerarenopleiders wiskunde degenen die de kennisbasis operationaliseren in hun onderwijs - betrokken bij itemconstructie, eindredactie en vaststelling van de kennistoetsen. Om deze redenen lijkt de extrapolatie naar het competentiedomein inderdaad in orde te zijn.

<sup>2</sup>. Zie COTAN <http://dare.uva.nl/document/2/79346> p. 33. Bij belangrijke beslissingen raadt COTAN aan om strengere normen te hanteren.

De laatste stap uit de validiteitsketen uit Figuur 1 betreft de cesuur en de standaard met het oog op de startbekwaamheid van de wiskundedocent. Inhoudelijk is het niveau van de kennistoets conform de gezette standaard in de kennisbasis. Toch valt in de vijfde kolom van Tabel 1 een zeer laag percentage geslaagden bij 1415T01 en 1415T02 op, in absolute zin maar ook in vergelijking met de voorgaande en volgende toetsen. Het verschil tussen 1314T03 en 1415T01 is bijvoorbeeld opmerkelijk: een vrijwel gelijk kennispercentage (55.2 resp. 56.6%) maar een slagingspercentage dat is gedaald van 68% naar 25%. Ook de prestaties op de ankeropgaven zijn vergelijkbaar. Mogelijk ligt de oorzaak van deze grote verschillen in een verschillende moeilijkheidsgraad van de items, of in het functioneren van de panels van experts bij de cesuurbepaling. Cesuurbepaling is een subtiele aangelegenheid: als er veel studenten rond het grensniveau scoren, zoals in 1415T01 en 1415T02 het geval was, heeft een kleine verandering van de cesuur grote invloed op de slagingspercentages.

De vraag is in hoeverre de resultaten van de LKT mede worden bepaald door praktische aspecten zoals de voorbereiding erop en de timing. Uit de gegevens van de vragenlijst aan hogescholen blijkt dat in het verleden weinig aandacht aan de voorbereiding van studenten op de kennistoets is besteed. Op dit moment zijn er tal van initiatieven gaande om deze voorbereiding beter ter hand te nemen. Verder is de timing van de LKT complex: als studenten in studiejaar 3 aan de LKT deelnemen, komen daarin onderwerpen aan de orde die in het reguliere programma al enige tijd zijn afgesloten. Juist in de eerste studie jaren ligt het accent immers op de vakinhoudelijke kennis. Er zijn opleidingen die hierom overwegen om vakinhoudelijke elementen te verplaatsen naar studiejaar 3. Een alternatief zou kunnen zijn om de LKT aan het einde van studiejaar 2 af te nemen.

### **De resultaten van de LKT wiskunde vergeleken met die van de instituutstentamens**

Stap 4 uit Figuur 1 betreft de vraag of de score op het competentiedomein kan worden geëxtrapoleerd naar het praktijkdomein. Aangenomen dat de opleidingen dicht bij de onderwijspraktijk staan en dat dit zich weerspiegelt in de instituutstoetsing, is het interessant om de kwantitatieve en kwalitatieve verschillen tussen kennistoetsen en instituutstentamens in kaart te brengen. Tenslotte zijn beide toetsen gericht op de opleidingsdoelen zoals geformuleerd in de kennisbasis. Overigens zijn de betrouwbaarheden van de verschillende instituutstentamens niet bijzonder hoog, maar dat wordt mede veroorzaakt door de veelal lage studentenaantallen en is dan ook geen reden tot zorg.

In Tabel 2 staan voor elk van de zes kennistoetsen de aantallen studenten, de gemiddelde cijfers LKT en de gemiddelde tentamencijfers over de vier instituutstentamens die in de ogen van de opleiders het meest bijdragen aan de kennis van de domeinen 1 - 4 van de kennisbasis. In de linker helft van Tabel 2 zijn deze data gegeven voor alle studenten van de tien opleidingen die aan de LKT hebben deelgenomen. In de rechter helft van Tabel 2 zijn studenten weggelaten van wie niet alle gegevens compleet zijn, bijvoorbeeld door vrijstellingen; van de overgebleven studenten is dus over alle vier instituutsvakken gemiddeld. Deze gemiddelde tentamencijfers betreffen overigens tentamens die per opleiding verschillen. Bovendien is de hogescholen gevraagd om van studenten die aan het betreffende tentamen meerdere keren hebben deelgenomen alleen het hoogste resultaat te vermelden. Als een student een of meer keren zakt voor het tentamen en vervolgens slaagt, is dus alleen het laatste, voldoende cijfer opgenomen.

