

BMJ Open Feasibility of peer assessment and clinical audit to self-regulate the quality of physiotherapy services: a mixed methods study

Marjo J M Maas,^{1,2} Maria W G Nijhuis-van der Sanden,² Femke Driehuis,² Yvonne F Heerkens,^{1,3} Cees P M van der Vleuten,⁴ Philip J van der Wees²

To cite: Maas MJM, Nijhuis-van der Sanden MWG, Driehuis F, *et al*. Feasibility of peer assessment and clinical audit to self-regulate the quality of physiotherapy services: a mixed methods study. *BMJ Open* 2017;**7**: e013726. doi:10.1136/bmjopen-2016-013726

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2016-013726>).

Received 2 August 2016
Accepted 12 January 2017



CrossMark

For numbered affiliations see end of article.

Correspondence to
Marjo J M Maas;
marjo.maas@han.nl

ABSTRACT

Objectives: To evaluate the feasibility of a quality improvement programme aimed to enhance the client-centredness, effectiveness and transparency of physiotherapy services by addressing three feasibility domains: (1) acceptability of the programme design, (2) appropriateness of the implementation strategy and (3) impact on quality improvement.

Design: Mixed methods study.

Participants and setting: 64 physiotherapists working in primary care, organised in a network of communities of practice in the Netherlands.

Methods: The programme contained: (1) two cycles of online self-assessment and peer assessment (PA) of clinical performance using client records and video-recordings of client communication followed by face-to-face group discussions, and (2) clinical audit assessing organisational performance. Assessment was based on predefined performance indicators which could be scored on a 5-point Likert scale. Discussions addressed performance standards and scoring differences. All feasibility domains were evaluated qualitatively with two focus groups and 10 in-depth interviews. In addition, we evaluated the impact on quality improvement quantitatively by comparing self-assessment and PA scores in cycles 1 and 2.

Results: We identified critical success features relevant to programme development and implementation, such as clarifying expectations at baseline, training in PA skills, prolonged engagement with video-assessment and competent group coaches. Self-reported impact on quality improvement included awareness of clinical and organisational performance, improved evidence-based practice and client-centredness and increased motivation to self-direct quality improvement. Differences between self-scores and peer scores on performance indicators were not significant. Between cycles 1 and 2, scores for record keeping showed significant improvement, however not for client communication.

Conclusions: This study demonstrated that bottom-up initiatives to improve healthcare quality can be effective. The results justify ongoing evaluation to inform nationwide implementation when the critical success features are addressed. Further research is

Strengths and limitations of this study

- This study evaluated the feasibility of a quality improvement programme based on what physiotherapists do in their day-to-day practice by assessing client records, client communication and management information.
- The quality improvement programme and implementation was theory-based and evidence-based; stakeholders and end-users were actively involved.
- The results provided meaningful information on the critical success features of the programme, relevant to more rigorous evaluation and nationwide implementation.
- In the beginning, some participants were unwilling or unable to expose their professional behaviours for online assessment; therefore, data on the improvements made are limited.

necessary to explore the sustainability of the results and the impact on client outcomes in a full-scale study.

INTRODUCTION

Healthcare professionals and provider organisations have an ethical and professional obligation to strive for continuous quality improvement of services. When healthcare professionals are able to self-regulate and account for the quality of their services, they perceive control of the quality improvement strategies and the outcome measures used, in contrast to external regulations. Professionals often resist external audits; they fear a deterioration of their professional identity and an increase of administrative burden. Moreover, external regulations can potentially be effective, but the evidence is not convincing regarding the sustainability of the results and the strategy might induce

unwanted consequences such as undertreatment of clients with multimorbidity and disparities in healthcare delivery.^{1–3}

