Ontwikkelen probleemoplossingsgerichte klinische colleges het diagnostisch denken en blijkt dit uit het extended-matching examen?¹

J. Beullens · B. van Damme · E. Struyf

Samenvatting Inleiding: Aan de K.U. Leuven besteden laatstejaars studenten geneeskunde twee maanden aan voorbereiding van en deelname aan een 70-tal probleemoplossingsgerichte klinische colleges (POKC's). Het klinisch redeneren staat daarbij centraal, met het accent op het stellen van differentiële diagnosen. De POKC's bereiden voor op een examen dat bestaat uit 200 extended-matching (EM) vragen. In dit onderzoek wordt nagegaan of deze POKC's het klinisch redeneren bevorderen en of in het EM-examen klinische beslissingen te meten zijn. Om een mogelijke toename van diagnostische bekwaamheid te meten, werd gebruik gemaakt van de Vragenlijst Diagnostisch Denken (VDD), een vertaling van de Diagnostic Thinking Inventory (DTI) van Bordage, Grant en Marsden. De onderzoekshypothesen zijn: (1) de VDD-scores zijn na de POKC's significant hoger dan daarvoor en (2) de correlaties tussen VDD-scores en EM-examenuitslagen zijn significant.

Methode: In het academisch jaar 2003-2004 werden drie reeksen van POKC's georganiseerd. Zowel in het begin als aan het einde van elke reeks vulden de studenten de VDD in. Deze vragenlijst meet twee aspecten van het diagnostisch denken: de graad van flexibiliteit in het denken en de graad van structurering van de kennis in het geheugen.

Resultaten: De VDD-scores waren na de POKC's significant hoger dan daarvoor. De Pearson correlaties tussen VDD-scores en EM-examenuitslagen waren laag maar significant, met uitzondering voor Structuur in de nameting (niet significant).

Discussie en conclusie: Twee maanden van intensieve probleemoplossingsgerichte klinische training leidt tot een verbetering in het diagnostisch denken, zoals gemeten met de VDD. De correlatie tussen VDD-scores en examenuitslagen toont aan dat via EM-vragen de

bekwaamheid van het klinisch redeneren te meten is. (Beullens J, Damme B van, Struyf E. Ontwikkelen probleemoplossingsgerichte klinische colleges het diagnostisch denken en blijkt dit uit het extended-matching examen? Tijdschrift voor Medisch Onderwijs 2007;26 (3):118–127.)

Inleiding

In het laatste jaar van de medische opleiding (vóór het artsexamen) aan de Katholieke Universiteit Leuven (K. U. Leuven) worden gedurende acht weken een 70-tal probleemoplossingsgerichte klinische colleges (POKC's) gegeven, met als doel het klinisch redeneren te oefenen.¹⁻³ Vóór elk college worden één of meer realistische casussen in kleine groep bestudeerd en worden de bijhorende opdrachten uitgevoerd. Individuele bestudering van de literatuur is vaak noodzakelijk. Deze zelfstudie en groepswork vormen de voorbereiding op het plenaire college dat door één of twee docenten wordt begeleid. Gedurende deze interactieve sessies presenteren de studenten de antwoorden van hun groep en de denkpaden die ze gevolgd hebben om deze te bereiken. De studenten ondervragen elkaar en discussiëren over de voorgestelde oplossingen. In de plenaire sessie staat dus het hypothetischdeductief redeneerproces centraal dat leidt naar verschillende plausibele oplossingen. Elke reeks POKC's wordt beëindigd met een examen dat bestaat uit 200

¹ Dit artikel is eerder gepubliceerd in het Engelstalige tijdschrift Medical Education: Beullens J, Struyf E, Van Damme B. Diagnostic ability in relation to clinical seminars and extended-matching examination. Med Educ 2006;40:1173-9.

