Acquisition of clinical skills in postgraduate training for general practice

Anneke W M Kramer, Koos J J M Jansen, Herman Düsman, Lisa H C Tan, Cees P M van der Vleuten, Richard P T M Grol

SUMMARY

Background: *Postgraduate training in general practice aims to develop clinical competence. However, little is known about its effect on trainees' development of clinical skills.*

Aim: *To assess the acquisition of clinical skills during a 3–year training programme and to evaluate whether a satisfactory level is achieved towards the end of training.*

Design of study: *Cross-sectional design.*

Setting: *Dutch postgraduate training in general practice from 1995 to 1998.*

Method: *Clinical skills were assessed using a written knowledge test of skills and by an Objective Structured Clinical Examination (OSCE). The written test was administered to trainees in all 3 years. Trainees at completion of their training took the OSCE. The results of both tests were compared with a standard of adequacy and to a reference group of general practitioner (GP) trainers.*

Results: *An increase in the level of clinical skills and pass rate was found throughout the training, being most prominent during the first 6 months. At completion of their training, trainees scored higher than the GPs on the written test (48% versus 39%) and on the OSCE (69% versus 63%). Eighty-eight per cent of the trainees would have passed the written test against 70% of the GPs and 94% would have passed the OSCE against 80% of the GPs.*

Conclusion: *A 3–year postgraduate training period in general practice contributes to a satisfactory acquisition of clinical skills. Further research into when and where acquisition takes place, and the role of the GP trainer, is needed.*

Keywords: *curriculum; education; graduate; educational measurement; clinical competence; family practice.*

A W M Kramer, MD, general practitioner; K J J M Jansen, MD, PhD, general practitioner; H Düsman, methodologist; and L H C Tan, PhD, educational researcher, National Centre for Evaluation of Postgraduate Training in General Practice (SVUH), Utrecht. C P M van der Vleuten, PhD, professor of educational development and research, Department of Educational Development and Research, University of Maastricht. R P T M Grol, PhD, professor of quality of care, Centre for Quality of Care Research, Universities of Nijmegen and Maastricht, The Netherlands.

Address for correspondence

Dr A Kramer, Mauritsstraat 92, 3583 HV Utrecht, The Netherlands. E-mail: Carol_ann@tiro.nl

Submitted: 3 July 2002; Editor's response: 16 January 2003; final acceptance: 24 February 2003.

Introduction

ASSESSMENT of clinical competence in postgraduate training for general practice is not only useful to evaluate trainees but may also serve to examine the quality of the training programme. Little is known about the effect of postgraduate training on the competence of trainees.[1-3] In a recent study, we demonstrated a more or less continuous increase of general practice knowledge during Dutch postgraduate training.[4] Although knowledge is fundamental to any kind of clinical competence,[1,5-7] insight into the development of other aspects of clinical competence is also necessary for a more complete evaluation of the effects of a training programme. As clinical skills (including history taking, physical examination, diagnosing and management) constitute a crucial part of the daily work of general practitioners (GPs),[8-10] it is one of the main tasks of postgraduate training for general practice to contribute to a satisfactory acquisition of these skills by trainees. Tan[11] and Sturmberg[10] investigated whether trainees have sufficient opportunity to learn and practice clinical skills relevant for general practice during medical school and postgraduate training. The findings reported in both studies suggest that, at completion of postgraduate training, trainees may have insufficient competence in clinical skills. However, the data were obtained by subjective ratings of trainees and trainers. More objective information about the level of clinical skills is needed to confirm this assumption.[12]

The Objective Structured Clinical Examination (OSCE) has been identified as a valid method for the assessment of clinical skills.[12,13] However, the number of skills that can be tested in an OSCE is relatively small, owing to its complex and costly organisation. As research into clinical competence has shown a large variability between tasks and cases, assessment of a large number of skills is necessary to obtain reliable results.[14] Therefore, a written test of skills can be a suitable alternative, as it can assess many skills per test. Although the drawback of such a test is that cognitive aspects of skills are assessed rather than hands-on performance, several studies have shown that it has sufficient reliability and predictive validity for performance on an OSCE, particularly when group results are concerned, such as in programme evaluation.[15-18] When combining the results of both an OSCE and a written test of skills, a higher validity and reliability can be achieved.[19,20]

In this study we assessed the acquisition of clinical skills during a 3–year postgraduate training programme for general practice and we evaluated whether a satisfactory level was achieved towards the end of training.

