

De Script Concordance Test als toetsmethode voor het onderwijskundig domein

■ Tom Adams

Tom Adams is docent bij Fontys Lerarenopleiding Tilburg/Alfa Cluster
Email: t.adams@fontys.nl

De kennis, vaardigheden en inzichten die de aanstaande leraar moet kennen en kunnen staat beschreven in een generieke kennisbasis. Een onderdeel hiervan is het kunnen kiezen van de juiste handelingen in een klaslokaal. Voor toetsontwikkelaars is het een uitdaging om passende en efficiënte toetsing te ontwerpen die hierop aansluiten. In een pilot is gekeken of de Script Concordance Test geschikt is waarbij deze toets vergeleken is met de tot dan toe gebruikte meerkeuzetoets. Dit onderzoek laat zien dat er potentie zit in deze wijze van toetsing.

Inleiding

Het opdoen en beheersen van onderwijskundige kennis, vaardigheden en inzichten is voor de aanstaande leraar een complexe leertaak. Het veronderstelt een mate van begrip van theoretische concepten die zich vooral in de praktijk van het klaslokaal tentoonspreiden. Bovendien vereist het van de aanstaande leraar het vermogen om de juiste handelingsafweging te maken om in de context toe te passen.

De inhoud van het onderwijskundig curriculum van de aanstaande leraar staat beschreven in

de generieke kennisbasis tweedegraads lerarenopleidingen (HBO raad, 2011). De generieke kennisbasis behelst het geheel aan theoretische inhoud dat de startbekwame leraar aan het einde van de opleiding moet kennen en beheersen en is geldig voor alle Nederlandse lerarenopleidingen. Op de Fontys Lerarenopleiding Tilburg (FLOT) wordt het onderwijskundig curriculum in het eerste en tweede jaar van alle opleidingen via het vak Algemene Professionele Vorming (APV) aangeboden. De laatste toets van het tweede jaar, APV 9, is de eindtoets van het onderwijskundig curriculum en omvat alle elementen van de generieke kennisbasis die eerder in voorgaande modules aan bod is gekomen. Aangezien het een groot aantal studenten betreft is er voor gekozen om de generieke kennisbasis te toetsen met meerkeuzevragen. De vragen in deze toets waren vooral gericht op de lagere orde cognitieve vaardigheden (kennis, inzicht) (Anderson & Krathwohl, 2001). De generieke kennisbasis veronderstelt echter dat studenten op basis van hun kennis moeten kunnen redeneren om te komen tot een juiste handeling in een specifieke context. Dit aspect laat zich slecht meten met een meerkeuzetoets. Het is een uitdaging om na te denken hoe de hogere orde kennisniveaus valide én efficiënt te toetsen zijn. Een blik over de grenzen van ons eigen domein heen, bracht ons bij het beoordelen van het klinisch redeneren. Om te komen tot een goede diagnose moeten artsen en verpleegkundigen hun medische kennis en inzicht

koppelen aan observaties bij een patiënt. Dit klinisch redeneren lijkt op het onderwijskundig redeneren van docenten in een klassensituatie: Om te komen tot een goede handeling in de klas moet de docent zijn kennis en inzicht koppelen aan zijn observaties in de klas. De vraag is dan ook of de toetsvorm die in het medisch domein gebruikt wordt voor klinisch redeneren, de Script Concordance Test (SCT), ook geschikt is voor het onderwijskundig domein. In dit artikel presenteren we hoe de SCT is ingezet en wat de bevindingen zijn van het gebruik van deze toets in het onderwijskundig domein.

De Script Concordance Test

De Script Concordance Test (SCT) is ontworpen om het klinisch redeneren in de keuze naar de best passende diagnose te meten. De SCT wordt voornamelijk binnen het domein van het medisch onderwijs gebruikt en is vooral

gericht op het besluitvormingsproces in een onzekere context. Dit gebeurt aan de hand van een beschrijving van een authentieke situatie. Op basis van die beschrijving moet de student gegevens interpreteren en een beslissing nemen. De beslissing van de student wordt vergeleken met de beslissing die experts in dezelfde situatie zouden hebben genomen (Charlin, Tardiff, & Boshuizen, 2000; Charlin & Van der Vleuten, 2004).