Research has shown that bottom-up quality improvement initiatives, such as communities of practice and professional networks focusing on collaborative learning, might hold better and more sustainable results than external, top-down regulations^{4–6} because shared social and professional norms are important predictors for behaviour change.^{7–8} Conditional to successful self-regulation is that professionals share the quality standards of their services and demonstrate the willingness and ability to critically appraise their own and their colleagues' performance.^{4–6–9} Literature showed that quality improvement programmes targeting self-regulation should not be limited to individual healthcare professionals, but also involve teams and provider organisations to align the desired processes and outcomes.¹⁰ Clinical governance has been introduced as a multilevel approach to quality improvement, bridging the gap between managerial and clinical approaches. The approach allows for early spotting of poorly performing clinicians, teams and organisations to support self-regulated quality improvement.^{4–11–12} Following this approach, we developed and tested the feasibility of a programme combining self-assessment, peer assessment (PA) and clinical audit as a strategy to improve the quality of physiotherapy services.

Self-assessment is the process whereby professionals reflect on their clinical and organisational performance according to quality indicators.¹³ In PA—also known as peer review—professionals evaluate or are being evaluated by their peers and provide each other with performance feedback.¹⁴ The aim of PA is to guide the self-directed quality improvement process towards desired mutual accepted performance standards.¹⁵ A cornerstone of the PA strategy is to raise awareness of clinical performance informing self-assessment¹⁶ and to develop a critical attitude towards the process and outcomes of healthcare by introducing professionals with an 'assessor' or 'auditor' perspective.¹⁷ The results of a systematic review by Fox *et al*¹⁸ on PA practices in healthcare demonstrated that peer review was associated with measurable performance improvement of healthcare professionals on several outcome levels and in a variety of competency domains. Clinical audit is a common strategy for quality improvement at the level of provider organisations. Aspects of the structure, processes and outcomes of care are selected and systematically evaluated against explicit criteria by trained colleagues. The method has proved its effectiveness in primary care.^{15–19} Although PA and clinical audit are well-studied strategies,^{20–22} our programme design is innovative because it focuses on assessment of authentic clinical behaviours and integrates clinical performance and organisational performance assessment.

We used the framework of the Medical Research Council 2015 to develop, test and implement the

programme.²³ The framework recommends a feasibility and piloting phase to allow for optimising the programme design and implementation prior to evaluating effectiveness in a larger study. Online supplementary appendix 1 shows the details of the design process including the development of performance indicators. This study addresses the evaluation of three feasibility domains: (1) acceptability of the programme for quality improvement purposes including strengths and weaknesses, (2) appropriateness of the implementation strategy to execute the programme as intended including barriers and facilitators and (3) impact of the programme on quality improvement and professional behaviour change.^{23–24}

AIM

The aim of this study was to evaluate the feasibility of a quality improvement programme aimed to enhance the effectiveness, client-centeredness and transparency of physiotherapy services to allow for optimising the programme design and implementation prior to more rigorous evaluation and nationwide implementation.

METHODS

Design

Programme feasibility was evaluated with mixed methods using qualitative and quantitative data.²⁴

Subjects and setting

We tested our programme with physiotherapists working in primary care clinics, organised in a regional network of communities of practice in the Netherlands. In communities of practice, professionals share the same interests, setting or specialisation. Formally registered physiotherapist networks were invited by the Royal Dutch Society for Physical Therapy (KNGF) via a digital newsletter. Participation was voluntary and was awarded with 30 accreditation points for the quality register. To facilitate the implementation process, we used two knowledge brokers (FT and HK) as the linking pin between researchers and participants,^{25–27} trained PA coaches (n=5) to support the PA process and trained auditors (n=3). The knowledge brokers were leaders of the professional network and took part of the stakeholder group. Coaches and auditors were members of the professional network recruited by the knowledge brokers.

Programme content

Assessment addressed three performance domains related to the three quality domains: client-centeredness, effectiveness (including evidence-based practice) and transparency of physiotherapy services: (1) record keeping, (2) client communication and (3) organisation and management. The programme contained two cycles of PA with an interval of 4–6 weeks, followed by one cycle of clinical audit. Assessment was based on

predefined performance indicators which could be scored on a 5-point Likert scale. Online supplementary appendix 1a–c presents the performance indicators and their relationships with the three quality domains. Online supplementary appendix 2 shows how all programme activities were scheduled.

