

Toetsing van technische vaardigheden van huisartsen : studies naar toepassingsmogelijkheden van vaardigheidstoetsing in deskundigheidsbevordering

Citation for published version (APA):

Jansen, K. (1998). *Toetsing van technische vaardigheden van huisartsen : studies naar toepassingsmogelijkheden van vaardigheidstoetsing in deskundigheidsbevordering*. [Doctoral Thesis, Maastricht University]. Universiteit Maastricht. <https://doi.org/10.26481/dis.19980604kj>

Document status and date:

Published: 01/01/1998

DOI:

[10.26481/dis.19980604kj](https://doi.org/10.26481/dis.19980604kj)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 16 Jan. 2025

Samenvatting

Technische vaardigheden vormen belangrijk gereedschap bij het dagelijkse handelen van de Nederlandse huisarts. Onder medisch-technische vaardigheden worden in dit proefschrift vaardigheden verstaan die de huisarts gebruikt bij het uitvoeren van patiëntgebonden diagnostische of therapeutische handelingen. In concreto gaat het daarbij om lichamelijk onderzoek, om aanvullend onderzoek in de vorm van functietests of laboratoriumonderzoek, en om therapeutische ingrepen. In de studies die in dit proefschrift beschreven worden, is onderzocht met welke methoden en instrumenten de deskundigheid van huisartsen op het gebied van technische vaardigheden getoetst kan worden in het kader van de deskundigheidsbevordering van huisartsen.

In hoofdstuk I worden de achtergronden van het onderzoeksproject geschetst. De professionalisering van de huisartsgeneeskunde en maatschappelijke behoefte aan meer systematische aandacht voor de kwaliteit van de huisartsgeneeskunde, hebben de basis gelegd voor de ontwikkeling van een kwaliteitssysteem voor de huisartsgeneeskunde. In dat verband worden richtlijnen voor effectieve zorg, toetsingsmethoden om vast te stellen in hoeverre de feitelijke zorg daarmee overeenstemt, en methoden voor kwaliteitsverbetering ontwikkeld. Technische vaardigheden zijn een van de aandachtsgebieden.

Er blijkt weinig informatie te zijn over de feitelijke vaardigheidsbeheersing van praktizerende huisartsen voor wat betreft de vaardigheden uit het Basistakenpakket. Bovendien hebben zich sinds het Basistakenpakket is vastgesteld ontwikkelingen voorgedaan, waardoor wellicht herziening van het pakket gewenst is.

Een tweede terrein waarop onduidelijkheid bestond gold de geschiktheid van diverse methoden van toetsing voor het meten van vaardigheidsbeheersing van praktizerende huisartsen.

Vanuit de noodzaak om toetsing in nascholing te integreren, was een derde punt van aandacht of toetsing ook de effectiviteit van nascholing ten goede zou komen. In het bijzonder waren we daarbij geïnteresseerd in de mogelijkheden en effecten van feedback naar de deelnemers. Tot slot moest het onderzoek meer duidelijkheid opleveren over de acceptatie en kosten van vaardigheidstoetsing en de organisatorische randvoorwaarden die daarbij gewenst zijn, teneinde een beter beeld te krijgen van de toepasbaarheid.

De volgende vraagstellingen werden geformuleerd:

1. *Welke medisch-technische vaardigheden vallen binnen het domein van de Nederlandse huisartsgeneeskunde? Welke vaardigheden dienen prioriteit te krijgen in het kader van deskundigheidsbevordering en toetsing?*
2. *Welke methoden en instrumenten zijn geschikt voor toetsing van technische vaardigheden van huisartsen en wat zijn hun meettechnische eigenschappen?*
3. *Wat is de effectiviteit van nascholing en toetsing van technische vaardigheden op het handelen in de praktijk? Heeft feedback over toetsingsresultaten effect op zelfbeoordeling?*
4. *Hoe is de acceptatie onder huisartsen van educatieve toetsing van technische vaardigheden? Welke vorm heeft de meeste voorkeur? Welke zijn de kosten en organisatorische randvoorwaarden voor toetsing van vaardigheden?*

In hoofdstuk 2 wordt het domein van medisch-technische vaardigheden voor de Nederlandse huisarts in kaart gebracht. Op basis van diverse huisartsgeneeskundige literatuurbronnen, werd een inventarisatie gemaakt, hetgeen uitmondde in een lijst met 263 vaardigheden, waarvan 85 % onder het Basistakenpakket gegroepeerd kan worden.

