

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/154010>

Please be advised that this information was generated on 2019-04-20 and may be subject to change.

M.J.M. Maas, MSc, Radboud University Nijmegen Medical Center, Scientific Institute for Quality of Healthcare (IQ Healthcare), PO Box 9101, 6500 HB Nijmegen, the Netherlands, and School for Physical Therapy, HAN University of Applied Sciences, Nijmegen, the Netherlands. Address all correspondence to Ms Maas at: m.maas@iq.umcn.nl or marjo.maas@han.nl.

P.J. van der Wees, PhD, Radboud University Medical Center, Scientific Institute for Quality of Healthcare (IQ Healthcare).

C. Braam, MSc, Radboud University Medical Center, Scientific Institute for Quality of Healthcare (IQ Healthcare).

J. Koetsenruijter, MSc, Radboud University Medical Center, Scientific Institute for Quality of Healthcare (IQ Healthcare).

Y.F. Heerkens, PhD, Research Centre for Rehabilitation, Work, and Sports, HAN University of Applied Sciences, and Dutch Institute of Allied Health Care, Amersfoort, the Netherlands.

C.P.M. van der Vleuten, PhD, Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, the Netherlands.

M.W.G. Nijhuis-van der Sanden, PhD, Radboud University Medical Center, Scientific Institute for Quality of Healthcare (IQ Healthcare).

[Maas MJM, van der Wees PJ, Braam C, et al. An innovative peer assessment approach to enhance guideline adherence in physical therapy: single-masked, cluster-randomized controlled trial. *Phys Ther.* 2015;95:600–612.]

© 2015 American Physical Therapy Association

Published Ahead of Print:
September 18, 2014

Accepted: August 28, 2014

Submitted: October 9, 2013

An Innovative Peer Assessment Approach to Enhance Guideline Adherence in Physical Therapy: Single-Masked, Cluster-Randomized Controlled Trial

Marjo J.M. Maas, Philip J. van der Wees, Carla Braam, Jan Koetsenruijter, Yvonne F. Heerkens, Cees P.M. van der Vleuten, Maria W.G. Nijhuis-van der Sanden

Background. Clinical practice guidelines (CPGs) are not readily implemented in clinical practice. One of the impeding factors is that physical therapists do not hold realistic perceptions of their adherence to CPGs. Peer assessment (PA) is an implementation strategy that aims at improving guideline adherence by enhancing reflective practice, awareness of professional performance, and attainment of personal goals.

Objective. The purpose of this study was to compare the effectiveness of PA with the usual case discussion (CD) strategy on adherence to CPGs for physical therapist management of upper extremity complaints.

Design. A single-masked, cluster-randomized controlled trial with pretest-posttest design was conducted.

Intervention. Twenty communities of practice ($n=149$ physical therapists) were randomly assigned to groups receiving PA or CD, with both interventions consisting of 4 sessions over 6 months. Both PA and CD groups worked on identical clinical cases relevant to the guidelines. Peer assessment focused on individual performance observed and evaluated by peers; CD focused on discussion.

Outcomes. Guideline adherence was measured with clinical vignettes, reflective practice was measured with the Self-Reflection and Insight Scale (SRIS), awareness of performance was measured via the correlation between perceived and assessed improvement, and attainment of personal goals was measured with written commitments to change.

Results. The PA groups improved more on guideline adherence compared with the CD groups (effect=22.52; 95% confidence interval [95% CI]=2.38, 42.66; $P=.03$). The SRIS scores did not differ between PA and CD groups. Awareness of performance was greater for the PA groups ($r=.36$) than for the CD groups ($r=.08$) (effect=14.73; 95% CI=2.78, 26.68; $P=.01$). The PA strategy was more effective than the CD strategy in attaining personal goals (effect=0.50; 95% CI=0.04, 0.96; $P=.03$).

Limitations. Limited validity of clinical vignettes as a proxy measure of clinical practice was a limitation of the study.

Conclusions. Peer assessment was more effective than CD in improving adherence to CPGs. Personal feedback may have contributed to its effectiveness. Future research should address the role of the group coach.



Post a Rapid Response to
this article at:
ptjournal.apta.org

Clinical practice guidelines (CPGs) are designed to facilitate evidence-based practice and to improve the quality of health care.¹ The purpose of guidelines is to enhance transparency of care, to reduce unwarranted variability in practice, and to increase accountability to external stakeholders.² Despite a multitude of implementation strategies, research has demonstrated unambiguously that CPGs are not readily implemented in everyday clinical practice.^{3,4} The main bottlenecks for practitioners are attributable to knowledge, attitudes, and factors concerning social, organizational, and societal support.⁵ Because education is assumed to be the first step to behavioral change in clinical practice, a variety of educational interventions have been designed to address knowledge, skills, and attitudes.⁶ Systematic reviews studying the effectiveness of educational strategies, however, have shown little to moderate effects in improving evidence-based practice.⁷

Rutten et al⁸ assessed the effectiveness of a quality improvement program aimed at professional and organizational behavioral change in physical therapist practice. Guideline adherence was assessed by clinical vignettes in a one-group pretest-posttest design. They found a 3.1% increase in adherence. Wensing et al⁶ reported a mean effect of 5% on different aspects of clinical practice, regardless of the type of educational intervention. Research showed that the effectiveness of educational strategies might improve when the intervention addresses small groups and allows for active participation and social interaction.⁹ In addition, change may be more likely if strategies are specifically chosen to address identified barriers to change.¹⁰ Bekkering et al¹¹ showed moderate improvement of adherence to CPGs by physical therapists in the Netherlands through active

educational strategies (discussion, role playing) compared with standard passive methods of guideline dissemination in physical therapy. Guideline adherence of physical therapists depends on levels of awareness of guideline-consistent behavior. Rutten et al¹² used clinical vignettes to compare self-reported and externally assessed adherence. Realistic perceptions of adherence to CPGs were found in 38.5% of the participants. Differences in levels of awareness interfered with other determinants of guideline adherence, such as motivation to change. Research showed that health care professionals have a limited ability to accurately assess their own level of competence,^{13,14} which they systematically overestimate or underestimate.^{15,16}

The development of adequate self-perception requires both internal and external information about one's professional performance as well as knowledge of appropriate performance standards.¹⁷ This finding is supported by studies showing that the effect of educational strategies on evidence-based practice increases when they are combined with other strategies, such as audit and feedback.^{3,18} Yet, audit and feedback have not consistently been found effective to change practice. A systematic review by Ivers et al¹⁹ showed mean improvements of adherence to desired practice of 4.3% for dichotomous outcomes and 1.3% for continuous outcomes. Whether feedback is accepted and used to change professional practice depends on a multitude of variables.^{20,21} Clinicians struggle with accepting feedback when it is incongruent with their self-assessment or threatens their self-confidence.^{17,22} Feedback appears to be more acceptable²⁰ when it is provided in an environment of trust and mutual respect, and it is likely to be rejected when the provider is not perceived to be a

credible and trustworthy source of information^{17,21} or when it conflicts with personal or group norms and values.²³ Acceptance may be enhanced when feedback is tailored to the stages of change as described by Prochaska et al,²⁴ and when it closely connects to the context of daily practice.^{5,25}

