

Der Fragebogen "SFDP26-German": Ein verlässliches Instrument zur Evaluation des klinischen Unterrichts?

Zusammenfassung

Zielsetzung: Die Evaluation klinischer Lehrkompetenz ist ein wichtiger Beitrag zur Qualitätssicherung medizinischer Lehrveranstaltungen. Die Evaluation sollte mit einem verlässlichen Instrument durchgeführt werden, um Status quo sowie Effekte von medizindidaktischen Schulungsmaßnahmen beurteilen zu können. Im „Stanford Faculty Development Program“ (SFDP) haben sich folgende Aspekte als sinnvoll erwiesen:

1. Etablierung des Lernklimas,
2. Leitung einer Lehreinheit,
3. Zielkommunikation,
4. Förderung von Verstehen und Behalten,
5. Evaluation,
6. Feedback und
7. Förderung selbstbestimmten Lernens.

Als Evaluationsinstrument ist seit 1998 der Fragebogen „SFDP26“ im englischsprachigen Raum zur Erfassung dieser Kategorien etabliert. Bisher existiert kein gleichwertiges deutschsprachiges Instrument zur Überprüfung von Lehrkompetenz im klinischen Kontext.

Fragestellung:

1. Entwicklung und testtheoretische Überprüfung einer deutschsprachigen Version des SFDP26 (SFDP26-German),
2. Überprüfung der Korrelation der Subskalen des SFDP26-German mit der Gesamtlehreffektivität.

Methodik: 19 Ärzte (w=7, m=12) der Klinik für Anästhesiologie an der Universität zu Lübeck wurden am Ende ihrer Lehrveranstaltungen zur Notfallmedizin durch Studierende mittels SFDP26-German evaluiert. Die Stichprobe setzte sich aus 173 Medizinstudierenden (w=119, bzw. 68.8%; m=54, 31.2%) des überwiegend fünften (6.6%) und sechsten Fachsemesters (80.3%) zusammen. Das durchschnittliche Alter der Studierenden betrug 23±3 Jahre.

Ergebnisse: Die Trennschärfen der Items lagen alle in einem guten bis sehr guten Bereich ($r_{it}=0.48-0.75$). Alle Subskalen zeigten gute interne Konsistenzen ($\alpha=0.69-0.92$) sowie signifikante positive Korrelationen ($r=0.40-0.70$) untereinander. Subskalen und „Gesamtlehreffektivität“ korrelierten signifikant ($p<0.001$) und am höchsten für die Subskala „Zielkommunikation“ ($r=0.61$).

Schlussfolgerungen: Die Analysen bestätigen dem „SFDP26-German“ eine hohe interne Konsistenz. Die Beleuchtung der Wirksamkeit der einzelnen Kategorien auf die Gesamtlehreleistung sowie eine Validierung anhand externer Kriterien sollte in Folgestudien erfolgen.

Schlüsselwörter: SFDP26, klinische Lehre, Lehreffektivität, Evaluation, Fragebogen

Peter Iblher^{1,2}

Michaela Zupanic²

Christoph Härtel³

Hermann Heinze¹

Peter Schmucker¹

Martin R. Fischer²

1 Universität zu Lübeck, Klinik für Anästhesiologie, Lübeck, Deutschland

2 Universität Witten/Herdecke, Fakultät für Gesundheit, Institut für Didaktik und Bildungsforschung im Gesundheitswesen (IDBG), Witten, Deutschland

3 Universität zu Lübeck, Klinik für Kinder- und Jugendmedizin, Lübeck, Deutschland

Einleitung

Einführung

Die Verbesserung der Lehre ist unbestritten eine zentrale Aufgabe in der deutschen medizinischen Hochschullandschaft [1]. Dabei spielt die Sensibilisierung der in der Lehre tätigen Institute und Kliniken durch Anreizprogramme wie beispielsweise die leistungsorientierte Mittelvergabe (LOM) und Lehrpreise eine wesentliche Rolle [2]. Feedbacksysteme, die mittels standardisierter Evaluationsinstrumente den Dozierenden und Abteilungen das Ausmaß der Lehrleistung rückmelden, sind eine wesentliche Voraussetzung für eine nachvollziehbare qualitative Bewertung der Lehrleistung. Die Diskussion, insbesondere in den Fachgremien, trägt dazu bei, dass die ehemals eher stiefmütterlich behandelte Lehre neben Forschung und Krankenversorgung mehr in den Vordergrund rückt und als Standortvorteil zur Teambotivierung und Nachwuchsrekrutierung wahrgenommen wird [1], [3], [4].

Unbestritten nimmt dabei neben der Etablierung von lernzielbasierten Curricula und der Verwendung adäquater Prüfungsformen [5] die Qualifizierung der Lehrenden eine entscheidende Schlüsselstellung ein [6], [7]. In einer Befragung von Klinikärztinnen und -ärzten an der Universität zu Lübeck zeigte sich, dass die Wahrnehmung von guter Lehre fast immer mit herausragenden Persönlichkeiten assoziiert war, die aus vielfältigen Gründen als gute Lehrende erlebt wurden [8]. Trotz der Vielfalt ließen sich diese Gründe unter der durch Sutkin et al. etablierten Definition der „nicht-kognitiven Eigenschaften“ eines guten Lehrers, wie Beziehungsfähigkeit, Gemütszustand und Persönlichkeit, zusammenführen [9]. Das wirft die Frage auf, welche Lehrkompetenzen insbesondere relevant sind für eine gute und effektive Lehrleistung. Anzahl und Akzeptanz von lokalen und überregionalen Schulungsangeboten haben sich in den letzten Jahren deutlich erhöht (Train-the-Trainer-Konzepte, beispielsweise das nationale Master of Medical Education-Programm des Medizinischen Fakultätentages (MFT) an der Universität Heidelberg [10]) und auch Eingang in viele Qualifizierungsprogramme von Habilitanden gefunden. Dennoch ist es häufig so, dass die Schulung von Lehrkompetenzen bei insbesondere im klinischen Unterricht tätigen Ärztinnen und Ärzten vernachlässigt wird und diese ihre Fähigkeiten autodidaktisch durch Vorbilder, Erfahrungen und Eigeninitiative erwerben [7], [11]. Es besteht also durchaus die Möglichkeit, die akademische Leistungsfähigkeit der Lernenden aufgrund methodischer, rhetorischer oder didaktischer Schwächen (trotz entsprechender fachspezifischer Kenntnisse zu Lehrinhalten) negativ zu beeinflussen [12]. Das bedeutet konsektiv, dass

1. spezifische Lehrkompetenzen der Dozierenden einer Überprüfung zugänglich gemacht werden sollten, um gegebenenfalls Schwächen zu erkennen und diese gezielt durch Schulungsmaßnahmen zu verbessern und

2. untersucht werden sollte, welche Kategorien der Lehrleistung in der studentischen Evaluation besonders relevant für eine gute Lehre sind.

Wünschenswert wäre also ein etabliertes Messinstrument, das entscheidende Kompetenzbereiche der Lehre abbildet, individuelle Stärken und Schwächen bei Lehrenden offenlegt und darüber hinaus auch im Rahmen von Lehrforschungsprojekten einsetzbar ist. Unseres Wissens nach gibt es bisher kein deutschsprachiges Instrument, das diese Erfordernisse erfüllt.

Stanford Faculty Development Program (SFDP)

Im Sinne eines Fakultätsentwicklungsprogramms wurde seit Anfang der achtziger Jahre an der Stanford University, Palo Alto, USA ein auf Lerntheorie gestütztes Schulungsprogramm (SFDP) etabliert. Im Rahmen dieses Programms wurden Dozierende im Unterricht beobachtet und deren Stärken und Schwächen in der Lehre in sieben Kernkompetenzen zusammengefasst:

1. Etablieren des Lernklimas (Learning climate),
2. Leitung einer Lehrinheit (Control of session),
3. Zielkommunikation (Communication of goals),
4. Fördern von Verstehen und Behalten (Understanding and Retention),
5. Evaluation,
6. Feedback,
7. Fördern vom Selbstbestimmten Lernen (Self-directed Learning).