Uit de lijst met vaardigheden werden 80 vaardigheden geselecteerd die verondersteld werden prioriteit te hebben voor toetsing en nascholing. Deze lijst werd voorgelegd aan een twintigtal huisarts-coördinatoren deskundigheidsbevordering, die een rangordening wat betreft prioriteit aanbrachten. Met name vaardigheden op het terrein van oogheelkundige diagnostiek, onderzoek van het bewegingsapparaat, onderzoek bij hart- en vaatziekten, en vaardigheden in verband met thuiszorg, bleken hoge prioriteit te hebben. Deze vaardigheden vormden het uitgangspunt bij de keuze van onderwerpen in de diverse experimenten.

In hoofdstuk 3 wordt de literatuur besproken over vaardigheidstoetsing. Vaardigheidstoetsing is in de zeventiger jaren in het medisch curriculum geïntroduceerd in de vorm van het zogenaamde 'objective structured clinical examination' of examenvormen gebaseerd op 'standardized-patients', vanuit de behoefte om, naast kennistoetsing, praktisch medisch handelen te toetsen in situaties die de werkelijkheid zo goed mogelijk nabootsen, maar niet het nadeel hebben van de willekeur van het traditionele patient-gebonden examen. Geleidelijk heeft de examenvorm ook zijn weg gevonden naar specialisten-opleidingen. Het onderzoek waarmee de introductie van deze examenvorm gepaard ging heeft goede ondersteuning voor de validiteit opgeleverd bij een redelijke betrouwbaarheid. Er is in de literatuur echter ook kritiek dat de examenvorm te rigide en/of triviaal van inhoud zou zijn om competentie van meer ervaren artsen te toetsen.

In verband met de complexiteit van de organisatie bij vaardigheidstoetsing is er ook beperkt onderzoek verricht naar mogelijke alternatieven die goedkoper zijn en eenvoudig om toe te passen. Uit dit onderzoek bleek dat een kennistoets over vaardigheden een goede voorspeller is voor vaardigheidsbeheersing bij gevorderde studenten. Ook zelf-beoordeling is onderzocht op de voorspellende waarde voor vaardigheidsbeheersing. Dit leverde wisselende resultaten op, waarbij echter werd opgemerkt dat zelf-beoordeling een vaardigheid betreft die geleerd moet worden.

Op grond van deze bevindingen uit de literatuur werden behalve de vaardigheidstoets, ook kennistoetsing over vaardigheden en zelfbeoordeling in de uitwerking van de experimenten opgenomen, waarover in de verdere hoofdstukken wordt gerapporteerd.

In hoofdstuk 4 wordt verslag gedaan van een experiment waarbij een vaardigheidstoets (bestaande uit acht stations), een kennistoets over vaardigheden (125 vragen) en een zelfbeoordelingslijst (41 items) met elkaar werden vergeleken. Deelnemers aan het experiment waren 49 huisartsen en 47 huisartsen-in-opleiding. De gemiddelde scores tussen huisartsen en huisartsen-in-opleiding bleken niet significant te verschillen op de vaardigheidstoets en de

kennistoets over vaardigheden, terwijl de huisartsen wel significant hoger scoorden op de zelfbeoordelingslijst. De betrouwbaarheid van rangordening van deelnemers op basis van de vaardigheidstoetsscore was niet erg hoog (0.48) in vergelijking met de kennistoetsscore (0.68) en zelfbeoordelings-score (0.92). De samenhang tussen vaardigheidstoetsscore en kennistoets-score bleek hoog te zijn, terwijl de samenhang met de zelfbeoordeling een stuk lager was. Op basis van dit onderzoek werd geconcludeerd dat de kennistoets over vaardigheden een redelijk alternatief vormt voor de vaardigheidstoets, met name voor screeningsdoeleinden en onderzoek.