Assessment of clinical performance included online self-assessment and PA of (1) client records and (2) video-recordings, followed by (3) face-to-face discussion of the results, supported by a trained group coach. Participants were assigned to upload one electronic client record and one video-recording of client communication—limited to the discussion of the diagnosis and treatment plan—in cycles 1 and 2. Before assessing their peers, the participants self-assessed their performance using the same indicators. Peers provided online scores and written improvement feedback if relevant. Discrepancies in scores were used as input for the subsequent discussions. After the first PA session, participants designed an online personal improvement plan; during the second session they reflected on the improvements made. Participants who objected to uploading videotapes were allowed to choose for role-playing instead; they simulated the client conversation and their performance was assessed on the spot.

PA coaches received a programme guide and two additional training sessions conducted by professional trainers (MJMM and PjvdW) using samples of client records, video-recordings and role-play to train the process of providing, receiving and using feedback.

Clinical audit was scheduled after all PA activities were completed. A convenience sample of four private clinics was invited to participate. These clinics provided online management information according to the programme guidelines and self-assessed their organisational performance using an online scoring sheet (see online supplementary appendix 1c). Clinical audit included inspection of the property, assessment of two at random selected client records and discussion of self-assessment scores for management and organisation. Afterwards a report was written according to a structured reporting format. Participants were invited to comment on the clinical audits report before it was finalised. Auditors received a programme guide and a training conducted by professional auditors (MO and MB) using worked samples and role-play to train the auditing process.

Website

We developed a web-based assessment system that allowed for (1) downloading programme guides and instruction manuals, (2) uploading assessment materials such as client records, video-recordings, management information and improvement plans, (3) online scoring, (4) downloading assessment results and (5) storing and exporting qualitative and quantitative data. See online supplementary appendix 1 for additional information.

Programme delivery

A research team member (FD) was the programme manager. She provided participants of a programme guide including a manual for uploading and downloading assessment material and guidelines for providing and using quality improvement feedback.^{16 22 28–31}

Ethical issues

All participants gave their online informed consent. Clients providing video-recordings and client records gave their written informed consent. This study was approved by the Medical Ethical Committee Arnhem Nijmegen (CMO): 2015–1797.

Evaluation of programme feasibility

All three feasibility domains were explored with focus groups, in-depth individual interviews and written coach reports aiming for saturated information from multiple perspectives to optimise the credibility and transferability of the results.³² In addition, we evaluated the impact on quality improvement quantitatively by comparing self-assessment and PA scores, and cycle 1 and cycle 2 scores.

Data sampling and analyses of interviews

We aimed to bring together coaches, clinic visitors and knowledge brokers in separate focus groups to explore their experiences in performing the same role. Individual participants were purposely sampled for in-depth interviews including all participants in clinic clinical audits. The website provided us with data to identify and select little, moderate and very active PA participants. They were approached by email. An interview guide was designed by the research team (MJMM, FD, PjvdW, MWGNvdS) addressing the three feasibility domains tailored to the participant role. Focus groups lasted 90–100 min and were conducted face-to-face by MJMM (MSc, educational scientist, performance assessment research) and PjvdW (PhD, movement scientist, evidence-based practice research) using open-ended questions allowing for group discussion and knowledge construction. In-depth interviews permitted us to explore thoughts and feelings that might not easily be shared with colleagues, but relevant to understanding participant's behaviour; they were conducted by MJMM or FD (MSc, health scientist, quality of healthcare research) using teleconferencing technology and lasted 50–60 min. Interviews of all participants, including verbal consent, were audiotaped and transcribed verbatim afterwards. The analytic process was guided by 'template analysis' that combines a-priori codes informed by the research questions with emerging codes from the interview data.³³ PjvdW and MJMM independently studied and coded five transcripts. Differences in coding were discussed, and a code book was created based on consensus. Subsequently, all transcripts were analysed line-by-line, using ATLAS-ti v.7 software. Codes were compared and some codes were merged into higher

order codes. Emerging themes were identified by constant comparison of codes and higher order codes. Finally, we summarised the results relevant to ongoing programme development and implementation.³⁴ To increase the credibility of the results, a peer debriefing and member checking procedure was conducted with research group members FD and MWGNvdS (PhD, allied healthcare scientist, healthcare quality research) and with knowledge brokers and stakeholders.