Es ist eines der frühen Entwicklungsprogramme [13], [14] und entsprang einem konzeptionellen Rahmen, der seit Ende der 1970er Jahre federführend von Dr. Kelley Skeff, einem Internisten der Stanford University und Dr. Georgette Stratos im Rahmen ihrer Untersuchungen zur Verbesserung der Lehre entwickelt wurde. Dieser filmte zunächst hunderte Stunden von Lehrveranstaltungen und korrelierte unter anderem bestimmte Lehrmethoden und Verhaltensweisen der Dozenten mit den Beurteilungen von Studierenden hinsichtlich ihrer Lehreffektivität. Im Rahmen von zahlreichen Studien entwickelte sich so ein Fakultätsentwicklungsprogramm, das es zum einen der Fakultät ermöglichte, seine Lehre zu analysieren und gleichzeitig die Lehrkompetenz der Lehrenden in standardisierten Veranstaltungen fortzubilden [15], [16], [17], [18], [19], [20], [21], [22], [23], [24]. Im Rahmen dieses Programms wurde ein reliables und verlässliches englischsprachiges Messinstrument der Lehrleistung auf der Grundlage von sieben Kategorien entwickelt (SFDP26), das breiten Eingang in Schulungsprogramme und medizinischdidaktische Forschungsvorhaben im englischsprachigen Raum gefunden hat [16], http://www.acgme.org/outcome/downloads/PBLI/15_pbli.pdf.

Fragestellung

Diese Arbeit untersucht zwei Aspekte hinsichtlich des Fragebogens SFDP26:

1. Entwicklung und testtheoretische Überprüfung der deutschsprachigen Version des SFDP26 (SFDP26-German) zur Überprüfung von Lehrkompetenzen mittels Item- und Reliabilitätsanalysen zur Ermittlung der internen Konsistenz (Cronbach's α);
2. Korrelationsanalysen zwischen den Subskalen und der Gesamtlehreffektivität zur Überprüfung der Relevanz der einzelnen Subskalen.

Methoden

Der Fragebogen SFDP26 besteht aus 25 Items in sieben Subskalen

1. Etablieren des Lernklimas,
2. Leitung einer Lehreinheit,
3. Zielkommunikation,
4. Fördern von Verstehen und Behalten,
5. Evaluation,
6. Feedback,
7. Fördern vom Selbstbestimmten Lernen und

einem Item „Gesamtlehreffektivität“ als globales Maß für die Lehrleistung.

Die Bewertung erfolgt auf einer fünfstufigen Likert Skala für Item 1-25 (1-5=Nein-Etwas-Ja) und für das Item „Gesamtlehreffektivität“ (1=sehr schwach bis 5=ausgezeichnet). Der englische Originalfragebogen SFDP26 wurde durch Muttersprachler zunächst aus dem Englischen ins Deutsche, und dann ins Englische zurück übersetzt (siehe Tabelle 1). Sieben Ärztinnen und zwölf Ärzte der Klinik für Anästhesiologie an der Universität zu Lübeck, die im studentischen Notfallkurs des Sommersemesters 2009 als Lehrende tätig waren, wurden jeweils am Ende ihrer Lehrveranstaltung durch Studierende mit dem Fragebogen SFDP26-German evaluiert. Die Stichprobe setzte sich dabei zusammen aus 173 Medizinstudierenden des überwiegend fünften (8.7%) und sechsten Fachsemesters (80.3%) (6.6% höhere Semester, 4.6% keine Angabe), die zufällig nach Kursgruppen den Dozenten zugeordnet wurden. Das durchschnittliche Alter der 119 Frauen und 54 Männer betrug 23 ± 3 Jahre. Von Seiten der Ethik-Kommission der Universität zu Lübeck bestanden keinerlei ethische Bedenken gegen diese Studie. Sowohl Studierende, als auch Dozenten wurden im Vorwege über diese Untersuchung informiert und hatten die Möglichkeit, die Teilnahme an dieser Studie zu verweigern. Die Einwilligung erfolgte durch konkludentes Handeln. Alle erhobenen Daten wurden pseudonymisiert ausgewertet. Zur Ermittlung der Itemkennwerte wurden statistische Analysen zu Mittelwert, Standardabweichung und Trennschärfe (nach Pearson) durchgeführt. Der Trennschärfekoeffizient ist dabei nach Lienert die Korrelation der Aufgabenbeantwortung mit dem Summenwert der Skala und Kennwert

dafür, in welchem Ausmaß die Differenzierung der Personen durch das Item mit derjenigen durch die Skala als Ganzes übereinstimmt [25]. Dabei gelten Trennschärfen r' von über 0.3 als gut, von 0.2-0.3 als akzeptabel, von 0.1-0.2 als marginal und Werte unter 0.1 als schlecht [26]. Zur Bestimmung der Skalenkennwerte und -interkorrelationen wurden Skalensummenwerte, Standardabweichung, Homogenitätskoeffizienten (Cronbach's α) und korrigierte Inter-Skalen-Korrelationen nach Pearson (r) mit dem SPSS-Softwarepaket Version 18 (SPSS Inc., Chicago, USA) berechnet. Für Gruppenvergleichstestungen der Reliabilität kann ein Cronbach's α ab 0.6 als ausreichend und ab 0.8 als gut bezeichnet werden [27]. Eine Irrtumswahrscheinlichkeit von $< 5\%$ ($p < 0.05$) wurde als signifikant, von $< 0.1\%$ ($p < 0.001$) als hochsignifikant bewertet.

Ergebnisse

Testtheoretische Überprüfung des SFDP26-German

Die Ergebnisse der Itemanalysen sind in Tabelle 1 dargestellt. Die größte Zustimmung fanden die Items 1 („Er/sie hörte den Lernenden zu.“) und 3 („Er/sie verhielt sich respektvoll den Lernenden gegenüber.“) der Skala „Lernklima“ mit Mittelwerten von $M=4.8 \pm 0.7$ bzw. 4.7 ± 0.7 . Den Items 25 („Er/sie bestärkte die Lernenden, außerhalb des Kurses nachzulesen.“) der Skala „Selbstbestimmtes Lernen“ und 15 („Er/sie überprüfte medizinischen Vorwissen der Lernenden.“) der Skala „Evaluation“ wurde am wenigsten zugestimmt mit Mittelwerten von $M=3.4 \pm 0.9$ bzw. 3.7 ± 1.2 .

Die Trennschärfen der Items lagen in keinem Fall unterhalb des kritischen Werts von 0.20, sondern ausnahmslos in einem guten bis sehr guten Bereich ($r \geq 0.60$).

Die Ergebnisse der Skalenanalysen sind in Tabelle 2 dargestellt. Die größte Zustimmung der Studierenden erfolgte in der Skala „Lernklima“ mit einem Mittelwert von 4.6 ± 0.6 . Die geringste Zustimmung wurde in der Skala „Selbstbestimmtes Lernen“ ersichtlich mit einem Mittelwert von 3.6 ± 1.1 . Bei nahezu allen Skalen zeigte sich eine hohe interne Konsistenz als Zeichen der Reliabilität, die als gut bewertet werden kann (Cronbach's α : 0.80-0.89). Die Skala „Verstehen und Behalten“ zeigte eine mittlere interne Konsistenz ($\alpha=0.69$), die Skala „Feedback“ lag im sehr guten Bereich ($\alpha=0.92$). Zwischen allen Skalen des SFDP26-German fanden sich hoch signifikante Korrelationen (r) von positiver Proportionalität (siehe Tabelle 3). Der höchste Zusammenhang zeigte sich zwischen den Skalen „Lernklima“ und „Zielkommunikation“ ($r=0.73$), sowie zwischen „Verstehen und Behalten“ und „Zielkommunikation“ ($r=0.72$). Die gemeinsame Varianz ($r^2 \times 100$) der beiden Skalen betrug 53% bzw. 52%. Die geringsten, aber noch signifikanten positiven Zusammenhänge wurden zwischen den Skalen „Selbstbestimmtes Lernen“ und „Leitung einer Lehreinheit“

Tabelle 1: Itemkennwerte für die Items des SFDP26-German dargestellt nach Subskalen (1. – 8.) (M = Mittelwert; SD = Standardabweichung, r_{it} = Trennschärfe nach Pearson).