De vragen in een SCT bestaan uit casussen uit de praktijk, met relatief vage informatie waarmee een expert een weloverwogen hypothese kan formuleren. Een hypothese wordt gepresenteerd en de student moet daarbij aangeven hoe waarschijnlijk deze hypothese is.

Daarbij is er niet een antwoord goed of fout, maar zijn er meerdere antwoordalternatieven verdedigbaar. In de praktijk zijn er immers ook

Een voorbeeld vanuit het medisch domein, gemaakt door de KU Leuven:

Casus: Je ziet een 60-jarige man op de spoedgevallendienst met een linkszijdige verlamming

Aanname: Je denkt hierbij aan een hersenabces.

Aanvullende informatie:

Je verneemt dat de patiënt 10 dagen geleden bij de tandarts een behandeling heeft gekregen.

Beïnvloeding hypothese:

Je veronderstelt dat er bij deze patiënt sprake is van een hersenabces wordt hiermee:

-2	-1	0	+1	+2
<i>Uitgesloten of bijna uitgesloten</i>	<i>Minder waarschijnlijk</i>	<i>Niet minder of niet meer waarschijnlijk</i>	<i>Meer waarschijnlijk</i>	<i>Zeker of bijna zeker juist</i>

meerdere oplossingen mogelijk die elk acceptabel en verdedigbaar kunnen zijn. Soms is er in de literatuur ook geen consensus over de te volgen overwogen strategie of handelswijze. De test beoordeelt didactische, pedagogische en/of onderwijskundige redeneringen die plaats kunnen vinden in werkelijke klassensituaties. De kwaliteit van het antwoord van de student wordt door de beoordelaar afgezet tegen het antwoord van een expertpanel. De score per antwoord wordt bepaald door de overeenkomst (i.e. concordance) tussen de leden van een expertpanel.

Om inzicht te krijgen of de SCT-vragen een effectieve bijdrage zouden kunnen leveren aan het huidig onderwijskundig toetskader zijn een aantal SCT-vragen opgenomen naast de standaard meerkeuzevragen. Het tentamen bestond uit 5 SCT-vragen en 25 meerkeuzevragen. Alle vragen zijn gebaseerd op de generieke kennisbasis.

Onderzoeksvraag

De onderzoeksvraag 'In hoeverre draagt de SCT bij aan de validiteit en betrouwbaarheid van de APV 9 toets?' is beantwoord aan de hand van de volgende vier deelvragen:

1. Gelden de theoretische voordelen, zoals toegepast in medische toetsen, ook voor het onderwijskundige domein?
2. Is er een relatie tussen de score van de SCT-vragen en de score op de meerkeuzevragen? Het is de verwachting (hypothese) dat studenten die goed scoren op de meerkeuzevragen ook goed scoren op de SCT-vragen, aangezien voor beide vraagvormen een gedegen kennisbasis noodzakelijk is. Aangezien de SCT-vragen echter hogere orde cognitieve vaardigheden meet, verwachten we dat hier gemiddeld lager op wordt gescoord dan de meerkeuzevragen.
3. Is er een verschil in de score tussen mannelijke en vrouwelijke studenten? Omdat we vermoeden dat talige intelligentie een

mogelijke rol speelt in het interpreteren en redeneren, verwachten we dat vrouwelijke studenten hier mogelijk hoger op scoren dan mannelijke studenten, omdat we verwachten dat de SCT vragen taliger en daarmee voor de studenten moeilijker te beantwoorden zijn dan meerkeuzevragen. Deze verwachting zou aansluiten bij eerder gevalideerd onderzoek op dergelijke intelligentieonderzoeksterreinen (Wechsler, 1958; Raiford & Coalson, 2014).