Data sampling and analyses of scores

Online scores for record keeping and client communication in the first and second assessment cycle were imported in IBM-SPSS Statistics 22. Indicators that were scored as 'not relevant' or 'not applicable' were treated as missing values.

The mean and median indicator scores for each performance indicator and for each performance domain were calculated for self-scores and peer scores as well as the percentages of missing values. We used the Wilcoxon Signed Rank test to calculate differences between self-assessment and PA scores, and between cycle 1 and cycle 2 scores, including p values for statistical significance.

RESULTS

In total 64 physiotherapists took part in the programme. Twelve peer groups were formed based on specialisation, each containing four to six participants. Eleven peer groups participated in online PA; one group used printed scoring sheets. Three group clinics and one solo clinic participated in clinical audits. Table 1 shows an overview of participants' characteristics.

Qualitative results

We conducted two focus groups and 10 in-depth interviews reaching data saturation. Results are discussed using predefined categories including references

to quotes labelled by number and participant's role (KB=knowledge broker, C=coach, V=visitor, P=participant). Quotes are presented in table 2. We identified strengths and weaknesses of the programme design, implementation barriers and facilitators, the impact on quality improvement and critical success features relevant to programme development and implementation. In table 3, the results are summarised.

Acceptability of the programme design

General perceptions

At the beginning, participants were sceptical regarding the feasibility of the programme aims and procedures. Frustrated by the quality demands of health insurers, they were not seeking for an extra administrative burden. However, their views changed along with the programme (Q1-P7). Looking back participants were positive about the programme, because it focused on their core-business and uncovered 'what happens behind closed doors' (Q2-P4).

Despite the guidelines for constructive feedback, providing and receiving feedback were not self-evident. Participants struggled with critically appraising their peers. Being insecure about their own performance, they were cautious in providing critical feedback resulting in 'halo marking' (too high) as supported by the quantitative data (Q3-P6). Experiencing a safe setting, allowing to make mistakes, was perceived as conditional to critical peer appraisal. Regarding receiving feedback, some participants faced difficulties with adequately responding to it (Q4-P9). Participants were unanimous in their view that feedback should be critical to enable meaningful improvement. Compliance to programme guidelines and shared responsibility for group learning was perceived as critical to programme efficacy (Q5-KB). Although participants generally accepted the programme for quality improvement purposes, some of them reported that client records and videotapes might present an overly optimistic picture of clinical practice because they were self-selected (Q6-P8).

Assessment of client communication

Some participants objected to making video-recordings, unwilling to put an unnecessary load on their clients, worrying about client privacy and assuming that they would not consent. Although clients rarely objected to it—in contrast to participant's preassumptions—online assessment activities in cycle 1 were limited (Q7-C). Initial reluctance disappeared when participants personally experienced the added value of video assessment by simply 'doing it', or by watching others 'doing it'; worked samples enhanced the acceptability. It became clear that peer groups needed time and deliberate practice to get used to video-assessment and to feel safe enough to expose their clinical performance. Participants who preferred video-assessment to client records argued that this instrument allows to observe what physiotherapists 'do' instead of what they 'say they

Table 1 Participant demographics and characteristics

Individual characteristics (n=64)		National (N=17.802)
Mean age in years (SD)	50 (10.1)	42
Gender: woman %	50%	56%
Communities of practice characteristics and number of participants		
General conditions	26	
Respiratory conditions	13	
Cardiovascular conditions	10	
Psychosomatic conditions	4	
Neurologic conditions	5	
Geriatric conditions	10	
Total	68*	

*Four physiotherapists participated in two groups. n, number of participants; N, Number of physiotherapists in the Netherlands working in primary care.