Thema:	Dubbelzien
Opties:	a. internucleaire oftalmoplegie b. monoc-ulaire diplopie c. myasthenia gravis d. myositis e. syndroom van de sinus cavemosus f. thyroïd oftalmopathie g. verlamming van de M. obliquus superior h. verlamming van de M. rectus externus i. verlamming van de N. oculomotorius j. verlamming van de N. trochlearis beiderzijds
Opdracht:	Selecteer de meest waarschijnlijke diagnose voor deze diplopie.
Ziektegeschiedenissen:	1. Een man van 52 jaar heeft sedert maanden last van dubbelzicht. Hij vertoont een licht vernauwde oogspleet links en een grotere pupil links. Tevens zijn adductie, abductie, elevatie en depressie beperkt aan dit oog. Er is ook een verminderde corneasensibiliteit. 2. Een man van 54 jaar heeft plots diplopie ontwikkeld. Hij heeft vooral last bij het afdalen van trappen en bij het lezen met zijn bifocale bril. Hij heeft de neiging het hoofd wat naar de linkerschouder te buigen. Hij lijdt aan diabetes.

extended-matching (EM) meerkeuzevragen. Tijdens de colleges worden voorbeelden van vragen besproken.

Een extended-matching set omvat vier componenten:

1. Een thema.
2. Een lijst met antwoordmogelijkheden.
3. Een inleidende bewering of vraag.
4. Ten minste twee itemstammen of casussen (zie box 1).⁴

Box 1. Een voorbeeld van een extended-matching set.

In dit onderzoek werd een variant gebruikt met slechts één itemstam en 7 à 26 antwoordmogelijkheden. Extended-matching vragen (EM-vragen) vormen een betrouwbare en valide toetsvorm.⁵⁻⁹ Wass en co-auteurs toonden aan dat EM-vragen het klinisch probleemoplossen meten omdat ze hoog correleerden met klinische toetsen (stationsproef en lange casussen) en probleemoplossingstoetsen (kort-antwoordvragen), matig met een feitenkennistoets (juist-fout vragen) en laag met een schriftelijke presentatie (essay).⁸ Coderre en medewerkers vroegen 20 gastroenterologen (experts) en 20 laatstejaars studenten geneeskunde (niet-experts) om vier meerkeuze- en vier EM-vragen over gastrointestinale problemen op te lossen. Na beantwoording van de acht vragen legden de respondenten uit hoe zij tot de diagnose zijn gekomen. Bij het oplossen van de EM-vragen rapporteerden de nietexperts meer een beroep te doen op hypothetisch-deductief redeneren (achterwaarts redeneren) en minder op patroonherkenning (voorwaarts redeneren). Bij het hypothetisch deductief redeneren worden, voordat de meest waarschijnlijke diagnose geselecteerd wordt, één voor één de antwoordalternatieven geanalyseerd. In patroonherkenning is er sprake van een enkele diagnose met hoogstens een vluchtige blik op de alternatieven.⁹ Beullens et al. onderzochten het redeneren van 20 laatstejaars studenten en 20 assistenten die zich specialiseerden in inwendige geneeskunde bij het hardop

oplossen van 20 EM-vragen. Tien EM-vragen hadden betrekking op interne geneeskunde en tien op andere medische disciplines. Binnen hun domein van expertise (inwendige geneeskunde) hanteerden de assistenten vaak het voorwaarts redeneren, terwijl ze buiten hun domein van expertise (andere medische disciplines) meestal overschakelden op het achterwaarts redeneren. Vergeleken met de studenten redeneerden zij bij het oplossen van de vragen uit inwendige geneeskunde meer voorwaarts en minder achterwaarts.¹⁰ Dit leidt tot de conclusie dat EM-vragen het klinisch redeneren meten hetgeen tot uitgangspunt komt in het nemen van klinische beslissingen.