To assess the level of skills, a written knowledge test of skills was administered to trainees in all 3 years and an OSCE was performed towards the end of training. The results were compared with standards of adequacy and to the results of a reference group of GP trainers.

HOW THIS FITS IN

What do we know?

Little is known about the effect of postgraduate training in general practice on the development of clinical skills.

What does this paper add?

This study shows that a 3–year postgraduate training programme contributes to a satisfactory acquisition of clinical skills.

Method

Context of the study

The curriculum under study was the Dutch postgraduate training in general practice from September 1995 to September 1998. Since 1988 this curriculum has been organised according to a standard blueprint (Figure 1).[21]

A 3–year training programme was employed, comprising 4 days a week of practical learning and 1 day of special training and reflection at the training institute. Eight institutes were responsible for the organisation of the training. The content of the programme was based on the basic job description for the GP, and generally aimed at the acquisition of knowledge relevant to general practice, clinical and communication skills, and attitudes. The course of the programme was structured around three blocks of 1 year, starting with general practice training, followed by rotation schemes in hospitals, clinics for chronically ill patients and psychiatric outpatient clinics, and completed by further general practice training.

Instruments

National test committees using national guidelines and consensus procedures developed the knowledge test of skills and the OSCE.[22] Members of the test committees were experienced GPs from the eight institutes with extensive teaching experience in postgraduate training, and test experts. The knowledge test of skills had a progress-testing format, so that the test was designed to reflect the cognitive final objectives of the curriculum.[23] It contained 180 items in the 'true/false/do not know' response format and was constructed according to a blueprint based on the International Classification for Primary Care (ICPC), and a list of clinical skills, representing a cross-section of the domain of clinical skills relevant to general practice. Knowledge of skills was defined as knowledge concerning the technical, procedural aspects of skills, and skill-related aspects (for example, materials and indications). The items were, as far as possible, embedded in vignettes representing general practice situations, to increase relevance for daily practice and assessment of crucial knowledge instead of trivial fact knowledge. An example is presented in Box 1.

The test score was calculated as the sum of the correct minus incorrect answers to discourage guessing (formula scoring) and was expressed as a percentage of the maximum score. The reliability (Cronbach's α) of the test was 0.76, which is satisfactory, in particular for the interpretation of group results used in this study.

The OSCE included 16 stations, again representing a cross-section of the clinical skills relevant for general practice, spread among 16 ICPC categories and reflecting the final objectives of the curriculum. The total testing time was 3 hours. Nine stations involved technical and procedural tasks; in seven stations — the so-called 'integral stations' — complete patient encounters were assessed. Two rating systems were used for each station: a task-specific checklist and a content-independent 10–point global scale.[16] Because evidence suggests that very detailed checklists do not reflect the approach to clinical problems of the more experienced postgraduate trainee,[24] the checklists of our study were less detailed and more expert-oriented than the checklists used in undergraduate medical education. The 10–point scale is the traditional Dutch marking system, with 1 representing highly unsatisfactory performance, 6 representing (barely) sufficient performance and 10 representing excellent performance.

Eighty-four examiners were involved, all experienced GPs as well as teachers at the institutes. Most of them were familiar with the OSCE format. Each examiner was trained in the rating of two different stations. The same examiners rated the same two stations. The purpose of the training was to reach consensus in scoring between raters.