Omdat vooraf onduidelijk was wat het effect was van deze vraagvorm op de resultaten van de studenten en om inzicht te krijgen in de wijze waarop de studenten het argumentatieproces van de SCT-vraag aanpakken is naast de reguliere scoremogelijkheden van de SCT ook gevraagd naar een toelichting. Bij elke SCT-vraag werd de student gevraagd om zijn of haar antwoord toe te lichten. In deze toelichting maakt de student de koppeling tussen de casus en de theorie. Dit leidde tot de vierde onderzoeksvraag:

4. Kan de tekstuele toelichting weg worden gelaten waardoor de toets meer efficiënt kan worden aangeboden?

Methode

De SCT-vraag vertrekt vanuit een casus waarin een onderwijskundige of pedagogische situatie centraal staat. Hierna volgt een hypothetische aanname die de student een bepaalde handelswijze of denkrichting op wijst. Vervolgens volgt er nieuwe informatie waarna het de vraag is hoe dat de aanvankelijke hypothese beïnvloedt. De student geeft een indicatie, op een 5-punts Likert-type schaal (variërend van -2 tot +2) van het effect dat hij of zij denkt dat de nieuwe informatie heeft op de aanvankelijke hypothese (Lubarsky, Charlin, Cook, Chalk & Van der Vleuten, 2011).

Als voorbeeld hieronder vraag 1 uit het tentamen:

Casus: Je geeft alle leerlingen van een klas een herhalingsoefening ter voorbereiding op een toets.

Aanname: Je denkt dat de oefening de leerlingen uitdaagt op het niveau van Vygotsky's 'zone van naaste ontwikkeling'.

Aanvullende informatie:

Je verneemt dat alle leerlingen de oefening gemakkelijk in staat waren om alle vragen goed te beantwoorden.

Beïnvloeding hypothese:

Je veronderstelt dat de leerlingen werden uitgedaagd in Vygotsky's 'zone van naaste ontwikkeling' wordt daarmee.

-2 <i>Uitgesloten of bijna uitgesloten</i>	-1 <i>Minder waarschijnlijk</i>	0 <i>Niet minder of niet meer waarschijnlijk</i>	+1 <i>Meer waarschijnlijk</i>	+2 <i>Zeker of bijna zeker juist</i>
---	--	---	--------------------------------------	---

Participanten

In totaal hebben 230 studenten het tentamen gemaakt; 104 mannen en 126 vrouwen. De populatie bestond uit tweedejaars studenten van de lerarenopleidingen Aardrijkskunde (13), Bèta (Biologie, Natuurkunde, Scheikunde; 38), Duits (11), Economie (13), Engels (43), Frans / Spaans (14), Geschiedenis (30), Gezondheidszorg en welzijn (14), Levensbeschouwing (6), Maatschappijleer (14), Nederlands (19) en Wiskunde (14).

Materialen en procedure

De studenten zijn in een extra college voorbereid op de nieuwe wijze van toetsing. Zij hebben een proeftentamen gemaakt zodat zij vertrouwd raakten met de stijl van vragen. Het tentamen is ontworpen in een werkgroep van 6 onderwijskundigen / docenten APV en de lector Kwaliteit van toetsen en beoordelen van FLOT. Bij het valideren hebben zij als experts de antwoorden gemaakt waarop de

SCT-scores zijn gebaseerd.. De makers zijn een brede afvaardiging vanuit het docententeam en lectoraat binnen het instituut.

Per meerkeuzevraag kon 1 punt verdiend worden en per SCT vraag 6 punten; maximaal 1,5 punt (en afhankelijk van aantal scores op alternatief door expertpanel ook variërende scores op 1, 0,5 of 0 punten) voor een antwoord dat overeenkomt met de meest voorkomende score bij de experts en maximaal 4,5 punten voor de toelichting (maximaal 1,5 punt voor de koppeling met de casus en max. 3 punten voor de theoretische onderbouwing).

Data-analyse

De data van de 230 tentamens zijn na afname eerst kwantitatief geanalyseerd. Voor de analyses is gekeken naar de gemiddelde moeilijkheid van de meerkeuzevragen en zijn alleen de eerste twee SCT-vragen meegenomen in de analyse.

