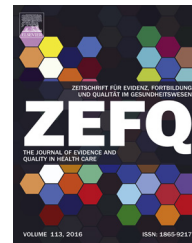




Available online at www.sciencedirect.com

ScienceDirect

journal homepage: <http://www.elsevier.com/locate/zefq>



SCHWERPUNKTREIHE / SPECIAL SECTION „WEITERBILDUNG IN DER ALLGEMEINMEDIZIN“

Adaptation, psychometric properties and feasibility of the Professionalism Scale Germany



Adaptation, psychometrische Eigenschaften und Anwendbarkeit der Professionalitäts-Skala Deutschland

Marco Roos^{1,*}, David Pfisterer², Doris Krug², Thomas Ledig², Jost Steinhäuser³, Joachim Szecsenyi², Katja Goetz³

¹ Institute of General Practice, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany

² Department of General Practice and Health Services Research, University Hospital Heidelberg, Heidelberg, Germany

³ Institute of Family Medicine, University Hospital Schleswig-Holstein, Campus Lübeck, Lübeck, Germany

Submitted/eingegangen 17 December 2015; revised/überarbeitet 14 April 2016; accepted/akzeptiert 20 April 2016

KEYWORDS

General practice;
postgraduate
training;
professionalism;
psychometric
properties;
feedback

Summary

Introduction: Pre- and postgraduate education is meant to be competency-based. Over the last two decades various competency frameworks have been published. An important aspect of competency is professionalism, being discussed widely in the literature while a clear-cut definition is still lacking. The purpose of this study was to translate the Nijmegen Professionalism Scale into the German language, to adapt the scale to the German setting and to examine the psychometric properties, test-retest reliability and feasibility of the culturally adapted instrument designed to assess professionalism in general practice, in addition to the validity of the concept of professionalism and to testify the transfer across linguistic, cultural and societal differences.

* Corresponding author: Dr. med. Marco Roos, MD, Institute of General Practice, Friedrich-Alexander University Erlangen-Nürnberg, Universitätsstraße 29, 91054 Erlangen, Germany.
E-Mail: marco.roos@uk-erlangen.de (M. Roos).

Method: After translating the Nijmegen Professional Scale into German, we conducted its cultural adaptation, the German Professionalism Scale (Pro-D). Its psychometric properties were assessed using Cronbach's α , descriptive statistics, and test-retest reliability. The validation of the construct was analysed by confirmatory factor analysis. Feasibility was confirmed in interviews with GP trainees and their trainers.

Results: A total of 133 trainees completed the Pro-D. The Pro-D showed high internal consistency (Cronbach's α 0.93) and good test-retest reliability (Spearman's rank correlation and Wilcoxon's matched-pairs test) for the different domains. Confirmatory factor analysis was unable to establish construct validity. Change in sensitivity of the instrument was good. Statements of interviews confirmed the feasibility of the new instrument.

Conclusions: We found good psychometric properties for the Pro-D. This might indicate transferability of the concept across linguistic, cultural and societal differences although the concept of professionalism was not replicated in a confirmatory factor analysis.

SCHLÜSSELWÖRTER

Allgemeinmedizin;
Weiterbildung;
Professionalität;
Psychometrische
Eigenschaften;
Rückmeldung

Zusammenfassung

Hintergrund: Die medizinische Ausbildung wird zunehmend kompetenzbasiert. In den letzten beiden Jahrzehnten wurden daher unterschiedliche Kompetenzmodelle veröffentlicht. Ein Kompetenzfeld ist das der Professionalität, wobei bisher noch keine einheitliche Definition für dieses Kompetenzfeld gefunden wurde. Ziel der hier berichteten Studie war die Übersetzung und Adaptation der „Nijmegen Professionalism Scale“, ein Instrument zur Beurteilung von Professionalität, auf die deutsche Weiterbildungssituation sowie die Überprüfung der psychometrischen Eigenschaften, der Test-Retest-Reliabilität und Anwendbarkeit des neuen Instruments in der Weiterbildung. Zusätzlich sollte die Validität des theoretischen Konstrukts des Instruments überprüft werden. Damit sollte ein Beispiel für den Transfer von Instrumenten zur Überprüfung des Kompetenzfelds über sprachliche und kulturelle Grenzen hinweg gezeigt werden.

Methoden: Es wurde eine Übersetzung und kulturelle Adaptation der „Nijmegen Professional Scale“ in die deutsche Sprache (Professionalitäts-Skala Deutschland, Pro-D) durchgeführt. Die psychometrischen Eigenschaften wurden mittels Cronbachs α , deskriptiver Statistik und der Betrachtung der Test-Retest-Reliabilität geprüft. Zur Validierung des theoretischen Konstrukts des Instruments wurde eine konfirmatorische Faktorenanalyse durchgeführt. Die Anwendbarkeit auf die deutsche Weiterbildungssituation wurde in Gruppeninterviews mit Ärzten in Weiterbildung und deren Weiterbildern evaluiert.

Ergebnisse: Insgesamt 133 Ärzte in Weiterbildung füllten das neue Instrument Pro-D aus. Die Ergebnisse zeigten ein hohes Maß an interner Konsistenz (Cronbach α 0,93) und eine gute Test-Retest-Reliabilität (Spearman's Rangsummen-Korrelation und der Wilcoxon-Vorzeichen-Rang-Test) für das neue Instrument an. Eine konfirmatorische Faktorenanalyse konnte das theoretische Konstrukt nicht bestätigen. Das Instrument zeigte eine gute Veränderungssensitivität. Die durchgeführten Interviews bestätigten die Anwendbarkeit in der deutschen Weiterbildung.

