



Behavioural elements of professionalism: Assessment of a fundamental concept in medical care

Fred Tromp, Myrra Vernooij-Dassen, Anneke Kramer, Richard Grol & Ben Bottema

To cite this article: Fred Tromp, Myrra Vernooij-Dassen, Anneke Kramer, Richard Grol & Ben Bottema (2010) Behavioural elements of professionalism: Assessment of a fundamental concept in medical care, Medical Teacher, 32:4, e161-e169, DOI: [10.3109/01421590903544728](https://doi.org/10.3109/01421590903544728)

To link to this article: <https://doi.org/10.3109/01421590903544728>



Published online: 30 Mar 2010.



Submit your article to this journal [↗](#)



Article views: 2153



View related articles [↗](#)



Citing articles: 1 View citing articles [↗](#)

WEB PAPER

Behavioural elements of professionalism: Assessment of a fundamental concept in medical care

FRED TROMP, MYRRA VERNOOIJ-DASSEN, ANNEKE KRAMER, RICHARD GROL & BEN BOTTEMA
Radboud University Nijmegen Medical Centre, The Netherlands

Abstract

Background: The Nijmegen Professionalism Scale, an instrument for assessing professional behaviour of general practitioner (GP) trainees, consists of four domains: professional behaviour towards patients, other professionals, society and oneself. The purpose of the instrument is to provide formative feedback.

Aim: The aim of this study was to examine the psychometric properties of the Nijmegen Professionalism Scale.

Methods: Both GP trainers and their GP trainees participated. Factor analysis was conducted for each domain. Factor structures of trainee and trainer groups were compared. Measure of congruence used was Tucker's phi. Cronbach's α was used to establish reliability.

Results: Factor structures of the instrument used by GP trainers and trainees were similar. Two factors for each domain were found: domain 1, *Respecting patient's interests* and *Professional distance*; domain 2, *Collaboration skills* and *Management skills*; domain 3, *Responsibility* and *Quality management*; and domain 4, *Reflection and learning* and *Dealing with emotions*. Congruence measures were substantial (>0.90). Reliability ranged from 0.78 to 0.95.

Conclusion: This study to validate the instrument represents one further step. To construct a sound validity argument, a much broader range of evidence is required. Nevertheless, this study shows that the Nijmegen Professionalism Scale is a reliable tool for assessing professional behaviour.

Background

Over the last decade, medical education has changed extensively. The focus has shifted from the acquisition of knowledge to the achievement of competence (Driessen et al. 2007). With the transformation into a competency-based programme, it is equally important that parallel strategies are chosen to assess these competencies. Competency-based teaching has been driven by competency frameworks such as the Accreditation Council for Graduate Medical Education and the American Board of Medical Specialties (ACGME/ABMS) competencies (Horowitz et al. 2004), and the Canadian Medical Education Directives for Specialists (CanMEDS) 2000, issued by Canada's Royal College of Physicians and Surgeons (Royal College of Physicians and Surgeons of Canada 2000). Competency-based education strives to assess the performance of residents. The basic essential elements consist of functional analysis of the occupational competencies, translation of these competencies into outcomes and assessment of the trainees' progress in these outcomes on the basis of their performance (Leung 2002). Assessments should be based on a set of clearly defined outcomes so that all parties concerned, including assessors and trainees, can make reasonably objective judgements about whether or not each trainee has achieved them.

Practice points

- Professionalism is often cited as an essential part of medical performance.
- The Nijmegen Professionalism Scale is a suitable tool to evaluate professional behaviour in general practice.
- This study to validate the Nijmegen Professionalism Scale represents one further step. However, to construct a sound validity argument, a much broader range of evidence is required.

One of the seven competencies that have been defined for the Dutch postgraduate training for family practice is professionalism. Professionalism is often cited as an essential part of medical performance and thus of medical training (Arnold 2002; Veloski et al. 2005; Cruess 2006; Joyner & Vemulakonda 2007; Tsai et al. 2007). However, professionalism has proved difficult to define (Arnold 2002; Lynch et al. 2004). Van de Camp et al. (2004) conducted a study to conceptualize professionalism, in which they reviewed the literature and proposed a multidimensional construct. The four domains they found within professionalism were: professional behaviour towards the patient, towards other professionals, towards society and towards oneself (van de Camp et al. 2004, 2006).

