

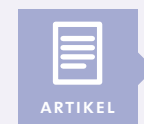
Verbetering van begeleiding door supervisors; evaluatie met feedback van aiossen

C.R.M.G. FLUIT, T.P.F.M. KLAASSEN, L. SANDER, M. DE VISSER-OOMEN, M.M. VAN DEN HURK

- ACHTERGROND** Om de kwaliteit van begeleiding van artsen in opleiding tot specialist (aiossen) te evalueren worden meestal vragenlijsten ingezet.
- DOEL** Wij onderzochten veranderingen van kwaliteitsscores van supervisors door aiossen middels de gevalideerde vragenlijst Evaluation and Feedback For Effective Clinical Teaching (EFFECT).
- METHODE** Supervisors (n = 89) van 9 disciplines werden geëvalueerd middels EFFECT. Gemiddelde overallscore (mos) en gemiddelde domeinscores werden vergeleken met t-toetsen voor gepaarde waarnemingen. Tien supervisors werden geïnterviewd om factoren die verbeteringen stimuleren of belemmeren te identificeren.
- RESULTATEN** De sterkste toename werd gezien bij supervisors met mos < 4. Motivatie om te veranderen en een dialoog over de feedback met de aiossen stimuleren supervisors om te verbeteren. Gebrek aan tijd en support op de werkplek werken belemmerend.
- CONCLUSIE** Evaluatie middels feedback door en dialoog met aiossen is geschikt om de begeleiding door supervisors te verbeteren.

TIJDSCHRIFT VOOR PSYCHIATRIE 61(2019)3, 182-187

TREFWOORDEN aiossen, evaluatie, supervisors, verbetering, vragenlijst



Om goede zorg te kunnen bieden is goed opleiden van professionals essentieel. Voor de medische vervolgoopleidingen ligt die verantwoordelijkheid voornamelijk bij artsen uit het betreffende specialisme, hierna supervisors genoemd. Een veelgebruikte strategie om begeleidingsvaardigheden van supervisors te bevorderen is feedback op basis van evaluaties middels vragenlijsten (Baker 2010). Onderzoek laat zien dat numerieke feedback en/of onduidelijke, weinig concreet beschreven feedback nauwelijks of geen effect hebben (Overeem e.a. 2010). Een dialoog kan het effect van feedback vergroten. Hierbij reflecteert de feedbackontvanger op het eigen gedrag, bespreekt de feedback en formuleert voornemens voor de praktijk (Fluit e.a. 2013a). Dit laatste draagt positief bij aan het effect van de feedback (Bing-You e.a. 2018).

Een veelgebruikte, gevalideerde vragenlijst in Nederland om de kwaliteit van begeleiding in kaart te brengen is de *Evaluation and Feedback For Effective Clinical Teaching* (EFFECT) (Fluit e.a. 2012). Kenmerkend is de terugkoppeling van resultaten middels een dialoog tussen arts in opleiding tot specialist (aios) en supervisor, begeleid door een moderator. De dialoog wordt afgerond met het formuleren van voornemens door de supervisor. Een pilotstudie liet zien dat van een deel van de supervisors de begeleidingsvaardigheden verbeteren na een EFFECT-evaluatie (Fluit e.a. 2013b). Onbekend is echter welke omstandigheden bijdragen aan het wel of niet verbeteren van de begeleiding. Binnen onderzoek naar transfer van het geleerde naar de praktijk maakt men onderscheid tussen: a. persoonsfactoren (o.a. motivatie), b. feedbackfactoren (o.a.

concrete feedback, voornemens formuleren), en c. werkplekfactoren (o.a. begeleiding en tijd) (Holton 2005). In dit onderzoek zoeken wij antwoord op de vraag in welke mate scores van supervisors verbeteren na een kwaliteits-evaluatie en welke factoren bijdragen aan het verbeteren van de begeleidingsvaardigheden.