Uit Tabel 2 blijkt ten eerste dat het gemiddelde tentamencijfer over de jaren heen weinig fluctueert (met uitzondering van 1314T01, maar dit cohort is erg klein). Dit suggereert dat de verschillende studentcohorten op de tentamens vrij constant scoren en dat dus ook de vaardigheid van verschillende cohorten wellicht vrij constant blijft, al ontbreken gegevens over de cesuurbepaling bij instituuttentamens. Als dit het geval is, dan is de grotere spreiding in de LKT-cijfers opmerkelijk.

Tabel 2 Vergelijking scores kennistoets en instituuttentamens

Kennistoets	Alle studenten			Studenten na listwise deletion		
	N	Gem. cijfer LKT	Gem. tentamencijfer	N	Gem. cijfer LKT	Gem. tentamencijfer
1314T01	38	6.5	7.6	12	5.7	7.4
1314T02	118	6.1	7.1	50	6.0	6.9
1314T03	149	5.9	7.2	79	5.9	7.1
1415T01	80	4.7	7.0	48	4.6	6.9
1415T02	152	4.8	7.0	73	4.8	7.1
1415T03	278	5.7	7.0	128	5.7	7.0

Ten tweede zien we dat de cijfers van de kennistoets beduidend lager liggen dan die van de tentamens. Het gewogen gemiddelde verschil is 1,55 cijferpunt voor de groep studenten van wie alle gegevens beschikbaar zijn. Voor 1415T01 en 1415T02 lopen deze verschillen zelfs op tot 2,3 cijferpunt. Deze verschillen kunnen veroorzaakt worden doordat per student alleen de hoogste tentamenresultaten zijn verwerkt. Een tweede verklaring is de tijd die is verstreken sinds de tentaminering. Immers, de vaktentamens vinden grotendeels in de studie jaren 1 en 2 plaats, terwijl het accent van de opleiding in de jaren 3 en 4, waarin de kennistoets wordt afgenomen, juist op de beroepsvoorbereiding ligt. Een derde mogelijke verklaring is het feit dat een tentamen veel specifiek is dan de kennistoets, die de hele kennisbasis overdekt. Binnen de opleidingen worden momenteel dan ook maatregelen genomen om studenten beter voor te bereiden op de kennistoets.

Wat kunnen we zeggen over de samenhang tussen de resultaten op de kennistoets en die van de instituuttentamens? Tabel 3 bevat in de kolom  $R_{geheel}$  de correlatiecoëfficiënten van cijfer kennistoets en gemiddeld tentamencijfer voor de zes kennistoetsen.  $R_{correctie}$  is hetzelfde maar dan gecorrigeerd voor attenuatie, dat wil zeggen voor de betrouwbaarheden binnen de toetsen, die voor de tentamencijfers overigens vrij laag. In de volgende vier kolommen staan de correlaties tussen de score op het betreffende domein in de kennistoets en die van het bijbehorende tentamen. Met uitzondering van de eerste kennistoets - waar van slechts 12 studenten alle gegevens beschikbaar waren zijn de waarden van  $R_{correctie}$  vrij goed te noemen. Dit suggereert dat wat men meet met de kennistoets voor een belangrijk deel overeen komt met wat de opleidingstentamens meten, en ondersteunt daarmee de validiteit van de kennistoets. Tabellen 4 en 5 bevestigen dit beeld. In tabel 4 staan de gemiddelde cijfers voor de tentamens van studenten bij een gegeven LKT-cijfer, met daarachter het aantal studenten waarover is gemiddeld. Weer met uitzondering van 1314T01 zien we dat de gemiddelde tentamencijfers stijgen naarmate het cijfer voor de kennistoets hoger is. In Tabel 5 zijn juist de gemiddelde