Correspondence: F. Tromp, Department of Primary and Community Care, University Medical Centre St. Radboud, PO Box 9101, 6500 HB Nijmegen, The Netherlands. Tel: +31 24 3610676; fax: +31 24 3619553; email: f.tromp@voaha.umcn.nl

These domains provided the framework for the development of an instrument to evaluate the professional behaviour of general practitioner (GP) trainees. In the development of this instrument, it was decided to assess professionalism by focusing on behaviour rather than on traits (van de Camp et al. 2006). Research showed that the key to valid assessment of professionalism lies in focusing in behaviour. Students do not identify themselves with abstract elements of professionalism, but they define professionalism in practical terms (Ginsburg et al. 2002; Ginsburg & Stern 2004). By framing professionalism in terms of behaviours rather than abstractions, we come much closer to a context bound, realistic framework for understanding professional behaviour (van de Camp et al. 2006). Furthermore, observable behaviour is the appropriate basis of providing feedback (Branch Jr. & Paranjape 2002; Tromp et al. 2007).

The primary aim of the Nijmegen Professionalism Scale is formative assessment. It allows GP trainers to systematically provide feedback about the professional behaviour of their GP trainees. The GP trainers complete the instrument to evaluate their GP trainees every 3 months, and the GP trainees evaluate themselves with the Nijmegen Professionalism Scale. To become a GP in The Netherlands, 3 years of training are required after graduation as a medical doctor. During these 3 years, the GP trainee spends the equivalent of 4 days a week with a GP during the 1st and 3rd years. These periods are usually completed in different general practices, where the GP then acts as a coach and a teacher. As the 1st and 3rd year are for practical training in a general practice, the 2nd year is dedicated to rotations through hospitals, clinics for chronically ill patients and psychiatric outpatient clinics. The assessments of the GP trainers and the self-assessments of the GP trainees are made independently. The results are discussed every 3 months in an interview of progress review. These evaluations are thus not one-time assessments, but cover a 3-month period of multiple observations. Both evaluations are compared in order to formulate 'professional behaviour learning points' for the GP trainee. These learning points are issues within professional behaviour selected to be improved systematically by setting goals for the future. In generating learning points, the GP trainees are encouraged to reflect on their strong and weak points in professional behaviour.

Instruments with good psychometric properties are needed for the evaluation of professional behaviour. Van de Camp et al. (2006) tested the content validity of the instrument in a qualitative study using the nominal group technique, which consists of a very structured procedure to gather information from relevant experts (Jones & Hunter 1995).

The goal of this study was to attain the best possible quality for the Nijmegen Professionalism Scale, which required the examination of its psychometric properties and further validation of the instrument by the assessment of its construct validity and reliability. Factor analyses were applied to determine its internal structure. The instrument is used by both GP trainees and trainers to evaluate professional behaviour. In this study, we compared the factor structure of these two groups. If the factor structure is the same for trainers and trainees using the instrument, then both GP trainers and trainees attach the same meaning to the construct of

professionalism, and this will contribute to the validity. Cronbach's α was used to establish reliability.

Method

Participants

As a part of the curriculum, GP trainers and their GP trainees associated with the Department of Postgraduate Training for General Practice of the Radboud University Nijmegen Medical Centre complete the Nijmegen Professionalism Scale every 3 months. Permission was asked to analyse the data for this study. All 119 trainers and 119 trainees consented. Due to practical reasons, we only made use of the data of one single 3-month period, dated from September until November 2005.

Measures

The Nijmegen Professionalism Scale is an instrument with 106 items, each representing an element of professional behaviour. The instrument consists of four parts, each of which addresses a different domain within professionalism: professionalism towards the patient, professionalism towards other professionals, professionalism towards society and professionalism towards oneself. Each domain consists of separate scales (varying in number from four to nine) that measure different elements of professional behaviour (van de Camp et al. 2006). Following each item is a four-point Likert scale on which the participants can indicate how often a GP trainee exhibited the specified behaviour, ranging from 'seldom or never' (1) to 'always' (4).

As already mentioned, the GP trainers complete the instrument to evaluate their GP trainees every 3 months, and the GP trainees assess themselves with the Nijmegen Professionalism Scale. The trainers were instructed to discuss both evaluations in a one-on-one tutorial within 3 weeks of completing the instrument. Before using the instrument, trainers and trainees were informed in a brief training session about the primarily formative purpose of the instrument. Furthermore, a written manual was provided.

Construct validity

To examine the construct validity of the instrument, the four domains were analysed separately, since each domain concerns a separate construct within professionalism (van de Camp et al. 2004). As a first step, a confirmatory factor analysis for each domain was performed to reproduce the original element structure. Whenever this analysis failed to replicate the original structure, an exploratory principal component analysis with varimax rotation was completed. Two criteria were used to determine the optimal number of factors to extract: the scree plot and interpretability of the factor loadings. Items were retained if they had a loading greater or equal than 0.40 (Nunnally & Bernstein 1994). Items with a factor loading less than 0.40 were discussed individually by researchers (FT, AK and RG) and GP trainers (MV and BB). If consensus was reached, items were retained or rejected. One of the criteria to reach consensus was face validity of the items.