Situated learning theory, based on studies by Lave and Wenger²⁶ and Li et al,²⁷ shows that professional knowledge acquired in a certain situation transfers only to similar situations. Their studies support the assumption that feedback provided within communities of practice (CoPs) has greater impact on the improvement of clinical practice than feedback provided by "outsiders." Moreover, the involvement of CoP participants in each other's professional development process may facilitate acceptance of feedback and alignment with personal learning needs and goals.²⁸⁻³⁰

Drawing on these considerations, we introduced peer assessment (PA) as a new implementation strategy for clinical guidelines within existing CoPs. Peer assessment is the process whereby professionals evaluate or are being evaluated by their peers and provide each other with performance feedback. The positive impact of PA on learning and change has been well researched in higher education³¹⁻³³ and health care professional education.³⁴⁻³⁷ However, Topping³⁸ argued that generalizations to professional practice should be made with caution because successful PA implementation depends on variables such as the context of peers, the nature of the PA intervention, and the outcomes assessed. Lack of specified knowledge about the PA practices impedes the transfer of results.³⁸

During the implementation of the Dutch guideline for physical thera-

pist management in patients with nonspecific low back pain,³⁹ PA showed promising results. In a randomized controlled trial conducted by van Dulmen and colleagues,⁴⁰ PA was significantly more effective in improving guideline adherence (measured using clinical vignettes) than the usual implementation strategy of case discussion (CD). We redesigned this PA program for the implementation of a newly developed guideline for complaints of the arm, neck, and shoulder⁴¹ and a new evidence statement for subacromial complaints.⁴² We also included the appraisal of patient records as a new element. Record keeping is an important quality indicator for physical therapy care, and patient records offer authentic assessment material that reflects clinical practice.^{43,44}

Peer assessment and CD are implementation strategies informed by several sometimes overlapping theoretical constructs concerning learning and behavior change: principles of social-constructivist learning theory,⁴⁵ such as contextual learning, collaborative learning, and active knowledge construction, and principles of self-regulated learning theory, such as conscious goal setting and reflection.^{29,46} In addition, the PA approach builds on principles of social-cognitive learning theory (concrete experience with and performance of desired behavior)⁴⁷ and stages of change theory (tailored feedback).^{29,30} Moreover, PA targets the development of a mutually accepted quality standard of performance by introducing peers to an “assessor” perspective.^{48,49}

The objective of this study was to compare the effectiveness of PA with the casual CD strategy on adherence to CPGs for physical therapist management of upper extremity conditions.

Following social-cognitive theory, our hypothesis was that the performance-based approach of PA, combined with giving and receiving personal performance feedback, would be a more powerful tool than the CD approach for uncovering areas in personal clinical practice that need improvement. Based on self-directed learning theory and stages of change theory, we also posited that PA would provide a stronger trigger for reflective practice, would develop greater awareness of guideline-consistent behavior in daily practice, and would be more effective in guiding self-directed change toward personal learning goals than CD. The effectiveness of PA and CD was tested on 4 outcome measures: (1) guideline adherence, (2) reflective practice, (3) awareness of performance, and (4) attainment of personal goals.

Method Design

This study was a single-masked, cluster-randomized controlled trial with a pretest-posttest design comparing the effectiveness of 2 implementation strategies.

Setting and Participants

Participants were physical therapists organized into CoPs, which are small groups of 5 to 15 professionals who share the same setting or the same interests and who work together on the improvement of the quality of care in postgraduate training programs provided yearly by the Royal Dutch Society for Physical Therapy (KNGF). Communities of practice can register with the KNGF to participate in such a program. The aim of the program under study was to implement 2 newly developed guidelines for physical therapist management in patients with upper extremity complaints. In November 2011, formal contact people of CoPs were invited by an electronic newsletter to a joint introduction meeting on

the training program. Communities of interest that showed interest in participating received an information letter containing details of the training program, randomization procedure, time investments, risks, and advantages. Participation was awarded with continuing education credits for the Dutch quality register. All CoPs that showed interest were eligible for inclusion. We conducted a sample size calculation based on an estimated difference between the 2 interventions of 5% (power=80%, $P=.05$), with an anticipated intraclass correlation coefficient (ICC) of .10 and 10% loss to follow-up. This calculation resulted in the required inclusion of 110 physical therapists in 22 clusters with at least 5 physical therapists per cluster.⁵⁰

Randomization

In December 2011, 22 CoPs showed interest in our study. Before randomization (January 2012), 2 CoPs withdrew because they felt the program would take too much time. A flowchart of the study sample is presented in the Figure. Because we expected that the size of the group would affect its learning,^{31,35,38} we aimed at a balanced distribution of large and small CoPs between CD and PA groups. The 20 CoPs were stratified by the number of participants into 2 blocks of groups with 5 to 10 and 11 to 15 participants and were randomly assigned to the intervention group or the control group using randomization software.⁵¹ This procedure resulted in 10 PA groups ($n=73$ physical therapists) and 10 CD groups ($n=76$ physical therapists). The CoPs were masked for the intervention because PA and CD were presented as alternative interventions. The primary researcher (M.M.) was not masked for the allocation of CoPs because she participated in conducting the intervention program. To reduce the risk of bias, she was masked for the outcomes until the data sampling was com-