SFDP26-German Subskalen	M	SD	r_{it}
1. Lernklima			
Er/sie hörte den Lernenden zu.	4.75	0.68	.647
Er/sie bestärkte die Lernenden, sich aktiv an der Diskussion zu beteiligen.	4.57	0.79	.678
Er/sie verhielt sich respektvoll den Lernenden gegenüber.	4.74	0.68	.631
Er/sie ermutigte die Lernenden, Probleme anzusprechen.	4.51	0.80	.619
2. Sitzungskontrolle			
Er/sie achtete auf den zeitlichen Rahmen.	4.58	0.94	.483
Er/sie vermied Abschweifungen vom Kursthema.	4.39	0.86	.633
Er/sie verhinderte äußere Störungen.	4.53	0.88	.654
3. Zielkommunikation			
Er/sie stellte die Lernziele des Kurses kurz und prägnant dar.	4.36	0.88	.685
Er/sie machte den Lernenden die Relevanz der Lernziele deutlich.	4.62	0.78	.735
Er/sie setzte inhaltliche Schwerpunkte (Prioritäten im Kurs).	4.47	0.83	.735
Er/sie wiederholte die Kurslernziele regelmäßig.	4.09	0.96	.700
4. Verstehen und Behalten			
Er/sie präsentierte den Lernstoff gut strukturiert.	4.37	0.82	.725
Er/sie machte die Zusammenhänge des Lernstoffes deutlich.	4.45	0.76	.740
Er/sie setzte visuelle Hilfen zur Veranschaulichung ein (Tafel, Modelle, etc.).	3.98	1.38	.557
5. Evaluation			
Er/sie überprüfte medizinisches Vorwissen der Lernenden.	3.69	1.15	.641
Er/sie überprüfte Fähigkeiten der Lernenden, medizinische Zusammenhänge herzuleiten.	3.90	1.07	.652
Er/sie überprüfte Fähigkeiten der Lernenden, medizinisches Wissen patientenspezifisch anzuwenden.	4.02	1.08	.681
Er/sie überprüfte praktische Fertigkeiten der Lernenden in der fallbezogenen Patientenversorgung.	4.02	1.20	.653
6. Feedback			
Er/sie gab den Lernenden negatives (korrigierendes) Feedback.	4.19	1.06	.672
Er/sie erklärte den Lernenden, was richtig und was falsch war.	4.44	0.93	.676
Er/sie machte den Lernenden Vorschläge, wie sie es besser machen könnten.	4.31	0.99	.692
Er/sie gab regelmäßig Feedback an die Lernenden.	4.24	1.00	.745
7. Selbstbestimmtes Lernen			
Er/sie bestärkte die Lernenden ausdrücklich, das Themengebiet weiter zu vertiefen.	3.76	1.18	.688
Er/sie motivierte die Lernenden, selbstständig zu lernen.	3.81	1.16	.665
Er/sie bestärkte die Lernenden, außerhalb des Kurses nachzulesen.	3.40	1.36	.510
8. Im Ganzen war die Lehrleistung des Dozenten ...	4.39	0.85	.676

($r=0.21$), bzw. „Lernklima“ ($r=0.28$) ersichtlich. Die gemeinsame Varianz war hier 4%, bzw. 8%.

Tabelle 2: Skalennennwerte für die SFDP26-German Subskalen (Σ = Skalensummenwerte, SD = Standardabweichung) und Interkorrelationen der Subskalen mit dem Global Rating „Gesamt-Lehreffektivität“ (r = Korrelation nach Pearson).

SFDP26-German Subskalen	Σ	SD	Cronbachs α	r
Lernklima	4.64	0.62	0.86	.511
Sitzungskontrolle	4.48	0.75	0.80	.443
Zielkommunikation	4.37	0.71	0.84	.611
Verstehen und Behalten	4.23	0.83	0.69	.578
Evaluation	3.87	0.95	0.85	.471
Feedback	4.25	0.96	0.92	.479
Selbstbestimmtes Lernen	3.64	1.13	0.89	.475

Tabelle 3: Skalen-Interkorrelationen der Subskalen des SFDP26-German

SFDP26-German Subskalen	1	2	3	4	5	6
1 Lernklima	-					
2 Sitzungskontrolle	.643	-				
3 Zielkommunikation	.730	.637	-			
4 Verstehen und Behalten	.556	.506	.722	-		
5 Evaluation	.472	.363	.576	.444	-	
6 Feedback	.598	.459	.522	.458	.608	-
7 Selbstbestimmtes Lernen	.278	.211	.488	.437	.567	.364

Korrelationen der Subskalen mit Global Rating (Gesamtlehreffektivität)

Die Korrelationen der Subskalen mit der globalen Gesamtbewertung der Lehrleistung (Gesamtlehreffektivität) sind in Tabelle 2 dargestellt. Alle Korrelationen waren hoch signifikant ($p<0.001$). Hier zeigten sich die höchsten Korrelationen mit der Gesamtlehreffektivität für die Subskalen „Zielkommunikation“ ($r=0.61$), „Fördern von Verstehen und Behalten“ ($r=0.58$) und „Lernklima“ ($r=0.51$).

Diskussion

Es handelt sich beim SFDP26 um ein gut untersuchtes Evaluationsinstrument [16], [28], so dass die Übertragung des englischen SFDP26 in den deutschsprachigen Raum sinnvoll und nahe liegend ist. Die Zuverlässigkeit eines solchen Prüfverfahrens ist dabei ein entscheidendes Gütekriterium, das durch die Reliabilität beschrieben wird. Die Ergebnisse dieser Studie bescheinigen dem SFDP26-German hohe interne Konsistenzen, die Reliabi-

lität dieses Fragebogens kann somit als gegeben angesehen werden. Die Trennschärfen lagen alle in einem guten Bereich, so dass die unterschiedlichen Aspekte der Lehrleistung der Dozierenden gut erfasst werden. Darüber hinaus zeigten alle Subskalen signifikante Korrelationen von positiver Proportionalität mit der Gesamtlehreffektivität. Dies kann dahingehend gewertet werden, dass die Subskalen relevante Aspekte der Gesamtlehreffektivität sind.

Beckman und Mitarbeiter untersuchten mit einem auf dem SFDP basierenden, aber durch neue 16 Items veränderten Fragebogen (Mayo Teaching Evaluation Form/MTEF-28), die Evaluation der Lehrleistung von Assistenzärzten in der klinischen Lehre der Mayo-Clinic [28]. Dort zeigten sich für die Subskalen „Leitung einer Lehrereinheit“, „Fördern von Verstehen und Behalten“ und „Feedback“ niedrige interne Konsistenzen (α : 0.147; 0.570; 0.648). Litzelman und Mitarbeiter, die den originären SFDP26 ebenfalls zur Evaluation der Lehrleistung durch Assistenzärzte verwendeten, fanden in allen Skalen hohe interne Konsistenzen ($\alpha > .85$) [17]. Dies weist darauf hin, dass der SFDP26 ein sehr verlässliches Messinstrument ist, dessen Reliabilität durch Hinzufügung weiterer Items, wie durch Beckmann und Mitarbeiter vorgenommen, durchaus vermindert werden kann. In einer Studie von Williams und Mitarbeitern wurden die Subskalen des SFDP26 mit der „Global Rating Scale“ der University of Michigan („Please rate the educational value of time with an attending physician.“) verglichen und zeigten dort hohe Korrelationen ($r: 0.86-0.98$) [29], am höchsten in den Subskalen „Zielkommunikation“, „Fördern von Verstehen und Behalten“, „Evaluation“, „Feedback“ und „Fördern vom Selbstbestimmten Lernen“ (0.94-0.98). Diese Ergebnisse sind erstaunlich, bedeuten sie ja, dass nahezu ein vollständiger positiver linearer Zusammenhang zwischen den betrachteten Skalen besteht, bzw. zwischen den Subskalen des SFDP26 und einer globalen „Single-Item“-Skala. Die Korrelationen der SFDP26 Subskalen mit der Gesamtlehreffektivität sowie zwischen der SFDP26 Gesamtlehreffektivität und der „Global Rating Scale“ der University of Michigan (Validitätsmaß) wurden leider nicht dargestellt.