Tabel 3 *Correlaties scores kennistoets en instituutstentamens*

Kennistoets	$R_{\text{geheel}}$	$R_{\text{correctie}}$	$R_{\text{analyse}}$	$R_{\text{meetkunde}}$	$R_{\text{algebra}}$	$R_{\text{statistiek}}$
1314T01	-0,26	-0,47	0,52	0,28	0,27	0,18
1314T02	0,46	0,71	0,36	0,37	0,42	0,19
1314T03	0,50	0,70	0,31	0,21	0,38	0,26
1415T01	0,44	0,71	0,21	0,21	0,29	0,28
1415T02	0,41	0,60	0,21	0,22	0,30	0,22
1415T03	0,47	0,78	0,36	0,24	0,21	0,15

Tabel 4 *Gemiddeld tentamencijfer afhankelijk van cijfer kennistoets*

Cijfer LKT	1314T01 (cijfer/freq)	1314T02 (cijfer/freq)	1314T03 (cijfer/freq)	1415T01 (cijfer/freq)	1415T02 (cijfer/freq)	1415T03 (cijfer/freq)	Totale rijfreq
1							
2				5.9 / 1			1
3		6.0 / 1	6.1 / 2	6.6 / 6	7.0 / 9	6.8 / 4	22
4	7.9 / 2	6.6 / 6	6.4 / 9	6.7 / 16	6.8 / 24	6.7 / 14	71
5	7.6 / 2	6.9 / 8	6.8 / 16	7.0 / 16	7.1 / 20	6.6 / 34	96
6	7.3 / 6	6.9 / 21	7.1 / 28	7.6 / 6	7.6 / 14	7.0 / 45	120
7	6.7 / 2	6.9 / 9	7.7 / 19	7.5 / 3	8.0 / 6	7.4 / 24	63
8		7.9 / 2	8.0 / 5			7.7 / 6	13
9		7.7 / 3				9.1 / 1	4
10							

Tabel 5 *Gemiddeld cijfer kennistoets afhankelijk van gemiddeld tentamencijfer*

Gem. tentamen-cijfer	1314T01 (cijfer/freq)	1314T02 (cijfer/freq)	1314T03 (cijfer/freq)	1415T01 (cijfer/freq)	1415T02 (cijfer/freq)	1415T03 (cijfer/freq)	Totale rijfreq
1							
2							
3							
4							
5			6.0 / 1				1
6		5.8 / 10	4.9 / 14	4.0 / 9	4.3 / 12	5.3 / 27	72
7	6.0 / 7	5.8 / 31	5.7 / 37	4.6 / 30	4.6 / 36	5.5 / 66	207
8	5.5 / 4	6.5 / 8	6.6 / 23	5.3 / 8	5.2 / 21	6.4 / 30	94
9	4.0 / 1	9.0 / 1	7.0 / 4	6.0 / 1	5.8 / 4	7.2 / 5	16
10							



LKT-cijfers gegeven voor de studenten met een gegeven afgerond gemiddelde tentamen-cijfer. Hier zien we dat het gemiddelde cijfer voor de kennistoets stijgt naarmate de gemiddelde tentamencijfers hoger zijn. Beide tabellen bevestigen de positieve samenhang tussen LKT-cijfer en gemiddeld tentamencijfer. Weer gaat het hier om het hoogste tentamencijfer dat een student voor de betreffende vakken heeft gehaald, zodat lage tentamencijfers ontbreken.

Als kanttekening bij het bovenstaande moeten we opmerken dat de verschillende opleidingen natuurlijk verschillende tentamens afnemen en dat de vakinhouden ook niet identiek zijn, aangezien elke opleiding een eigen vakindeling heeft. Ondanks deze verschillen hebben de instituutstentamens de volgende kenmerken gemeenschappelijk die afwijken van de kennistoets.