METHODE

Wij verrichtten een kwantitatieve en kwalitatieve studie waarbij we resultaten van supervisors van 2 opeenvolgende evaluaties vergeleken. Vervolgens interviewden we 10 supervisors aan de hand van een semigestructureerde vragenlijst.

De EFFECT-vragenlijst bestaat uit 58 items verdeeld over zeven domeinen: rolmodel, zorg dragen voor leerzaam werk, planning, feedback, doceervaardigheden, beoordeling en persoonskenmerken. De items worden gescoord op een vijfpuntslikertschaal lopend van 1 = zeer slecht, kan niet doorgaan zo, 2 = onvoldoende, moet opgepakt worden, 3 = matig, verbeteringen zijn nodig, 4 = voldoende, verbetering op onderdelen, 5 = goed (Fluit e.a. 2012).

De evaluatieprocedure

De evaluatieprocedure begon met een introductiebijeenkomst op de afdeling met supervisors en aiossen. Hierin werd het formatieve doel en de procedure van EFFECT uitgelegd. Vervolgens vulden aiossen anoniem de webbased vragenlijst in over de stafleden met wie zij het afgelopen jaar hadden gewerkt. De supervisors vulden de vragenlijst in als zelfevaluatie. De resultaten werden verwerkt in een rapportage en aan de individuele supervisor schriftelijk teruggekoppeld. Daarna bespraken twee aiossen de resultaten met iedere supervisor onder begeleiding van een onafhankelijke moderator (bijv. onderwijskundige).

Dataverzameling

Het kwantitatieve deel was gebaseerd op de resultaten van supervisors die tweemaal een EFFECT-procedure hadden doorlopen en beoordeeld waren door minimaal 3 aiossen. Tussen de twee metingen lag een periode van twee jaar. Doordat veel aiossen gedurende hun opleiding van stageplek wisselen, waren het nagenoeg altijd andere aiossen die de vragenlijst invulden op meetmoment twee. Voor het kwalitatieve onderzoek werden 10 supervisors (9 mannen en 1 vrouw) geselecteerd die een hogere, lagere of gelijkblijvende gemiddelde score bij de tweede evaluatie hadden. Zij werkten gemiddeld 16,5 jaar als specialist en hadden gemiddeld 14,6 jaar ervaring in het begeleiden van aiossen. Semigestructureerde interviews van een uur werden afgenomen met een vooraf opgestelde topiclijst

AUTEURS

LIA FLUIT, onderwijskundige en arts n.p., universitair hoofddocent en hoofd Research in Learning and Education, Radboudumc Health Academy, Nijmegen.

TIM KLAASSEN, onderwijskundige en beleidsmedewerker onderwijs, Radboudumc Health Academy, Nijmegen.

LAURA SANDER, onderwijskundige, Hogeschool van Arnhem en Nijmegen.

MARIEKE DE VISSER-OOMEN, onderwijskundige en beleidsmedewerker onderwijs, Radboudumc Health Academy, Nijmegen.

MARIANNE VAN DEN HURK, docent, Radboud Universiteit, Nijmegen.

CORRESPONDENTIEADRES

Dr. C.R.M.G. Fluit, 51 Radboudumc Health Academy, Radboudumc, Postbus 9101, 6500 HB Nijmegen.
E-mail: lia.fluit@radboudumc.nl

Geen strijdige belangen meegedeeld.

Het artikel werd voor publicatie geaccepteerd op 23-1-2019.

(achtergrondinformatie, sterke punten, verbeterpunten, persoonsfactoren, feedbackfactoren en werkplekfactoren).

Data-analyse

De gemiddelde score van alle items samen (MOS) en de gemiddelde score per domein van beide metingen werden berekend. T-toetsen voor gepaarde steekproeven werden uitgevoerd om de verschillen tussen 2 opeenvolgende metingen te meten. Om relevante veranderingen op de vragenlijst te kunnen bepalen, gingen we uit van een effectgrootte van 0,4, wat correspondeert met een score van 0,2 op de MOS en domeinen (Fluit e.a. 2013b).