Die Skala „Zielkommunikation“ in dieser Arbeit zum SFDP26-German korreliert am höchsten mit der Gesamtlehreffektivität ($r=0.61$), 37% der Varianz wurden aufgeklärt. Das bedeutet, dass bei Betrachtung aller Subskalen des SFDP26-German zwischen dieser Skala und der Gesamtlehreffektivität der stärkste Zusammenhang besteht. Die Frage der Kausalität lässt sich in dieser Studie nicht klären, allerdings ist die hohe Korrelation ein Hinweis darauf, dass die Skala „Zielkommunikation“ eine relevante Einflussgröße ist und eine wesentliche Bedeutung für die Gesamtlehreffektivität haben könnte. Die Frage zur Bedeutung der einzelnen Subskalen des SFDP26-German auf die effektive Lehrkompetenz sollte also in Folgestudien noch näher beleuchtet werden, um beispielsweise gezielte Schulungsmaßnahmen zur Lehreffektivität entwickeln zu können. Hier wäre es günstig, zusätzlich auch auf den tatsächlich messbaren Kompetenzzuwachs der

Studierenden zu fokussieren, der allein mit diesem Fragebogen nicht zu beantworten ist. Der Vergleich des Prüfungserfolgs der Studierenden und ihrer Leistungsfähigkeit im klinischen Kontext als externe Kriterien mit den Ergebnissen des SFDP26-German würde dann auch Aussagen zur Validität des Fragebogens ermöglichen. Kritisch angemerkt werden muss, dass der SFDP26 keine Vorkehrungen trifft, um zu verhindern, dass Antworttendenzen bei den Studierenden auftreten, z.B. immer nur die gleiche Zahlenreihe angekreuzt wird. Wir haben in unseren Untersuchungen zwar keine Hinweise darauf, dass dies eine Rolle spielen könnte, dennoch muss dieser Punkt in Analysen kritisch gewürdigt werden. Auch die Tatsache, dass die Datenerhebung nur mittels quantitativer Elemente erfolgt, muss kritisch diskutiert werden. Zwar bescheinigen die Analysen den Skalen des SFDP26 eine hohe Reliabilität, dennoch wären möglicherweise für detailliertere Beurteilungen die Ergänzung mit qualitativen Evaluationselementen, wie der Option zur Freitextantwort, sinnvoll. Auch die Anordnung der Stufen auf der Likert-Skala von eins bis fünf sind nicht unproblematisch. Zwar dient die Beschriftung der Gesamtskala („Nein – Etwas – Ja“) der Orientierung, trotzdem ist es möglich, dass die Studierenden unbewusst der im deutschsprachigen Raum etablierten Schulnotengliederung folgen und es zu einer versehentlichen „Umkehrung“ beim Ankreuzen kommt (1=sehr gut, usw.). Dies könnte durch die Umkehr der Beschriftung verhindert werden. Trotz der guten Ergebnisse dieser Studie muss einschränkend festgestellt werden, dass es sich hier um die Untersuchung eines speziellen Kollektivs im Medizinstudium handelt. Untersuchungen zur Anwendung des SFDP26-German in verschiedenen Fächern und zu verschiedenen Zeitpunkten sind daher erforderlich, um die generelle Einsetzbarkeit zu überprüfen.

Schlussfolgerung

Die hier vorgestellte Arbeit untersucht erstmalig die deutschsprachige Version des SFDP26 im klinischen Unterricht von Medizinstudierenden und bescheinigt dem Verfahren eine ausreichende Zuverlässigkeit für den Einsatz in der studentischen Lehrevaluation. Folgestudien in weiteren klinischen Feldern sind notwendig, um die Bedeutung der Subskalen und die Validität des Fragebogens durch Vergleich mit externen Kriterien näher zu beleuchten.

Danksagung

Die Autoren danken allen Dozierenden und Studierenden, die sich an dieser Studie beteiligt haben und insbesondere Herrn Joe Letkeman für die Übersetzung des Fragebogens.

Interessenkonflikt

Die Autoren erklären, dass keinerlei finanzielle oder sonstige Interessenkonflikte bestehen.

Literatur

- Wissenschaftsrat. Empfehlungen zur Qualitätsverbesserung von Lehre und Studium. Berlin: Wissenschaftsrat; 2008; Zugänglich unter/available from: <http://www.wissenschaftsrat.de/download/archiv/8639-08.pdf>
- Wissenschaftsrat. IV.2. Anerkennung von besonderen Leistungen in der Lehre. In: Empfehlungen zur Qualitätsverbesserung von Lehre und Studium. Köln: Geschäftsstelle des Wissenschaftsrates; 2008.
- Hahn EG, Fischer MR. Nationaler Kompetenzbasierter Lernzielkatalog Medizin (NKLM) für Deutschland: Zusammenarbeit der Gesellschaft für Medizinische Ausbildung (GMA) und des Medizinischen Fakultätentages (MFT). GMS Z Med Ausbild. 2009;26(3):Doc35. DOI: 10.3205/zma000627
- Lammerding-Köppel M, Fabry G, Hofer M, Ochsendorf F, Schirlo C. Hochschuldidaktische Qualifizierung in der Medizin: I. Bestandsaufnahme: Ein Positionspapier des GMA-Ausschusses Personal- und Organisationsentwicklung für die medizinische Lehre der Gesellschaft für Medizinische Ausbildung sowie des Kompetenzzentrums für Hochschuldidaktik in Medizin Baden-Württemberg. GMS Z Med Ausbild. 2006;23(4):Doc73. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2006-23/zma000292.shtml>
- Fischer MR, Gesellschaft für Medizinische Ausbildung, GMA-Ausschuss Prüfungen, Kompetenzzentrum Prüfungen Baden-Württemberg. Leitlinie für Fakultätsinterne Leistungsnachweise während des Medizinstudiums: Ein Positionspapier des GMA-Ausschusses Prüfungen und des Kompetenzzentrums Prüfungen Baden-Württemberg. GMS Z Med. 2008;25(1):Doc74. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2008-25/zma000558.shtml>
- Geraci SA, Kovach RA, Babbott SF, Hollander H, Buranosky R, Devine DR, Berkowitz L. AAIM Report on Master Teachers and Clinician Educators Part 2: faculty development and training. Am J Med. 2010;123(9):869-872 e6.
- McLean M, Cilliers F, Van Wyk JM. Faculty development: yesterday, today and tomorrow. Med Teach. 2008;30(6):555-584. DOI: 10.1080/01421590802109834
- Iblher P. Lernen zu Lehren – Implementierung eines vorklinischen Wahlfaches zur frühen Förderung von Lehrkompetenzen an der Universität zu Lübeck-Unveröffentlichte Projektarbeit. Lübeck: Universität zu Lübeck; 2008.
- Sutkin G, Wagner E, Harris I, Schiffer R. What makes a good clinical teacher in medicine? A review of the literature. Acad Med. 2008;83(5):452-466. DOI: 10.1097/ACM.0b013e31816bee61
- Jünger J, Fischer MR, Duelli R, Putz R, Resch F. Implementierung und Evaluation eines interfakultären Master of Medical Education Programms. Z Evid Fortbild Qual Gesundheitsw. 2008;102(10):620-627.
- Busari JO, Prince KJ, Scherpbier AJ, Van Der Vleuten CP, Essed GG. How residents perceive their teaching role in the clinical setting: a qualitative study. Med Teach. 2002;24(1):57-61. DOI: 10.1080/00034980120103496
- Morrison EH, Hafler JP. Yesterday a learner, today a teacher too: residents as teachers in 2000. Pediatrics. 2000;105(1 Pt 3):238-241.
- Searle NS, Thompson BM, Friedland JA, Lomax JW, Drutz JE, Coburn M, Nelson EA. The prevalence and practice of academies of medical educators: a survey of U.S. medical schools. Acad Med. 2010;85(1):48-56. DOI: 10.1097/ACM.0b013e3181c4846b
- Steinert Y, Mann K, Centeno A, Dolmans D, Spencer J, Gelula M, Prideaux D. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. Med Teach. 2006;28(6):497-526. DOI: 10.1080/01421590600902976
- Litzelman DK, Stratos GA, Marriott DJ, Lazaridis EN, Skeff KM. Beneficial and harmful effects of augmented feedback on physicians' clinical-teaching performances. Acad Med. 1998;73(3):324-332. DOI: 10.1097/00001888-199803000-00022
- Litzelman DK, Stratos GA, Marriott DJ, Skeff KM. Factorial validation of a widely disseminated educational framework for evaluating clinical teachers. Acad Med. 1998;73(6):688-695. DOI: 10.1097/00001888-199806000-00016
- Litzelman DK, Westmoreland GR, Skeff KM, Stratos GA. Factorial validation of an educational framework using residents' evaluations of clinician-educators. Acad Med. 1999;74(10 Suppl):S25-S27.
- Skeff KM. Evaluation of a method for improving the teaching performance of attending physicians. Am J Med. 1983;75(3):465-470. DOI: 10.1016/0002-9343(83)90351-0
- Skeff KM, Campbell M, Stratos G, Jones HW, 3rd, Cooke M. Assessment by attending physicians of a seminar method to improve clinical teaching. J Med Educ. 1984;59(12):944-950.
- Skeff KM, Stratos G, Campbell M, Cooke M, Jones HW. 3rd. Evaluation of the seminar method to improve clinical teaching. J Gen Intern Med. 1986;1(5):315-322. DOI: 10.1007/BF02596211
- Skeff KM, Stratos GA, Bergen MR, Regula DP, Jr. A pilot study of faculty development for basic science teachers. Acad Med. 1998;73(6):701-704. DOI: 10.1097/00001888-199806000-00018
- Skeff KM, Stratos GA, Bergen MR, Sampson K, Deutsch SL. Regional teaching improvement programs for community-based teachers. Am J Med. 1999;106(1):76-80. DOI: 10.1016/S0002-9343(98)00360-X
- Skeff KM, Stratos GA, Berman J, Bergen MR. Improving clinical teaching. Evaluation of a national dissemination program. Arch Intern Med. 1992;152(6):1156-1161.
- Skeff KM, Stratos GA, Mygdal W, DeWitt TA, Manfred L, Quirk M, Roberts K, Greenberg L, Bland CJ. Faculty development. A resource for clinical teachers. J Gen Intern Med. 1997;12(Suppl 2):S56-S63. DOI: 10.1046/j.1525-1497.12.s2.8.x
- Lienert GA, Raatz U. Berechnung von Schwierigkeitsindex, Trennschärferkoeffizient und Aufgabeninterkorrelation. In: Lienert GA, Raatz U, editors. Testaufbau und Testanalyse. Weinheim: Beltz, Psychologie-Verl.-Union; 1994. S.73-113.
- Möltner A, Schellberg D, Jünger J. Grundlegende quantitative Analysen medizinischer Prüfungen. GMS Z Med Ausbild. 2006;23(3):Doc53. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2006-23/zma000272.shtml>
- Bortz JD, Döring N. Hypothesengewinnung und Theoriebildung. In: Bortz JD, Döring N (Hrsg). Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Berlin, Heidelberg, New York, Tokio: Springer; 2006. S.355-936.
- Litzelman DK, Westmoreland GR, Skeff KM, Stratos G. Student and resident evaluations of faculty-how dependable are they? Acad Med. 1999;74(Suppl.):s25-s27.