Voortbouwend op het onderzoek naar de evolutie van het klinisch redeneren van novice tot expert hebben Bordage, Grant en Marsden (1990) een vragenlijst ontworpen die het klinisch redeneren in kaart brengt: de Diagnostic Thinking Inventory (DTI).¹¹ Twee aspecten van het diagnostisch denken worden gemeten: de flexibiliteit in het denken en de structuur van de kennis in het geheugen. De Vragenlijst Diagnostisch Denken (VDD) is een letterlijke vertaling van de DTI. Bordage, Grant en Marsden vergeleken de scores op de DTI van twee groepen studenten geneeskunde met zeven groepen ervaren klinici. Ze vonden slechts twee significante verschillen. De eerste- en derdejaars studenten hadden lagere scores dan de overige groepen. Binnen deze laatste zeven groepen verschilden slechts twee groepen van elkaar. Dit betekent dat alleen verschillen werden gevonden tussen groepen die tenminste één trainingsjaar van elkaar verschilden.¹¹ Round liet de DTI invullen door vierdejaars studenten geneeskunde die al (experimentele groep) dan niet (controlegroep) deelnamen aan een interventie van twee uur rond klinisch redeneren.¹² De interventiegroep scoorde significant hoger op de DTI. De DTI was echter vóór de interventie niet afgenomen. Het is dus niet uitgesloten dat de experimentele groep initieel reeds een hoger redeneerniveau had. Indien de POKC's het

klinisch redeneren aanscherpen, verwachten wij een toename van de score op de VDD nadien in vergelijking met voordien.

Sobral onderzocht bij studenten die een stage inwendige geneeskunde liepen het verband tussen klinisch redeneren en kennis.¹³ De DTI-score correleerde 0.19 met de examenuitslag en 0.34 met de score op een probleemoplossingstoets. Deze coëfficiënten waren weliswaar zwak maar significant. Het verband was sterker met de probleemoplossingstoets dan met de examenuitslag en gold meer voor Structuur dan voor Flexibiliteit.¹³ Als het EM-examen het klinisch redeneren meet, dan verwachten wij een verband tussen VDD-scores en examenuitslag.

De onderzoekshypothesen zijn dus:

1. De VDD-scores zijn significant hoger na de POKC's dan daarvoor.
2. De correlaties tussen VDD-scores en examenuitslagen zijn significant positief.

Methode

Deelnemers

In het academisch jaar 2003-2004 werden drie reeksen POKC's georganiseerd voor drie verschillende groepen studenten: in september – oktober, in november – december en in februari – maart. Elke reeks startte met een introductiesessie waarin de studenten een boek met ziektegeschiedenissen en de vragenlijst kregen. Hen werd verzocht de vragenlijst in te vullen en te retourneren. Tijdens een informatiesessie over het examen, die een week vóór het examen plaatsvond, werd dezelfde vragenlijst opnieuw uitgedeeld met het verzoek deze ingevuld in te leveren op de dag van het examen.

Instrument

De Diagnostic Thinking Inventory (DTI) bestaat uit 41 items: 21 items meten de graad van flexibiliteit in het denken en 20 items meten de graad van kennisstructuur in het geheugen. Het instrument is gebaseerd op onderzoek over de ontwikkeling van het klinisch redeneren van novice tot expert.¹⁴⁻¹⁷ De DTI is een betrouwbare en valide test. Bordage, Grant en Marsden vonden Cronbach alfa's van respectievelijk 0.72 voor Flexibiliteit, 0.74 voor Structuur en 0.83 voor de Totaalscore.¹¹

Elk item omvat een stam, twee vergezellende beweringen en een meetschaal in de vorm van zes hokjes. De respondent moet een kruisje zetten in het hokje dat het best zijn standpunt op het continuüm tussen de beweringen weergeeft.

Flexibiliteit betreft de mate waarin betekenissen of processen toegepast kunnen worden tijdens het diagnostisch proces. Voorbeelden van items zijn opgenomen in tabel 1.