Schlussfolgerungen: Die Pro-D weist gute psychometrische Eigenschaften auf. Die Bestätigung des theoretischen Konstrukts in der konfirmatorischen Faktorenanalyse schlug fehl. Dennoch kann diese Studie als Hinweis für den Transfer eines Instruments zur Überprüfung des Kompetenzfeldes über sprachliche und kulturelle Grenzen hinweg angesehen werden.

Introduction

For more than two decades, professionalism has emerged as a substantial and sustained theme within the medical society [1–3]. Health delivery systems worldwide are facing the same challenges because of shifting priorities including patients' demands, societal requirements, financial struggles and governance [4]. Concepts, future demands and ideas regarding a definition of professionalism are changing [5–7]. Considering that a great number of studies have addressed the topic of professionalism, a definition remains complex and general best practices approaches for assessment even more so [8–11]. In current discussions, professionalism is understood as a complex and multi-dimensional construct. Further ideas on assessment

of professionalism therefore require considerations of its individual, inter-personal, societal and cultural dimensions [7,12,13].

In 2004, a group of researchers in the Netherlands conceptualized professionalism for general practice and developed an instrument for assessing professional behaviour in general practitioner trainees, the Nijmegen Professionalism Scale [14–16]. The instrument should provide both a possibility for trainees' self-assessment as well as an instrument for formative trainee assessment. The instrument consists of 93 items and conceptualizes professionalism as professional behaviour within four domains: professionalism towards the patient (25 items), professionalism towards other professionals (19 items), professionalism towards society (17 items) and professionalism

towards oneself (32 items). All domains showed good internal consistency with Cronbach's alpha coefficients ranging from 0.72 to 0.95 and reliability from 0.78 to 0.95. Nonetheless, this construct was not replicated in confirmatory factor analysis so far [14].

The goal of this study was to translate and adapt this instrument to the German situation and examine the psychometric properties, test-retest reliability and the feasibility of the cultural adapted German instrument. Another goal was to examine the validity of the theoretical construct of professional behaviour in a confirmatory factor analysis. Finally, this study is an attempt to transfer a concept across linguistic, cultural and societal differences.

Methods

We performed an observational study within the program 'Verbundweiterbildung^{plus}' (a vocational training program for general practice in Baden-Wuerttemberg, a federal state of Germany, www.weiterbildung-allgemeinmedizin.de) [17]. We invited all GP trainees within the program to participate. The study was funded by the young scientist programme of the German network 'Health Services Research Baden-Württemberg' of the Ministry of Science, Research and Arts in collaboration with the Ministry of Employment and Social Order, Family, Woman and Senior Citizens, Baden-Württemberg, Germany.

Translation and Cultural Adaptation

To adapt the Nijmegen Professionalism Scale we followed the Principles of Good Practice for the Translation and Cultural Adaptation Process by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) task force [18] as follows: We obtained permission from the authors of the Nijmegen-Professional-Scale, Tromp et al. from Radboud University Nijmegen Medical Centre, to translate and adapt a German version of the instrument [14]. Two linguistic experts independently translated the Nijmegen-Professionalism-Scale (93 items) into German. Divergent translations were discussed in a consensus meeting with a GP trainee, a GP trainer and a researcher. The cultural adaptation of the translated items (93 items) was proven in a think-aloud technique with two GP trainers and two GP trainees. They were asked to go through the German instrument and thinking aloud on anything coming into their minds as they completed the items. In a second step they were asked to evaluate all items for their relevance in a German general practice setting [19]. Items with less than three votes for relevance were removed from the questionnaire: For example, items like 'able to influence specialist care (e.g. during consultation at hospital visits)', not transferable to the German situation. At the end of that process, the German questionnaire Professionalism-Scale-Deutschland (Pro-D) consisted of 67 items.

Recruitment and Data Collection

As a target population we defined all GP trainees within the vocational training program 'Verbundweiterbildung^{plus}'

Table 1 Participant Characteristics.

Participants Characteristics	Respondents (n = 133)
Age	
years, mean (range)	33 (25-53)
Gender	
female, n (%)	88 (66.2)
male, n (%)	45 (33.8)
Duration of vocational training	
years, mean (range)	3 (1-8)
Sample	
web-based, n (%)	62 (46.6)
paper-based, n (%)	71 (53.4)

(266 at date 01/2013). Recruitment took place in two ways. First, an invitation to a web-based version of the questionnaire was sent per email, followed by two reminders being sent each following week. In total, the web-based version was available for four weeks, yet the response rate was low. So, secondly, we asked GP trainees to fill out a paper-based questionnaire at different teaching sessions, regular offered in this vocational training program (T0). Written informed consent was obtained from each participant. Of 266 trainees invited to participate in the study, 133 (50.0%) returned a completed questionnaire at T0. These 133 participants were invited to take part in the second measurement, where a postage free enveloped questionnaire was sent out four weeks later (T1) for test-retest reliability. Another 50 (37.6%) returned a completed questionnaire at T1. Two reminders for test-retest were sent out per email after 7 and 14 days. All questionnaires were depersonalized using an individual, reproducible coding. Two thirds were female with a mean age of 33 years. The characteristics of our study population are presented in Table 1.