29. Williams BC, Litzelman DK, Babbott SF, Lubitz RM, Hofer TP. Validation of a global measure of faculty's clinical teaching performance. *Acad Med.* 2002;77(2):177-180. DOI: 10.1097/00001888-200202000-00020
30. Beckman TJ, Lee MC, Rohren CH, Pankratz VS. Evaluating an instrument for the peer review of inpatient teaching. *Med Teach.* 2003;25(2):131-135. DOI: 10.1080/0142159031000092508

Bitte zitieren als

Iblher P, Zupanic M, Härtel C, Heinze H, Schmucker P, Fischer MR. Der Fragebogen "SFDP26-German": Ein verlässliches Instrument zur Evaluation des klinischen Unterrichts? *GMS Z Med Ausbild.* 2011;28(2):Doc30.

DOI: 10.3205/zma000742, URN: urn:nbn:de:0183-zma0007429

Artikel online frei zugänglich unter

<http://www.egms.de/en/journals/zma/2011-28/zma000742.shtml>

Korrespondenzadresse:

Dr. med. Peter Iblher, MME
Universität zu Lübeck, Klinik für Anästhesiologie,
Ratzeburger Allee 160, 23538 Lübeck, Deutschland, Tel.:
+49 (0)451/500-2766, Fax: +49 (0)500-3405
peter.iblher@uk-sh.de

Eingereicht: 30.11.2010

Überarbeitet: 07.02.2011

Angenommen: 23.03.2011

Veröffentlicht: 16.05.2011

Copyright

©2011 Iblher et al. Dieser Artikel ist ein Open Access-Artikel und steht unter den Creative Commons Lizenzbedingungen (<http://creativecommons.org/licenses/by-nc-nd/3.0/deed.de>). Er darf vervielfältigt, verbreitet und öffentlich zugänglich gemacht werden, vorausgesetzt dass Autor und Quelle genannt werden.

The Questionnaire "SFDP26-German": a reliable tool for evaluation of clinical teaching?

Abstract

Aims: Evaluation of the effectiveness of clinical teaching is an important contribution for the quality control of medical teaching. This should be evaluated using a reliable instrument in order to be able to both gauge the status quo and the effects of instruction. In the Stanford Faculty Development Program (SFDP), seven categories have proven to be appropriate:

1. Establishing the Learning Climate,
2. Controlling a Teaching Session,
3. Communication of Goals,
4. Encouraging Understanding and Retention,
5. Evaluation,
6. Feedback and
7. Self-directed Learning.

Since 1998, the SFDP26 questionnaire has established itself as an evaluation tool in English speaking countries. To date there is no equivalent German-language questionnaire available which evaluates the overall effectiveness of teaching.

Question:

1. Development and theoretical testing of a German-language version of SFDP26 (SFDP26-German),
2. Check the correlation of subscale of SFDPGerman against overall effectiveness of teaching.

Methods: 19 anaesthetists (7 female, 12 male) from the University of Lübeck were evaluated at the end of a teaching seminar on emergency medical care using SFDP-German. The sample consisted of 173 medical students (119 female (68.8%) and 54 male (31.2%), mostly from the fifth semester (6.6%) and sixth semester (80.3%). The mean age of the students was 23±3 years.

Results: The discriminatory power of all items ranged between good and excellent (r_{it} =0.48-0.75). All subscales displayed good internal consistency (α =0.69-0.92) and significant positive inter-scale correlations (r =0.40-0.70). The subscales and "overall effectiveness of teaching" showed significant correlation, with the highest correlation for the subscale "communication of goals ($p < 0.001$; $r = 0.61$).

Conclusion: The analysis of SFDP26-German confirms high internal consistency. Future research should investigate the effectiveness of the individual categories on the overall effectiveness of teaching and validate according to external criteria.

Keywords: SFDP26, clinical teaching, teaching effectiveness, evaluation, questionnaire

Peter Iblher^{1,2}

Michaela Zupanic²

Christoph Härtel³

Hermann Heinze¹

Peter Schmucker¹

Martin R. Fischer²

1 Universität zu Lübeck, Klinik für Anästhesiologie, Lübeck, Deutschland

2 Universität Witten/Herdecke, Fakultät für Gesundheit, Institut für Didaktik und Bildungsforschung im Gesundheitswesen (IDBG), Witten, Deutschland

3 Universität zu Lübeck, Klinik für Kinder- und Jugendmedizin, Lübeck, Deutschland

Introduction

Introduction of a new method

The improvement of teaching undoubtedly has a central role in German medical universities [1]. Raising awareness in the teaching institutes and clinics through incentive programs, such as performance-based funding (LOM) and teaching awards play essential role to this end [2]. Feedback systems, which give feedback to the teaching staff and departments on the teaching performance using standardised evaluation tools are an essential prerequisite for reproducible qualitative assessment of teaching performance. The discussion, especially in technical committees, contributes to the fact that in addition to research patient care, the formerly somewhat neglected field of teaching moves the fore and is perceived as a location advantage for team motivation and recruitment of young talent [1], [3], [4].