In hoofdstuk 5 wordt verslag gedaan van een tweede experiment, waarbij toetsing werd geïntegreerd in nascholing voor huisartsen. In dit experiment werd met name de construct-validiteit van de vaardigheidstoets nader onderzocht, alsmede de onderlinge samenhang tussen kennis en beheersing voor afzonderlijke vaardigheden. Daartoe werd een vaardigheidstraining (bestaande uit vier verschillende onderwerpen) gegeven aan 71 huisartsen. Het effect van de training werd gemeten met een vaardigheidstoets (vier stations) en een kennistoets over vaardigheden (60 vragen). De vaardigheidstoets liet een duidelijk trainingseffect zien voor alle onderwerpen, terwijl de kennistoets over vaardigheden dat voor drie onderwerpen liet zien. De samenhang tussen kennisscore en vaardigheidstoetsscore bleek echter laag te zijn voor al de afzonderlijke onderwerpen. Op grond hiervan werd geconcludeerd dat de vaardigheidstoets een goede construct-validiteit heeft, terwijl de samenhang tussen kennis over afzonderlijke vaardigheden en beheersing van die vaardigheden gering is.

In hoofdstuk 6 wordt het bij vaardigheidstoetsen veel gebruikte scoringssysteem - scoringslijsten die worden ingevuld door observatoren - nader beschouwd. Dit gebeurt aan de hand van het station 'basale reanimatie', één van de vaardigheden uit het tweede experiment. De resultaten van het oordeel van een observator werden vergeleken met de resultaten van mechanische registratie verkregen tijdens het reanimeren. Beide scoringssystemen lieten een duidelijk trainingseffect zien bij de deelnemers. De observatoren bleken een goede onderlinge overeenstemming te hebben over het scoren van de diagnostische handelingen (0.77), terwijl de overeenstemming over de reanimatie handelingen beduidend lager was (0.56). Ook de correlatie tussen de scoringslijst en de mechanische registratie was laag (0.45) voor de reanimatie handelingen. Opvallend was ook dat de correlatie tussen de score voor diagnostische en die voor reanimatie handelingen laag was (0.22) voor beide scoringssystemen. Op basis van deze studie kan geconcludeerd worden dat de mechanische registratie een beter beeld geeft van de reanimatie handelingen dan de score door de observator. Omdat de score voor de reanimatie handelingen een slechte voorspelling geeft voor de diagnostische handelingen wordt echter een scoringssysteem aanbevolen dat zowel de diagnostische fase (middels een observator) als de reanimatie handelingen (middels mechanische registratie) omvat, zoals overigens ook veelal wordt gebruikt.

In hoofdstuk 7 wordt onderzocht in hoeverre een vaardigheidstraining, waarbij toetsing een integraal onderdeel van de cursus vormt, behalve een effect op de competentie ook een effect op het handelen in de praktijk heeft. De vaardigheidstraining omvatte een viertal vaardigheden: schouder injectie, cervixuitsrijk, fluor diagnostiek en fundoscopie bij diabetes mellitus. De deelnemende huisartsen ($n=59$) werden verdeeld in een interventie- en controlegroep. Alle deelnemers vulden een schriftelijke kennistoets over vaardigheden in en registreerden in hun praktijk gedurende twintig werkdagen hoe vaak zij de vier vaardigheden uitvoerden. De interventiegroep ontving vervolgens de vaardigheidscursus. De scores op de vaardigheidstoets na afloop van de training wezen op een redelijke tot goede vaardigheidsbeheersing. Daarop werd de schriftelijke toets herhaald voor alle deelnemers en volgde een tweede registratieperiode van twintig werkdagen.

De kennistoetsscores lieten een duidelijk trainingseffect zien van de vaardigheidscursus. Voor twee onderwerpen werd ook een trainingseffect op het handelen in de praktijk gevonden (schouder injectie en cervixuitsrijk), terwijl voor fluordiagnostiek en fundoscopie bij diabetes mellitus geen effect werd gevonden. Geconcludeerd wordt dat de effectiviteit van vaardigheidstraining om veranderingen in de praktijk te bewerkstelligen afhankelijk is van het onderwerp. Voor sommige vaardigheden is een training geschikt en voldoende om gewenste veranderingen te bewerkstelligen, terwijl voor andere vaardigheden waarschijnlijk meer complexe interventies nodig zijn.