**Wij vonden een positieve samenhang tussen LKT-cijfer en gemiddeld tentamen-cijfer.**

- ▶ De instituutstentamens bevatten veel clusteropgaven, opgaven waarin meerdere vragen worden gesteld aan de hand van eenzelfde probleemsituatie. Dergelijke vragen komen bij de kennistoets niet voor;
- ▶ De instituutstentamens bevatten contextopgaven waarin gemodelleerd moet worden (domein 0.4 uit de kennisbasis), wat bij de LKT niet of nauwelijks voorkomt;
- ▶ De instituutstentamens bevatten vragen naar bewijzen, die in de kennistoets niet of nauwelijks voorkomen (domein 0 uit de kennisbasis);
- ▶ In de instituutstentamens op het gebied van meetkunde worden constructies en tekeningen gevraagd, wat bij de kennistoets niet gebeurt (domein 2.1 uit de kennisbasis);
- ▶ De instituutstentamens bevatten open vragen, die studenten beantwoorden in termen van (series van) formules en afleidingen. Bij de kennistoets hebben studenten dergelijke formules en afleidingen ook nodig om antwoorden te vinden, maar zijn deze niet zichtbaar omdat de LKT alleen meerkeuzevragen en numerieke antwoordvragen bevat;
- ▶ De instituutstentamens bevatten meerstapsopgaven waarbij beoordeling in deelscores plaatsvindt, terwijl een opgave bij de kennistoets geheel goed of geheel fout wordt gerekend.

De kennistoets wijkt dus qua karakter in een aantal opzichten af van de gebruikelijke toetsen in de lerarenopleiding wiskunde. Bepaalde onderdelen van de kennisbasis, zoals construeren en bewijzen, komen in de instituutstentamens meer aan de orde dan in de kennistoets. Door het gebruik van contexten en clusteropgaven kan modelleren in de instituutstentamens beter worden getoetst dan in de kennistoets. Ook maakt het gebruik van pen en papier het toekennen van deelscores voor bijvoorbeeld algebraïsche technieken beter mogelijk dan de digitale toetsvorm van de kennistoets. Afhankelijk van de visie op wat het praktijkdomein inhoudt (zie stap 4 van Figuur 1) vormt dit karakterverschil een beperking van de validiteit van de kennistoets: het overgrote deel van de wiskundetoetsen, die op dit moment in de praktijk van de startbekwame leraar bij leerlingen worden afgenomen, lijkt op de genoemde verschilpunten meer op de instituutstentamens dan op de kennistoets.

### **De digitale toetsvorm van de LKT wiskunde**

Over het geheel genomen is de wiskundige en toetsinhoudelijke kwaliteit van de items van de kennistoets meer dan voldoende. Een steekproefsgewijze inhoudelijke analyse van de


kennistoetsen heeft geleid tot de signalering van een aantal inhoudelijke beperkingen van de digitale realisatie van de toets, die wordt geconstrueerd en gemaakt in QuestionMark Perception (QMP).

- 1 Met uitzondering van het criterium veiligheid en betrouwbaarheid voldoet de LKT in de QMP omgeving niet aan de in paragraaf 2 genoemde criteria voor een digitale toetsomgeving voor wiskunde.
- 2 De wiskundige notaties en presentaties laten te wensen over. De layout en uitlijning van formules zijn niet goed; voor eenzelfde wiskundige object worden verschillende lettertypes gebruikt; notaties zijn niet consequent..
- 3 Sommige vragen zijn gekunsteld om de beperkingen van de digitale toetsomgeving (alleen meerkeuze of enkelvoudig numeriek vraagtype) te omzeilen; denk aan vragen naar een benadering waar een exact antwoord meer voor de hand ligt.

Als illustratie van de drie bovengenoemde punten wordt in Figuur 2 een opgave uit het oefenmateriaal van de LKT wiskunde afgebeeld. Daarin zien we dat de formules voor  $\cos(a)$  en voor  $BC$  niet goed zijn uitgelijnd. Ook zijn de haken om  $a$  te groot. Voor hetzelfde lijnstuk  $BC$  en voor dezelfde variabele  $k$  worden verschillende lettertypes gebruikt. Belangrijker is dat de vraagstelling onnatuurlijk is: het zou in deze situatie veel meer voor de hand liggen om naar  $BC$  te vragen en niet naar het kwadraat  $k$  daarvan.  $BC = \sqrt{k}$  is kunstmatig. Dat de vraag toch op deze manier wordt gesteld, wordt veroorzaakt door de beperking van QMP om geen wortel te kunnen invoeren.