Extended-matching examen

Een extended-matching set bestaat uit vier componenten: een thema of probleem, een lijst met opties of antwoordalternatieven, een inleidende bewering of vraag en itemstammen in de vorm van ziektegeschiedenissen.⁴ Om correctie voor het raden te vermijden is het minimum aantal opties gesteld op 7; het maximum aantal is 26. Voor eenzelfde thema, optielijst en vraag kunnen verschillende casussen aangeboden worden. Wij hebben echter vragen gebruikt met slechts één itemstam of casus. Hoewel ook andere aspecten aan bod komen zoals behandeling, beleid, onderzoek of preventie betreffen de meeste examenvragen de diagnose.

Statistische analyse

De VDD-score werd berekend door getallen van 1 tot 6 toe te kennen aan de hokjes. Voor de gehele vragenlijst en de subschalen Flexibiliteit en Structuur werden gemiddelden en standaarddeviaties berekend. Voordat begonnen werd met het analyseren van de data, werd de interne consistentie gemeten met behulp van de Cronbach alfa. Verschillen tussen voormeting en nameting werden getoetst met de Wilcoxon signed rank test. Het verband tussen VDD-scores en examenuitslagen werd nagegaan met de Pearson correlatiecoëfficiënt. Bij de beoordeling van verschillen werd het 5%-significantieniveau gebruikt.

Resultaten

De eerste stap bij de berekening van de examenuitslagen was een itemanalyse. Items die negatief correleerden met de totaalscore werden geschrapt. In de drie examens werden respectievelijk 18, 13 en 18 items geschrapt. De interne consistentie, gemeten met de Cronbach alfa, bedroeg 0.88 in de eerste reeks, 0.87 in de tweede en 0.82 in de derde reeks. Het gemiddelde percentage correcte antwoorden bedroeg respectievelijk 74.92 (met een SD van 8.14), 71.64 (SD 8.03) en 76.21 (SD 6.65). Aangezien de gemiddelden significant verschilden werden deze teruggebracht tot een gemeenschappelijk gemiddelde van 73%.

De respons was hoog. Aan de eerste reeks POKC's namen 110 studenten deel; 107 daarvan namen deel aan het extended-matchingexamen (97%), 109 aan de VDD-voormeting (99%) en 104 aan de nameting (95%). In de tweede reeks waren de aantallen respectievelijk: 120

Flexibiliteit

4. Wanneer ik een patiënt ondervraag, lijkt er vaak één idee in mijn hoofd te blijven hangen over wat fout gaat		vind ik het gewoonlijk gemakkelijk verschillende mogelijke diagnoses te exploreren
5. Tijdens het gehele interview, dreig ik mijn eigen rode draad te verliezen indien ik de gedachtegang van de patiënt volg		kan ik mijn eigen ideeën helder houden zelfs als ik de gedachtegang van de patiënt volg
6. Wanneer ik een diagnoseconclusie moet trekken, vind ik het niet erg mijn diagnostische beslissingen over een geval uit te stellen		voel ik mij verplicht voor één of andere diagnose te kiezen zelfs als ik niet erg zeker ben

Structuur

1. Wanneer de patiënt zich aanmeldt met zijn symptomen, denk ik aan de symptomen in de exacte bewoording van de patiënt		denk ik aan de symptomen in meer abstracte termen dan de werkelijk gebruikte uitdrukkingen (b.v. "duur van 4 dagen" wordt "acuut"; "twee handen" wordt "bilateraal")
13. Wanneer ik informatie verzamel over een patiënt, lijken de verschillende stukjes informatie zich gewoonlijk vanzelf te groeperen		vind ik het dikwijls moeilijk om in te zien hoe de stukjes informatie zich tot elkaar verhouden
18. Wanneer mij een mogelijke diagnose voor de geest komt, betrap ik mezelf erop vooruit te denken aan mogelijke abnormale tekenen en symptomen waarmee deze diagnose gepaard gaat		helpt dit mij zelden te beslissen wat vervolgens aan de patiënt te vragen

Structuur betreft de beschikbaarheid van de in het geheugen opgeslagen kennis tijdens het diagnostisch proces. Itemvoorbeelden zijn vermeld in tabel 1.

studenten in de colleges, 118 bij het examen (98%), 114 bij de voormeting (95%) en 116 bij de nameting (97%). In de kleinere derde reeks bedroegen de aantallen: 39 studenten in de POKC's, 39 legden het examen af (100%), 35 vulden de voormeting in (90%) en 36 de nameting (92%).