<table>
<tr><td>Content</td><td colspan="4">Basic job description for the general practitioner</td></tr>
<tr><td rowspan="4">Structure</td><td colspan="2">Three blocks of 1 year</td><td colspan="2">Throughout the curriculum</td></tr>
<tr><td>Block 1</td><td>General practice training</td><td rowspan="3">One day per week for special training and reflection, in groups of trainees, at the training institute</td><td rowspan="3">Half-day per week for self-directed learning</td></tr>
<tr><td>Block 2</td><td>Rotations through hospitals, clinics for chronically ill patients and psychiatric outpatient clinics</td></tr>
<tr><td>Block 3</td><td>General practice training</td></tr>
<tr><td rowspan="3">Learning objectives</td><td>Block 1</td><td colspan="3">Acquisition of knowledge, skills and attitudes with emphasis on common problems</td></tr>
<tr><td>Block 2</td><td colspan="3">Acquisition of additional knowledge, skills and attitudes that cannot be learned in general practice itself</td></tr>
<tr><td>Block 3</td><td colspan="3">Integration of the new knowledge, skills and attitudes with emphasis on management of complex situations</td></tr>
</table>

Figure 1. Dutch postgraduate training in general practice at the time of the study (1995–1998).

Mr Monsma, aged 62 years, visits his GP for an annual control of diabetes mellitus type 2 and hypertension. A physical examination of the vascular system is performed by the GP, which comprises, among other tests, the auscultation of the carotid arteries for evidence of stenosis.

- The best place for auscultation is low in the neck (just above the clavicle). (True/False?)

The GP also wants to examine the possible existence of an aneurysm of the aorta.

- It is better to examine an aneurysm of the aorta above the navel than under it. (True/False?)

Bouters EJ, Eikelboom BC, Lanson JA. Perifere vaataandoeningen. Utrecht: Wetenschappelijke Uitgeverij Bunge, 1988.

Box 1. Example of two vignette-based knowledge of skills items.

The checklist score of a station was defined by the percentage of correctly performed items on the checklist. The total test score was calculated by averaging the 16 station scores. The global rating was used for standard setting. Based on the checklist score, the reliability (Cronbach's α) of the test was 0.63, which is satisfactory for the interpretation of group results used in this study.

The correlation (corrected for attenuation, assuming perfect reliability) between the knowledge test of skills and the OSCE was 0.54 for the total OSCE, 0.64 for the technical stations, and 0.36 for the integral stations.

Procedure and subjects

The knowledge test of skills was administered in spring 1998 to all Dutch GP trainees in all 3 years of training (cross-sectional comparison), under supervision of the eight institutes. Six groups of trainees with training times varying from 0 to 30 months were included (Table 1). Owing to practical reasons 755 of the total number of 1029 trainees took the test. It was only possible to investigate whether participants differed from non-participants for the group of trainees in their final stage of training. Out of the 191 trainees of this group, 130 took the test. The participants did not significantly differ from the non-participants in age, time between graduation and start of postgraduate training, entry level of general knowledge, and knowledge of skills (*t*-test, $P>0.05$). The percentage for males was somewhat higher (46% versus 43%). Seven of the eight institutes were represented with a minimum of 10 and a maximum of 30 trainees. In addition, 78 GP trainers, coming from one training institute, volunteered to serve as a reference group.

One and a half months after the knowledge test of skills, the OSCE was performed by a national sample of trainees in their final stage of training. A maximum of 12 trainees were randomly selected from each institute and 73 of them were included in total. The sample did not significantly differ from the non-participants in time between graduation and start of postgraduate training, and in entry level of general knowledge and knowledge of skills (*t*-test, $P>0.05$). The sample was a little younger and the percentage of males was somewhat higher (48% versus 43%). Seven of the eight institutes were represented with a minimum of six and a maximum of 13 trainees. In addition, 35 GP trainers, coming from several training institutes, volunteered to serve as a reference group.