We felt that face validity was important because items with great face validity make the instrument conceptually clear and more straightforward for GP trainers and GP trainees to use.

Construct equivalence

In many cases, self-assessment and external assessment show only moderate agreement (Kramer et al. 2007). Self-assessment, however, can be accurate under certain conditions (Gordon 1991, 1992; Ginsburg & Stern 2004), namely, when learners are expected to gather and interpret data about their performance and, at the same time, when they are required to reconcile their self-assessments with credible external evaluations. These conditions appear to be met with the Nijmegen Professionalism Scale. In testing the feasibility of the instrument, van de Camp et al. found very good agreement in the ratings of professional behaviour as observed by the GP trainer and the GP trainee. (van de Camp et al. 2006) Consequently, both the data sets from the GP trainers and from the GP trainees were used and compared for this analysis.

Tucker's phi coefficients were computed for each factor. Phi values of 0.90 or more provide evidence of construct equivalence of both groups (Van de Vijver & Leung 2001), which shows that both GP trainers and trainees attach the same meaning to the construct of professionalism.

Internal consistency

Cronbach's α was used to determine the internal consistency of items within each factor. We calculated Cronbach's α to provide additional evidence that the items within a factor were measuring the same underlying construct.

Results

Sample

The GP trainers and the GP trainees provided 116 lists that were eligible for inclusion. Three GP trainers and three GP trainees returned incomplete lists. The sample consisted of 60 1st-year and 56 3rd-year GP trainers and their GP trainees.

No indication of leniency, halo or ceiling effects were found as the scores ranged from 1 to 4 and showed sufficiently variance.

Construct validity and equivalence

Confirmatory factor analysis failed to replicate the original structure in all four domains. We, therefore, conducted an exploratory principal component analysis with varimax rotation. Examination of the scree plots after the principal component analysis indicated that there were two factors in each domain that best described the data.

Domain 1: Professional behaviour towards the patient

A two-factor solution was derived for both GP trainers and GP trainees. The results are shown in Table 1.

The factor structure of the instrument used by the two groups is very similar. The only item that loads on a different factor in the two groups is 'does not give patient false hope'.

The first factor was labelled *respecting patient's interests*, since it comprised such behaviours as showing sympathy, adjusting language to communicate with patients with little education, taking gender-specific differences into account and dealing correctly with legislative rules. The second factor was labelled *professional distance*, since its items concerned such behaviours as 'taking care not to become too involved in the emotions of the patient' and 'not becoming too intimate'.

We discussed the item 'does not give the patient false hope' and reached the consensus that it should be assigned to the first factor *respecting patient's interests*, following the structure found by the GP trainees. This decision was based on face validity.

Almost all items had a factor loading of at least 0.40, except the items: 'looks clean and tidy and dresses according to current norms' and 'has difficulty taking decisions regarding diagnosis and treatment policy'. We weighed the removal of these items against their educational significance. In our view, the educational significance of the item 'looks clean and tidy and dresses according to current norms' is considerable. Educators informed us that this item helped them raise an otherwise very difficult subject. The item was, therefore, retained despite its low factor loading and assigned according to its highest loading (0.30 on factor 1). The other item, 'has difficulty taking decisions regarding diagnosis and treatment policy', was removed from the list. In our judgement, this item did not fit in either of the two factors. The two factors yielded Tucker's phi values of 0.95 and 0.94.

Domain 2: Professional behaviour towards other professionals

Here, also a two-factor solution was derived for both GP trainers and GP trainees. The factor loading matrix is shown in Table 2.

Again, the factor structure of the instrument used by both groups looks approximately the same. The item 'is able to manage the mutual demarcation of tasks between GP and specialists' loads on different factors. The item 'chooses the correct time and place for comments about functioning' has greater factor loadings in the trainee group, but it loads on both factors. The items 'conducts structural consultations with support personnel', 'is able to provide emotional support for colleagues' and 'shirks tasks' display higher factor loadings in the trainee group.

The items of the first factor included such behaviours as 'complying with multidisciplinary working agreements' and 'being able to motivate support personnel'. These behaviours were considered relevant to the relational part of collaboration with other healthcare workers; this factor was, therefore, interpreted as *collaboration skills*. The second factor included items related to management, such as 'being able to take policy decisions' and 'dealing constructively with conflicts'. This factor was labelled *management skills*. Five items (Table 2) had a loading of less than 0.40 in the trainer group. These items proved to be of little educational

Table 1. Factor loading matrix domain 1: Professional behaviour towards the patient.