De gemiddelde moeilijkheid van de meerkeuzevragen was .62 (sd = .10). De SCT-items scoorden respectievelijk SCT-vraag 1 (gem. score = .41; sd = .26) en SCT-vraag 2 (gemiddelde score = .46; sd = .30).

Resultaten en conclusie

In dit onderzoek stond de volgende onderzoeksvraag centraal: 'In hoeverre draagt de SCT bij aan de validiteit en betrouwbaarheid van de APV 9 toets?'.

1. Gelden de theoretische voordelen, zoals toegepast in medische toetsen, ook voor onderwijskundige toetsing?

Deelvraag 1 is beantwoord door een kwalitatieve analyse van de items. Was het docententeam in staat items te ontwerpen die passen binnen de principes van de SCT-vraag? Voor dit onderzoek heeft het expert panel bestaande uit 7 onderwijskundigen onafhankelijk van elkaar de SCT vragen gescoord. Deze expert-antwoorden zijn naar de onderzoekers gestuurd die de gegevens hebben samengevoegd. In Figuur 1 staat de verdeling van de expert-antwoorden voor de eerste 2 SCT-vragen.

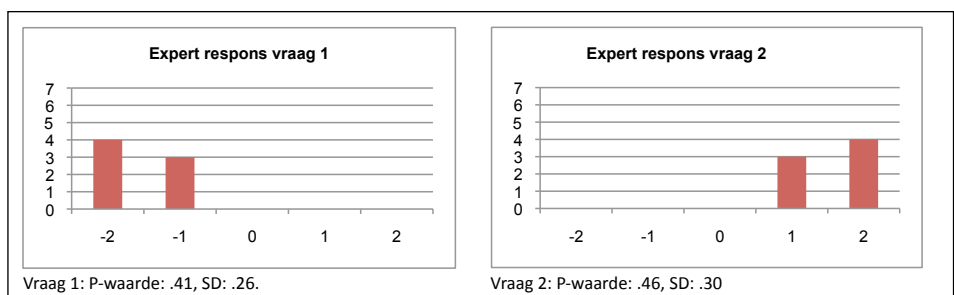
De gemiddelde scores liggen hoger bij de meerkeuzevragen dan bij de SCT-vragen. De SCT-vragen worden dus als moeilijker ervaren door de studenten dan de meerkeuzevragen. Wanneer we dit in het licht zetten van ons aanvankelijke doel, het beogen van het meten

van hogere kennisniveaus, zijn we mogelijk in die opzet geslaagd. De SCT-vragen zijn moeilijker waardoor het van de student een andere cognitieve capaciteit vereist in het beantwoorden van deze vragen. Het opstellen van goede SCT-vragen was echter niet eenvoudig. Inhoudelijke analyses achteraf door collega's bevestigden dat de items een hoger beheersingsniveau meten. Dit laatste impliceert dat de theoretische voordelen in het klinische domein in deze situatie ook gelden voor het onderwijskundige domein.

2. Is er een relatie tussen de score van de SCT vragen en de score op de meerkeuzevragen?

Deelvraag 2 is beantwoord door te kijken naar de correlatie tussen meerkeuzevragen en de SCT-items. De score op de meerkeuzevragen correleert significant positief met de gemiddelde scores op de SCT-items ($r = .28$, $p < 0.00$). De gepaarde t-toets laat zien dat de gemiddelde score op de meerkeuzevragen significant hoger is dan de gemiddelde score op respectievelijk SCT-vraag 1 ($t(229) = 11,44$, $p < .01$) en SCT-vraag 2 ($t(226) = 8,20$, $p < .01$). De SCT-vragen zijn dus moeilijker dan (het gemiddelde van) de meerkeuzevragen.

De verwachting dat studenten die goed scoren op de meerkeuzevragen ook goed scoren op de SCT-vragen werd bevestigd. Dit betekent dat de consistentie vergelijkbaar is tussen de SCT-vraag en de meerkeuzevragen. Er werden



Figuur 1. Variabele respons van de verschillende experts met betrekking tot de SCT vragen.

conform de verwachting lagere scores op de SCT-vragen behaald dan op de meerkeuzevragen.