Standard setting procedures

To set a pass mark for the knowledge test of skills, the Cohen–Schotanus method was applied.[25] This absolute method uses the best performing candidates (90th percentile score, P90) as a reference to correct for quality and test difficulty variations, under the assumption that the best performing candidates will always do well regardless of the conditions. Subsequently, an arbitrary but conventional percentage of 55% (i.e. the borderline for the Dutch marking system) of the P90 score is taken as a pass/fail point. The pass mark was thus defined as the score obtained by calculating the 55% score of the P90 score (corrected for guessing) of the sample of trainees in their final stage of training.

To establish the pass mark for the OSCE, the borderline regression method was applied, an absolute norm based on the empirical approach.[26] By using the global ratings of the overall performance (on the 10–point scale) the pass/fail borderline was defined at 5.5 on the scale. For each station the corresponding pass mark on the checklist score scale was obtained by regression analysis of the checklist scores and the global ratings, and then calculating the checklist score on the regression line for the global rating set at 5.5. The corresponding borderline regression pass mark of the total test was defined by taking the mean of the 16 station pass marks.

Analysis

For the knowledge test of skills, mean scores and 95% confidence intervals were calculated for each of the trainees with the same training time and for the GP trainers. Training group 1 consisted of trainees who had just started postgraduate training, training group 2 consisted of trainees with 6 months of training, training group 3 consisted of trainees with 12 months of training, and so on. A statistically significant difference was inferred from non-overlapping 95% confidence intervals. For the OSCE, mean scores and standard deviations were calculated for the trainees and the GP trainers.

The frequency of passing for trainees and GP trainers was calculated using the standard setting procedures described.

Results

Table 1 represents the mean scores, corresponding 95% confidence intervals, and pass rates of the knowledge test of skills for the six training groups and the GP trainers.

The results show an increase in knowledge level of skills and pass rate over the 3 years of training. The increase is the most noticeable in the first 6 months (10.6%), whereas stagnation is observed at the end of the second year of training. The knowledge level was the highest towards the end (48.2%), although this did not differ significantly from the one attained at 18 months. Eighty-eight per cent of the trainees in their final stage of training would have passed. These trainees performed better than the GP trainers, on both their test score (48.2 % versus 39.3 %) and pass rate (88.1% versus 70.5%).

The results of the OSCE are shown in Table 2. For the total OSCE the level of skills towards the end of training was

Table 1. Knowledge test of skills: mean of correct-minus-incorrect score in percentage of total, 95% confidence interval (95% CI), and pass rate for trainees and GP trainers (cross-sectional data).

Training group	Training time (months)	*n*	Correct – incorrect (mean)	95% CI	Pass rate (%)[a]
1	0	114	30.3	28.3–32.2	37.7
2	6	139	40.9	39.0–42.9	67.6
3	12	137	44.1	42.5–45.8	83.2
4	18	133	47.5	45.7–49.3	89.5
5	24	102	45.0	43.2–46.8	85.3
6[b]	30	130	48.2	46.3–50.1	88.1
GP trainers	–	78	39.3	37.3–41.4	70.5

[a]Pass mark = 34.1%. [b]Training level at completion of training.

Table 2. OSCE: mean scores in percentage of total and standard deviation (SD), pass marks, and pass rates for trainees and GP trainers.