S. No.	Items	Factors and item factor loadings			
		GP trainer factor 1* eigenvalue 5.42	GP trainer factor 2* eigenvalue 2.61	GP trainee factor 1* eigenvalue 5.91	GP trainee factor 2* eigenvalue 2.09
The GP trainee:					
1.	Deals carefully with professional secrecy when talking to colleagues or acquaintances		0.44		0.51
2.	Deals correctly with legislative rules regarding informed consent	0.59		0.57	
3.	Is able to bring up difficult subjects	0.51		0.59	
4.	Respects the right of patients to inspect their medical records	0.41		0.46	
5.	Does not give patients false hope		0.52	0.52	
6.	Is able to show sympathy	0.49		0.48	
7.	Takes care not to become part of the patient's system		0.72		0.44
8.	Takes care not to become too involved in the Patient's emotions		0.86		0.59
9.	Takes care not to become too intimate		0.72		0.46
10.	Takes patient's opinions seriously	0.51		0.40	
11.	Takes patients' embarrassment, shyness and reluctance into account	0.63		0.46	
12.	During physical examinations, explains the aim of the procedures and what is expected of the patient	0.46		0.40	
13.	Approaches patients with a different frame of reference (e.g. religion) openly	0.58		0.53	
14.	Looks clean and tidy and dresses according to current norms	–	–	–	–
15.	Adjusts language to communicate with patients with little education	0.47		0.56	
16.	Takes sex-specific differences into account	0.43		0.61	
17.	Is able to deal with cultural differences in the presentation and experience of complaints	0.53		0.59	
18.	Takes care not to be influenced by patients of high social status		0.59		0.40
19.	Is able to cope with the different expectations that patients have of their GP	0.40		0.53	
20.	Involves the previous history of the patient in the provision of care	0.48		0.49	
21.	Pays attention to the consequence of the treatment policy on the daily functioning of the patient	0.69		0.57	
22.	Involves relevant aspects of the patient's home and environment in the provision of care	0.55		0.66	
23.	Retains insight into the medical history of patients in order to act proactively if necessary	0.62		0.53	
24.	If necessary, takes action after life events	0.65		0.52	
25.	Respects patients' self-determination	0.51		0.40	
26.	<i>Has difficulty taking decisions regarding diagnosis and treatment policy</i>	–	–	–	–
Total variance explained (%)		30.9		30.8	

Notes: Factor loadings less than 0.40 not shown.

Italicized items have been removed from the instrument.

$n = 116$.

*Factors: 1 = respecting patient's interests; and 2 = professional distance.

significance according to the trainers, as these behaviours were seldom observed in practice. The items were, therefore, removed from the list. Tucker's phi values of the two factors were computed as 0.90 and 0.96.

Domain 3: Professional behaviour towards society

Table 3 shows the results. Two factors were determined. There are no considerable differences in the factor structure of the instrument used by the GP trainers and the GP trainees. Two items 'has perceptions about how form can be given to means

of contact (telephone services, diabetes, surgery hours, etc.)' and 'is able to justify indications for making house calls' load on different factors.

The items 'deals meticulously with moral requests for care (e.g. abortion, euthanasia)' and 'has perceptions about how repeat prescriptions can be written in a responsible way' had greater factor loadings in the trainee group.

The first factor was composed of such behaviours as 'bearing the consequences of his/her own conduct' and 'not hiding behind others' and was labelled *responsibility*. The content of the second factor included items representing

Table 2. Factor loading matrix in domain 2: Professional behaviour towards other professionals.

S. No.	Items	Factors and item factor loadings			
		GP trainer factor 1* eigenvalue 7.22	GP trainer factor 2* eigenvalue 1.88	GP trainee factor 1* eigenvalue 7.88	GP trainee factor 2* eigenvalue 1.84
The GP trainee:					
1.	Is able to mediate with other care providers in the interests of the patient	0.48		0.51	
2.	Consults other care providers with targeted questions	0.43		0.53	
3.	Discusses bottlenecks in cooperation with others directly		0.58		0.48
4.	Complies with multidisciplinary working agreements	0.63		0.44	
5.	Ensures structured information transfer with other care providers	0.57		0.50	
6.	Deals correctly with targeted questions from other care providers	0.77		0.64	
7.	Is able to write a good referral letter with direct questions	0.47		0.57	
8.	Is able to discuss a difference of opinion with a specialist directly		0.65		0.53
9.	Is able to manage the mutual demarcation of tasks between GP and specialists		0.49	0.57	
10.	Is able to influence specialist care (e.g. during consultation at hospital visits)		0.48		0.72
11.	Ensures coherence in first and second line medical care		0.40		0.57
12.	Is able to motivate support personnel	0.58		0.61	
13.	<i>Chooses the correct time and place for comments about functioning</i>	–	–	0.47	0.45
14.	<i>Conducts structural consultations with support personnel</i>				0.57
15.	Makes clear agreements with support personnel	0.56		0.61	
16.	Listens to the contributions of support personnel	0.77		0.64	
17.	Is able to distinguish between professional and personal interests in negotiations		0.63		0.53
18.	Is able to deal constructively with conflicts		0.55		0.54
19.	Is able to take policy decisions		0.67		0.73
20.	Is able to conduct job evaluations		0.63		0.60
21.	<i>Is able to help seek solutions if a colleague has too high a workload</i>	–	–	–	–
22.	<i>Is able to provide emotional support for colleagues</i>	–	–	0.56	
23.	Transfers services correctly	0.69		0.62	
24.	<i>Shirks tasks</i>	–	–	0.59	
Total variance explained (%)		37.9		40.4	