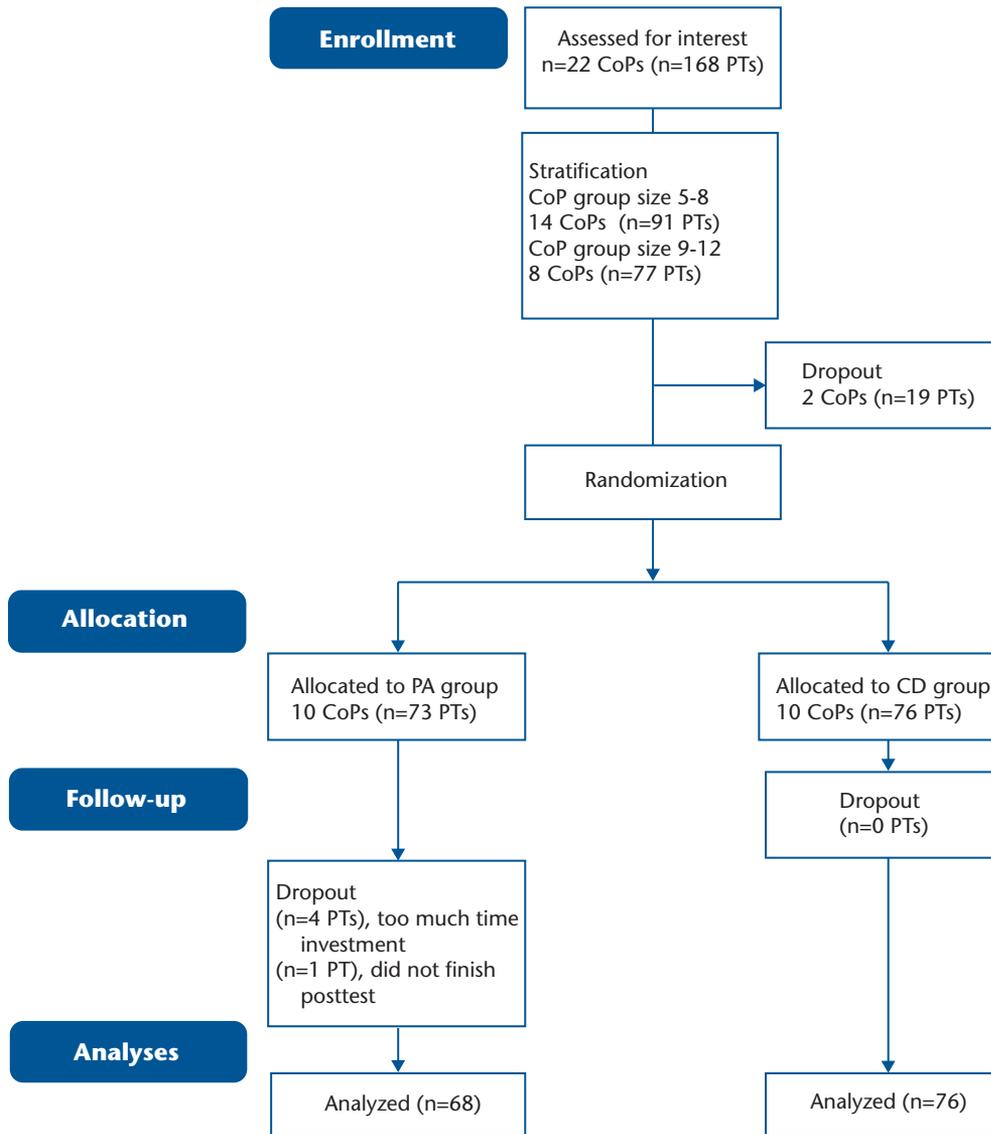


Figure.

Sample flowchart of participants in the study. CoP=community of practice, PT=physical therapist, PA=peer assessment, CD=case discussion.

pleted and the pretest-posttest differences were calculated.

Interventions

Before the start of the program, both PA and CD groups received a link to the KNGF guidelines and a link to the pretest questionnaires. All participants received by e-mail a program guide tailored to the intervention providing detailed information about learning objectives, learning content, training schedule, didactic for-

mat, and procedure. The program for both groups consisted of four 3-hour sessions and was launched in February 2012. Table 1 shows a detailed program overview and time schedule. In sessions 1, 2, and 4, the participants worked on written cases that fully covered the patient profiles described in the guidelines. Session 3 consisted of a review of patient records using a set of quality indicators derived from the KNGF guidelines on record keeping.⁵²

The main difference between the 2 interventions is that in the PA approach the tasks are structured, with a focus on performance rather than discussion, and roles are predefined. Each participant performed 3 roles: physical therapist, assessor, and simulated patient. Because the therapists were complete novices in the PA method, the process was supervised by a group coach. In the CD approach, tasks are less structured, with ample opportunity for

Table 1.

Intervention Programs in Both Groups

Measure	Period	Peer Assessment (PA)	Case Discussion (CD)
Pretest	February 2012	Online test based on 4 clinical vignettes (TCV)	
		Online questionnaire: Commitment to Change Statements (CTCS)	
	February 2012	Online questionnaire: Self-Reflection and Insight Scale (SRIS)	
	February 2012	Program manual by e-mail	
Session 1	February 2012	PA of individual performance	Case-based group discussion
Session 2	March 2012	PA of individual performance	Case-based group discussion
Session 3	April 2012	Review of personal patient records	Review of personal patient records
Session 4	June 2012	PA of individual performance	Case-based group discussion
	July 2012	Answering key to clinical cases	
Posttest	July 2012	TCV	
		CTCS	
		SRIS	
	September 2012	Personal knowledge of results by e-mail	

in-depth elaboration and discussion, and participant roles are not defined. In both PA and CD groups, participants worked on identical cases concerning problem content, but for the PA group, these cases were adjusted to allow for performance of participants in different roles. In the PA group, written cases were not known in advance but were presented by a coach on the spot, simulating daily practice. Participants were provided with ground rules for providing and receiving constructive feedback and for creating a safe learning environment. In the role of physical therapist, they analyzed the case by reasoning aloud and demonstrated (hands-on) diagnostic and treatment skills. Peer performance was assessed by using a global scoring sheet designed to support peer assessors in giving constructive feedback. It contained 3 performance categories (planning, performance, and evaluation) that were scored on a 5-point Likert scale (from 1=much improvement needed to 5=no improvement needed). Accordingly, qualitative oral improvement feedback was given. The complete PA program guide, including assessment criteria, is accessible online.⁵³

Three group coaches (H.E., H.N., and V.V.) were trained by the primary researcher in the PA procedure, supported by a coaching manual. They were experienced tutors in problem-based learning, and they were instructed to encourage the group in providing tailored performance feedback and not to serve as an information source themselves. To reduce the risk of bias, the group coaches were not involved in the development of clinical vignettes. For CD groups, written cases were included in the program guide to allow for proper preparation, along with instructions and written questions to guide the discussion process. After completion of the program in July 2012, and before the posttest, all participants received an e-mail with model answers to all of the cases that were discussed during the program to control for unintended differences in knowledge development between and within groups due to the influence of the group coach.

Outcome Measures

Guideline adherence. Participants completed an online test based on 4 clinical vignettes 1 week before

the start of the program and within 2 weeks after completion of the program. A previous study by Rutten et al⁵⁴ showed that vignettes have acceptable validity to measure physical therapists' adherence to CPGs, and these results were consistent with studies by Peabody and colleagues.⁵⁵⁻⁵⁷

Clinical vignettes require factual knowledge of CPGs as well as clinical reasoning consistent with CPGs in the context of a clinical problem. Four clinical vignettes were based on upper extremity disorders in the context of direct physical therapy access.⁵⁸ Three vignettes adequately covered the patient profiles described in the guidelines, and the fourth vignette did not because of "red flags." The vignettes and test items were constructed by a team containing 2 physical therapy scientists involved with guideline development, 5 physical therapy practitioners specializing in upper extremity conditions, and 1 physical therapy education scientist specializing in assessment development. Each vignette was accompanied by 11 response categories derived from the guidelines: (1) clinical pattern, (2) impairments and disabilities, (3) onset risk factors, (4) impeding recovery factors, (5) patient profile according to guidelines, (6) measurement instruments, (7) diagnostic clinical tests, (8) main treatment goals, (9) treatment approach, (10) information and advice, and (11) expected recovery time. Each response category contained a set of test items in the form of statements. Vignettes 1, 2, and 3 each contained 119 items; vignette 4 consisted of fewer items (n=31) because quality indicators addressing additional diagnosis and treatment were not applicable.