Betrouwbaarheid

De interne consistentie bedroeg 0.64 voor de subschaal Flexibiliteit, 0.71 voor de subschaal Structuur en 0.80 voor de gehele schaal.

Vergelijking voormeting – nameting

De scores van de drie afnames samen waren na de POKC's significant hoger dan voordien (zie tabel 2).

Bestudering van de afzonderlijke afnames liet zien dat de scores na de POKC's significant hoger waren dan voordien behalve voor Flexibiliteit in de eerste reeks (alleen een tendens) en in de tweede reeks (niet significant; zie tabel 2).

Correlatie met EM-examen

De Pearson correlaties tussen VDD-scores en examen-uitslagen waren aan de lage kant maar toch significant (met uitzondering van Structuur in de nameting) als de drie afnames samen worden genomen (zie tabel 3).

Bij elk van de afnames afzonderlijk waren de correlaties tussen VDD-scores en het examen laag tot matig (0.03–0.48). Ze waren in elke reeks wat hoger dan in de voorgaande: 0.03–0.12 in de eerste reeks, 0.15–0.25 in de tweede en 0.26–0.48 in de derde. Er waren geen

Tabel 2. Gemiddelde VDD-scores vóór en na de POKC's (in het totaal en de drie reeksen afzonderlijk).

	Voormeting	Nameting	P
In het totaal (n = 249)			
<i>Flexibiliteit 1</i>	85.89	87.00	<0.05
<i>Structuur 2</i>	82.24	85.10	<0.001
<i>Totaalscore 3</i>	168.13	172.10	<0.001
Eerste reeks (n = 104)			
<i>Flexibiliteit</i>	86.39	87.85	0.05 < P < 0.1
<i>Structuur</i>	82.66	85.35	<0.001
<i>Totaalscore</i>	169.06	173.19	<0.001
Tweede reeks (n = 112)			
<i>Flexibiliteit</i>	85.72	85.86	n.s.
<i>Structuur</i>	82.25	85.12	<0.001
<i>Totaalscore</i>	167.96	170.98	<0.005
Derde reeks (n = 33)			
<i>Flexibiliteit</i>	84.85	88.18	<0.05
<i>Structuur</i>	80.91	84.24	<0.05
<i>Totaalscore</i>	165.76	172.42	= 0.001

1 maximale score = 126; 2 maximale score = 120; 3 maximale score = 246.

Tabel 3. Correlaties tussen VDD-scores en examenuitslagen (in het totaal en de drie reeksen afzonderlijk).

	Voormeting	P	Nameting	P
In het totaal	(n = 239)		(n = 246)	
<i>Flexibiliteit</i>	0.14	<0.05	0.18	<0.05
<i>Structuur</i>	0.20	<0.01	0.13	n.s.
<i>Totaalscore</i>	0.19	<0.01	0.17	<0.05
Eerste reeks	(n = 100)		(n = 101)	
<i>Flexibiliteit</i>	0.09	n.s.	0.07	n.s.
<i>Structuur</i>	0.11	n.s.	0.03	n.s.
<i>Totaalscore</i>	0.12	n.s.	0.06	n.s.
Tweede reeks	(n = 108)		(n = 111)	
<i>Flexibiliteit</i>	0.15	n.s.	0.18	n.s.
<i>Structuur</i>	0.25	<0.05	0.15	n.s.
<i>Totaalscore</i>	0.22	<0.05	0.18	n.s.
Derde reeks	(n = 31)		(n = 34)	
<i>Flexibiliteit</i>	0.26	n.s.	0.48	<0.01
<i>Structuur</i>	0.30	n.s.	0.34	<0.05
<i>Totaalscore</i>	0.31	n.s.	0.45	<0.01

significante correlaties tussen de VDD-scores en de examenuitslag behalve die in de voormeting van de tweede reeks (uitgezonderd voor Flexibiliteit) en die in de naming van de derde reeks (zie tabel 3).