	Correct, Mean (SD)			Pass rate (%)	
	Trainees (*n* = 73)	GP trainers (*n* = 35)	Pass mark	Trainees	GP trainers
Total test	68.7 (6.1)	63.3 (7.0)	57.6	94.5	80.0
Technical stations	68.2 (8.2)	59.2 (9.3)	57.4	89.0	60.0
Insertion of nasogastric tube	57.2 (17.8)	42.6 (19.5)	48.7	74.0	41.2
Resuscitation	74.5 (15.1)	57.6 (18.5)	63.3	75.3	37.1
Urinary catheterisation	65.5 (18.0)	60.2 (20.6)	49.3	87.5	71.4
Painful micturition	68.6 (15.3)	61.6 (17.8)	55.5	84.7	74.3
Intravenous cannulation	77.6 (16.1)	46.8 (25.7)	59.1	83.1	32.4
Injection into the shoulder	64.3 (20.3)	68.9 (20.6)	55.9	73.6	79.4
Wound suturing	83.1 (7.5)	83.8 (7.0)	73.0	83.6	91.4
Compression bandage	55.7 (20.0)	56.6 (21.5)	52.7	52.1	60.0
Investigation of vaginal discharge	67.5 (19.0)	53.6 (23.1)	59.1	61.6	31.4
Integral stations	69.4 (5.8)	68.6 (5.9)	57.8	95.9	94.3
Impaired vision	69.2 (11.7)	67.1 (14.1)	56.1	87.7	80.0
Painful knee	78.6 (11.5)	76.9 (12.3)	56.8	100.0	94.3
Fishbone throat	65.5 (20.1)	65.5 (18.8)	53.7	78.1	77.1
Backache	70.8 (11.2)	64.5 (11.1)	61.1	80.8	57.1
Pulmonary reversibility	84.7 (10.2)	81.2 (10.9)	73.0	89.0	80.0
Examination of female breast	65.6 (10.3)	67.7 (14.1)	57.1	80.8	80.0
Annual control of diabetes mellitus type 2	51.0 (11.3)	57.0 (10.5)	46.7	70.8	88.6

68.7%, with a pass rate of 94.5%. For the GP trainers this level was 63.3%, with a pass rate of 80%.

The mean score for the technical stations was somewhat lower (68.2%) than for the integral stations (69.4%). Eighty-nine per cent of the trainees would have passed the technical stations assessments, while 95.1% would have passed the integral stations assessments. For the GP trainers the mean score of the technical stations was also lower (59.2%) than for the integral stations (68.6%). Sixty per cent of them would have passed the technical stations assessments, whereas 94.3% of them would have passed the integral stations assessments.

The mean scores and pass rates of the stations varied considerably both between stations and between groups of candidates, indicating that the competence in skills was not uniform (Table 2).

Discussion

In this study, the acquisition of clinical skills during a 3–year postgraduate training programme for general practice is evaluated. Mastery of clinical skills was one of the essential learning objectives of the curriculum.

The increase in clinical skills over the 3 years was 18% (from 30% to 48%). The greatest acquisition took place in the first 6 months. Thereafter, a continuous but less rapid increase was observed, except for training level 5 in which the increase stagnated. At this level, trainees completed their second year in which they rotated through hospitals, clinics for chronically ill patients, and psychiatric outpatient clinics (Figure 1). Towards the end of training, trainees achieved a satisfactory level of both knowledge and performance of skills. They performed better than the reference group of GP trainers and about 90% of them would have passed according to a standard of adequacy. Moreover, their final level of knowledge of skills was considerably higher than the final level of undergraduates at the end of medical school, as the mean final score of four medical schools on a similar knowledge test of skills varied from 18% to 39%.[27]

Before we come to a discussion of our findings, some methodological issues should be considered. Firstly, the acquisition of clinical skills during training was assessed by a written test of skills in a cross-sectional design. This design requires a careful interpretation of the results. Moreover, the test was not performed by all trainees in all 3 years. Since we had no data with which to compare par-

ticipants and non-participants, we could not control for selection bias. However, since the number of participants was substantial (about 75%), selection effects are less likely to have occurred. Secondly, not all trainees at completion of their training participated both in the written test and the OSCE. Drop out was for practical reasons (written test) and random selection (OSCE). It transpired that only 47 trainees took both tests. For both tests, participants did not significantly differ from non-participants on several aspects. Therefore, and because the number of participants in both tests was relatively large, we think the results of this study are representative for the total group of trainees. Thirdly, the standard-setting procedure for the knowledge test of skills is less well validated than the one for the OSCE.[26] The pass rates of the knowledge test are therefore less convincing than the pass rates of the OSCE. Finally, we compared the trainees to experienced GPs who were not randomly selected from all Dutch GPs. As they volunteered, (and were thus highly motivated) and were GP trainers as well, they probably were not representative of the population of Dutch GPs, and had more knowledge and skills than the average GP.