The statements were scored on a 3-point scale: D=disagree, D/A=neither disagree nor agree, and

A=agree. Because clinical evidence is limited and guidelines cannot inform all clinical decisions, the option D/A was offered to reflect the way information is processed in the context of uncertainty.⁵⁹ The group of 8 experts evaluated and adjusted the vignettes and test items. All experts completed the final test informed by the guidelines. The scoring method took variability of reasoning among experts into account as long as differences were limited to 2 alternatives (D and D/A, or D/A and A). Items with contradictory answers (D and A) were reviewed. The alternative that was chosen by the majority (>4) was assigned 2 points, and equal distribution was assigned 1 point for each alternative. A majority opting for the alternative D/A did not occur. The final scoring key was discussed among 4 experts until consensus was reached. The maximum score was 737 points (some answers received 1 point). The Appendix shows an example of a test item and its scoring key. The scores for each vignette were added, and mean total scores on the 4 clinical vignettes were perceived as a measure of guideline adherence.

Reflective practice. Participants completed the validated questionnaire, the Self-Reflection and Insight Scale (SRIS), developed by Grant et al.⁶⁰ It aims to measure the readiness for purposeful behavior change and has been shown responsive to change in the context of continuing professional education.⁶¹ The SRIS has been validated by Roberts and Stark⁶² and modified for the medical education context. It contains 3 subscales: the *engagement* with reflection, the *need* for reflection, and the *insights* obtained by reflection. Engagement and need refer to the practice of inspecting and evaluating one's own thoughts, feelings, and behavior; insight refers to understanding them. Sum scores for each

subscale were computed, and mean total scores were conceived of as a measure of reflective practice.

Awareness of performance. Awareness was conceived of as the association between perceived improvement and assessed improvement. At posttest, participants were asked to indicate how much guideline knowledge they had at pretest and how much at posttest on a scale from 1 (no knowledge) to 5 (much knowledge). The pretest-posttest difference was conceived of as a measure of perceived improvement. Assessed improvement was the difference between pretest and posttest scores on clinical vignettes.

Attainment of personal goals. At pretest, all participants were asked to formulate 3 learning goals, ordered on personal importance according to the concept of Commitment to Change Statements (CTCS).^{29,63} Conscious goal setting belonged to the intervention strategy to enhance self-directed learning and progression through the stages of change.³⁰ They also served as an outcome measure.⁶³ Before the posttest, all participants were e-mailed a reminder of their personal goals at pretest. At posttest, they were asked to indicate the extent to which their goals were achieved on a 3-point scale (1=not achieved, 2=partly achieved, and 3=achieved). Achievement scores for each personal goal were added, and mean total scores were conceived of as a measure of goal attainment.

Data Analysis

IBM SPSS, version 20 (IBM Corp, Armonk, New York) was used for statistical analysis. For baseline characteristics (age, sex, clinical setting, and specialization), pretest scores on clinical vignettes and SRIS of physical therapists were described and tested for differences between the PA and CD groups using chi-square tests and unpaired *t* tests. Internal

consistency of the clinical vignettes was tested by Cronbach alpha. Outcome differences between the PA and CD groups were described and tested by multilevel linear regression to adjust for clustering within CoPs. For each outcome measure, the ICC was calculated to test clustering at the CoP level. Baseline characteristics were included as covariates when differences between groups were statistically significant.

Pretest and posttest sum scores and mean total scores were calculated for each vignette. The intervention effect for guideline adherence was estimated with posttest scores on vignettes as the dependent variable and intervention and pretest scores as covariates. In the same way, mean pretest and posttest SRIS scores were calculated. The intervention effect for reflective practice was tested with posttest scores as the dependent variable and intervention and pretest scores as covariates. Mean posttest sum scores were calculated for each personal objective and total scores. Differences in attainment of personal goals were tested with scores on personal goals as the dependent variable and intervention as covariate. Mean assessed improvement scores and mean perceived improvement scores on clinical vignettes were calculated, and correlations were computed with assessed improvement as the dependent variable and perceived improvement as the independent variable. Differences in awareness were estimated with assessed improvement as the dependent variable and the interaction between the variables intervention and perceived improvement as covariate.

Role of Funding Source

This was a study initiated by researchers and funded by the Royal Dutch Society for Physical Therapy (KNGF). The KNGF had no role in the conduct of this study, analysis or

Table 2.
Physical Therapists' Characteristics^a

Variable	PA Group (n=73)	CD Group (n=76)
Age (y), \bar{X} (SD)	45.15 (11.03)	44.76 (9.74)
Working experience (y), \bar{X} (SD)	20.42 (11.37)	20.86 (9.71)
Sex, male/female ^b	39/34	27/49
Specialization, n (%)		
No specialization	47 (64.4)	46 (61.3)
Manual therapist	7 (9.6)	8 (10.7)
Other specialization	19 (26.0)	22 (28.0)
Clinical setting		
Primary care	56 (76.7)	58 (76.3)
Hospital or nursing home	17 (23.3)	18 (23.7)

^a PA=peer assessment, CD=case discussion.
^b Significant difference, $P<.05$.

interpretation of data, or preparation of the manuscript.

Results

The pretest response was 100%. The posttest response was 93,2% (n=68) for PA and 100% (n=76) for CD. Baseline characteristics of the participating physical therapists are presented in Table 2. We found differences between PA and CD groups for sex ($P=.028$), so we controlled for this confounder in multilevel lin-

ear regression. Internal consistency between scores across clinical vignettes (n=4) was good (pretest $\alpha=.82$, posttest $\alpha=.86$).

Table 3 presents the results of the outcome measures of guideline adherence, reflective practice, and attainment of personal goals. Results of awareness of performance are presented separately.

Concerning guideline adherence, Table 3 shows that mean pretest scores on vignettes were comparable between PA and CD groups. At posttest, the PA and CD groups showed significant improvement: PA groups=29.82 (SD=63.97), $P<.001$, and CD groups=9.49 (SD=40.52), $P<.001$. Percent improvement was 5.8% for the PA groups and 2.0% for the CD groups. Multilevel linear regression analysis, controlling for sex, showed that the difference between the PA and CD groups was statistically significant in favor of the PA groups (estimated effect=22.52 points; 95% CI=2.38, 42.66; $P=.031$).

Mean pretest scores on the SRIS showed no difference between the PA and CD groups. At posttest, scores were significantly improved in both PA and CD groups: PA groups=2.34 (SD=8.69), $P<.001$, and CD groups=1.85 (SD=7.05), $P<.001$. Percent improvement was 2.8% for the PA groups and 2.2% for the CD groups. The difference between the PA and CD groups was not statistically significant (estimated effect=-0.06 points; 95% CI=

Table 3.
Multilevel Analyses for Guideline Adherence, Reflective Practice, and Attaining Personal Goals^a

Measure	PA Group		CD Group		Estimated Difference	ICC	95% CI	P
	Pretest (n=73)	Posttest (n=68)	Pretest (n=76)	Posttest (n=76)				
Clinical vignettes ^b					22.52	.079	2.38, 42.66	.03 ^c
\bar{X}	474.26	501.99	472.54	482.03				
SD	65.85	65.07	56.09	62.19				
SRIS					-0.06	.048	-2.79, 2.65	.96
\bar{X}	82.71	85.10	83.47	85.33				
SD	9.80	9.32	8.08	8.90				
CTCS					0.50	.002	0.04, 0.96	.03 ^c
\bar{X}		7.44		6.90				
SD		1.44		1.29				

^a PA=peer assessment, CD=case discussion, ICC=intraclass correlation coefficient, 95% CI=95% confidence interval, SRIS=Self-Reflection and Insight Scale (reflective practice), CTCS=Commitment to Change Statements (attainment of personal goals).
^b Clinical vignettes measured guideline adherence.
^c Significant difference, $P<.05$.