The establishment of learning objectives-based curricula and the use of appropriate forms of examination [5], the training of teachers undoubtedly plays a crucial role [6], [7]. A survey of doctors at the clinic of the University of Lübeck found that the perception of good teaching was almost always associated with exceptional individuals who were perceived as good teachers for a wide variety of reasons [8]. Despite this diversity, using definitions of "non-cognitive attributes" of a good teacher, such as relationship skills, mood and personality, established by Sutkin et al it was possible to structure these [9]. This raises the question of which educational competencies are particularly relevant for good and effective teaching performance. Number and acceptance of local and national training courses have increased significantly in recent years (Train-the-trainer concepts such as the National Master of Medical Education Programme of the Day of Medical Faculties (MFT) at the University of Heidelberg [10]) and has found its way into many post-doctoral training programs. However, it is often the case that providing training in teaching skills is particularly neglected in case of physicians working in clinical teaching, leading them to acquire these on their own through role models, experience and personal initiative [7], [11]. This means there are indeed opportunities to negatively impact the academic performance of students due to methodological, rhetorical or didactic weaknesses (in spite of subject-specific knowledge regarding course content) [12]. Consequently this means that

1. specific teaching skills of staff should be made accessible to review to identify weaknesses if necessary and improve such through training and
2. to study which categories of teaching performance were particularly relevant to good teaching in student evaluation.

It would therefore be desirable to have an established instrument which maps crucial areas of teaching competence, revealing individual strengths and weaknesses in teaching and which also can be used in the context of

educational research projects. To our knowledge, to date there is no German-language instrument which meets these requirements.

Stanford Faculty Development Program (SFDP)

In terms of a faculty development program, Stanford University, Palo Alto, U.S. established the training program (SFDP), based on learning theory, in the early eighties. Under this program, teaching staff were observed in class and their strengths and weaknesses in teaching grouped together under seven core competencies:

1. Establishing the learning climate,
2. Control of session,
3. Communication of goals,
4. Facilitating understanding and retention,
5. Evaluation,
6. Feedback,
7. Promoting self-directed learning.

This is one of the early development programs [13], [14] and hails from a conceptual framework which was developed from the late 1970s onwards under Dr. Kelley Skeff, an internal specialist at Stanford University and Dr. Georgette Stratos as part of their research on improving teaching. He initially filmed hundreds of hours of classroom teaching and correlated certain teaching methods and teacher behaviours with student assessments regarding their teaching effectiveness. In the course of numerous studies, a faculty development program was developed which made it possible for the faculty to analyse its teaching and at the same time improve the teaching competence of teachers in through standardised events [15], [16], [17], [18], [19], [20], [21], [22], [23], [24]. As part of this program, a reliable and dependable English-language instrument tool of teaching performance was developed on the basis of seven categories (SFDP26) and which has been widely introduced to training programs and medical educational research in the English speaking world [16], http://www.acgme.org/outcome/downloads/PBLI/15_pbli.pdf.

Problem

This study investigates two aspects of the SFDP26 questionnaire:

1. Developing and theoretical testing of the German language version of SFDP26 (SFDP26-German) to review teaching skills using item and reliability analyses to determine internal consistency (Cronbach's α);
2. Correlation analyses between the subscales and overall teaching effectiveness to verify the relevance of the individual subscales.

Methods

The SFDP26 questionnaire consists of 25 items in seven subscales

1. Establishing the learning climate,
2. Control of session,
3. Communication of goals,
4. Facilitating understanding and retention,
5. Evaluation,
6. Feedback,
7. Promoting self-directed learning and

one item on "Overall teaching effectiveness" as a global measure for teaching.

They are measured on a five-point Likert scale for item 1-25 (1-5=No-Somewhat-Yes) and for the item "Overall teaching effectiveness" (1=very weak to 5=excellent). The original English questionnaire SFDP26 was first translated from English into German by native speakers and then translated back into English (see Table 1). Seven female and twelve male physicians of the Clinic of Anaesthesiology at the University of Lübeck who were actively teaching in the emergencies course for students of the summer semester 2009 were evaluated at the end of their course by students using the questionnaire SFDP26-German. The sample was composed of 173 medical students, mostly from the fifth (8.7%) and sixth semester (80.3%) (6.6% higher semesters, 4.6% unspecified) who were randomly assigned to the instructors in course groups. The average age of the 119 women and 54 men was 23 ± 3 years. The Ethics Committee of the University of Lübeck had no ethical concerns regarding this study. Both students and teachers were informed in advance about the study and had the opportunity to refuse participation in this study. Consent was taken as implied through action. All data was analysed anonymously. To determine the item parameters, statistical analyses of mean, standard deviation and selectivity (after Pearson) were performed. The selectivity coefficient is, after Lienert, the correlation of tasks answered with the total value of the scale and a parameter for the degree of differentiation between an individual's agreement with the item with that of the scale as a whole [25]. In this, separation efficiencies are r' over 0.3 are seen as good, of 0.2-0.3 as acceptable, of 0.1-0.2 as marginal and as bad if below 0.1 [26]. To determine the scale parameters and inter-correlations, scale sums, standard deviation, uniformity coefficients (Cronbach's α) and corrected inter-scale correlations according to Pearson (r) were calculated using SPSS software Version 18 (SPSS Inc., Chicago, USA). For group comparison tests of reliability, Cronbach's α above 0.6 can be described as adequate and from 0.8 as good (27). Error probability of <5% ($p < 0.05$) was taken as significant, <0.1% ($p < 0.001$) as highly significant.

Results

Theoretical Review of SFDP26-German

The results of the item analyses are shown in Table 1. The most support was awarded to items 1 ("He/she listened to the students") and 3 ("He/she was respectful toward students") of the "Learning environment" scale with average values of $M=4.8 \pm 0.7$ and 4.7 ± 0.7 . Items 25 ("He/she encouraged students to read up on topics outside of class") of the "self-directed learning" scale and 15 ("He/she checked prior medical knowledge of students") of the "Evaluation" scale least highly rated with mean values of $M=3.4 \pm 0.9$ and 3.7 ± 1.2 .

In no case the discriminatory power of the items was below the critical value of 0.20, but invariably in the good to excellent range ($r \geq 0.60$).

The results of the scale analyses are shown in Table 2. Students gave their biggest endorsement on the "Learning environment" scale with a mean of 4.6 ± 0.6 . The endorsement was lowest on the "Self-directed learning" scale with a mean of 3.6 ± 1.1 . On almost all scales, high internal consistency was evident as a sign of reliability which can be evaluated as good (Cronbach's α : 0.80-0.89). The scale "Understanding and retention" showed an average internal consistency ($\alpha=0.69$) and the "Feedback" scale was in the very good range ($\alpha=0.92$). Highly significant correlations (r) of positive proportionality (see Table 3) were found between all scales of the German-language SFDP. The highest correlation was observed between the scales "Learning environment" and "Communication of goals" ($r=0.73$), and between "Understanding and retention" and "Communication of goals" ($r=0.72$). The shared variance ($r^2 \times 100$) of the two scales was 53% and 52%. The smallest but still significant positive correlations were found between the scales "Self-directed learning" and "Control of session" ($r=0.21$) and "Learning climate" ($r=0.28$). The shared variance here was 4% and 8%.

Table 2: Scale values for the German-language SFDP26 subscales (Σ = sum values of scales, SD = standard deviation) and correlations of subscales with the Global Rating "Overall teaching effectiveness" (r = correlation after Pearson).

German-language SFDP26 Subscales	Σ	SD	Cronbach's α	r
Learning environment	4.64	0.62	0.86	.511
Control of session	4.48	0.75	0.80	.443
Communication of goals	4.37	0.71	0.84	.611
Understanding and retention	4.23	0.83	0.69	.578
Evaluation	3.87	0.95	0.85	.471
Feedback	4.25	0.96	0.92	.479
Self-directed learning	3.64	1.13	0.89	.475

Table 3: Scale inter-correlations of the subscales of the German-language SFDP26

German-language SFDP26 Subscales	1	2	3	4	5	6
1 Learning environment	-					
2 Control of session	0.643	-				
3 Communication of goals	0.730	0.637	-			
4 Understanding and retention	0.556	0.506	0.722	-		
5 Evaluation	0.472	0.363	0.576	0.444	-	
6 Feedback	0.598	0.459	0.522	0.458	0.608	-
7 Self-directed learning	0.278	0.211	0.488	0.437	0.567	0.364

Table 1: Item parameters for the items of the German-language SFDP26 by subscales (1. - 8.) (M = mean, SD = standard deviation r_{it} = selectivity after Pearson). Measurements were done on a 5-point Likert scale (see text).