Notes: Factor loadings less than 0.40 not shown.

Italicized items have been removed from the instrument.

$n = 116$.

*Factors: 1 = collaboration skills; and 2 = management skills.

behaviour such as 'being able to signal suboptimal care within the practice' and 'being able to set priorities in the choice of topics for quality management'. This factor was labelled *quality management*. Two items had a factor loading of just less than 0.40 (Table 3) in the trainer group and were removed after reaching consensus. Although the item 'deals meticulously with moral requests for care (e.g. abortion, euthanasia)' had a factor loading less than 0.40 in the trainer group, it was retained because of educational significance and assigned to the first factor *responsibility*. Tucker's phi values were 0.91 for the factor *quality management* and 0.94 for *responsibility*.

Domain 4: Professional behaviour towards oneself

The results are shown in Table 4. A two-factor solution for both groups was indicated. In the fourth domain, professional behaviour towards oneself, again no major differences in the

factor structures between GP trainers and trainees were found. The items 'discusses one's shortcomings and failures without losing belief in one's own competence', 'makes a realistic estimation of one's own strong and weak points', 'is able to mention aspects of work that increase satisfaction', 'is able to cope with feelings of powerlessness in the care process', and 'learns from one's own mistakes' load on different factors.

The second factor in the GP trainer group has approximately the same items as the factor of the GP trainees. The only exception is the item 'learns from one's mistakes'.

The first factor included items that reflect behaviours such as 'being able to name thoughts and feelings that patients evoked in oneself', 'being able to analyse one's own behaviour in specific situations' and 'being able to figure things out by oneself' and was labelled *reflection and learning*. The second factor consisted of items such as 'being able to cope after making a mistake' and 'being able to deal with the possibility

Table 3. Factor loading matrix in domain 3: Professional behaviour towards society.

S. No.	Items	Items factors and item factor loadings			
		GP trainer factor 1* eigenvalue 5.49	GP trainer factor 2* eigenvalue 2.16	GP trainee factor 1* eigenvalue 6.75	GP trainee factor 2* eigenvalue 1.87
The GP trainee:					
1.	Bears the consequences of his/her own conduct	0.67		0.61	
2.	Is able to justify deviations from rules and guidelines	0.53		0.62	
3.	Keeps promises and agreements	0.62		0.57	
4.	Does not hide behind others (give others the blame or responsibility)	0.64		0.61	
5.	Is aware of how his/her own norms regarding disease influence disease management	0.51		0.59	
6.	<i>Does not impose his/her own norms and values upon others</i>	–	–	–	–
7.	Deals meticulously with moral requests for care (e.g. abortion, euthanasia)	–	–	0.58	
8.	Is able to set priorities in the choice of topics for quality improvement		0.65		0.76
9.	Is able to signal suboptimal care within the practice		0.67		0.72
10.	Is able to work out a quality-improvement project		0.58		0.80
11.	Is able to estimate which problems are suitable for a quality-improvement project		0.76		0.79
12.	Is able to name the tasks to be, or that he/she would like to be, delegated to the assistant		0.64		0.49
13.	Has perceptions about how form can be given to means of contact (telephone services, diabetes surgery hours, etc.)		0.76	0.60	
14.	Is able to justify indications for making home visits		0.49	0.66	
15.	<i>Has perceptions about how repeat prescriptions can be written in a responsible manner</i>	–	–	0.65	
16.	Is aware of the meaning and relative value of scientific evidence in decision-making	0.55		0.48	
17.	In decision-making, weighs scientific evidence against factors related to the patient or the circumstances	0.67		0.45	
18.	Is able to justify choices made on the basis of scientific evidence	0.63		0.46	
19.	Is able to explain his/her own norms and values regarding the application of scientific evidence	0.52		0.56	
Total variance explained (%)		40.3		46.0	

Notes: Factor loadings less than 0.40 not shown.