-2.79, 2.65; $P=.96$). The results related to attainment of personal goals showed that scores were significantly higher for the PA groups than for the CD groups (estimated effect=0.50; 95% CI=0.04, 0.96; $P=.03$).

At posttest, participants in the PA groups showed greater awareness of their professional performance. The correlation between perceived improvement and assessed improvement was $r=.36$ ($P=.002$) for the PA groups and $r=.08$ ($P=.50$) for the CD groups. The difference was statistically significant (estimated effect=14.73; 95% CI=2.78, 26.68; $P=.01$).

Discussion

This study evaluated the effects of 2 strategies for the implementation of Dutch physical therapy guidelines. It showed that PA was more effective than CD in improving guideline adherence as measured by clinical vignettes. Moreover, the PA groups were more effective in attaining personal goals and showed higher levels of awareness of performance. The strength of this study is that we offered the PA and CD groups high-quality programs. Program evaluation showed that the perceived instructional value of PA and CD was comparable between PA and CD groups (results not presented). The outcome measures were equally facilitated by both interventions. First, the PA and CD groups had equal access to the guidelines, worked on solving identical clinical problems, and had equal access to the model answers of each problem. Second, neither of the interventions included tasks such as writing reflection reports and improvement plans that explicitly aimed to facilitate the outcomes of reflective practice, awareness of performance, and attainment of personal goals. Any pretest effect of the SRIS or the

CTCS would have applied to both interventions.

We showed that a tailored, multifaceted intervention that addresses specific barriers to change,¹⁰ such as “awareness of performance” as identified by Rutten et al,¹² is effective, and these findings are in line with the existing research evidence on implementation strategies.^{4,10,19,64} We observed high baseline scores and moderate, but statistically significant, improvement scores for continuous outcomes of clinical vignettes (PA groups=5.8%, CD groups=2.0%). High baseline scores can be attributed to the fact that participants received the guidelines before the pretest and were allowed to study them beforehand. Studies have shown that the intervention effect on desired practice increases when baseline performance is low.^{19,65}

Rutten et al⁸ observed a 3.1% guideline adherence increase for the low back pain guideline using clinical vignettes that assessed the effectiveness of their program. This program, however, involved interventions addressing professional as well as organizational determinants of guideline adherence, so the results cannot be compared. We did not find studies that assessed comparable content and constructs concerning the improvement of the uptake of CPGs except for the study by van Dulmen et al,⁴⁰ which showed that PA was more effective than CD in the implementation of the low back pain guideline, and that result is in line with our findings.

Given the notion that intervention programs aimed at enhancing the transfer of research evidence to clinical practice are very heterogeneous and the generalizability of the effects is limited,^{18,66} we explored the key differences between PA and CD informed by theory, which may con-

tribute to the generalizability of the results. First, the PA task is highly structured and necessitates strong involvement of each participant. Individual contributions in learning groups may vary widely when conditions such as shared responsibility, interdependency, mutual trust, and psychological safety are not met.^{32,67} Discussion tasks do aim at active participation, but the task structure does not control for individual contributions to group learning.

Second, in contrast to CD, PA focuses on performance that can be observed and evaluated. The PA group participants performed in predefined roles that forced the transfer of knowledge and skills in order to fulfill this role convincingly. In the role of physical therapist, participants needed to make the transfer from implicit reasoning to explicit reasoning and from intentional behavior to observable behavior. The transferred knowledge and skills became transparent, and this new information became accessible for group review.⁶⁸ The variety of feedback that PA group participants obtained about their performance may have helped them to become aware of areas in professional practice that need improvement and may have supported them in attaining personal goals. In the assessor role, participants needed to make a transfer from implicit appraisal to explicit appraisal. Supported by predefined performance criteria, peer assessors revealed their personal norms about the quality of the observed behavior.⁶⁹ Personal standards could be compared with group standards. Research has revealed that the availability of both internal and external data about an individual's performance is conditional on the development of correct self-perceptions (awareness),^{49,50} which may explain why PA groups outperformed CD groups in this respect. A different perspective on why PA groups

showed more improvement on guideline adherence is the testing effect. Recent insights in cognitive psychology show that tested information is better stored and retrieved from memory than information that is not.^{70,71} Because PA is based on assessment (unlike CD), PA group participants were repeatedly challenged to reproduce and apply newly acquired knowledge of CPGs. That may have strengthened awareness of deficiencies and facilitated retrieval of information from memory at posttest.

Although PA was more effective in 3 outcome measures, we could not explain these results by differences in reflective practice. Both the PA and CD groups showed comparable improvement scores on the SRIS. These scores reflect perceptions of conscious reflective practice,^{60,62} and conscious reflective practice was apparently enhanced by both interventions. Professional behavioral change, however, does not necessarily depend on conscious reflection but also might occur spontaneously through informal learning, such as concrete experience, role modeling,⁷² and action observation.⁷³ Peer assessment involved concrete experience with guideline recommendations, including hands-on clinical skills. This approach might have prompted spontaneous (unintended) learning experiences more than the cognitive directed approach of CD. A study by Bandura and Locke⁷⁴ showed that experience is the strongest source of information for the development of self-efficacy beliefs and that self-efficacy beliefs contribute significantly to motivation for behavioral change.

A third difference between the PA and CD groups was the presence of the group coach. Peer groups contained experienced health care practitioners, but they were absolute

novices in the peer assessment method. Research has revealed that the acceptability of peer feedback highly depends on its perceived reliability^{32,68} and that reliability and validity of peer feedback improve by training and experience.³¹ It is possible that therapists have used the group coach as a tool to justify feedback because they did not fully rely on their peers' judgment. We assume that the effect of PA may increase when groups have more training in giving and receiving peer feedback and when standards for the quality of physical therapy care are internalized and mutually shared.^{48,49} We also assume that successful PA practices depend on commitment of physical therapists to the PA procedure. The role of the group coach might be important in this respect. On the other hand, it should be noted that the CD groups might have performed better when guided by a coach.