3. Is er een verschil in de score tussen mannelijke en vrouwelijke studenten?

De gemiddelde toetsresultaten van de meerkeuze- en SCT-vragen zijn verdeeld op geslacht (zie tabel 1). Bij de meerkeuzevragen is er geen significant verschil tussen de gemiddelde scores van de mannelijke ($M = .61$; $sd = .10$) en vrouwelijke studenten ($M = .63$; $sd = .11$). Bij de SCT vragen is er onderling wel een significant verschil tussen de gemiddelde score van de mannelijke ($M = .62$; $sd = .16$) en vrouwelijke studenten ($M = .70$; $sd = .18$). Dit verschil is significant.

Dit verschil is mogelijk te verklaren doordat vrouwelijke studenten op het terrein van talige intelligentie beter scoren op het interpreteren en beredeneren van informatie (Wechsler, 1958; Raiford & Coalson, 2014).

4. Kan de tekstuele toelichting weg worden gelaten waardoor de toets meer efficiënt kan worden aangeboden?

Deelvraag 4 is beantwoord met een inhoudelijke analyse van de gegeven toelichtingen om meer inzicht te krijgen in met name de antwoorden die foutief zijn gerekend. Daarin is gekeken naar alternatieve, wellicht zelfs verdedigbare alternatieven die de partici-



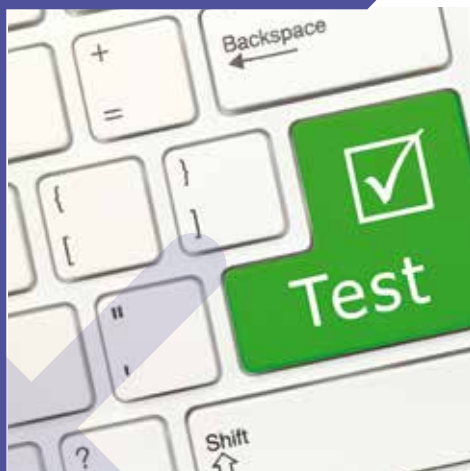
Opdoen en beheersen van onderwijskundige kennis, vaardigheden en inzichten is voor aanstaande leraren een complexe leertaak

panten hebben aangedragen. Ook is gezocht naar patronen in het beargumenteren van de SCT-vragen.

Voor het beantwoorden van de vraag of de toelichting weggelaten kan worden zodat de toets meer efficiënt kan worden aangeboden

Tabel 1. P-waarden van meerkeuze- en SCT toetsscores verdeeld op basis van geslacht

Geslacht	N	P-Waarde MK-vragen	Standaard- deviatie	P-Waarde SCT	Standaard- deviatie
Man	104	.61	.10	.62	.16
Vrouw	126	.63	.11	.70	.18
Totaal	230	.62	.10	.66	.18



De Script Concordance Test is ontworpen om het klinisch redeneren in de keuze naar de best passende diagnose te meten

zijn 50% van de gemaakte tentamens van de verschillende studentgroepen kwalitatief geanalyseerd. Concreet is er per vraag gekeken naar welke alternatieve, eventueel zelfs verdeigbare, antwoorden de student geformuleerd heeft. Specifiek is gekeken naar items waarbij studenten wel een SCT-score kregen en geen punten voor de inhoudelijke toelichting en andersom, waarbij studenten dus geen punten kregen voor de SCT-score, maar wel voor de inhoudelijke toelichting. Beide situaties kwamen even vaak voor.

Vervolgens is gekeken naar de alternatieve toelichtingen. Daarbij viel een aantal aspecten op:

- Het merendeel van de studenten met een SCT-score, geeft ook een juiste inhoudelij-

ke toelichting. Het merendeel ging daarin helder en voldoende onderbouwd in op de casus en de relevante theorie.