Samengevat kent de huidige digitale toetsvorm van de LKT wiskunde een aantal beperkingen qua vraagtypen en invoermogelijkheden, die de kwaliteit van de items niet verhogen en de verschillen met de instituu(t)stentamens vergroten.

10  
voor  
de  
leraar



---

Aangemeld als: deelnemer\_oefentoets

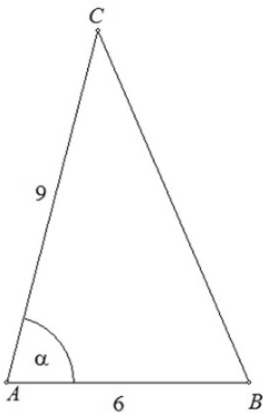
---

9 van 10  
WK-020401-02a-v01

---

Gegeven

In een driehoek  $ABC$  geldt  $AB = 6$ ,  $AC = 9$  en  $\cos(\alpha) = \frac{1}{4}$  Er geldt  $BC = \sqrt{k}$



Gevraagd  
Wat is de waarde van  $k$ ?

Figuur 2. Opgave uit de oefentoets LKT wiskunde (<https://www.10voordeleraar.nl/toetsen/oefenmateriaal>).

## Conclusie

### Conclusies

De eerste onderzoeksvraag was in hoeverre de LKT wiskunde valide is in termen van de extrapolatie van de resultaten naar de beheersing van de kennisbasis en de praktijk van het wiskundeonderwijs. De conclusie is dat de validiteit van de huidige kennistoets wiskunde in orde is waar het gaat om (1) de duidelijke relatie met de kennisbasis, (2) de betrokkenheid van de opleiders bij itemconstructie, redactie en vaststelling, (3) de psychometrische behandeling van items na afname, (4) de betrouwbaarheid, en (5) de correlatie van cijfers met tentamencijfers. Twijfels bestaan over de lage slagingspercentages bij de afnames 1415T01 en 1415T02.

Verder wijkt het karakter van de kennistoets af van dat van de instituutstentamens wat betreft een aantal te toetsen vaardigheden (construeren, bewijzen, modelleren), het gebruik van clusteropgaven en contexten, en het toekennen van deelscores, terwijl deze aspecten wel deel uitmaken van het praktijkdomein van de startbekwame leraar.

De tweede onderzoeksvraag was in hoeverre de toetsomgeving van de LKT wiskunde voldoet aan de criteria die men aan een digitale wiskundetoets kan stellen. De conclusie is dat dit slechts gedeeltelijk het geval is. Voor de hogere orde vaardigheden is een schriftelijke toetsing wellicht passender. Met een overgang naar een meer geavanceerde toetsomgeving valt mogelijk winst te behalen op het terrein van rijkere vraagtypes en interacties die beter aansluiten bij de praktijk van wiskunde met pen en papier, en de eventuele beoordeling van tussenstappen.

### Aanbevelingen

De conclusies laten zich eenvoudig vertalen in aanbevelingen. De eerste is om voor de landelijke kennistoets wiskunde over te stappen op een digitale toetsomgeving die meer vragen en interactietypen mogelijk maakt. Daarnaast lijkt het verstandig om de LKT wiskunde te beperken tot de domeinen 1-4 en daarbij de nadruk te leggen op procedurele vaardigheden en parate kennis. De ambities van de LKT worden dan wat realistischer en de hogere orde vaardigheden kunnen via de instituutstentamens worden getoetst. Tevens is de inbedding van de LKT binnen de opleidingen voor verbetering vatbaar. De timing van de LKT kan bijvoorbeeld beter worden afgestemd op de vakinhoudelijke opleiding en studenten kunnen beter worden voorbereid op de toets. Verder zou bij het bepalen van de cesuur wellicht beter gebruik kunnen worden gemaakt van de ankeropgaven.