Measures

The Pro-D consists of 67 items, each representing an element of professional behaviour. Following the Nijmegen-Professionalism-Scale, the instrument consists of four domains addressing professionalism: professionalism towards the patient (21 items), professionalism towards other professionals (14 items), professionalism towards society (10 items) and professionalism towards oneself (22 items). Each item is rated by a 4-point Likert-scale ranging from 'seldom or never' to 'always'. Additionally, there is the possibility to mark 'leave blank' for items that do not apply to the individual level of training (at time of measurement). In addition to the Pro-D we measured sociodemographic data to describe the sample. This included questions regarding age, gender and duration of vocational training.

Statistical analysis

All data was analysed using SPSS 20.0 (IBM Corp., New York, USA) and R statistics 2.15.2 software (The R Project

for Statistical Computing, www.r-project.org). For descriptive analyses items were encoded with 1 equals 'leave blank', 2 equals 'seldom or never' to 5 equals 'always'. The reliability was assessed by using Cronbach alpha which indicates whether an item of scale is appropriate for assessing the underlying concept of its scale [20]. Values for Cronbach alpha range from 0 to 1. The closer to 0, the least related are the items to one another. Values above 0.6 indicate a satisfying internal consistency, values above 0.8 indicate a high internal consistency. Additionally, Guttman-split-half coefficient for reliability was tested.

For test-retest reliability we chose the nonparametric Spearman rank order correlation coefficient (r) to determine the stability of the questionnaire. This criterion refers to the likelihood that a test will yield the same description of a phenomenon if the test is repeated and the phenomenon is unchanged [21]. Retest reliability is defined as the correlation between two tests ratings. Spearman rank scores range from -1 to 1, where a score of 1 indicates the highest correspondence, r values often range between 0.2 and 0.6, rarely above; correlations between 0.4 and 0.6 are considered an acceptable correlation and to be very reliable [22]. However, reliability also depends on the expected stability of the investigated construct. The non-parametric Wilcoxon matched paired test was used to test for differences between T0 and T1. If no significant differences were detected, the stability of the construct could be assumed. For changes in sensitivity correlation between level of training (duration of vocational training, Table 1) and sum score of items is reported (Pearsons correlation coefficients and Spearman rho). The level of significance was $p \leq 0.05$.

To examine the construct validity of the theoretical framework we performed a confirmatory factor analysis based on the model of the Nijmegen-Professionalism-Scale [14,15]. We defined a model with Professionalism as a latent variable over four latent variables (the four domains mentioned above). These are represented by the amount of observable variables (items). Afterwards we performed a recommended model fitting [23,24]. Second, we reported different (recommended) fit-indices and the development of fit-indices by model fitting. We used absolute fit-indices like the Chi-square (χ^2) with degrees of freedom (df) and relative Chi-square (χ^2/df), the Adjusted Goodness of Fit Index (AGFI), the Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR). The Chi-square value should be as low as possible and the relative Chi-square should show a relation of 2:1 [25,26]. Values for the AGFI above 0.8 are acceptable, above 0.95 represent good model fit [24]. RMSEA and SRMR indicate good model fit with values below 0.07 and 0.08 respectively [24,27]. Additionally, we reported Bentler's Comparative Fit Index (CFI) and the Non-Normed Fit Index (NNFI) as relative fit-indices. Both indicate good model fit with values above 0.95 [28].

Feasibility

Feasibility was tested with qualitative group interviews with pairs of GP trainees and their GP trainer. Both were asked to

fill out the questionnaire before the interview. GP trainees should use the questionnaire for self-assessment, whereas GP tutors use the questionnaire for observing their trainees. Interviews were guided by questions about the relevance of the questionnaires' content, feasibility of the questionnaire in daily practice, ideas for improvement and an overall impression of the questionnaire. All interviews were recorded and transcribed. The interviews were analysed via content analysis according to Mayring [29] supported by the software Atlas.ti 5.2.17 (Scientific Software Developing GmbH). Three independent researchers coded the interviews. The assigned codes and categories were matched in a consensus meeting.

Results

In general, internal consistency (α) of all 67 items was high with a score of 0.93. We found five items (items 1.8 and 1.17 in domain "professional behaviour towards the patient"; items 4.10, 4.12 and 4.14 on the domain "professionalism towards oneself", Table 2) which show ceiling effects (kurtosis between 7.90 and 15.90, skew between -2.37 and -3.41). No bottom effects were found.

Psychometric properties

Table 3 shows the results for the different domains. The internal consistency was determined to be above 0.80 for all of them. Guttman-split-half reliability coefficients showed good results around 0.75, except domain 2 with an acceptable result of 0.45. Results of test-retest reliability with Spearman rho and Wilcoxon matched pair test showed good results.