Finally, it should be noted that research has shown that improved guideline adherence is associated with improved process of care but not always with improved patient outcomes.^{5,11,75}

Limitations

First, clinical vignettes remain a proxy measure of clinical practice. Direct observation or audio or video recording might be measures that better reflect authentic practice, but a systematic review by Hrisos et al⁷⁶ suggests that such measures may lack reliability and validity as well because the behavior of interest cannot be standardized beforehand, and generalizations of the inferences are hard to make. Standardized (simulated) patients are generally considered to be an acceptable substitute, but these measures are costly and were not feasible given the sample size. Moreover, standardized patients do not provide a sufficiently broad case mix compared with clinical

vignettes. Based on these considerations and the existing validity evidence,⁵⁵⁻⁵⁷ we opted for clinical vignettes.

A second limitation is the involvement of the primary researcher (M.M.) in conducting the intervention program. To reduce risk of bias, the primary researcher was masked for the outcomes until pretest-posttest scores had been described and between-group differences were calculated. The primary researcher was involved in additional multilevel analyses, supervised by another researcher (J.K.) who was masked for the intervention.

Third, the involvement of the group coaches should be addressed. We controlled for differences in knowledge development between and within groups by e-mailing to each participant, before the posttest, the model answers for all the clinical cases. Outcomes on all outcome measures did not show a significant difference between group coaches (M.M., H.E., H.N., or V.M.) (results not presented). However, we could not control for implicit effects of the group coaches on motivation to change, such as role modeling effects, increased self-efficacy beliefs, improved attitudes toward guidelines,^{24,32} and shared quality standards of performance.⁴⁹

Fourth, the reliability of the test scores should be considered. The test contained a considerable number of test items (N=388). Although each participant fully completed the test within time limits (2 hours) at pretest and posttest, cognitive overload caused by time on task may have biased test results. The effect, however, applied to both the PA and CD groups, so it does not affect the validity of the inferences made about between-group differences.

Finally, we address the generalizability of our results. Studies have demonstrated cultural differences in attitudes toward PA, such as reluctance of peers in giving face-to-face feedback.^{28,32} External validity might be limited because the sample contained only Dutch physical therapists.

In conclusion, PA is more effective in guideline implementation than CD. The PA group participants showed higher improvement scores on clinical vignettes, showed more awareness of guideline-consistent behavior, and were more successful in attaining personal goals. The focus on individual performance, allowing for concrete experience with the guideline, and obtaining personal performance feedback probably contributed to its effectiveness. Moreover, performance in the assessor role necessitates critical appraisal of the observed behavior as well as critical self-appraisal.

We recommend PA for guideline implementation within CoPs. Further research should address the role of the group coach on the intervention effect and should explore the feasibility of replacing the group coaches by trained CoP members. They could play an important role in future bottom-up quality improvement initiatives addressing evidence-based practice and unwarranted variability in physical therapy care.

The authors thank all participating physical therapists for their commitment to this research project. The authors thank Femke Atsma for verification of data analyses; Henk van Enck (H.E.), Henk Nieuwenhuijzen (H.N.), and Volcmar Visser (V.V.) for their contribution as group coaches; and all physical therapists and experts who contributed to the development of clinical vignettes. They also thank Moira Jackson and John Gabbay for reviewing for English language usage.

Ethical approval for the study was given by the Medical Ethical Committee on Research Involving Human Subjects, Arnhem-

Nijmegen, the Netherlands (CMO registration number: 2013/036).

This study was funded by the Royal Dutch Society for Physical Therapy (KNGF) (registration number: 8203).

This trial is registered at ISRCTN (registration number: ISRCTN69003553).

DOI: 10.2522/ptj.20130469

References

- Field MJ, Lohr KN, eds. *Guidelines for Clinical Practice: From Development to Use*. Washington, DC: National Academies Press; 1992.
- Burgers J, Smolders M, Wollersheim H, Grol R. Richtlijnen als hulpmiddel bij de verbetering van de zorg [Guidelines as a tool to improve patient care]. In: *Implementatie: Effectieve verbetering van de patiëntenzorg [Implementation: Effective Improvement of Patient Care]*. 4th ed. Amsterdam, the Netherlands: Reed Business; 2011:155-189.
- van der Wees PJ, Jamtvedt G, Rebeck T, et al. Multifaceted strategies may increase implementation of physiotherapy clinical guidelines: a systematic review. *Aust J Physiother*. 2008;54:233-241.
- Grimshaw J, McAuley LM, Bero LA, et al. Systematic reviews of the effectiveness of quality improvement strategies and programmes. *Qual Saf Health Care*. 2003;12:298-303.
- Grol R, Wensing M, Bosch M, et al. Theorieën over implementatie. In: *Implementatie: Effectieve verbetering van de patiëntenzorg [Theories about implementation. In: Implementation: Effective Improvement of Patient Care]*. 4th ed. Amsterdam, the Netherlands: Reed Business; 2011:43-68.
- Wensing M, Grol R, Fluit C. Educatieve strategieën. In: *Implementatie: Effectieve verbetering van de patiëntenzorg [Educational strategies. In: Implementation: Effective Improvement of Patient Care]*. 4th ed. Amsterdam, the Netherlands: Reed Business; 2011:326-340.
- Forsetlund L, Bjørndal A, Rashidian A, et al. Continuing education meetings and workshops: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev*. 2009;2:CD003030.
- Rutten GM, Harting J, Bartholomew LK, et al. Evaluation of the theory-based quality improvement in physical therapy (QUIP) programme: a one-group, pre-test post-test pilot study. *BMC Health Serv Res*. 2013;13:194.
- Li LC, Grimshaw JM, Nielsen C, et al. Use of communities of practice in business and health care sectors: a systematic review. *Implement Sci*. 2009;4:27.
- Cheater F, Baker R, Gillies C, et al. Tailored interventions to overcome identified barriers to change: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev*. 2009;4:CD005470.
- Bekkering GE, Hendriks HJ, van Tulder MW, et al. Effect on the process of care of an active strategy to implement clinical guidelines on physiotherapy for low back pain: a cluster randomised controlled trial. *Qual Saf Health Care*. 2005;14:107-112.
- Rutten G, Kremers S, Rutten S, Harting J. A theory-based cross-sectional survey demonstrated the important role of awareness in guideline implementation. *J Clin Epidemiol*. 2009;62:167-176.
- Adams AS, Soumerai SB, Lomas J, Ross-Degnan D. Evidence of self-report bias in assessing adherence to guidelines. *Int J Qual Health Care*. 1999;11:187-192.
- Davis DA, Mazmanian PE, Fordis M, et al. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA*. 2006;296:1094-1102.
- Eva KW, Regehr G. "I'll never play professional football" and other fallacies of self-assessment. *J Contin Educ Health Prof*. 2008;28:14-19.
- Regehr G, Eva KW. Self-assessment, self-direction, and the self-regulating professional. *Clin Orthop Relat Res*. 2006;449:34-38.
- Sargeant J, Eva KW, Armson H, et al. Features of assessment learners use to make informed self-assessments of clinical performance. *Med Educ*. 2011;45:636-647.
- Brehaut JC, Eva KW. Building theories of knowledge translation interventions: use the entire menu of constructs. *Implement Sci*. 2012;7:114.
- Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev*. 2012;6:CD000259.
- Eva KW, Armson H, Holmboe E, et al. Factors influencing responsiveness to feedback: on the interplay between fear, confidence, and reasoning processes. *Adv Health Sci Educ Theory Pract*. 2012;17:15-26.
- Mann K, van der Vleuten CP, Eva KW, et al. Tensions in informed self-assessment: how the desire for feedback and reticence to collect and use it can conflict. *Acad Med*. 2011;86:1120-1127.
- Bandura A. Self-efficacy: toward a unifying theory of behavioral change. *Psychol Rev*. 1977;84:191-215.
- Ajzen I. Nature and operation of attitudes. *Annu Rev Psychol*. 2001;52:27-58.
- Prochaska J, DiClemente CC, Norcross JC. In search of how people change: applications to addictive behaviors. *Am Psychol*. 1992;47:1102-1114.
- Burke L, Hutchins H. Training transfer: an integrative literature review. *Hum Res Dev Rev*. 2007;6:263-296.
- Lave J, Wenger E. *Situated Learning: Legitimate Peripheral Participation*. Cambridge, United Kingdom: Cambridge University Press; 1991.
- Li LC, Grimshaw JM, Nielsen C, et al. Evolution of Wenger's concept of community of practice. *Implement Sci*. 2009;4:11.