Our findings highlight some important aspects that need further discussion. Firstly, the stagnation in growth of clinical skills at the end of the second year may possibly indicate that training in a general practice (first and third year) contributes more to the acquisition of clinical competence than training during clinical rotations (second year). This careful conclusion is supported by a study into the acquisition of knowledge during Dutch postgraduate training in general practice.[4] In this study, a stagnation in growth was also found at the end of the clinical rotations. As it had a mixed longitudinal design, which is considered an optimal design for the measurement of change over time, its results are more robust. However, the present study does not necessarily indicate that trainees do not acquire new knowledge and skills in the second year. Perhaps our evaluation system fails to assess the acquired competence. Therefore, a closer examination of the contribution of the second year to the development of clinical competence is recommended. Secondly, the results of the OSCE show a remarkable distinction between the technical and integral stations. More trainees would have passed the integral than the technical stations, indicating that integral skills are better mastered than technical skills. Furthermore, the results of the integral stations are more or less similar for the trainees and GPs, while for the technical stations the GPs performed less well than the trainees. The implication of this finding for the teaching of technical skills needs to be further investigated. Finally, in several studies it has been found that GPs perform less well than trainees at the end of training.[1] Our results are consistent with this finding. The fact that the difference is mainly found in the knowledge test of skills and the technical stations, and not in the integral stations, may indicate that the tests of our study are more appropriate for the assessment of trainees than experienced doctors.

We conclude that a 3–year postgraduate training programme for general practice, which emphasises the acquisition of clinical skills, contributes to a satisfactory acquisition of these skills. Integral skills dealing with complete patient encounters seem to be better acquired than the more technical skills. The evidence also suggests that the second year of the curriculum, in which trainees rotate through hospitals, clinics for chronically ill patients, and psychiatric outpatient clinics, contributes less to the acquisition of clinical skills than the first and third years of training in a general practice. Further research into when and where the acquisition of clinical skills takes place, and into the role of the GP trainer in that, is required. Moreover, a closer examination of the contribution of the second year to the development of clinical competence is recommended.

References

1. Van Leeuwen YD. *Growth in knowledge of trainees in general practice*. [Dissertation] Maastricht: Universitaire Pers Maastricht, 1995.
2. Yudkowsky R. Can resident evaluations demonstrate increase in residents' skills over time? *Acad Med* 1999; **74(10 suppl):** S108-S110.
3. Marel GM, Lyon PM, Field MJ, *et al.* Clinical skills in early postgraduate medical trainees: patterns of acquisition of confidence and experience among junior doctors in a university teaching hospital. *Med Educ* 2000; **34:** 1013-1015.
4. Kramer AWM, Düsman H, Tan LHC, *et al.* Effect of extension postgraduate training in general practice on the acquisition of knowledge of trainees. *Fam Pract* 2003; **20(2):** 207-212.
5. Schmidt H, Norman G, Boshuizen H. A cognitive perspective on medical expertise: theory and implications. *Acad Med* 1990; **65:** 611-621.
6. Ramsey PG. Does clinical performance correlate with physicians' scores on written examinations? In: Scherpbier AJJA, van der Vleuten CPM, Rethans JJ, van de Steegh AFW (eds). *Advances in Medical Education*. *Proceedings of the Seventh Ottawa Conference.* Dordrecht: Kluwer Academic Publishers; 1997.
7. Ram P, van der Vleuten C, Rethans J, *et al.* Assessment in general practice: the predictive value of written knowledge tests and a multiple station examination for actual medical performance in daily practice. *Med Educ* 1999; **33:** 197-203.
8. Lamberts H, Brouwer H, Mohrs J. *Reason for encounter-, episode- and process-oriented standard output from the Transition Project*. Amsterdam: Department of General Practice, University of Amsterdam, 1991.
9. Kelly MH, Campbell LM, Murray TS. Clinical skills assessment. *Br J Gen Pract* 1999; **49:** 447-450.
10. Sturmberg J. Learning relevant procedural skills. *Aust Fam Physician* 1997; **10:** 1163-1165.
11. Tan LHC. *Tekorten in de opleiding van huisartsen*. (*Deficiencies in postgraduate training for general practice*.) [Dissertation with English summary]. Amsterdam: Universiteit van Amsterdam, 1989.
12. Friedlich M, MacRae H, Oandasan I, *et al.* Structured assessment of minor surgical skills (samss) for family medicine residents. *Acad Med* 2001; **76:** 1241-1246.
13. Van der Vleuten CPM, Swanson DB. Assessment of clinical skills with standardized patients: state of the art. *Teach Learn Med* 1990; **2:** 58-76.
14. Van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ* 1996; **1:** 41-67.
15. Van der Vleuten CPM, van Luyk SJ, Beckers HJM. A written test as an alternative to performance testing. *Med Educ* 1989; **23:** 97-107.
16. Jansen JJM, Tan LHC, van der Vleuten CPM, *et al.* Assessment of competence in technical clinical skills of general practitioners. *Med Educ* 1995; **29:** 247-253.
17. Kramer AWM, Zuithoff P, Düsman H, *et al.* Predictive validity of a written knowledge test of skills in postgraduate training for general practice. *Med Educ* 2002; **36:** 812-819.
18. van Dalen J, Kerkhofs E, Verwijnen GM, *et al.* Predicting communication skills with a paper-and-pencil test. *Med Educ* 2002; **36:** 148-153.
19. Swanson DB, Norman GR, Linn RL. Performance-based assessment: lessons from the health professions. *Educational Researcher* 1995; **24:** 5-11, 35.
20. Verhoeven BH, Hamers JGHC, Scherpbier AJJA, *et al.* The effect on reliability of adding a separate written assessment component to an objective structured clinical examination. *Med Educ* 2000;