Table 2 Quotes of participants**Acceptability of the programme design***General perceptions*

- Q1-P7 "Some of my colleagues were very critical, but now their views are changed. In particular because the program was 'again' about quality standards that we must meet (...). First the health insurers with their audits and now this (...). We don't want more paperwork."
- Q2-P4 "I think the system is appropriate. In fact nobody evaluates you this way. No one comes so close; (...) no one comes into your room and that's how it felt somehow. Yes, perhaps a trainee, or you consult a colleague to look at your patient's problem, but you never ask your colleague to evaluate you; (...) we are very much loners in this respect."
- Q3-P6 "I think it (critical appraisal of peer performance) needs time to develop. I think it will come by doing it a number of times (...). You need to feel safe enough to trust."
- Q4-P9 "In the beginning—I think we needed to get used to it—I saw some participants instantly responding by defending themselves. But I also observed—probably because there was enough empathy and respect for each other—that there was no need to. I saw that (responding to feedback) gradually improved."
- Q5-KB "Because there are always two or three early adopters and the rest is lagging behind. (...) I think it's that sense of responsibility that you need as colleagues to get these things right."
- Q6-P8 "Well, I would do the same. I always say: 'When someone comes to have a look into your kitchen, you make sure that it is cleaned'."

Assessment of client communication

- Q7-C "The first time we (the group majority) chose to role-play, but we also watched two videotapes. After that, everybody said: 'we'll go for the videotapes the next time, these are far more interesting'. And by doing that, we already made some improvements."
- Q8-P6 "Well, actually it was enjoyable. (...). It is quite surprising to see how your colleagues, that's how you know them, how they interact with their client. That provides a lot of information to reflect on. I think attitudes are very important and client records are such a long stories—of course important—but in particular those videos were interesting. Although I also noticed that everyone had more trouble with it."
- Q9-P9 "We saw a COPD-patient [on the video]. It didn't become clear how long she intended to treat this patient?' You often miss some kind of timeline in chronic disease management. I understand that it is not easy, but you can help yourself by setting an evaluation moment."
- Q10-P1 "When they saw my videotape they commented that it was 'big'... I am a bit wordy, that's what I am, that's what I have been doing for thirty years now. Some said it was OK, but one said: 'actually, I scored a 2' [for patient centeredness], meaning that much improvement was needed. I was shocked, because I scored myself with a 4, I thought I was not bad (...). The patient [on the video] said 'yes...yes' all the time. I thought the patient was agreeing, but I should have asked. That was really confronting for me."
- Q11-P2 "...it was great, it was fun, but the tapes were pretty short (...), snapshots of six minutes. I think an electronic client record provides much more information when it comes to critical appraisal."

Assessment of record keeping

- Q12-KB "I think we need to address [clinical practice] guidelines. Put them on the table, show them. We know that they exist, but little effort has been made to applying them (...). Everybody has them on their book shelf, but no one knows the content, well...that might be an exaggeration."
- Q13-P3 "Yes, it was insightful, it confirms what you do and what you don't do. Nobody ever taught me how to keep my records, yes...I once took a course, but that was twenty years ago...you keep records according to your best knowledge; you don't receive feedback until you are audited. Now your colleagues can guide you, I perceived that as very helpful."

Assessment of management and organisation

- Q14-P4 "Well, I think it's very appropriate. In this way—unlike the health insurer—your colleagues come to visit you, it feels more like feedback...because it allows you to create real improvements for yourself."
- Q15-P5 "It is very important that people really feel that they can improve, instead of being challenged. And that's the basis on which people dare to do this."

Appropriateness of the implementation strategy

- Q16-P7 "The minister [Public Health] has cooked up all this and meanwhile the health insurers laugh themselves to death... I think it was fun, we had a nice group and we definitely learnt from each other, but it feels incredibly bad that we only invest and never get something back."
- Q17-P10 "We all struggled with the website somehow. For me, I'm not skilled in computer work. Therefore, I asked my colleague for help in the beginning. But once you start working with it, it becomes more clear. For instance, I succeeded better in uploading the second case than the first one."
- Q18-KB "I never realised that when you want to run such a project, that it takes so much effort to keep PTs focused."
- Q19-C "This morning Karen told me: 'In my view, the most difficult thing of coaching was to activate my group', so, looking back, I think that we should have addressed this issue more extensively during our training'."