In hoofdstuk 8 wordt nader ingegaan op zelfbeoordeling als methode om vaardigheidsbeheersing vast te stellen. Doel van het onderzoek waarover in dit hoofdstuk wordt gerapporteerd was na te gaan in hoeverre persoonlijke feedback over de scores op de kennistoets over vaardigheden en vaardigheidstoets de nauwkeurigheid van zelfbeoordeling verbeterde. De deelnemende huisartsen van het experiment, beschreven in hoofdstuk 7, werden aan het begin van de onderzoeksperiode, na drie maanden en na zes maanden getoetst op kennis (60 vragen) en vaardigheden (4 stations) en vulden een zelfbeoordelingslijst (22 items) in. Na verwerking van de resultaten ontvingen de deelnemers steeds persoonlijk feedback over hun scores. Na drie maanden bleken de kennistoets score en de zelfbeoordeling in de interventiegroep sterker gestegen in vergelijking met de controlegroep. Na zes maanden - nadat de controlegroep eveneens de training had gevolgd - waren de scores van beide groepen weer gelijk. De correlaties tussen de zelfbeoordeling en objectieve test scores waren vrij laag, zonder toename van de verklaarde variantie bij de achtereenvolgende meetmomenten.

Op basis van deze bevindingen lijken zelfbeoordelings scores in zekere mate in staat om op groepsniveau veranderingen in vaardigheidsbeheersing te meten. Echter, op individueel niveau, vormt zelfbeoordeling geen valide bron van informatie omtrent de beheersing van vaardigheden, en dit verbetert ook niet na feedback.

In hoofdstuk 9 wordt de waardering van de deelnemers voor vaardigheidstraining en -toetsing

beschreven, en wordt een overzicht gegeven van de kosten. Bij alle drie in eerdere hoofdstukken beschreven experimenten werden de deelnemers met een anonieme schriftelijke enquête gevraagd om hun mening over en waardering van de toetsing. De deelnemers waardeerden de toetsing als positief. De waardering was hoger bij huisartsen in vergelijking met huisartsen-in-opleiding, was hoger voor vaardigheidstoetsing in vergelijking met toetsing van kennis over vaardigheden, en was hoger bij experimenten waarbij de toetsing onderdeel uitmaakte van nascholing.

De kosten van vaardigheidstoetsing bedroegen ongeveer NLG 40,- per persoon per vaardigheid. Voor de combinatie van training en toetsing waren de kosten ongeveer NLG 100,-. De kosten zijn daarmee vergelijkbaar met de kosten van reguliere nascholing.

In hoofdstuk 10 worden de belangrijkste resultaten uit het onderzoek samengevat en van kanttekeningen voorzien. De volgende bevindingen worden als de meest belangrijke beschouwd. Ten eerste blijkt de vaardigheidstoets een goede validiteit te hebben voor het meten van technische vaardigheidsbeheersing bij huisartsen. Voor afzonderlijke vaardigheden kan een betrouwbare individuele beoordeling gegeven worden, maar voor het meten van het algemene vaardigheidsniveau zijn lange toetstijden nodig vanwege het inhoudsspecifieke karakter van vaardigheidsbeheersing. De kennistoets over vaardigheden vormt een goed alternatief voor een betrouwbare beoordeling van het individuele algemene vaardigheidsniveau, maar is niet geschikt voor het beoordelen van het vaardigheidsniveau voor afzonderlijke vaardigheden. Individuele zelfbeoordeling lijkt geen valide methode om vaardigheidsbeheersing vast te stellen.

Een tweede bevinding vormt het duidelijke effect van vaardigheidstraining op de competentie van de deelnemers, en de goede retentie die werd gevonden enkele maanden na de cursus. Bovendien werd bij twee van de vier vaardigheden waarbij dat was onderzocht, ook een effect van de nascholing op het handelen in de praktijk gevonden.