Italicized items have been removed from the instrument.

$n = 116$.

*Factors: 1 = responsibility; and 2 = quality management.

that a treatment is unsuccessful'. This factor was labelled *dealing with emotions*. After discussion, we decided to assign the items 'discusses one's shortcomings and failures without losing belief in one's own competence', 'makes a realistic estimation of one's strong and weak points', 'is able to mention aspects of work that increase satisfaction' and 'learns from one's own mistakes' to the scale *reflection and learning*. The item 'is able to cope with feelings of powerlessness in the care process' was assigned to the scale *dealing with emotions*. These decisions were based on face validity.

Six items had factor loadings less than 0.40 (Table 4). The items 'set priorities in learning' and 'be able to balance work and private life' have an important educational significance, so they were retained and assigned to the factor *reflection and learning*. The other items were removed from the list.

Tucker's phi values were computed for both factors; they were 0.90 and 0.91.

e166

Internal consistency

Table 5 shows the Cronbach's α coefficients for the two groups of participants in our study (GP trainers and GP trainees).

The Cronbach's α coefficients for the GP trainer sample ranged from 0.79 (*dealing with emotions*) to 0.95 (*reflection and learning*), which indicates good to excellent internal consistency within each factor (Nunnally & Bernstein 1994). Good to excellent internal consistency was also found in the trainee group, with values ranging from 0.72 (*professional distance*) to 0.91 (*reflection and learning*).

Discussion

The results of this study provide psychometric support for the Nijmegen Professionalism Scale. Previous results (van de Camp et al. 2006) supported the content validity of the

Table 4. . Factor loading matrix in domain 4: Professional behaviour towards oneself.

S. No.	Items	Factors and item factor loadings			
		GP trainer factor 1* eigenvalue 12.07	GP trainer factor 2* eigenvalue 2.62	GP trainee factor 1* eigenvalue 12.52	GP trainee factor 2* eigenvalue 2.40
The GP trainee:					
1.	Is able to name reactions, thoughts and feelings that patients evoke	0.70		0.75	
2.	Asks questions about his/her own role in relationships (patient, group, gp trainer, etc.)	0.74		0.72	
3.	Uses specific practical situations as starting points for critical self-reflection	0.78		0.82	
4.	In a specific situation with a patient, is able to analyse his/her own behaviour and adjust it if necessary	0.74		0.61	
5.	<i>Dares to express and act upon his/her own point of view</i>	–	–	–	–
6.	Is able to mention differences of opinion	0.60		0.47	
7.	Discusses his/her shortcomings and failures without losing belief in his/her own competence	0.50			0.57
8.	Makes a realistic estimation of his/her own strong and weak points	0.63			0.60
9.	Is able to balance work and private life	–	–		0.53
10.	Is able to mention aspects of work that increase satisfaction	0.54		0.48	
11.	Is able to seek timely professional help when experiencing personal problems	0.52		–	–
12.	Is able to cope with feelings of powerlessness in the care process	0.43			0.57
13.	Is able to give positive as well as negative feedback	0.62		0.47	
14.	Is open about feelings provoked by feedback	0.79		0.58	
15.	Attaches importance to what others think about his/her behaviour	0.68		0.60	
16.	Is able to make feedback concrete and specific	0.70		0.56	
17.	Adheres to agreements made during feedback	0.67		0.65	
18.	Sets priorities in learning	–	–	–	
19.	Does not resist being judged	0.59		0.66	
20.	Has an enquiring mind (asks questions and takes initiative)	0.59		0.66	
21.	Is able to figure things out for him/herself	0.53			0.41
22.	<i>Is able to adapt when patients change their minds about a treatment</i>	–	–	–	–
23.	Is able to adapt and keep control of the situation if patients unexpectedly need to be seen during other activities		0.51		0.66
24.	Recovers rapidly after an unpleasant consultation		0.60		0.64
25.	Is able to admit his/her own mistakes	0.48		0.44	
26.	Takes action to rectify his/her own mistakes	0.52		0.47	
27.	Withdraws from the consequences of his/her own mistakes	0.41		0.46	
28.	Is able to cope after making a mistake		0.66		0.63
29.	Learns from mistakes	0.51			0.65
30.	Is able to let a mild disorder (e.g. tiredness) run its own course even though the correct diagnosis is a mystery		0.58		0.61
31.	Makes rational deliberations about whether it is necessary to request specialist or other advice		0.62		0.44
32.	Is able to deal with the possibility that a treatment decision may be unsuccessful		0.63		0.49
33.	<i>Expresses strong emotions by means of words instead of actions</i>	–	–	–	–
34.	Is able to deal with difficult or angry patients		0.59		0.59
35.	Is able to conduct interventions that lead to a decrease in aggression from the patient		0.52		0.60
36.	Is able to formulate his/her own opinion in a clear and inoffensive manner		0.61		0.53
37.	<i>Keeps an eye on his/her own safety adequately</i>		–		0.52
Total variance explained (%)		39.7		39.2	

Notes: Factor loadings less than 0.40 not shown.