- Sommige studenten van de studenten met een SCT-score beschreven geen of een vage relatie met de theorie of pasten deze verkeerd toe. Daarbij werd het niet helder of de student de kwantitatieve keuze had gegokt of deze niet specifiek kon beschrijven.
- Sommige studenten gaven geen of een onbegrijpelijke toelichting. Soms vervielen ze in drogredensaties waarbij er feiten uit de casus in eigen woorden werden herhaald. Een duidelijke theoretische toelichting ontbrak veelal.

Sommige studenten betrokken casus op zichzelf en gaven een alternatieve oplossing wat zij in een dergelijk geval zouden doen. Daarbij werd er vaak geen koppeling gemaakt naar theoretische concepten en was de kern van de redenering onhelder. Het zoeken naar verdeigbare, alternatieve antwoordmogelijkheden leverde inhoudelijk geen nieuwe of andere visies op. Daarmee kunnen we concluderen dat het antwoordmodel van het expertpanel voldoende volledig was. Omdat de ruimte voor het geven van een kwalitatief antwoord vrij groot was bood dit studenten de mogelijkheid tot het geven van lange antwoorden die inhoudelijk weinig meer opleverde dan de SCT-antwoordmogelijkheid. De tekstuele toelichting kan dan ook op termijn worden weggelaten.

In het onderzoek naar het ontwerpen van betrouwbare en valide SCT-vragen binnen het onderwijskundig domein kan de tekstuele toelichting echter wel de bevestiging geven dat de student de keuze van de vraag met de passende theorie onderbouwt. Omdat dit ook korter kan is het voor een volgend tentamen wellicht aan te raden de ruimte voor de toelichting te kaderen tot enkele woorden zodat de student het antwoord in enkele sleutelbegrippen kan onderbouwen.

Samengevat concluderen we dat de SCT-vraag ook voor het onderwijskundig domein interessant is en valide hogere orde vaardigheden kan meten. Het goed formuleren van de vragen vraagt wel extra scholing voor de docenten. Het vragen om een extra toelichting op het antwoord levert geen aanvullende informatie op over het vaardigheidsniveau van de studenten. De resultaten lijken erop te wijzen dat dit in de toekomst weggelaten kan worden. ■

Referenties

- Anderson, L. W. & Krathwohl, D. R. (Eds.) (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives (Complete edition). New York: Longman.
- Charlin, B., Tardif, J. & Boshuizen, H.P.A. (2000). Scripts and medical diagnostic knowledge: Theory and applications for clinical reasoning instruction and research. *Academic Medicine* 75: 182–90.
- Charlin, B. & Van der Vleuten, C.P.M. (2004). Standardized assessment of reasoning in concordance approach. *Evaluation and the Health Professions* 27, no. 3: 304–19.
- HBO raad. (2011). Generieke kennisbasis tweedegraads lerarenopleidingen. Den Haag: HBO raad, vereniging van hogescholen.
- KU Leuven. (z.d.). Oefenexamen Script Concordance Test. Geraadpleegd op 4 december 2015, van <http://docslide.nl/documents/oefenles-examen-bloedsomloop-script-concordance-test-inleiding.html>
- Lubarsky S., Charlin B., Cook D.A., Chalk, C. & van der Vleuten C. (2011). Script concordance testing: A review of published validity evidence. *Medical Education*, 45, 329-338.
- Raiford, S. E., & Coalson, D. L. (2014). *Essentials of WPPSI-IV Assessment*. New Jersey: Wiley & Sons Inc.
- Wechsler, D. (1958). *The Measurement and Appraisal of Adult Intelligence* (4th ed.). Baltimore (MD): Williams & Witkins. Geraadpleegd op 5 november 2015, van <https://archive.org/details/measurementandap001570mbp>

Gesignaleerd

Chinees officieel examenvak

Al jaren kan op een aantal scholen het vak Chinees gevolgd worden. Vanaf schooljaar 2017-2018 wordt het ook mogelijk om het vak Chinees op de cijferlijst te krijgen. Vooralsnog zal het vak worden afgesloten met een schoolexamen. (nieuws.eindexamens.nu 15 december 2015)