Table 2 showed more detailed information of each item. Within the domain "professional behaviour towards the patient" correlation coefficients (r) for single items showed a range from 0.37 and 0.75, except item 8 and 10 with correlation coefficients of 0.18 and 0.23, respectively. Item 6 showed significant difference in matched pair test ($p=0.01$). Within domain "professional behaviour towards other professionals" item 3 showed a correlation coefficient of -0.05. Other correlation values ranged from 0.33 to 0.65. Matched pair test found no significant differences between items. At domain "professional behaviour towards society" items 3 and 4 showed values of 0.10 and 0.21 for test-retest reliability, item 5 had also a low value of 0.28, the p -value was tendentious significant ($p=0.05$). Additionally, item 8 showed a significant difference in the matched pair test. At domain "professional behaviour towards oneself" item 5 showed a correlation coefficient of 0.16. The other correlation coefficients for test-retest reliability ranged from 0.28 and 0.71. Matched pair test found no significant differences between items.

Construct validity/Sensitivity

The confirmatory factor analysis for a model based on all 67 items showed a poor model fit as presented in Table 4. After model fitting we found a noticeably better model fit with a shortened questionnaire consisting of 25 items (Table 2-5).

Table 2 Results of items.

Item	Mean (SD) T0 (n = 133)	Mean (SD) T1 (n = 50)	Skew T0	Kurtosis T0	Test-retest reliability: Spearman rho		Wilcoxon matched pair test	
					r	p-value ^a	p-value ^a	
<i>The GP trainee...</i>								
Domain: professionalism towards the patient								
1.1	<i>Deals correctly with legislative rules regarding informed consent</i>	4.23 (0.86)	4.52 (0.65)	-0.91	0.03	0.47	<0.01	0.20
1.2	<i>Is able to bring up difficult subjects</i>	4.47 (0.61)	4.48 (0.58)	-0.68	-0.48	0.52	<0.01	0.44
1.3	<i>Respects the right of patients to inspect their medical records</i>	4.75 (0.45)	4.80 (0.40)	-1.43	0.72	0.63	<0.01	0.71
1.4	<i>Is able to show sympathy</i>	4.61 (0.52)	4.72 (0.45)	-0.78	-0.69	0.40	<0.01	0.41
1.5	<i>Takes patients' embarrassment, shyness and reluctance into account</i>	4.30 (0.70)	4.36 (0.56)	-1.17	3.21	0.59	<0.01	0.78
1.6	<i>During physical examinations, explains the aim of the procedures and what is expected of the patient</i>	4.20 (0.89)	4.54 (0.58)	-0.98	0.55	0.63	<0.01	0.01
1.7	<i>Approaches patients with a different frame of reference (e.g. religion) openly</i>	4.29 (0.66)	4.28 (0.57)	-0.56	0.06	0.39	<0.01	0.83
1.8	<i>Looks clean and tidy and dresses according to current norms</i>	4.71 (0.64)	4.70 (0.46)	-3.41	15.90	0.18	0.22	0.97
1.9	<i>Adjusts language to communicate with patients</i>	4.40 (0.70)	4.36 (0.56)	-1.41	3.81	0.75	<0.01	0.37
1.10	<i>Takes sex specific differences into account</i>	4.44 (0.67)	4.40 (0.70)	-1.42	4.11	0.23	0.11	0.37
1.11	<i>Is able to cope with the different expectations that patients have of their GP</i>	3.86 (1.05)	3.70 (1.11)	-1.49	2.13	0.37	<0.01	0.89
1.12*	<i>Involves the previous history of the patient in the provision of care</i>	4.37 (0.77)	4.30 (0.58)	-1.65	4.29	0.44	<0.01	0.65
1.13*	<i>Pays attention to the consequence of treatment policy on the daily functioning of the patient</i>	3.35 (1.10)	3.08 (1.23)	-0.53	-0.35	0.52	<0.01	0.34
1.14*	<i>Involves relevant aspects of the patient's home and environment in the provision of care</i>	4.41 (0.75)	4.32 (0.65)	-1.70	4.85	0.54	<0.01	0.98
1.15*	<i>Retains insight into the medical history of the patients in order to act proactively if necessary</i>	4.23 (1.17)	3.96 (1.37)	-1.84	2.59	0.57	<0.01	0.72
1.16*	<i>If necessary, takes action after life events</i>	4.18 (0.93)	4.16 (0.91)	-1.81	4.15	0.37	<0.01	0.11
1.17*	<i>Respects patients' self-determination</i>	4.71 (0.65)	4.70 (0.46)	-3.35	14.92	0.49	<0.01	0.80
1.18	<i>Deals carefully with professional secrecy when talking to colleagues or acquaintances</i>	4.66 (0.56)	4.62 (0.49)	-1.69	3.31	0.63	<0.01	0.32
1.19	<i>Does not give patients false hope</i>	4.18 (0.68)	4.26 (0.72)	-0.97	2.99	0.49	<0.01	0.85
1.20	<i>Takes care not to become too involved in the patient's emotions</i>	3.88 (0.62)	3.96 (0.49)	-0.32	0.62	0.52	<0.01	0.17
1.21	<i>Takes care not to be influenced by patients of high social status</i>	4.02 (0.68)	4.02 (0.65)	-0.18	-0.34	0.55	<0.01	0.25
Domain: professionalism towards other professionals								
2.1	<i>Consults other care providers with targeted questions</i>	4.20 (0.89)	4.42 (0.91)	-1.01	0.97	0.34	0.02	0.15

Table 2 (Fortsetzung)