Interviews werden na transcriberen aan iedere participant voorgelegd als member-check. Een eerste codeboek werd opgesteld op basis van het *transfer of learning model* van Holton (2005). Twee onderzoekers (CF en IS) codeerden onafhankelijk van elkaar de interviews en vergeleken deze steeds met elkaar en vulden ze waar nodig aan. Dit leidde tot een definitief codeboek waarmee alle interviews werden gecodeerd.

Ethische toetsing

Voor dit onderzoek werd goedkeuring verkregen van de Ethische toetsingscommissie van de Nederlandse Vereniging voor Medisch onderwijs (NvMO).

TABEL 1 Toename gemiddelde EFFECT-score voor hoog en laag scorende supervisors

Gem. overall score supervisor	supervisors (N)	subscales sign. verbeterd	subscales sign. verslechterd	Gem. toename	Uitersten	Supervisors met	
						toename score > 0,20	afnamescore > 0,20
< 4,0	18	3,67	0,61	0,4	-0,17 - +1,5	12 (66,66%)	-
4,0-4,5	38	2,13	1,11	0,1	-0,67 - +0,61	11 (28,95%)	3 (7,89%)
> 4,5	33	0,91	1,30	0,01	-0,51 - +0,46	4 (12,12%)	5 (15,15%)

*De subscales zijn: rolmodel, zorgdragen voor leerzaam werk, planning, feedback, doceervaardigheden en support.

EFFECT: Evaluation and Feedback For Effective Clinical Teaching

RESULTATEN

Kwantitatieve bevindingen

De resultaten van 89 supervisors, uit drie verschillende ziekenhuizen (waarvan één academisch) en van 9 verschillende afdelingen werden geanalyseerd. Het aantal beoordelingen per supervisor varieerde van 3 tot 14.

TABEL 1 laat zien dat 12 van de 18 supervisors (ruim 66%) met een lage MOS (< 4) bij de eerste meting een relevante vooruitgang lieten zien bij meting twee (gemiddelde vooruitgang 0,4; uitersten: -0,17-1,5). 11 van de 38 supervisors (29%) met een MOS tussen de 4 en de 4,5, vertoonden een relevante toename van de MOS van > 0,2 en 3 (8%) een relevante daling in de MOS van > 0,2 (gemiddelde vooruitgang 0,1; uitersten: -0,67-0,61). Van de 33 supervisors met een hoge MOS (> 4,5) vertoonden 4 (12%) verdere vooruitgang in de MOS van > 0,2, terwijl 5 supervisors (15%) een

daling vertoonden in hun MOS van > 0,2 (gemiddelde vooruitgang 0,01; uitersten: -0,51-0,46). Supervisors lieten significante vooruitgang zien op de domeinen rolmodel, zorg dragen voor leerzaam werk, feedback en doceervaardigheden (**TABEL 2**).

Kwalitatieve resultaten

Tijdens de interviews gaven de supervisors aan zich na de evaluatie te richten op het behouden van genoemde sterke punten, expliciteren van denkstappen en eigen opleidingsstijl, volgen van scholing over begeleiden van aiossen, oefenen van gesprekstechnieken, verbeteren van feedback, aanpassen van de visite en feedback vragen aan de aiossen over de kwaliteit van hun begeleiding. Een deel had een voornemen geformuleerd, maar was er nog niet mee gestart, sommigen waren gestart en voor anderen was toepassen van het voornemen al afgerond.