- 28 Le May A. Introducing communities of practice. In: Le May A, ed. *Communities of Practice in Health and Social Care*. Oxford, United Kingdom: Wiley-Blackwell; 2008:3-16.
- 29 Mazmanian PE, Mazmanian PM. Commitment to change: theoretical foundations, methods, and outcomes. *J Contin Educ Health Prof*. 1999;19:200-207.
- 30 Prochaska JO, Redding CA, Evers KE. Health behavior and health education. In: Glanz K, Rimer B, Viswanath K, eds. *Health Behavior and Health Education: Theory, Research, and Practice*. 4th ed. New York, NY: John Wiley & Sons Inc; 2008:97-121.
- 31 Van Zundert M, Sluijsmans D, van Merriënboer J. Effective peer assessment processes: research findings and future directions. *Learn Instr*. 2010;20:270-279.
- 32 Van Gennip NA, Seger MS, Tillema HH. Peer assessment as a collaborative learning activity: the role of interpersonal variables and conceptions. *Learn Instr*. 2010;20:280-290.
- 33 Strijbos JW, Sluijsmans D. Unravelling peer assessment: methodological, functional, and conceptual developments. *Learn Instr*. 2010;20:265-269.
- 34 Arnold L, Shue CK, Kalishman S, et al. Can there be a single system for peer assessment of professionalism among medical students? A multi-institutional study. *Acad Med*. 2007;82:578-586.
- 35 Dannefer EF, Henson LC, Bierer SB, et al. Peer assessment of professional competence. *Med Educ*. 2005;39:713-722.
- 36 Norcini JJ. Peer assessment of competence. *Med Educ*. 2003;37:539-543.
- 37 Wenghofer EF, Way D, Moxam RS, et al. Effectiveness of an enhanced peer assessment program: introducing education into regulatory assessment. *J Contin Educ Health Prof*. 2006;26:199-208.
- 38 Topping KJ. Methodological quandaries in studying process and outcomes in peer assessment. *Learn Instr*. 2010;20:339-343.
- 39 Staal BJ, Hendriks EJ, Heijmans M, et al. KNGF Richtlijn Lage-Rugpijn voor fysiotherapie en manuele therapie [Guideline low back pain for physical therapy and manual therapy]. *R Dutch Soc Phys Ther*. 2010. Available at: <http://www.fysionet-evidencebased.nl/index.php/component/kngf/richtlijnen>. Accessed September 10, 2012.
- 40 van Dulmen SA, Maas M, Staal JB, et al. Effectiveness of peer-assessment for implementing a Dutch physical therapy low back pain guideline: cluster randomized controlled trial. *Phys Ther*. 2014;94:1396-1409.
- 41 Heemskerk M, Staal J, Bierma-Zeinstra S, et al. KNGF-richtlijn Klachten aan de arm, nek en/of schouder (KANS) [KNGF-guideline complaints of arm, neck and/or shoulder (CANS)]. 2010;(1). Available at: <http://www.fysionet-evidencebased.nl/index.php/component/kngf/richtlijnen>. Accessed December 16, 2011.
- 42 Jansen M, Brooijmans F, Geraets J, et al. KNGF Evidence Statement Subacromiale klachten [Evidence statement subacromial complaints]. 2011:1-14. Available at: <http://www.fysionet-evidencebased.nl/index.php/component/kngf/richtlijnen>. Accessed December 16, 2011.
- 43 Oostendorp RA, Plumiers DJ, Nijhuis-van der Sanden MW. Fysiotherapeutische verslaglegging: de Achilleshiel voor Evidence-based Practice (EBP)? [Record keeping in physical therapy: the Achilles heel for evidence based practice (EBP)?]. *Ned Tijdschr voor Fysiother*. 2006;116:56.
- 44 Van Dulmen SA, Calsbeek H, Cruisberg J, et al. Kwaliteitsindicatoren Eerstelijns fysiotherapie [Quality indicators physical therapy primary care]. Available at: [http://www.iqhealthcare.nl/nl/kennisbank/rapporten/k/kwaliteitsindicatoren-eerstelijns-fysiotherapie-\(kwaliefy\)/](http://www.iqhealthcare.nl/nl/kennisbank/rapporten/k/kwaliteitsindicatoren-eerstelijns-fysiotherapie-(kwaliefy)/). Published April 2012. Accessed September 10, 2013.
- 45 Norman GR, Schmidt HG. The psychological basis of problem-based learning: a review of the evidence. *Acad Med*. 1992;67:557-365.
- 46 Greene J, Azevedo R. A theoretical review of Winne and Hadwin's model of self-regulated learning: new perspectives and directions. *Rev Educ Res*. 2007;77:334-372.
- 47 Bandura A. *Self-Efficacy: The Exercise of Control*. New York, NY: WH Freeman; 1997:604.
- 48 Epstein RM, Siegel DJ, Silberman J. Self-monitoring in clinical practice: a challenge for medical educators. *J Contin Educ Health Prof*. 2008;28:5-13.
- 49 Pronovost PJ, Hudson DW. Improving healthcare quality through organisational peer-to-peer assessment: lessons from the nuclear power industry. *BMJ Qual Saf*. 2012;21:872-875.
- 50 Maas CJ, Hox JJ. Sufficient sample sizes for multilevel modeling. *Methodology*. 2005;1:86-92.
- 51 Randomization.com, 2008. Available at: <http://www.randomization.com>. Accessed February 2, 2012.
- 52 Heerkens YF, Hendriks H, De Graaf-Peters VB. KNGF-richtlijn Fysiotherapeutische verslaglegging [KNGF-guideline record keeping in physical therapy]. 2011. Available at: <https://www.fysionet-evidencebased.nl/index.php/richtlijnen/richtlijnen/fysiotherapeutische-verslaglegging-2011>. Accessed December 16, 2012.
- 53 HAN. Projecten bij lectoraat Arbeid and Gezondheid. Available at: <http://www.han.nl/onderzoek/kennismaken/revalidatie-arbeid-sport/lectorat/arbeid-en-gezondheid/projecten/>. Accessed July 12, 2014.
- 54 Rutten GM, Harting J, Rutten ST, et al. Measuring physiotherapists' guideline adherence by means of clinical vignettes: a validation study. *J Eval Clin Pract*. 2006;12:491-500.
- 55 Peabody JW, Luck J, Glassman P, et al. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *JAMA*. 2000;283:1715-1722.
- 56 Peabody JW, Dresselhaus TR, Luck J, Bertenthal D. Improving patient care measuring the quality of physician practice by using clinical vignettes: a prospective validation study. *Ann Intern Med*. 2004;141:813-814.
- 57 Dresselhaus TR, Peabody JW, Luck J, Bertenthal D. An evaluation of vignettes for predicting variation in the quality of preventive care. *J Gen Intern Med*. 2004;19:1013-1018.
- 58 Childs JD, Whitman JM, Sizer PS, et al. A description of physical therapists' knowledge in managing musculoskeletal conditions. *BMC Musculoskelet Disord*. 2005;6:32.
- 59 Charlin B, van der Vleuten C. Standardized assessment of reasoning in contexts of uncertainty: the script concordance approach. *Eval Health Prof*. 2004;27:304-319.
- 60 Grant AM, Franklin J, Langford P. The self-reflection and insight scale: a new measure of private self-consciousness. *Soc Behav Pers*. 2002;30:821-836.
- 61 Grant AM. Personal life coaching for coaches-in-training enhances goal attainment, insight and learning. *Coach Int J Theory Res Pract*. 2008;1:54-70.
- 62 Roberts C, Stark P. Readiness for self-directed change in professional behaviours: factorial validation of the Self-Reflection and Insight Scale. *Med Educ*. 2008;42:1054-1063.
- 63 Wakefield J, Herbert CP, Maclure M, et al. Commitment to change statements can predict actual change in practice. *J Contin Educ Health Prof*. 2003;23:81-93.
- 64 Grol R, Bosch M, Wensing M. Ontwikkeling of selectie van strategieën voor verandering [Development or selection of strategies for change]. In: *Implementatie: Effectieve verbetering van de patiëntenzorg [Implementation: Effective Improvement of Patient Care]*. 4th ed. Amsterdam, the Netherlands: Reed Business; 2011:281-323.
- 65 Jamtvedt G, Young J, Kristoffersen D, et al. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev*. 2007;4:CD000259.
- 66 Scott SD, Albrecht L, O'Leary K, et al. Systematic review of knowledge translation strategies in the allied health professions. *Implement Sci*. 2012;7:70.
- 67 Van Gennip NA, Segers MS, Tillema HH. Peer assessment for learning from a social perspective: the influence of interpersonal variables and structural features. *Educ Res Rev*. 2009;4:41-54.
- 68 Sargeant JM, Mann KV, van der Vleuten CP, Metsemakers JF. Reflection: a link between receiving and using assessment feedback. *Adv Health Sci Educ Theory Pract*. 2009;14:399-410.