Een derde bevinding geldt de waardering en kosten van vaardigheidstoetsing in het kader van deskundigheidsbevordering voor huisartsen. De waardering blijkt hoog, en de meerkosten van toetsing zijn relatief bescheiden. De organisatie is relatief complex, maar goed realiseerbaar indien voldaan wordt aan een aantal randvoorwaarden. Daarom kan gesproken worden van een goede haalbaarheid van vaardigheidstoetsing als onderdeel van nascholing voor huisartsen. Vervolgens worden een aantal methodologische kanttekeningen geplaatst bij de wijze waarop het onderzoek vorm heeft gekregen, en meer in het bijzonder bij verschillende aspecten van vaardigheidstoetsing.

Afgesloten wordt met aanbevelingen voor verder onderzoek, enerzijds gericht op een verdere onderbouwing van validiteit en educatieve waarde van vaardigheidstoetsing in nascholing voor huisartsen, anderzijds gericht op onderzoek naar een goede normstelling voor vaardigheidsbeheersing, met het oog op selectief gebruik van toetsing. De aanbeveling voor de praktijk betreft de noodzaak om goed doordachte ondersteuning te bieden wil men vaardigheidstoetsing in de praktijk ook ingang doen vinden.

Summary

Performance of technical clinical procedures constitute an important part of the work of general practitioners in the Netherlands. In this thesis technical clinical procedures of general practitioners are defined as patient related diagnostic and therapeutic procedures performed by a general practitioner. Examples of such procedures are physical examination, laboratory tests and minor surgery. In this thesis various methods for assessment of competence in technical clinical skills of general practitioners are explored in the context of continuing medical education.

In chapter 1 the background of the research project is described. Developments within the profession of general practitioners, as well as an increasing public demand for greater accountability with respect to quality of care, have provided the basis for the development of a quality system for Dutch general practice. The necessary elements for the quality system, such as national practice guidelines, assessment tools to ascertain how actual care compares with the guidelines, and methods for quality improvement for different dimensions of competence, are being developed. Technical clinical skills was considered one of the dimensions of interest.

Little information is available concerning the proficiency of practicing general practitioners with respect to the relevant technical procedures in general practice as agreed on by the profession. With respect to more recent technical developments in care which require training of new skills it is not known if and how general practitioners acquire these skills.

Related to the question which technical skills should be considered as essential to general practitioners, was the question how competence in technical clinical skills of general practitioners could best be measured, and how these methods could be integrated into the continuing medical education system. We hypothesized that inclusion of assessment and feedback to participants of continuing medical education would increase its efficacy. And finally the research project had to assess the feasibility of assessment, with regard to acceptability, cost and organisational requirements.

The research questions were:

1. *Which technical clinical skills are relevant for the general practitioner? Which skills should receive priority in continuing medical education and assessment?*
2. *Which methods are appropriate for assessment of technical clinical skills of general practitioners. What are the psychometric characteristics of these methods?*
3. *What is the effect of training and assessment of technical clinical skills on performance in practice? Does personal feedback enhance accuracy of self-assessment?*
4. *What is the acceptability among general practitioners of formative assessment of technical clinical skills? Which format has preference? What are the costs and organisational requirements for assessment of technical clinical skills?*

The subject of chapter 2 is which technical clinical skills are actually relevant for the general practitioner in the Netherlands. Based on a search in general practice literature a list of 263 technical skills was compiled. Of these 85% can be considered essential to the general practitioner.

From this list 80 skills were selected which were considered as having priority for assessment and continuing medical education. Subsequently twenty general practitioners involved as coordinators in continuing medical education rankordered these skills in terms of priority. Ophthalmological diagnostic skills, physical examination of the locomotor system, diagnostic skills of the cardiovascular system, and skills related to intensive care at home, were categorized as having most priority. This priority list was used to select topics for the various experiments.