TABEL 2 T-toets voor gepaarde steekproeven op EFFECT-domeinen

	N	MOS t1	MOS t2	df	T-score	p
Rolmodel	89	3,65	4,07	88	-3,74	0,00*
Zorg dragen voor leerzaam werk	89	3,69	4,07	88	-2,07	0,04*
Planning onderwijsmomenten	89	3,91	4,28	88	-1,27	0,21
Feedback	89	3,46	3,89	88	-2,55	0,01*
Doceervaardigheden	89	3,59	4,10	88	-4,29	0,00*
Persoonskenmerken	89	3,71	3,99	88	-1,12	0,27

* Significant op 0,05 (2-zijdig)

EFFECT: Evaluation and Feedback For Effective Clinical Teaching

PERSOONSKENMERKEN

Supervisors benoemden dat hun enthousiasme voor het opleiden en een sterke motivatie om te veranderen het gedragsveranderingsproces positief hadden beïnvloed. Zij konden zich goed identificeren met hun rol. Zij zagen zichzelf als sterk betrokken bij hun werk, behulpzaam, open en vriendelijk, met een groot vertrouwen in eigen kunnen. Supervisors zagen het oppakken van de verbeterpunten als een eigen verantwoordelijkheid.

FEEDBACKFACTOREN

De meeste supervisors stonden positief tegenover de aios-feedback. Eén supervisor (met een afgenomen MOS-score) gaf aan geen vertrouwen in de evaluatie door de aios te hebben. De narratieve feedback was informatiever dan itemscores omdat deze concrete informatie opleverde. De dialoog met de aios was het meest informatief: deze vergrootte het begrip van de feedback en hielp om na te denken over hoe te verbeteren, behalve voor de supervisor die geen vertrouwen in de evaluatie door de aios had.

WERKPLEKFACTOREN

Tijdgebrek bleek de belangrijkste belemmerende werkplekfactor wat betreft gedragsverandering. Door de grote hoeveelheid taken ontbrak het aan tijd om actief met de voornemens aan de slag te gaan. Supervisors onderling spraken elkaar niet aan op begeleidingsvaardigheden. In één afdeling was dit wel het geval en werden resultaten gedeeld. Er was weinig support vanuit de formele opleider; supervisors verwachtten dit ook niet. Er was weinig support vanuit de aiossen, deels doordat niet alle aiossen actief bij de EFFECT-evaluatie waren betrokken en/of geïnformeerd waren over de inhoud van de feedbackgesprekken. Opleiden werd als een vanzelfsprekende taak gezien binnen afdelingen en minder gewaardeerd dan onderzoek, zoals bleek uit de beperkte expliciete beloning voor opleiden.

DISCUSSIE

Participatie in een kwaliteitsevaluatie met EFFECT was geassocieerd met toename van de scores. Deze verbetering was het grootst bij supervisors die bij de eerste meting het laagst scoorden. De grootste vooruitgang werd gevonden op de subschalen rolmodel, zorg dragen voor leerzaam werk, feedback en docevvaardigheden. Supervisors die bij de eerste meting al hoog scoorden, vertoonden bij de tweede meting nauwelijks progressie of vertoonden een terugval. Persoons-, feedback- en werkplekfactoren speelden een belangrijke rol in het realiseren van geformuleerde voornemens tijdens het EFFECT-gesprek.

De resultaten van de kwantitatieve studie komen overeen met een eerdere pilotstudie (Fluit e.a. 2013b). Mogelijk zijn de supervisors met de laagste scores het meest gemotiveerd

om hun gedrag te verbeteren (Reilly e.a. 1996). Bij hoog scorende supervisors zien we nauwelijks nog toename van de scores, mogelijk doordat verdere groei niet mogelijk is door een plafondeffect (Bercken & Voeten 2002). Dit betekent dat hoewel deze supervisors hun begeleidingsvaardigheden mogelijk nog verder verbeterd hebben, dat niet meer meetbaar is omdat ze al de hoogste score hadden behaald bij de eerste meting.