Italicized items have been removed from the instrument.

n = 116.

*Factors: 1 = reflection and learning; and 2 = dealing with emotions.

Table 5. Cronbach's α associated with each factor.

Domain: Professional behaviour towards	Factor	Number of items	Cronbach's α GP trainers	Cronbach's α GP trainees
The patient	Respecting patient's interests	20	0.88	0.87
The patient	Professional distance	5	0.82	0.72
Other professionals	Collaboration skills	10	0.91	0.86
Other professionals	Management skills	9	0.94	0.87
The society	Responsibility	10	0.78	0.82
The society	Quality management	7	0.89	0.87
Oneself	Reflection and learning	23	0.95	0.91
Oneself	Dealing with emotions	9	0.79	0.86
	Total	93		

instrument as well as its feasibility as a tool to educate for professionalism.

The original structure, based on consensus and face validity alone, was not replicated (van de Camp et al. 2006). Instead, a much simpler structure, with two scales for each domain, was found. In our view, this new structure makes the instrument conceptually clearer and more straightforward for GP trainers and GP trainees to use.

In contrast with the traditional approach, competency-based medical education can potentially lead to individualized flexible training, transparent standards and increased public accountability. If applied inappropriately, it can also result in demotivation, focus on minimally acceptable standards, increased administrative burden and a reduction in the educational content. Higher-order competencies, such as professionalism, need to be defined and developed more robustly (Leung 2002). Professional behaviour is a complex construct to define, and without consensus of this construct, teaching and assessing professional behaviour are problematic. We compared the factor structure of self-assessment and external evaluations. No considerable differences in the four domains were found. This indicates that GP trainers and trainees attach the same meaning to the construct of professional behaviour, which creates a solid foundation for effective teaching and assessing of this essential part of medical performance.

Feasibility is the most common limitation, since assessment tools often take a lot of time. Some concern remains about the feasibility of the Nijmegen Professionalism Scale; the final list contains 93 items and may be too time consuming for most GP trainers. However, users were asked to comment on the instrument (van de Camp et al. 2006). They appreciated the valuable input it provided during the tutorial. No one criticized the length of the list. The Nijmegen Professionalism Scale is designed to guide professional growth. The specific details of the instrument enable trainers not only to assess, but also to encourage and monitor specific behaviour. As professional behaviour should be observed from the beginning of the training, feedback in an early stage of the training allows GP trainees to remedy possible lack of professional behaviour. Nevertheless, exploring whether the number of items could be reduced would be worthwhile. In addition, as the Nijmegen Professionalism Scale consists of four domains, it is possible for users to administer one domain at the time.

GP trainer and trainee work in close cooperation for an extended period of time and this may undermine the independence of the scores. Although it is relatively time consuming to complete the instrument, the use of instruments like the Nijmegen Professionalism Scale have to be considered as a very important element in the development of their trainees. GP trainers must be aware that only by using the Nijmegen Professionalism Scale in the appropriate way, they formatively support the development of professional behaviour of their trainees in a integrated, coherent and longitudinal fashion. This issue underscores the importance of rater training in the accurate use of assessment instruments before implementing them.

This study to validate the Nijmegen Professionalism Scale represents one further step. Van de Camp et al. (2006) already tested the content validity of the instrument in a qualitative study. However, to construct a sound validity argument, a much broader range of evidence is required. It has been argued that we cannot infer validity from a single analysis (Schuwirth & van der Vleuten 2006). Further information supporting the validity of the Nijmegen Professionalism Scale, for instance, would be data suggesting that it accurately identifies trainees with performance deficits and that the instrument is able to measure the professional growth of the trainees. Since we did not detect ceiling effects growth can in principle be assessed, but we did not test this.

Conclusion

Meaningful, reliable and valid assessment is crucial in the promotion of professionalism in GP trainees. On the basis of this study, we can conclude that the Nijmegen Professionalism Scale is a reliable tool to assess their professional behaviour. The results of this study show that GP trainers and trainees agree on the definition and meaning of professional behaviour. We consider the Nijmegen Professionalism Scale to be a promising tool for assessing and enhancing the professional behaviour of GP trainees.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

Notes on contributors

FRED TROMP is a researcher at the Department of Postgraduate Training for General Practice, Radboud University Nijmegen Medical Centre. He is currently working on his PhD project on assessing competencies in general practice.