German-language SFDP26 subscales	M	SD	r_{it}
1. Establishing the learning environment			
He/she listens to the students.	4.75	0.68	.647
He/she encouraged students to actively participate in discussions.	4.57	0.79	.678
He/she was respectful towards students.	4.74	0.68	.631
He/she encouraged students to raise issues.	4.51	0.80	.619
2. Control of session			
He/she kept to the time plan.	4.58	0.94	.483
He/she avoided digressing from the course topic.	4.39	0.86	.633
He/she prevented external intrusions.	4.53	0.88	.654
3. Communication of goals			
He/she presented the course goals concisely and to the point.	4.36	0.88	.685
He/she drew the students' attention to the relevance of the learning goals.	4.62	0.78	.735
He/she had topic priorities in the course.	4.47	0.83	.735
He/she repeated the course learning goals frequently.	4.09	0.96	.700
4. Facilitating understanding and retention			
He/she presented the content in a well-structured manner.	4.37	0.82	.725
He/she presented the context of the material clearly.	4.45	0.76	.740
He/she used visual aids for demonstration purposes (blackboard, models, etc.).	3.98	1.38	.557
5. Evaluation			
He/she checked prior medical knowledge amongst students.	3.69	1.15	.641
He/she checked the students ability to deduct medical connections.	3.90	1.07	.652
He/she checked the students ability to apply medical knowledge to the specific patient in question.	4.02	1.08	.681
He/she check the practical skills of the students in case-based patient care.	4.02	1.20	.653
6. Feedback			
He/she gave negative (corrective) feedback to the students.	4.19	1.06	.672
He/she explained to the students what was right and what was wrong.	4.44	0.93	.676
He/she made suggestions to the students as to how they could improve.	4.31	0.99	.692
He/she regularly gave feedback to the students.	4.24	1.00	.745
7. Promoting self-directed learning			
He/she explicitly encouraged the students to look deeper into the topic.	3.76	1.18	.688
He/she motivated the students to learn independently.	3.81	1.16	.665
He/she encouraged the students to read up on the topic outside the course.	3.40	1.36	.510
8. Overall, the teaching performance of the teacher was ...	4.39	0.85	.676

Correlations of subscales with global rating (Overall teaching effectiveness)

The correlations of the subscales with the global overall assessment of teaching performance (Overall teaching effectiveness) are shown in Table 2. All correlations were highly significant ($p < 0.001$). Here, the highest correlations with overall teaching effectiveness were found for the subscales "Communication of goals" ($r = 0.61$), "Facilitating understanding and retention" ($r = 0.58$) and "Learning climate" ($r = 0.51$).

Discussion

SFDP26 is in fact a well-researched evaluation instrument [16], [28], so transfer of the English SFDP26 to the German-speaking countries makes sense and is an obvious step. The reliability of such an investigative tool is a crucial quality criterion which is described by the reliability. The results of this study confirms the high internal consistency of the German SFDP26, so the reliability of this questionnaire can be regarded as given. Selectivity

throughout was good so it was possible to capture the different aspects of staff teaching performance well. Furthermore, all subscales showed significant correlations of positive proportionality to the overall teaching effectiveness. This can be interpreted as meaning that the subscales are relevant aspects of overall teaching effectiveness. Beckman and colleagues, using a questionnaire based on SFDP but with new 16 items (Mayo Teaching Evaluation Form/MTEF-28), investigated the evaluation of teaching performance of junior doctors in clinical teaching at the Mayo Clinic [30]. For the subscales "Control of session," "Facilitating understanding and retention" and "Feedback," low internal consistencies (α : 0.147, 0.570, 0.648) were found. Litzelman and colleagues, who also used the original SFDP26 to evaluate teaching performance by junior doctors found high internal consistencies ($\alpha > .85$) for all scales [17]. This indicates that SFDP26 is a very reliable evaluation instrument, whose reliability may indeed be reduced by adding additional items, as with Beckmann and colleagues. In a study by Williams and colleagues, the subscales of SFDP26 were compared with the "Global Rating Scale" of the University of Michigan ("Please rate the educational value

of time with an attending physician.”), showing high correlations ($r:0.86-0.98$) [29], with the highest in the subscales “Communication of goals”, “Facilitating understanding and retention”, “Evaluation”, “Feedback” und “Promoting self-directed learning” (0.94-0.98). These results are surprising as this means that there is an almost totally positive linear correlation between the scales; and between the subscales of SFDP26 and a global “single-item” scale. The correlations of the SFDP26 subscales with the total teaching effectiveness, and between the SFDP26 overall teaching effectiveness and the “Global Rating Scale” of the University of Michigan (validity measure) are unfortunately not shown.

The scale “Communication of goals” in this study on the German-language SFDP26 most highly correlated with the overall teaching effectiveness ($r=0.61$), 37% of the variance was explained. This means that when considering all subscales of the German-language SFDP26 the strongest association is between this scale and the overall teaching effectiveness. The issue of causality cannot be clarified in this study, however, the high correlation indicates that the scale “Communication of goals” is a relevant determinant and could have a major impact on the overall teaching effectiveness. The question about the importance of the individual subscales of the German-language SFDP26 on effective teaching ability should therefore be examined in more detail in future studies, for example to enable the development of targeted training measures for teaching effectiveness. It would be beneficial, in addition, to focus on the actual measurable gains in student competence which cannot be answered solely using the questionnaire. The comparison of the examination success of students and their performance in the clinical context as external criteria with the results of the German-language SFDP26 would then also allow conclusions about the validity of the questionnaire.

It should be mentioned critically that SFDP26 has no arrangements to prevent that response tendencies from students occur, e.g. that only the same number sequence is ticked. In our research we have found no indication that this could play a role but this issue needs to be critically evaluated in analyses. Also, the fact that the data will be collected only by means of quantitative elements must be discussed critically. Although the analyses confirm the high reliability of the SFDP26 scales, adding qualitative evaluation elements, such as the option of free text answers, could make sense to allow more detailed assessments. The arrangement of the stages on the Likert scale from one to five is also not without problems. While the captions of the full scale (“No - Somewhat - Yes”) provide some guidance, it is nonetheless possible that German students will unconsciously follow the school grading system used in German-speaking countries, leading to an accidental reversal when checking (i.e. 1=very good, etc.). This could be avoided by reversing the labelling. Despite the good results of this study, we should qualify that this is an investigation of a special collective amongst students of medicine. Studies on the application of the German-language SFDP26 in different subjects

and at different points in time are therefore required to verify its general applicability.

Conclusions

This study for the first time investigates the German-language version of SFDP26 in clinical teaching of medical students and certifies that the procedure is sufficiently reliability for use in student evaluation of teaching. Follow-up studies in other clinical areas are necessary to illuminate the importance of the subscales and validity of the questionnaire by comparison with external criteria.

Acknowledgements

The authors would like to thank all the teaching staff and the students who participated in this study, in particular Joe Letkeman for the translation of the questionnaire.

Competing interests

The authors declare that there are no financial or other types of conflict of interest.