Item	Mean (SD) T0 (n = 133)	Mean (SD) T1 (n = 50)	Skew T0	Kurtosis T0	Test-retest reliability: Spearman rho		Wilcoxon matched pair test	
					r	p-value*	p-value*	
<i>The GP trainee...</i>								
2.2	<i>Ensures structured information transfer with other care providers</i>	3.86 (1.02)	3.76 (1.13)	-0.64	0.04	0.35	0.01	0.93
2.3	<i>Deals correctly with targeted questions from other care providers</i>	4.18 (0.94)	4.00 (1.16)	-1.77	3.89	-0.05	0.72	0.85
2.4	<i>Is able to motivate support personnel</i>	4.20 (0.69)	4.16 (0.65)	-0.56	0.27	0.50	0.01	0.54
2.5	<i>Makes clear agreements with support personnel</i>	4.26 (0.75)	4.30 (0.58)	-1.35	3.89	0.41	<0.01	0.82
2.6	<i>Listens to the contributions of support personnel</i>	4.54 (0.62)	4.70 (0.46)	-1.79	6.59	0.33	0.02	0.13
2.7	<i>Transfer services correctly</i>	4.23 (1.01)	4.32 (0.87)	-2.08	4.49	0.43	0.02	0.31
2.8	<i>Discusses bottlenecks in cooperation with others directly</i>	3.98 (0.72)	4.04 (0.70)	-0.46	0.27	0.59	<0.01	0.20
2.9	<i>Is able to deal constructively with conflicts</i>	4.05 (0.64)	3.92 (0.67)	-0.39	0.68	0.61	<0.01	1.00
2.10*	<i>Is able to manage the mutual demarcation of tasks between GP and specialists</i>	3.74 (1.32)	3.84 (1.38)	-1.17	0.20	0.47	<0.01	0.70
2.11*	<i>Ensures coherence in first and second line medical care</i>	3.74 (1.37)	3.68 (1.38)	-1.06	-0.10	0.38	<0.01	0.93
2.12*	<i>Is able to distinguish between professional and personal interests in negotiations</i>	3.56 (1.40)	3.70 (1.45)	-0.90	-0.48	0.49	<0.01	0.23
2.13*	<i>Is able to take policy decisions</i>	3.11 (1.53)	2.76 (1.53)	-0.32	-1.40	0.58	<0.01	0.66
2.14*	<i>Is able to conduct job evaluations</i>	3.07 (1.52)	2.50 (1.43)	-0.23	-1.44	0.65	<0.01	0.67
Domain: professionalism towards society								
3.1*	<i>Bears the consequences of his/her own conduct</i>	4.44 (0.63)	4.36 (0.75)	-1.43	5.18	0.36	<0.01	0.83
3.2*	<i>Is able to justify deviations from rules and guidelines</i>	3.92 (0.96)	3.78 (0.98)	-1.34	2.33	0.74	<0.01	0.69
3.3	<i>Is aware of his/her own norms regarding disease influence disease management</i>	3.94 (0.92)	3.66 (1.22)	-1.36	2.67	0.21	0.16	0.51
3.4*	<i>Is aware of the meaning and the relative value of scientific evidence in decision-making</i>	3.98 (0.89)	3.56 (1.25)	-1.29	2.77	0.10	0.51	0.48
3.5*	<i>In decision-making, weighs scientific evidence against factors related to the patient or the circumstances</i>	3.65 (1.18)	3.60 (1.25)	-0.90	0.20	0.28	0.05	0.46
3.6*	<i>Is able to justify choices made on the basis of scientific evidence</i>	3.79 (0.99)	3.66 (0.87)	-1.11	1.44	0.33	0.02	0.13
3.7*	<i>Is able to explain his/her own norms and values regarding the application of scientific evidence</i>	3.69 (1.12)	3.38 (1.21)	-0.99	0.58	0.40	<0.01	0.65
3.8*	<i>Is able to estimate which problems are suitable for a quality-improvement project</i>	3.77 (1.21)	3.36 (1.40)	-1.17	0.64	0.46	<0.01	0.02
3.9*	<i>Is able to work out a quality-improvement project</i>	3.11 (1.34)	2.68 (1.41)	-0.23	-1.08	0.52	<0.01	0.70
3.10*	<i>Is able to justify indications for making home visits</i>	3.35 (1.59)	3.40 (1.65)	-0.56	-1.32	0.67	<0.01	0.12
Domain: professionalism towards oneself								
4.1	<i>Is able to name reactions, thoughts and feelings that patients evoke</i>	4.23 (0.61)	4.22 (0.71)	-0.97	4.83	0.32	0.02	0.97

Table 2 (Fortsetzung)