Als inderdaad het toetsen van hogere orde vaardigheden via instituutstentamens plaatsvindt, ligt het voor de hand om ook op dat punt landelijke samenwerking te bevorderen. Denk aan het uitwisselen van tentamens, aan peer review tussen opleidingen, aan tweede correctie van tentamens, waardoor ook dit deel van de kennisbasis landelijke borging kent. Een structurele vorm van intercollegiale samenwerking tussen de hogescholen op dit gebied kan bijdragen aan een beter niveau van toetsing en op termijn wellicht zelfs tot verdere afstemming van de plaatselijke curricula.

### Discussie

Wat leren we nu van deze casus? Ten eerste is de landelijke toetsing van de beheersing van een kennisbasis geen eenvoudige zaak; dat zal niet alleen voor het vak wiskunde gelden. De invoering van een dergelijke toets kan alleen succesvol zijn als er een breed draagvlak voor is

en de invoering gepaard gaat met ontwikkelingen in het onderwijs en de toetsing op de hogescholen. Ten tweede vraagt een dergelijke toets om fine-tuning waar het gaat om de moeilijkheidsgraad, de omvang, de tijdsduur en de timing in de opleiding. Voor de LKT wiskunde is op deze punten al een aantal maatregelen genomen, zoals het beperken van het aantal items, die mogelijk de resultaten ten goede komen. Daarnaast zijn ook de inbedding in de opleidingen, de aansluiting van LKT op de toetspraktijken op de opleidingen, het digitale platform, en de voorbereiding van de student op de kennistoets cruciaal. Samengevat bieden de hierboven geformuleerde conclusies en aanbevelingen handvatten voor het verbeteren van het functioneren van landelijke kennistoetsen bij andere vakken en voor de borging van het kennisniveau van de lerarenopleiding in het algemeen.

**LKT vraagt een breed draagvlak, ontwikkelingen binnen de hogescholen en een fine-tuning van de toetsen zelf.**

## Dankwoord

Dank gaat uit naar *10voordeleraar* met name naar Arian van Staa en Monika Vaheoja en naar de opleiders van de verschillende lerarenopleidingen voor het beschikbaar stellen van de gegevens en de inhoudelijke bijdragen aan dit artikel.

## Referenties

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing* (2014 ed.). Washington: American Psychological Association.
- Cito (2015a). *Prestaties op papieren en digitale examens: wat is het verschil? Verslag van een literatuurstudie*. Eindrapport, mei 2015. Arnhem: Stichting Cito. [https://www.hetcvte.nl/item/overige\\_publicaties](https://www.hetcvte.nl/item/overige_publicaties)
- Cito (2015b). *De mogelijkheden en consequenties van het werken met deelscores*. Eindrapport, mei 2015. Arnhem: Stichting Cito. [https://www.hetcvte.nl/item/overige\\_publicaties](https://www.hetcvte.nl/item/overige_publicaties)
- Drijvers, P., Ball, L., Barzel, B., Heid, M.K., Cao, Y., & Maschietto, M. (2016). *Uses of technology in lower secondary mathematics education; A concise topical survey* (down te loaden op <http://www.springer.com/us/book/9783319336657>). New York: Springer.
- HBO-raad /Vereniging van Hogescholen (2009). *Kennisbasis docent wiskunde bachelor*. <https://www.10voordeleraar.nl/toetsen/oefenmateriaal>
- Kane, M.T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-
- 64). Westport: American Council on Education and Praeger Publishers.
- Kane, M.T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73.
- Stacey, K., & Wiliam, D. (2013). Technology and assessment in mathematics. In M. A. Clements, A. Bishop, C. Keitel, J. Kilpatrick, & F. Leung (Eds.), *Third International Handbook of Mathematics Education* (pp. 721-751). New York: Springer.
- Wools, S. (2015). *All About Validity. An evaluation system for the quality of educational assessment*. Dissertation. Enschede: Universiteit Twente.
- Wools, S., Eggen, T., & Sanders, P. (2010). Evaluation of validity and validation by means of the argument-based approach. *CADMO*, 8, 63-82.