References

1. Wissenschaftsrat. Empfehlungen zur Qualitätsverbesserung von Lehre und Studium. Berlin: Wissenschaftsrat; 2008; Zugänglich unter/available from: <http://www.wissenschaftsrat.de/download/archiv/8639-08.pdf>
2. Wissenschaftsrat. IV.2. Anerkennung von besonderen Leistungen in der Lehre. In: Empfehlungen zur Qualitätsverbesserung von Lehre und Studium. Köln: Geschäftsstelle des Wissenschaftsrates; 2008.
3. Hahn EG, Fischer MR. Nationaler Kompetenzbasierter Lernzielkatalog Medizin (NKLM) für Deutschland: Zusammenarbeit der Gesellschaft für Medizinische Ausbildung (GMA) und des Medizinischen Fakultätentages (MFT). GMS Z Med Ausbild. 2009;26(3):Doc35. DOI: 10.3205/zma000627
4. Lammerding-Köppel M, Fabry G, Hofer M, Ochsendorf F, Schirlo C. Hochschuldidaktische Qualifizierung in der Medizin: I. Bestandsaufnahme: Ein Positionspapier des GMA-Ausschusses Personal- und Organisationsentwicklung für die medizinische Lehre der Gesellschaft für Medizinische Ausbildung sowie des Kompetenzzentrums für Hochschuldidaktik in Medizin Baden-Württemberg. GMS Z Med Ausbild. 2006;23(4):Doc73. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2006-23/zma000292.shtml>
5. Fischer MR, Gesellschaft für Medizinische Ausbildung, GMA-Ausschuss Prüfungen, Kompetenzzentrum Prüfungen Baden-Württemberg. Leitlinie für Fakultätsinterne Leistungsnachweise während des Medizinstudiums: Ein Positionspapier des GMA-Ausschusses Prüfungen und des Kompetenzzentrums Prüfungen Baden-Württemberg. GMS Z Med. 2008;25(1):Doc74. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2008-25/zma000558.shtml>
6. Geraci SA, Kovach RA, Babbott SF, Hollander H, Buranosky R, Devine DR, Berkowitz L. AAIM Report on Master Teachers and Clinician Educators Part 2: faculty development and training. Am J Med. 2010;123(9):869-872 e6.

7. McLean M, Cilliers F, Van Wyk JM. Faculty development: yesterday, today and tomorrow. *Med Teach.* 2008;30(6):555-584. DOI: 10.1080/01421590802109834
8. Iblher P. Lernen zu Lehren – Implementierung eines vorklinischen Wahlfaches zur frühen Förderung von Lehrkompetenzen an der Universität zu Lübeck-Unveröffentlichte Projektarbeit. Lübeck: Universität zu Lübeck; 2008.
9. Sutkin G, Wagner E, Harris I, Schiffer R. What makes a good clinical teacher in medicine? A review of the literature. *Acad Med.* 2008;83(5):452-466. DOI: 10.1097/ACM.0b013e31816bee61
10. Jünger J, Fischer MR, Duelli R, Putz R, Resch F. Implementierung und Evaluation eines inter fakultären Master of Medical Education Programms. *Z Evid Fortbild Qual Gesundhwes.* 2008;102(10):620-627.
11. Busari JO, Prince KJ, Scherpbier AJ, Van Der Vleuten CP, Essed GG. How residents perceive their teaching role in the clinical setting: a qualitative study. *Med Teach.* 2002;24(1):57-61. DOI: 10.1080/00034980120103496
12. Morrison EH, Hafler JP. Yesterday a learner, today a teacher too: residents as teachers in 2000. *Pediatrics.* 2000;105(1 Pt 3):238-241.
13. Searle NS, Thompson BM, Friedland JA, Lomax JW, Drutz JE, Coburn M, Nelson EA. The prevalence and practice of academies of medical educators: a survey of U.S. medical schools. *Acad Med.* 2010;85(1):48-56. DOI: 10.1097/ACM.0b013e3181c4846b
14. Steinert Y, Mann K, Centeno A, Dolmans D, Spencer J, Gelula M, Prideaux D. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Med Teach.* 2006;28(6):497-526. DOI: 10.1080/01421590600902976
15. Litzelman DK, Stratos GA, Marriott DJ, Lazaridis EN, Skeff KM. Beneficial and harmful effects of augmented feedback on physicians' clinical-teaching performances. *Acad Med.* 1998;73(3):324-332. DOI: 10.1097/00001888-199803000-00022
16. Litzelman DK, Stratos GA, Marriott DJ, Skeff KM. Factorial validation of a widely disseminated educational framework for evaluating clinical teachers. *Acad Med.* 1998;73(6):688-695. DOI: 10.1097/00001888-199806000-00016
17. Litzelman DK, Westmoreland GR, Skeff KM, Stratos GA. Factorial validation of an educational framework using residents' evaluations of clinician-educators. *Acad Med.* 1999;74(10 Suppl):S25-S27.
18. Skeff KM. Evaluation of a method for improving the teaching performance of attending physicians. *Am J Med.* 1983;75(3):465-470. DOI: 10.1016/0002-9343(83)90351-0
19. Skeff KM, Campbell M, Stratos G, Jones HW, 3rd, Cooke M. Assessment by attending physicians of a seminar method to improve clinical teaching. *J Med Educ.* 1984;59(12):944-950.
20. Skeff KM, Stratos G, Campbell M, Cooke M, Jones HW. 3rd. Evaluation of the seminar method to improve clinical teaching. *J Gen Intern Med.* 1986;1(5):315-322. DOI: 10.1007/BF02596211
21. Skeff KM, Stratos GA, Bergen MR, Regula DP, Jr. A pilot study of faculty development for basic science teachers. *Acad Med.* 1998;73(6):701-704. DOI: 10.1097/00001888-199806000-00018
22. Skeff KM, Stratos GA, Bergen MR, Sampson K, Deutsch SL. Regional teaching improvement programs for community-based teachers. *Am J Med.* 1999;106(1):76-80. DOI: 10.1016/S0002-9343(98)00360-X
23. Skeff KM, Stratos GA, Berman J, Bergen MR. Improving clinical teaching. Evaluation of a national dissemination program. *Arch Intern Med.* 1992;152(6):1156-1161.
24. Skeff KM, Stratos GA, Mygdal W, DeWitt TA, Manfred L, Quirk M, Roberts K, Greenberg L, Bland CJ. Faculty development. A resource for clinical teachers. *J Gen Intern Med.* 1997;12(Suppl 2):S56-S63. DOI: 10.1046/j.1525-1497.12.s2.8.x
25. Lienert GA, Raatz U. Berechnung von Schwierigkeitsindex, Trennschärfenkoeffizient und Aufgabeninterkorrelation. In: Lienert GA, Raatz U, editors. *Testaufbau und Testanalyse.* Weinheim: Beltz, Psychologie-Verl.-Union; 1994. S.73-113.
26. Möltner A, Schellberg D, Jünger J. Grundlegende quantitative Analysen medizinischer Prüfungen. *GMS Z Med Ausbild.* 2006;23(3):Doc53. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2006-23/zma000272.shtml>
27. Bortz JD, Döring N. *Hypothesengewinnung und Theoriebildung.* In: Bortz JD, Döring N (Hrsg). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler.* Berlin, Heidelberg, New York, Tokio: Springer; 2006. S.355–936.
28. Litzelman DK, Westmoreland GR, Skeff KM, Stratos G. Student and resident evaluations of faculty-how dependable are they? *Acad Med.* 1999;74(Suppl.):s25-s27.
29. Williams BC, Litzelman DK, Babbott SF, Lubitz RM, Hofer TP. Validation of a global measure of faculty's clinical teaching performance. *Acad Med.* 2002;77(2):177-180. DOI: 10.1097/00001888-200202000-00020
30. Beckman TJ, Lee MC, Rohren CH, Pankratz VS. Evaluating an instrument for the peer review of inpatient teaching. *Med Teach.* 2003;25(2):131-135. DOI: 10.1080/0142159031000092508

Corresponding author:

Dr. med. Peter Iblher, MME
 Universität zu Lübeck, Klinik für Anästhesiologie,
 Ratzeburger Allee 160, 23538 Lübeck, Deutschland, Tel.:
 +49 (0)451/500-2766, Fax: +49 (0)500-3405
 peter.iblher@uk-sh.de

Please cite as

Iblher P, Zupanec M, Härtel C, Heinze H, Schmucker P, Fischer MR. The Questionnaire "SFDP26-German": a reliable tool for evaluation of clinical teaching? *GMS Z Med Ausbild.* 2011;28(2):Doc30. DOI: 10.3205/zma000742, URN: urn:nbn:de:0183-zma0007429

This article is freely available from

<http://www.egms.de/en/journals/zma/2011-28/zma000742.shtml>

Received: 2010-11-30

Revised: 2011-02-07

Accepted: 2011-03-23

Published: 2011-05-16

Copyright

©2011 Iblher et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by-nc-nd/3.0/deed.en>). You are free: to Share – to copy, distribute and transmit the work, provided the original author and source are credited.