34: 525-529.
21. Dubois V, Everwijn S, van Geldorp G, *et al. The construction of a new curriculum of postgraduate training for general practice in The Netherlands*. Utrecht: Royal Dutch Medical Association (KNMG), 1987.
22. Grol RPTM. National standard setting for quality of care in general practice: attitudes of general practitioners and response to a set of standards. *Br J Gen Pract* 1990; **40:** 361-364.
23. van der Vleuten CPM, Verwijnen GM, Wijnen WHFW. Fifteen years of experience with progress testing in a problem-based learning curriculum. *Med Teacher* 1996; **18(2):** 103-109.
24. Hodges B, Regehr G, McNaughton N, *et al.* OSCE checklists do not capture increasing levels of expertise. *Acad Med* 1999; **74:** 1129-1134.
25. Cohen-Schotanus J, van der Vleuten CPM. Een betere cesuur bij tentamens: de beste studenten als referentiepunt (A better norm for examinations: the best students as reference point). In: Ten Cate ThJ, Dikkers JH, Houtkoop E, *et al.* (eds). *Proceedings Gezond Onderwijs Congres 1995*. Houten/Diegem 1996.
26. Kramer AWM, Muijtjens AMM, Jansen JJM, *et al.* Comparison of a rational and an empirical standard setting procedure for an OSCE. Accepted for publication *Med Educ* 2003; **37(2):** 132-139.
27. Remmen R, Scherpbier AJJA, Denekens J, *et al.* Effectiveness of basic clinical skills training programmes: a cross-sectional comparison of four medical schools. *Med Educ* 2001; **35:** 121-128.

Acknowledgements

We would like to thank all trainees and GPs who took part in the study. We are grateful for the support from the staff of the eight Dutch training institutes for postgraduate training in general practice. The members of the national committees for the knowledge test of skills and the OSCE are thanked for their contribution to the development of the tests. Peter Zuithoff is thanked for his contribution to the analysis of the data. IJsbrand Kramer polished the English. This study was initiated by the Registration Committee of Postgraduate Training in General Practice (HVRC), financially supported by the Foundation of Postgraduate Training in General Practice (SBOH), and executed by the National Centre for Evaluation of Postgraduate Training in General Practice (SVUH).