Item	Mean (SD) T0 (n = 133)	Mean (SD) T1 (n = 50)	Skew T0	Kurtosis T0	Test-retest reliability: Spearman rho		Wilcoxon matched pair test	
					r	p-value*	p-value*	
<i>The GP trainee...</i>								
4.2	Asks questions about his/her own role in relationships (patient, team, gp, trainer, etc.)	4.44 (0.62)	4.42 (0.64)	-1.40	5.58	0.28	0.05	0.43
4.3	Uses specific practical situations as starting points for critical self-reflection	4.05 (1.10)	4.06 (1.04)	-1.30	1.34	0.65	<0.01	0.08
4.4*	Discusses his/her own shortcomings and failures without losing belief in his/her own competence	3.89 (0.94)	3.86 (0.64)	-1.20	2.02	0.35	0.01	0.29
4.5*	Makes a realistic estimation of his/her own strong and weak points	4.05 (0.68)	4.10 (0.68)	-0.66	2.10	0.16	0.27	0.24
4.6	Is able to balance work and private life	3.87 (0.87)	4.00 (0.81)	-0.44	-0.11	0.71	<0.01	0.83
4.7	Is able to mention aspects of work that increase satisfaction	4.24 (0.73)	4.40 (0.67)	-0.88	1.63	0.43	<0.01	0.24
4.8	Is able to deal with the possibility that a treatment decision may be unsuccessful	3.73 (0.98)	3.86 (0.83)	-1.01	1.25	0.37	<0.01	0.22
4.9	Adheres to agreements made during feedback	3.80 (1.45)	3.90 (1.54)	-1.07	-0.29	0.38	<0.01	0.72
4.10	Attaches importance to what others think about his/her behaviour	4.63 (0.65)	4.76 (0.85)	-2.40	8.40	0.34	0.02	0.71
4.11	Does not resist being judged	4.16 (0.94)	4.02 (1.12)	-1.25	1.58	0.54	<0.01	0.06
4.12	Has an enquiring mind (asks questions and takes initiatives)	4.65 (0.63)	4.80 (0.50)	-2.37	8.28	0.48	<0.01	0.13
4.13	Is able to admit his/her own mistakes	4.46 (0.57)	4.50 (0.51)	-0.46	-0.75	0.65	<0.01	0.32
4.14	Takes action to rectify his/her own mistakes	4.56 (0.73)	4.76 (0.48)	-2.40	7.90	0.35	<0.01	0.47
4.15	Withdraws from the consequences of his/her own mistakes	4.49 (0.70)	4.60 (0.57)	-1.69	4.36	0.37	<0.01	0.97
4.16	Is able to adapt and keep control of the situation if patients unexpectedly need to be seen during other activities	4.20 (0.84)	4.42 (0.54)	-1.95	5.749	0.54	<0.01	0.80
4.17	Recovers rapidly after an unpleasant consultation	4.08 (0.68)	3.92 (0.73)	-0.85	2.81	0.56	<0.01	0.30
4.18	Is able to let a mild disorder (e.g. tiredness) run its own course even though the correct diagnosis is a mystery	3.64 (1.06)	3.66 (1.26)	-0.97	0.74	0.32	0.02	0.38
4.19*	Is able to cope after making a mistake	4.04 (0.76)	4.10 (0.71)	-0.69	1.11	0.56	<0.01	0.28
4.20*	Is able to deal with difficult or angry patients	3.96 (0.74)	4.06 (0.55)	-0.73	1.48	0.51	<0.01	0.35
4.21*	Is able to conduct interventions that lead to decrease in aggression from the patient	3.77 (0.90)	3.84 (0.71)	-0.81	1.18	0.35	0.01	0.22
4.22*	Is able to formulate his/her own opinion in a clear and inoffensive manner	4.18 (0.68)	4.28 (0.64)	-0.53	0.35	0.41	<0.01	0.18

* Statistical significance of differences: $p \leq 0.05$

* Items of the solution with 25 items in confirmatory factor analysis (greyed cells)

Table 3 Results of domains.

	Domain	Cronbach alpha	Guttman-split-half reliability coefficient	Test-retest reliability Spearman rho		Wilcoxon matched pair test
				r	p-value*	p-value*
1	<i>Professional behaviour towards the patient</i>	0.81	0.75	0.78	<0.01	0.21
2	<i>Professional behaviour other professionals</i>	0.80	0.45	0.72	<0.01	0.26
3	<i>Professional behaviour towards society</i>	0.84	0.76	0.72	<0.01	0.82
4	<i>Professional behaviour towards oneself</i>	0.82	0.73	0.66	<0.01	0.09

* Statistical significance of differences: $p \leq 0.05$

Table 4 Fit indices of confirmatory factor analysis.

	Solution with 67 items	Solution with 25 items
Chi square (χ^2)	4327.57	591.45
Degrees of freedom (df)	2141	271
χ^2/df	2.02	2.18
p-value*	<0.01	<0.01
AGFI	0.43	0.69
RMSEA (90% CI)	0.09	0.08
SRMR	0.11	0.09
CFI	0.40	0.75
NNFI	0.38	0.72

Changes in sensitivity showed good correlation coefficients between level of training and sum score of the 67 items; spearman rho was 0.49 and Pearson correlation was 0.48.

Feasibility

We conducted 7 group interviews with pairs of GP trainees and their GP trainer. The answers to the question “How do you evaluate the feasibility of the Pro-D?” produced 616 codes which were assigned to four main categories: (a) overall impression (66 codes), (b) content of the Pro-D (126 codes), (c) applicability of the Pro-D (336 codes) and (d) usefulness of the Pro-D (88 codes). In general, all participants appreciated the usage of the Pro-D. Some even recommended a regular usage, although there were some critical voices on the length. In summary, the application in daily practice seemed to be easy. Especially the differences between self-assessment and observation (external assessment) by the GP trainer. Some even asked for more instruments like the Pro-D to guide feedback sessions between GP trainees and their tutors. More detailed results of the feasibility study are published elsewhere [30].

Discussion

The presented study provides psychometric support for the Professionalism-Scale-Germany (Pro-D). The results of our study confirm high internal consistency and reliability of the questionnaire and good test-retest reliability. We also could not replicate the original structure of the theoretical model given by the Nijmegen-Professionalism-Scale in a confirmatory factor analysis. Although, model fit may indicate a shortened version of the questionnaire. Feasibility of Pro-D in daily practice is easy and useful.