MYRRA VERNOOIJ-DASSEN is a professor in psychosocial aspects of care for frail elderly, senior lecturer at Scientific Institute of Quality of Healthcare (IQ healthcare), coordinator of the Alzheimer Centre of the University Medical Centre Nijmegen and coordinator of qualitative research of the Dutch Research School CARE.

ANNEKE KRAMER is a senior researcher at the Department of Postgraduate Training for General Practice, Radboud University Nijmegen Medical Centre and also works as a GP in Utrecht.

RICHARD GROL holds a chair on Quality of Care at the Radboud University Nijmegen, a honorary chair at Maastricht University and guest professorships at the University of Louvain, Belgium and Manchester University, UK. He is a director of the Scientific Institute of Quality of Healthcare (IQ healthcare), one of the leading research centres on quality and safety of care in the world.

BEN BOTTEMA is a director of the Department of Postgraduate Training for General Practice, Radboud University Nijmegen Medical Centre. He supervises several research projects about education and disease management of asthma and COPD.

References

- Arnold L. 2002. Assessing professional behavior: Yesterday, today, and tomorrow. *Acad Med* 77:502–515.
- Branch Jr WT, Paranjape A. 2002. Feedback and reflection: Teaching methods for clinical settings. *Acad Med* 77:1185–1188.
- Cruess RL. 2006. Teaching professionalism: Theory, principles, and practices 19. *Clin Orthop Relat Res* 449:177–185.
- Driessen E, van TJ, van dV, Wass V. 2007. Portfolios in medical education: Why do they meet with mixed success? A systematic review. *Med Educ* 41:1224–1233.
- Ginsburg S, Regehr G, Stern D, Lingard L. 2002. The anatomy of the professional lapse: Bridging the gap between traditional frameworks and students' perceptions. *Acad Med* 77:516–522.
- Ginsburg S, Stern D. 2004. The professionalism movement: Behaviors are the key to progress. *Am J Bioeth* 4:14–15.
- Gordon MJ. 1991. A review of the validity and accuracy of self-assessments in health professions training. *Acad Med* 66:762–769.
- Gordon MJ. 1992. Self-assessment programs and their implications for health professions training. *Acad Med* 67:672–679.
- Horowitz SD, Miller SH, Miles PV. 2004. Board certification and physician quality. *Med Educ* 38:10–11.
- Jones J, Hunter D. 1995. Consensus methods for medical and health services research. *BMJ* 311:376–380.
- Joyner BD, Vemulakonda VM. 2007. Improving professionalism: Making the implicit more explicit. *J Urol* 177:2287–2290.
- Kramer AW, Zuihoff P, Jansen JJ, Tan LH, Grol RP, van dV. 2007. Growth of self-perceived clinical competence in postgraduate training for general practice and its relation to potentially influencing factors. *Adv Health Sci Educ Theory Pract* 12:135–145.
- Leung WC. 2002. Competency based medical training: Review. *BMJ* 325:693–696.
- Lynch DC, Surdyk PM, Eiser AR. 2004. Assessing professionalism: A review of the literature. *Med Teach* 26:366–373.
- Nunnally JC, Bernstein IR. 1994. *Psychometric theory*. 3rd ed. New York: McGraw-Hill.
- Royal College of Physicians and Surgeons of Canada 2000. *CanMEDS 2000: Extract from the CanMEDS 2000 project societal needs working group report*. *Med Teach* 22:549–554.
- Schuwirth LW, van der Vleuten CPM. 2006. A plea for new psychometric models in educational assessment. *Med Educ* 40:296–300.
- Tromp F, Rademakers JJ, ten Cate TJ. 2007. Development of an instrument to assess professional behaviour of foreign medical graduates. *Med Teach* 29:150–155.
- Tsai TC, Lin CH, Harasym PH, Violato C. 2007. Students' perception on medical professionalism: The psychometric perspective. *Med Teach* 29:128–134.
- van de Camp K, Vernooij-Dassen M, Grol R, Bottema B. 2006. Professionalism in general practice: Development of an instrument to assess professional behaviour in general practitioner trainees. *Med Educ* 40:43–50.
- van de Camp K, Vernooij-Dassen MJ, Grol RP, Bottema BJ. 2004. How to conceptualize professionalism: A qualitative study. *Med Teach* 26:696–702.
- Van de Vijver FJ, Leung K. 2001. Personality in cultural context: Methodological issues. *J Pers* 69:1007–1031.
- Veloski JJ, Fields SK, Boex JR, Blank LL. 2005. Measuring professionalism: A review of studies with instruments reported in the literature between 1982 and 2002. *Acad Med* 80:366–370.