Supervisors waren zich bewust van hun voorgenomen acties, sommigen waren er actief mee bezig en een enkeling had zijn gedrag reeds aangepast. Deze fasen van voornemens zijn goed te relateren aan de fasen van gedragsverandering beschreven door Prochaska en DiClemente (Prochaska & Velicer 1997). Zij onderscheiden hierin vijfstadia: voorbeschouwing, overpeinzing, voorbereiding, actieve verandering en stabilisatie. De meeste mensen die gedrag willen veranderen, doorlopen de cirkel meerdere keren. Er is kans op terugval, al is die vanaf de fase van stabilisatie kleiner. Supervisors zaten met name in de laatste drie fasen. Om de stabilisatiefase te bereiken is wederom tijd nodig en ook hier kan de tijdsdruk belemmerend zijn.

Het moeilijk of geen tijd vrij kunnen maken voor opleiden komt als de meest belemmerende werkplekfactor naar voren. Wellicht verklaart dit waarom de scores op de subschaal planning niet toenamen. De oorzaak van het tijdgebrek lag mogelijk in de combinatie van het gedragsveranderingsproces met de grote hoeveelheid (reguliere) taken (Prideaux e.a. 2000). Opvallend is dat het leren van elkaar en de support van een leidinggevende vaak ontbreken of matig aanwezig zijn. De verantwoordelijkheid van gedragsveranderingen legt men sterk bij zichzelf waardoor support vragen van een collega of leidinggevende weinig gebeurt. Dit staat een duurzame ontwikkeling van de begeleidingsvaardigheden in de weg.

Sterktes en beperkingen

Sterke aspecten van het kwantitatieve onderzoek zijn het hoge aantal supervisors dat is geïncludeerd en het gebruik van een gevalideerd instrument (Fluit e.a. 2012).

Een beperking van het onderzoek is het ontbreken van een controlegroep. Ook is niet uit te sluiten dat veranderingen in scores bij de tweede meting veroorzaakt worden door regressie naar het gemiddelde (Barnett e.a. 2005). Op basis hiervan is de kans groot dat een slechte prestatie wordt gevolgd door een goede, zonder een direct causaal verband. In deze studie betekent het dat de verbetering niet is toe te schrijven aan de EFFECT-evaluatie. Omdat de EFFECT-scores een gemiddelde score zijn, afkomstig van verschillende aiossen, wordt de kans hierop verkleind. Daarnaast is de terugval van de scores bij hoog scorende supervisors veel kleiner. Regressie naar het gemiddelde zou alle scores

moeten betreffen, hetgeen wij niet gevonden hebben. De metingen waren bovendien onafhankelijk van elkaar en tussen de metingen lag een lange periode (2 jaar), waardoor er geen beïnvloeding door eerder gegeven scores was. Meer follow-upmetingen kunnen inzicht geven of de toename een groei weergeeft, dan wel toegeschreven moet worden aan toeval.

Bij zeer hoog scorende supervisors zien we weinig verandering in scores. Hierbij kan het plafondeffect een rol spelen. Immers, als de maximale score is behaald, kan verbetering niet gemeten worden, terwijl die mogelijk wel plaatsvindt. Tot slot, bij constructies van rapportages gaan we uit van minimaal 3 beoordelingen. Dit kan een vertekend beeld geven van de kwaliteit van de begeleiding, alhoewel tijdens de feedbackgesprekken nagenoeg altijd bleek dat de resultaten herkenbaar zijn voor zowel de aios als de supervisor.

CONCLUSIE

Evaluatie middels feedback door aiossen is geschikt om de begeleiding door supervisors te verbeteren. De dialoog tussen supervisor en de aios is hierbij essentieel. Het is zinvol als een opleidingsteam elkaars sterke en verbeterpunten bespreekt en tips uitwisselt. Dit is ook nuttig omdat stafleden elkaar nauwelijks in actie zien. Een stap verder is om supervisors als coach voor elkaar te laten fungeren. Randvoorwaarden zijn een open en veilige feedbackcultuur waarin men elkaar aanspreekt op begeleidingsgedrag, openlijk evaluatiegegevens kan delen, steun ervaart van collega's en aiossen, en de leidinggevende bij het leren (of afleren) van gedrag wordt betrokken. Dit alles kan de effectiviteit van een EFFECT-evaluatie vergroten.