Chapter 3 reviews the literature on performance-based assessment. Performance-based assessment was introduced into the medical curriculum in the nineteen seventies as the so called 'objective structured clinical examination' or 'standardized patient-based' examination. This development was caused by the dissatisfaction with knowledge testing as dominant mode of competence assessment, and the subjectivity involved in most existing clinical assessments. Evaluation of performance of clinical tasks resembling real practice as close as possible, without the disadvantages of the traditional viva exam, was considered as a useful complement. The performance-based assessment format gradually also found it's way to postgraduate education. The research accompanying the introduction of this assessment method has provided support for good validity and reasonable reliability. However, the method has also been criticised for rigidity and/or trivialization of content, claiming it to be less suitable for assessment of competence of more experienced physicians.

Because performance-based assessment requires considerable resources some researchers have looked for alternative assessment methods which are less costly and easy to apply. A written knowledge test of skills showed good predictive validity for competence in technical clinical skills of advanced medical students. Self-assessment of technical clinical skills produced mixed results, with various authors commenting that self-assessment seemed to be a skill which has to be mastered. Based on the findings in the literature the performance-based test, the knowledge test of skills and self-assessment were included as assessment methods in the various studies reported in the following chapters.

Chapter 4 describes an experiment investigating the psychometric characteristics of three different methods for assessment of competence in technical clinical skills for general practitioners. A performance-based test (8 stations), a written knowledge test of skills (125 items) and a self-assessment questionnaire (41 items) on technical clinical skills were administered to 49 GPs and 47 trainees in general practice. The mean scores on the performance-based test and the written knowledge test of skills showed no substantial differences between general practiti-

oners and trainees, whereas the general practitioners scored higher on the self-assessment questionnaire. Norm-referenced reliability of the performance-based test was moderate (0.48) compared to the knowledge test of skills (0.68) and the self-assessment questionnaire (0.92). While the correlation of the score on the knowledge test of skills with the score on the performance-based test was moderately high, the score on the self-assessment questionnaire showed a rather low correlation with the performance-based test.

It was concluded that, although performance-based testing is obviously the best method to assess proficiency in hands-on skills, a written test can serve as a reasonable alternative, particularly for screening and research purposes.

Chapter 5 reports the results of a study with assessment integrated into continuing medical education for general practitioners. The study focussed on demonstrating construct-validity of the performance-based test for technical clinical skills, and exploring the correlation between performance and knowledge of specific skills. A one-day skills training was given to 71 general practitioners, covering four different technical clinical skills.

The effect of the training on performance was measured with a performance-based test (4 stations) using a randomized controlled trial design, while the effect on knowledge was measured with a written test (60 items) administered one month before and directly after the training. A training effect was found with the performance-based test for all four clinical skills. The written test also demonstrated a training effect for all but one skill. However, correlations between scores on the written test and on the performance-based test were low for all skills. It is concluded that construct validity of a performance-based test for technical clinical skills of general practitioners was demonstrated, while the knowledge test score showed to be a poor predictor of competence for specific technical skills.

In chapter 6 the scoring system which is predominantly used in performance-based assessment - with raters marking checklists while observing performance - is further explored. For cardiopulmonary resuscitation - one of the skills trained in the study reported in chapter 5 - checklist-based scores and mechanical recording scores were compared. Both checklist and recording strip based scores showed significant improvement after instruction, but only 37% were judged proficient according to the American Heart Association standards (checklist scoring), and 47% according to the recording print based scoring system, while raters judged 97% as satisfactory by general impression. Interrater reliability between observers was high for the diagnostic procedures (0.77) but much lower for CPR-performance (0.56). Comparison of checklist and recording print showed that the checklist was specific but not very sensitive in identifying poor performance for cardiac compression rate, since observers overestimated performance. The correlation for CPR-performance between checklist score and recording strip score was low (0.45), indicating that candidates were ranked differently. The correlation between diagnosis and performance score was low for checklist as well as recording print

(0.22), indicating that the score on diagnosis was a poor predictor for the score on performance of CPR. These results support the use of the recording manikin as compared with the use of a checklist for formative evaluation of basic life support skills. However, as proficiency in diagnosis and performance in CPR are poorly correlated, assessment of diagnosis using a checklist must be included. Therefore the combination of assessment by observers using a checklist for diagnostic procedures and the recording strip of the manikin for performance of CPR, as employed in most evaluation schemes, is recommended.