Discussie

De respons was uitstekend: 90% of meer. De betrouwbaarheid van de gehele schaal was hoog en de alfa's van de subschalen volstonden voor onderzoeksdoeleinden.

De voormetingscores in dit onderzoek (totaalscore 168.1, Flexibiliteit 85.9 en Structuur 82.2) waren hoger

dan die van de derdejaars studenten geneeskunde in de studie van Bordage, Grant en Marsden (totaalscore 158.3, Flexibiliteit 81.6 en Structuur 76.7).¹¹ Onze nametingscores (172.1, 87.0 en 85.1) waren gelijk aan die van huisartsen in opleiding in de studie van Bordage (172.2, 88.1 en 84.1).¹¹ Onze studenten, die allen zouden starten met een opleiding huisartsgeneeskunde of een andere medische specialistenopleiding, hadden even hoge scores als huisartsen in opleiding.

De eerste hypothese werd bevestigd: de meeste naming VDD-scores waren significant hoger dan die van de voormeting. Deelname aan de POKC's resulteerde tot een meer flexibel denken en een beter gestructureerde

kennis in het geheugen. Bordage en medewerkers vergeleken de DTI-scores van negen groepen van studenten en afgestudeerden en vonden slechts twee significante verschillen tussen groepen die minstens één studiejaar van elkaar verschilden.¹¹ Een significante toename van de scores na twee maanden in onderhavig onderzoek is bijgevolg relevant. De intensieve training tijdens de probleemoplossingsgerichte klinische colleges heeft hiertoe bijgedragen. Immers, gedurende acht weken volgden de studenten elke dag twee klinische colleges, waarin van de gepresenteerde casussen het klinisch redeneren, resulterend in plausibele oplossingen, werd besproken. Gezien het ontbreken van een controlegroep kan echter niet uitgesloten worden dat het effect een tijdseffect is.

Als elke reeks echter afzonderlijk wordt beschouwd, blijkt de toename alleen te gelden voor Structuur, niet voor Flexibiliteit. De subschaal Flexibiliteit heeft echter een lage betrouwbaarheid en dit zou de niet-significante resultaten kunnen verklaren. De POKC's vormen dus een efficiënte onderwijsleeractiviteit die op korte termijn de structuur van het geheugen kan verbeteren. Het verschil tussen voor-en nameting is echter niet spectaculair. Andere onderzoeken bevestigen dat kennisherstructurering plaats vindt tijdens de medische opleiding.¹⁸⁻¹⁹ Twee wijzen van herstructurering kunnen worden onderscheiden: 'kennisinkapseling' en 'ziektescrpt'-vorming. Volgens de kennisinkapselingshypothese bij de vorming van een medisch expert, verminderen de biomedische begrippen in het klinisch redeneren omdat deze omgezet worden in concepten van een hoger niveau, die wel toenemen. In het ziektescriptmodel is een script een netwerk van relevante kennis en ervaring, die 'enabling conditions' bevat (kenmerken verbonden met de verwerving van de aandoening), 'faults' (misfuncties) en 'consequences' (tekenen en symptomen). Boshuizen heeft aangetoond dat over een stageperiode van twee jaar de klinische ervaring van studenten geneeskunde resulteerde in een inkapseling van de gedetailleerde biomedische begrippen.¹⁸ Uit een onderzoek van Boshuizen en Schmidt blijkt dat vierdejaars studenten geneeskunde uitgewerkte biomedische kennis toepassen bij pogingen de diagnose te stellen bij een casus, terwijl vijfdejaars studenten op het einde van hun stageperiode hele ziektescripts activeren in plaats van biomedische concepten.¹⁹ In het vijfde jaar bevorderde de klinische ervaring dus de herstructurering van kennis in het geheugen. De resultaten in onderhavig onderzoek suggereren dat een uitgebreide bestudering van papieren casussen ook de herstructurering van de kennis initieert.