- 69 Ramsey PG, Wenrich MD, Carline JD, et al. Use of peer ratings to evaluate physician performance. *JAMA*. 1993;269:1655-1660.
- 70 Roediger HL, Karpicke JD. Test-enhanced learning: taking memory tests improves long-term retention. *Psychol Sci*. 2006;17:249-255.
- 71 Kromann CB, Jensen ML, Ringsted C. The effect of testing on skills training. *Med Educ*. 2009;43:21-27.
- 72 Simons P, Ruijters M. Learning professionals: towards an integrated model. In: Boschhuizen HP, Bromme R, Gruber H, eds. *Professional Learning: Gaps and Transitions on the Way From Novice to Expert*. Vol 2. Amsterdam, the Netherlands: Springer; 2004:207-229.
- 73 Iacoboni M. *Mirroring People: The New Science of How We Connect With Others*. New York, NY: Macmillan; 2009.
- 74 Bandura A, Locke EA. Negative self-efficacy and goal effects revisited. *J Appl Psychol*. 2003;88:87-99.
- 75 Bekkering GE, van Tulder MW, Hendriks EJ, et al. Implementation of clinical guidelines on physical therapy for patients with low back pain—randomized trial comparing patient outcomes after a standard and active implementation strategy. *Phys Ther*. 2005;85:544-555.
- 76 Hrisos S, Eccles MP, Francis JJ, et al. Are there valid proxy measures of clinical behaviour: a systematic review. *Implementation Sci*. 2009;4:37.

Appendix.

Example of a Clinical Vignette With Exemplary Test Items

Vignette 3

Part 1: case history

Personal information: female, 38 years of age, married, 2 daughters (ages 21 and 14 years)

Work: nurse 3 days a week

Hobbies: reading, playing cards

Sports: nothing specific except for taking long walks

“Marianne” visits the physical therapist without referral of a physician. She reports having had intense right shoulder pain for a week. She reports existing neck complaints and having had vague shoulder pain for a few months. Shoulder complaints were conspicuous in activities such as changing a drip or moving patients in their bed. She also was hindered in driving her car because of limited rotation of the cervical spine. These problems, however, did not have a serious impact on her daily life. Complaints became seriously aggravated when she went swimming with her 2 nieces, aged 5 and 7 years. The girls repeatedly climbed on her shoulders to subsequently dive into the water. Shoulder pain gradually developed the same evening and increased the following days. Because of the pain, she is currently unable to lift her right arm without the support of her left arm, which hinders her in activities such as dressing, undressing, and other activities involving raising her right arm. She locates the pain at the anterolateral aspect of her right shoulder and upper arm. Currently, she is unable to perform her nursing tasks, which she perceives as very annoying because there is much failure of nursing staff lately. At night, Marianne has difficulty sleeping. She easily wakes up when she rests on her painful shoulder. She uses paracetamol and ibuprofen as medications, and that relieves the pain but not enough to perform her daily tasks.

Marianne still smokes after repeated attempts to quit, and her youngest daughter uses every opportunity to comment on her smoking habit. Besides a hypofunction of the thyroid, for which she receives medication, Marianne has no other health problems. Apart from the last few months, she never had shoulder problems before. Marianne is right handed.

Example of a response category and related test items						
Which risk factors may have played a role in the onset of the symptoms in this health problem according to the guideline? D=disagree, D/A=neither disagree nor agree, A=agree	Distribution of expert scores (n=8)			Rewarded points		
	D	D/A	A	D	D/A	A
Age	0	0	8	0	0	2
Sex	0	4	4	0	1	1
Medication use	7	1	0	2	1	0
Neck complaints	0	0	8	0	0	2
Postures related to work	0	4	4	0	1	1
Movements related to work	0	0	8	0	0	2
Amount of working hours	8	0	0	2	0	0
Arm dominance	0	2	6	0	1	2
Coping with stress	0	2	6	0	1	2
Smoking	6	2	0	2	1	0
Hormonal changes	8	0	0	2	0	0