Continued

Table 2 Continued

Appropriateness of the implementation strategy

Q20-V "It was difficult to see that C. was very insecure. (...). I thought: 'I am not a monster? Shall I comfort her by saying that she faces the same problems as I do, and that making mistakes is human?' You actually want to help, by saying 'do this, or try that' (...). It's difficult to stay neutral."

Impact on quality improvement and professional behaviour change

Q21-P2 "Yes, according to the guidelines, an EPR includes a baseline-measurement, an in-between measurement, and a final evaluation. That is how it should be, but I don't always meet that standard. I promised to improve that."

Q22-P9 "I need to make clear *what* we are going to do [treatment] and *when* are we going to stop. Especially since I am dealing with clients with a chronic condition."

Q23-P1 "...my record keeping was alright, but regarding patient communication...I explained a lot, but I didn't check to see if my message was understood. That's an improvement I need to make. I try now to ask my patient: 'What did you learn about what I explained just now?' I consciously address that issue. Moreover, I pay more attention to their personal goals. I can have a plan, but that plan might not be in line with their expectations...I might be too dominant in this respect because I think I know what they need, but I should not think for them."

Q24-KB "You know, we saw clinical practice and science bump into each other, and now I think that we have finally come together, pushing, pulling and rolling the same cart."

Q, quote; P, participant; C, Coach; KB, knowledge broker; V, visitor.

do'. They agreed that 'taking a look inside' provided valuable information, such as attitudes becoming observable (Q8-P6). On the one hand, video-recordings allowed for modelling professional behaviours of skillful colleagues, on the other hand, unwanted behaviour became transparent triggering suggestions for alternative behaviours, especially regarding the efficiency of chronic disease management (Q9-P9).

Although online peer scores did not always reflect the ability or willingness of participants to critically appraise their peers, during the sessions feedback quality increased by comparing self-perceptions with peer perceptions and by discussing quality standards of performance. Participants who consciously selected their best videotape could be confronted with different views on quality indicators (Q10-P1).

Participants who preferred assessment of client records to video-recordings argued that they felt uncomfortable with the knowledge that their conversation was recorded ('audience effect') or that a 'snapshot' poorly represents the process of patient management (Q11-P2).

PA of client records

Assessment of record keeping was valued because client records present the process of patient management unlike videotapes, allowing to assess clinical reasoning and decision-making such as the application of clinical practice guidelines, the use of client-reported outcome measures and performance outcome measures. Here again face-to-face discussions were critical to an in-depth understanding of quality indicators, for example, for evidence-based practice. For example, one of the knowledge brokers noticed that 'clicking on the guideline button' in the electronic record system, indicating the use of a particular guideline, was no guarantee for adequately 'applying' the guideline in the specific context of the patient problem (Q12-KB).

In contrast to feedback provided by professional auditors, peer feedback was perceived as a good vehicle to self-direct improvement (Q13-P3).

Clinical audit

Participating private clinics all appreciated clinical audits. They reported that they were 'pretty nervous' in advance, but valued the safe setting allowing for discussion of strengths and weaknesses, providing them feedback to guide improvement of management and organisation towards its quality standards and giving them back responsibility and ownership (Q14-P4; Q15-P5). They all agreed with their audit reports, providing minor comments, although reports were perceived as more formal than audits.

Appropriateness of the implementation strategy**Motivational issues**

At baseline, participants were poorly informed about the programme aims, intended outcomes and consequences. Frustrated by the dominant role of insurers in quality control, participants were suspicious about whose interests were being served by their extra efforts affecting their motivation to participate. Alignment of expectations might have prevented false cognitions and enhanced motivation to invest time and effort (Q16-P7).

Communication technology support

Although the website has been improved continuously throughout the programme, it was not perceived as user-friendly causing feelings of frustration (Q17-P10). Despite the supply of a user manual, peer support and learning by doing turned out to be more effective.