Het denken van de expert wordt ook gekenmerkt door flexibiliteit. Inderdaad, ziektescripts zijn dynamische structuren, die aangepast worden na elk onderhoud met een nieuwe patiënt.²⁰ In onderhavige studie vonden wij

geen toename van flexibiliteit binnen de reeks. Mogelijk hebben echte gesprekken met patiënten meer invloed op de flexibiliteit van het denken dan papieren casussen. Op basis van de onderzoeksresultaten kunnen wij ons de vraag stellen of de verwerving van ziektescripts een voorwaarde is voor de verwerving van flexibiliteit of omgekeerd. Meer onderzoek is vereist om deze vraag op te lossen. Het vergelijken van de VDD-scores in dit onderzoek met de scores van studenten met meer of minder klinische ervaring kan de rol van de flexibiliteit van het denken en van de kennisstructuur in de ontwikkeling van het klinisch redeneren onthullen.

Indien de scores over de drie reeksen colleges werden samengenomen, vonden wij dat de correlaties met een probleemoplossingsgericht EM-examen laag maar significant waren (0.13-0.20). Sobral vond eveneens significante maar zwakke correlaties van 0.14-0.21 tussen DTI-score en examenuitslag en van 0.27-0.35 met de score op een probleemoplossingstoets.¹³ De lage correlaties in beide onderzoeken suggereren dat de DTI-metingen veeleer perceptie van probleemoplossen betreffen dan puur probleemoplossen. Indien de scores van de drie collegereeksen echter afzonderlijk beschouwd werden, waren de correlaties tussen VDD en examen zelden significant. De correlaties waren enigszins hoger in elke reeks. De studenten in de laatste reeks hadden intussen hun klinisch co-assistentenschap volbracht. De toename van de correlaties zou dus een toename in klinische ervaring kunnen weerspiegelen. Over het geheel genomen werd de tweede hypothese eveneens bevestigd.

Dit onderzoek vertoont enkele methodologische tekorten. De retrospectieve perceptie van het proces van het stellen van een diagnose zou kunnen verschillen van het reële denkproces. Verder missen de VDD-scores een hoge betrouwbaarheid.

Conclusie

De bevindingen van de eerste hypothese tonen aan dat de probleemoplossingsgerichte klinische colleges gepaard gaan met een toename in VDD-scores. De resultaten van de tweede hypothese suggereren dat de EM-vragen een aspect van competentie meten dat verbonden is met het klinisch redeneren.

Belangenconflict: geen gemeld.

Financiële ondersteuning: geen gemeld.

Summary

Introduction: At the Katholieke Universiteit Leuven, Belgium, final year medical students participate in 70 problemsolving clinical seminars over a period of two months, concentrating on clinical reasoning for formulating differential diagnoses. The clinical seminars end in an examination consisting of 200 extended-matching questions (EMQs). This study asks whether problem-solving

clinical seminars improve clinical reasoning and whether the EMQ examination measures clinical decision making. In order to detect an increase in diagnostic ability, the Diagnostic Thinking Inventory (DTI) of Bordage and Grant and Marsden was applied. The research hypotheses were: (1) DTI scores will be higher after the seminars than before, and (2) the correlations between DTI scores and EMQ examination score will be significant.

Method: In the academic year 2003–2004 three series of problem-solving clinical seminars were held. At the beginning and end of each series the students filled in the DTI. This questionnaire measures two aspects of diagnostic thinking: the degree of flexibility in thinking, and how knowledge is structured in the memory.

Results: For all data together, the DTI scores after the clinical seminars were significantly higher than before. Pearson correlations between DTI scores and EMQ examination scores were low but significant, with the exception of posttest Structure (not significant).