Compared with the Nijmegen-Professionalism-Scale, the Pro-D as an adapted German version shows comparable psychometric properties. We found almost the same scores for excellent internal consistency [14]. Additionally, the test-retest reliability showed stability in the construct and supported the original structure of four domains of professionalism, based on consensus and face validity [15,16]. To replicate the original (theoretical) structure of the concept of professionalism we performed a confirmatory factor analysis. The initial model fit including all 67 items of the questionnaire was poor. We found noticeable model fit for a shortened version of the questionnaire consisting of 25 items. Although we could not replicate the original structure, good results in psychometric properties might indicate that definitions and concepts of professionalism are comparable and instruments are transferable across linguistic, cultural and societal contexts [12,31,32]. Changes in sensitivity over time (level of training) emphasize these results.

Feasibility of the instrument is another crucial point, because assessment instruments need acceptance and practicability [10]. Professional behaviour is a complex construct and a final consensus has not been found so far [7,11,16]. Differing understanding of professionalism makes assessment and teaching problematic. In our qualitative results we found high concordance of the concept of professionalism. GP trainers and trainees associated the same meaning to the construct of professionalism and agreed on the content of the instrument and the response to it, which is an indispensable foundation for effective teaching and assessing. The initial version of the Pro-D consists of 67 items. The interviewees were mostly approving its length and content. They appreciated the input of the instrument as it provides more

specific information for feedback. Especially for personal improvement of professional behaviour, it is important to equip trainers with instruments to help review professional growth of their trainees for more specific feedback instead of leaving it up to a hidden curriculum [12]. However, some comments on the lengths of the instrument and some items' relevance to daily reality indicate the need for a shortened version of the instrument. The results of the confirmatory factor analysis might be a first step. A long and a short form of the instrument could meet these requirements. The long form supports professional growth of the trainee by self-assessment and allows the trainer to monitor and encourage specific professional behaviour. We recommend its usage in a regular longitudinal approach. However, the questionnaire might be a good instrument for screening and assessment of professional behaviour because correlation between level of training (duration of vocational training) and sum score of items were given.

Our results are informative for those interested in how to assess professional behaviour. The Pro-D is a questionnaire for assessment of professional behaviour in general practice. It can be used in a large number of practices without previous teaching and is one the one hand useful for trainees' self-assessment as well as an external formative assessment by the trainer. Therefore it supports formative information and individual development of trainees. The current version consists of 67 items, which each provides formative information on a specific behavioural aspect of professionalism. The usage of this instrument as a summative assessment tool is still limited. Although, changes in sensitivity of the sum score correlates highly with stage of training, a final replication of the original structure has been missing. But we found evidence in an improved model fit [23,24] of a shortened version. Therefore, it will be a further step to focus on research to replicate the original structure.

Our study included a convenient sample of GP trainees within the program 'Verbundweiterbildung^{plus}' (a vocational training program for general practice in Baden-Wuerttemberg, a federal state of Germany). Our results have to be interpreted considering a potential selection bias due to the program and moderate participation rate. Moderate participation rates in electronic and paper-based questionnaires are very common especially without (financial) incentives for the participants [33]. However, sample and distribution of age, gender and duration of vocational training are comparable to the study of assessment of the Nijmegen Professional Scale [14]. We conducted a classic qualitative method for testing feasibility [33]. In seven interviews with GP trainees and their trainer, we were able to achieve saturation of themes and ideas. Hence, we did not carry out more interviews. Interpretative validity was optimized and researcher bias was minimized by coding of three independent researchers [34].

Conclusion

Qualitative high standard development of competencies needs meaningful, reliable and valid instruments for assessment. Today these instruments should be transferable across linguistic and cultural contexts. In this study, we have adapted the Nijmegen-Professionalism-Scale for Germany.

The German version - Professionalism-Scale-Germany (Pro-D) - showed good psychometric properties and feasibility. However, confirmatory analysis did not replicate the original concept of Professionalism. The Pro-D serves as a tool to support self-assessment and formative assessment in trainees' daily routine in general practice. This study can serve as an example for comparable interventions to work on a future concept and definition of professionalism.

Funding

The study was funded by the young scientists programme of the German network 'Health Services Research Baden-Württemberg' of the Ministry of Science, Research and Arts in collaboration with the Ministry of Employment and Social Order, Family, Woman and Senior Citizens, Baden-Württemberg, Germany.

Conflict of Interest

All authors have read and approved the submission of this manuscript to your journal. All authors on this publication contributed to the study. The authors have no financial interests to disclose directly or indirectly related to the research in the manuscript.

Ethical approval

The study was fully approved by the ethics committee of the Medical Faculty of the University of Heidelberg (approval number S-513/2011).

Acknowledgements

The authors thank Mrs. Friederike Böhlen and Mrs. Sanne Custers for the translation of the Nijmegen Professional Scale to German language. Furthermore, the authors thank Daniel Nittka, Henrik Lamers, Peter Engeser and Thomas Kühlein for their participation in the Think-aloud technique. Finally, the authors thank all participants of the study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.zefq.2016.04.002](https://doi.org/10.1016/j.zefq.2016.04.002).