LITERATUUR

- Baker MD. Clinical teaching improves with resident evaluation and feedback. *Anesthesiology* 2010; 113: 693-703.
- Barnett AG, Pols JC van der, Dobson AJ. Regression to the mean: what it is and how to deal with it. *Int J Epidemiol* 2005; 34: 215-20.
- Bercken JHL, Voeten MJM. Variantieanalyse. De GLM benadering. Groningen/Houten: Noordhoff; 2002.
- Bing-You R, Varaklis K, Hayes V, Trowbridge R, Kemp H, McKelvy D. The feedback tango: an integrative review and analysis of the content of the teacher-learner feedback exchange. *Acad Med* 2018; 93: 657-63.
- Fluit CRMG, Bolhuis S, Grol R, Ham M, Feskens R, Laan R, e.a. Evaluation and feedback for effective clinical teaching in postgraduate medical education: validation of an assessment instrument incorporating the CanMEDS roles. *Med Teach* 2012; 34: 893-901.
- Fluit C, Bolhuis S, Klaassen T, Visser M de, Grol R, Wensing M. Residents provide feedback to their clinical teachers: Reflection through dialogue. *Med Teach* 2013a; 35: e1485-92.
- Fluit C, Feskens R, Bolhuis S, Grol R, Wensing M, Laan R. Repeated evaluations of the quality of clinical teaching by residents. *PME* 2013b; 2: 87-94.
- Holton III EF. Holton's Evaluation Model: new evidence and construct elaborations. *Adv Dev Hum Res* 2005; 7: 37-54.
- Overeem K, Lombarts MJMH, Sarah OA, Klazinga NS, Grol RPTM, Wollersheim HC. Three methods of multisource feedback compared: a plea for narrative comments and coworkers' perspectives. *Med Teach* 2010; 32: 141-7.
- Prideaux D, Alexander H, Bower A, Dacre J, Haist S, Jolly B, e.a. Clinical teaching: maintaining an educational role for doctors in the new health care environment. *Med Educ* 2000; 34: 820-6.
- Prochaska JO, Velicer W. The transtheoretical model of health behavior change. *Am J Health Promot* 1997; 12: 38-48.
- Reilly RR, Smither JW, Vasilopoulos NL. A longitudinal study of upward feedback. *Pers Psychol* 1996; 49: 599-612.

SUMMARY

Improving the quality of supervision of residents: an evaluation of supervisors using feedback by residents

C.R.M.G. FLUIT, T.P.F.M. KLAASSEN, L. SANDER, M. DE VISSER-OOMEN, M.M. VAN DEN HURK

- BACKGROUND** To evaluate the quality of the supervision of residents, questionnaires are frequently used.
- AIM** Changes in resident ratings were assessed using the validated Evaluation and Feedback for Effective Clinical Teaching (EFFECT) questionnaire.
- METHOD** Supervisors (n=89) of nine medical specialities were evaluated using a validated instrument (EFFECT). Mean overall scores (mos) and mean scale scores were compared using paired t-tests. Ten supervisors were interviewed to determine factors that stimulate or hinder improvement after an evaluation.
- RESULTS** The strongest increase was seen in supervisors with an mos <4.0. Motivation to improve and dialogue about the feedback stimulated improvement. Lack of time and support hindered improvement.
- CONCLUSION** Evaluating supervisors by providing feedback and a dialogue is suitable to improve their supervision of residents.

TIJDSCHRIFT VOOR PSYCHIATRIE 61(2019)3, 182-187

KEY WORDS evaluation, improvement, questionnaire, residents, supervisors