Discussion and conclusion: Two months of intensive problem-solving clinical training was accompanied by an improvement in diagnostic thinking, as measured by DTI scores. Correlation between DTI scores and the examination indicates that the EMQs measure an aspect of student achievement that is related to clinical reasoning. (Beullens J, Van Damme B, Struyf E. Diagnostic ability in relation to clinical seminars and extended-matching questions examinations. *Dutch Journal of Medical Education* 2007;26 (3):118–127.)

Literatuur

- Beullens J, Struyf E, Jaspaert H, Aubry C, Janssen PJ, Damme B van. Evaluaties door studenten van een reeks probleemoplossingsgerichte klinische colleges in een traditioneel curriculum. *Tijdschrift voor Medisch Onderwijs* 2003;22:28–39.
- Beullens J, Struyf E, Jaspaert H, Damme B van, Aubry C, Janssen PJ. De evolutie in de waardering van studenten en docenten van probleemoplossingsgerichte klinische colleges. *Tijdschrift voor Medisch Onderwijs* 2004;23:80–90.
- Struyf E, Beullens J, Damme B van, Janssen P, Jaspaert H. A new methodology for teaching clinical reasoning skills: problem-solving clinical seminars. *Med Teacher* 2005;27:364–8.
- Case SM, Swanson DB. Constructing written test questions for the basic and clinical sciences. Philadelphia: National Board of Medical Examiners; 1996.
- Case SM, Swanson DB, Ripkey DR. Comparison of items in five-option and extended-matching formats for assessment of diagnostic skills. *Acad Med* 1994;69(10 Suppl):S1–S3.
- Fenderson BA, Damjanov I, Robeson MR, Veloski JJ, Rubin E. The virtues of extended matching and uncued tests as alternatives to multiple choice questions. *Hum Pathol* 1997;28:526–32.
- Kreiter CD, Ferguson K, Gruppen LD. Evaluating the usefulness of computerized adaptive testing for medical in-course assessment. *Acad Med* 1999;74:1125–8.
- Wass V, McGibbon D, Van der Vleuten C. Composite undergraduate clinical examinations: how should the components be combined to maximize reliability? *Med Educ* 2001;35:326–30.
- Coderre SP, Harasym P, Mandin H, Fick G. The impact of two multiple-choice question formats on the problem-solving strategies used by novices and experts. *BMC Med Educ* 2004;4:23.
- Beullens J, Struyf E, Van Damme B. Do extended matching multiple-choice questions measure clinical reasoning? *Med Educ* 2005;39:410–7.
- Bordage G, Grant J, Marsden P. Quantitative assessment of diagnostic ability. *Med Educ* 1990;24:413–25.
- Round AP. Teaching clinical reasoning: a preliminary controlled study. *Med Educ* 1999;33:480–3.
- Sobral DT. Diagnostic ability of medical students in relation to their learning characteristics and preclinical background. *Med Educ* 1995;29:278–82.
- Grant J, Marsden P. The structure of memorized knowledge in students and clinicians: an explanation for diagnostic expertise. *Med Educ* 1987; 21:92–8.
- Grant J, Marsden P. Primary knowledge, medical education and consultant expertise. *Med Educ* 1988;22:171–2.
- Lemieux M, Bordage G. Structuralisme et pédagogie médicale: étude comparative des stratégies cognitives d'apprentis-cliniciens. *Recherches Sémiotiques* 1986;6:143–79.
- Bordage G, Lemieux M. Some cognitive characteristics of students with and without diagnostic reasoning difficulties. In: *Proceedings of the 25th Annual Conference in Medical Education of the Association of American Medical Colleges*. New Orleans; 1987. p. 185–90.
- Boshuizen HPA. Cognitive effects of practical experience in high- and low-achieving medical students. *Learn Instruct* 1994;4:313–29.
- Boshuizen HPA, Schmidt HG. On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cogn Sci* 1992;16:153–84.
- Charlin B, Tardif J, Boshuizen HPA. Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research. *Acad Med* 2000;75:182–190.