References

- [1] Foundation ABIM. European Federation of Internal Medicine, Medical professionalism in the new millennium: a physicians' charter. *Lancet* 2002;359:520–2.
- [2] Arnold L. Assessing professional behaviour: yesterday, today, and tomorrow. *Acad Med* 2002;77:502–15.
- [3] Collier R. Professionalism: what is it? *Can Med Assoc J* 2012;184:1129–30.
- [4] Borgstrom E, Cohn S, Barclay S. Medical professionalism: conflicting values for tomorrow's doctors. *J Gen Intern Med* 2010;25:1330–6.

- [5] Bryden P, Ginsburg S, Kurabi B, Ahmed N. Professing professionalism: are we our own worst enemy? Faculty members' experiences of teaching and evaluating professionalism in medical education at one school. *Acad Med* 2010;85:1025–34.
- [6] Cruess RL, Cruess SR, Johnston SE. Professionalism: an ideal to be sustained. *Lancet* 2000;356:156–9.
- [7] Martimianakis MA, Maniate JM, Hodges BD. Sociological interpretations of professionalism. *Med Educ* 2009;43:829–37.
- [8] Collier R. Professionalism: assessing physician behaviour. *Can Med Assoc J* 2002;184:1349–50.
- [9] Cruess RL, Cruess SR. Teaching professionalism: general principles. *Med Teach* 2006;28:205–8.
- [10] Veloski JJ, Fields SK, Boex JR, Blank LL. Measuring professionalism: a review of studies with instruments reported in the literature between 1982 and 2002. *Acad Med* 2005;80:366–70.
- [11] Wilkinson TJ, Wade WB, Knock LD. A blueprint to assess professionalism: results of a systematic review. *Acad Med* 2009;84:551–8.
- [12] Hodges BD, Ginsburg S, Cruess R, et al. Assessment of professionalism: recommendations from the Ottawa 2010 Conference. *Med Teach* 2011;33:354–63.
- [13] West CP, Shanafelt TD. The influence of personal and environmental factors on professionalism in medical education. *BMC Med Educ* 2007;7:29.
- [14] Tromp F, Vernooij-Dassen M, Kramer A, Grol R, Bottema B. Behavioural elements of professionalism: assessment of a fundamental concept in medical care. *Med Teach* 2010;32:e161–9.
- [15] Van de Camp K, Vernooij-Dassen M, Grol R, Bottema B. Professionalism in general practice: development of an instrument to assess professional behaviour in general practitioner trainees. *Med Educ* 2006;40:43–50.
- [16] Van de Camp K, Vernooij-Dassen M, Grol R, Bottema B. How to conceptualize professionalism: a qualitative study. *Med Teach* 2004;26:696–702.
- [17] Steinhäuser J, Roos M, Haberer K, et al. Report from general practice: the composite graduate education(plus) program of the Baden-Württemberg General Practice Competence Center - development, implementation and prospects. *Z Evid Fortbild Qual Gesundhwes* 2011;105:105–9.
- [18] Wild D, Grove A, Martin M, et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health* 2005;8:94–104.
- [19] Boren MT. Thinking aloud: reconciling theory and practice. *IEEE T Prof Commun* 2000;43:261–78.
- [20] Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull* 1955;52:281–302.
- [21] Roussin V, Gasser T, Seifert B. Assessing intrarater, interrater and test-retest reliability of continuous measurements. *Stat Med* 2002;21:3431–46.
- [22] Nunnally J. *Psychometric Theory*. New York: McGraw-Hill; 1994.
- [23] Hooper D, Coughlan J, Mullen M. Structural Equation Modelling: Guidelines for Determining Model Fit. *EJBRM* 2008;6:53–60.
- [24] Hu L, Bentler P. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Modeling* 1999;6:1–55.
- [25] Hayduk L, Cummings GG, Boadu K, Pazderka-Robinson H, Boulianne S. Testing! Testing! One, Two, Three - Testing the theory in structural equation models! *Pers Indiv Differ* 2007;42:841–50.
- [26] Tabachnik BG, Fidell LS. *Using Multivariate Statistics*. New York: Allyn and Bacon; 2007.
- [27] Steiger JH. Understanding the limits of global fit assessment in structural equation modeling. *Pers Indiv Differ* 2007;42:893–8.
- [28] Sharma S, Mukherjee S, Kumar A, Dillon WR. A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. *J Bus Res* 2005;58:935–43.
- [29] Mayring P. *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Basel: Beltz; 2008.
- [30] Roos M, Krug D, Pfisterer D, Joos S. Professionalism in general practice in Germany - a qualitative approximation. *Z Evid Fortbild Qual Gesundhwes* 2013;107:475–83.
- [31] Cruess SR, Cruess RL, Steinert Y. Teaching professionalism across cultural and national borders: lessons learned from an AMEE workshop. *Med Teach* 2010;32:371–4.
- [32] Cruess SR, Cruess RL, Steinert Y. Linking the teaching of professionalism to the social contract: a call for cultural humility. *Med Teach* 2010;32:357–9.
- [33] Edwards PJ, Roberts I, Clarke MJ, et al. Methods to increase response to postal and electronic questionnaires. *Cochrane Database Syst Rev* 2009. MR000008.
- [34] Britten N, Jones R, Murphy E, Stacy R. Qualitative research methods in general practice and primary care. *Fam Pract* 1995;12:104–14.