Chapter 7 reports the results of a study investigating whether a short course of technical clinical skills with performance-based assessment integrated into the course has an effect on performance in practice. The course covered four different technical clinical skills (shoulder injection technique, PAP-smear, laboratory examination of *fluor vaginalis*, ophthalmoscopic control in diabetes mellitus). Subjects were self-selected general practitioners ($n=59$), who were assigned to the intervention group ($n=31$) or control group ($n=28$) according to their preference for date of a course. The intervention group received the course three months after enrollment in the study, while the control-group received the training after the study period. Main outcome measures used were pre- and post-training scores on a knowledge test (60 items) and pre- and post-training performance of procedures in practice using a log-diary covering 20 days. Competence as measured with the knowledge test improved significantly as a result of the training, and skills test scores were satisfactory after training. A significant effect on performance was found for two out of four skills (shoulder injection and PAP-smear) whereas no effect could be demonstrated for the two other skills. It is concluded that a good degree of competence is a necessary but not always a sufficient condition for a physician to alter his performance in daily practice. While for some skills training seems adequate and sufficient to bring about desired changes, for other skills more complex interventions are needed.

Chapter 8 focusses on self-assessment as a method to determine competence in technical clinical skills. The purpose of the study reported was to ascertain if repeated personal feedback on knowledge test and performance-based test scores would enhance accuracy of self-assessment with regard to competence of technical clinical skills. Participants of the study described in chapter 7 completed a self-assessment questionnaire (22 items) covering four technical clinical skills and were assessed on relevant knowledge (60 items multiple choice test) at the start of the study period, again after three months and after six months. A performance-based test (four stations) was administered after three and six months. After every assessment participants received personal feedback on their scores. At three months mean scores on the self-assessment questionnaire and knowledge test had increased significantly more in the intervention group compared to the control group, while after six months, after the control group had also received the training - no differences remained. Correlations between self-assessment rating and objective scores were low to moderate, with little overall improvement

of explained variance.

It is concluded that while self-assessment scores at the group level can be useful to some extent in measuring perceived changes in competence, individual self-assessment scores are an invalid source of information concerning competence of practicing physicians, and this does not improve significantly with regular feedback.

Chapter 9 reports on the acceptability of performance-based assessment and cost of the training and assessment applied in the various studies. Participants were asked to give their opinion on the assessment procedures and content of the course with an anonymous questionnaire. Participants valued the assessment as positive. Acceptability was higher among practicing physicians compared to trainees in general practice, higher for performance-based assessment compared to knowledge assessment, and higher when assessment was integrated into continuing medical education. The cost of performance-based assessment amounted to NLG 40,- per person per skill, and combined with training to NLG 100,- per person per skill. These costs are comparable to costs of regular CME in the Netherlands.

In chapter 10 the main findings are summarized and discussed. First, the performance-based test demonstrated good validity for assessment of technical skills of general practitioners. A reliable individual assessment is possible for specific skills. However for a reliable assessment of individual general competence in technical skills many stations are required and consequently long testing time is needed. This is due to the case-specificity of competence in technical clinical skills. The knowledge test of skills is considered a reasonable alternative for assessment of individual general competence in skills, whereas it is not suitable for individual assessment of separate skills. Individual self-assessment does not provide valid results for competence in technical clinical skills. A second finding is the positive effect of skillstraining, including assessment, on competence with good retention various months after the course. Moreover, an effect of the course on performance in practice could be demonstrated for two out of four skills in which this was investigated.

A third finding concerns the acceptability and cost of assessment as part of continuing medical education. Acceptability is high and costs are reasonable. Organisation is complex, but feasible if certain requirements are met. Therefore feasibility of assessment of technical clinical skills in continuing medical education of general practitioners is favorable.

Subsequently some methodological problems are highlighted with respect to how the research was undertaken. This chapter is concluded with recommendations for research on validity and efficacy of performance-based testing in continuing medical education, and on developing standards for proficiency of technical clinical skills in general practice for selective assessment. Finally it is recommended to develop a sound strategy if implementation of performance-based assessment in CME is to be successful.