The role of knowledge brokers, group coaches and auditors

Although the knowledge brokers were involved in writing the programme guide, they did not succeed in adequately inform the participants. Their role as linking pin between

Table 3 Summary of findings

Acceptability of the programme design		
Strengths	Weaknesses	Critical success features
<i>General perceptions</i>		
Focus on the core-business of physiotherapists. Uncovers 'what happens behind closed doors'.	Limited validity client records and videotapes because they are self-selected. Limited validity of online scores due to unwillingness or incompetence to adequately apply performance indicators.	Training in critical performance appraisal to support self-directed quality improvement. Time to build a safe setting allowing to make mistakes. Face-to-face discussions of discrepancies in online scores to compare self-perceptions with peer perceptions and to discuss quality standards of clinical performance. Active participation. Compliance to the programme guidelines. Safe setting.
<i>Assessment of client communication</i>		
Shows what physiotherapists 'do' instead of what they 'say they do'. Uncovers undesired attitudes. Allows for modelling desired behaviour.	Reluctance to expose clinical performance to an 'audience'. Snapshot, poorly representing the process of patient management.	Using worked samples of video-assessment to enhance its acceptability. Extended engagement with video-assessment.
<i>Assessment of record keeping</i>		
Presents the process of patient management allowing to assess clinical reasoning and evidence-based practice.		
<i>Assessment of management and organisation</i>		
Provides guidance to self-direct improvement.		
Appropriateness of the implementation strategy		
Barriers	Facilitators	Critical success features
Programme aims, expected efforts and desired outcomes insufficiently clarified at baseline. Dominant role of insurers in quality control causing doubts about the stakeholders in the quality improvement programme. Poor programme efficacy beliefs. Absence of financial incentives. Complex website design.	Learning by 'doing' or by watching others doing it (role models). Emphasis on learning and improvement instead of judgement. Awarding efforts with credits. Peer coaching in using communication technology.	Discussion of programme aims, desired results and consequences on the long term to clarify and align expectations. Shared responsibility for group learning and quality improvement programme outcomes. User friendly website design. Competent group coaches.
Limited skills of group coaches to enhance shared responsibility for group learning and results.		
Impact on quality improvement and professional behaviour change		
Individual level	Organisational level	Network level
Awareness of clinical performance.	Awareness of organisational performance.	Increased self-efficacy beliefs and motivation for ongoing PA activities.
New insights in the application of clinical practice guidelines, the use of client reported outcomes and performance measures. Improved client involvement in goal setting and treatment planning. Improved peer assessment skills.	Increased self-efficacy and programme-efficacy beliefs.	Commitment to ongoing PA activities and clinical audits.
PA, peer assessment.		

researchers and clinicians required advanced communication and leadership skills (Q18-KB).

The role of the coach was perceived as crucial in facilitating critical reflection and an in-depth understanding

of quality standards. However, some group coaches had to deal with the 'wait and see' attitude of some participants who did not provide online materials in time. They lacked the coaching skills to support active

participation and shared responsibility for group learning (Q19-C).

Some clinic auditors struggled with their role identity. They were trained to communicate what they observed regarding the quality indicators; as such they felt competent to provide information on 'what' can be improved (feedback), but not on 'how' (feed forward) appealing to a counsellor role rather than an auditor role (Q20-V).

Impact on quality improvement and professional behaviour change

The programme impacted on different levels of professional practice, providing feedback to individuals, peer groups and clinics. Regarding professional development, positive feedback enhanced self-efficacy beliefs and motivation to participate in continuing PA activities. Intentions to behaviour change focused on guideline adherence, performance measurement and client-reported outcome measurement (Q21-P2; Q22-P9; Q23-P1). On the level of organisation and management, participants reported improved awareness of strengths and weaknesses and increased beliefs in the change capacity of the programme (Q14-P4; Q15-P5).

The collaboration between the research team and the network of participating physiotherapists resulted in context-specific knowledge, relevant to ongoing quality improvement activities. The network committed to continue with PA and clinical audits, intending to address its critical success features (Q24-KB).

Quantitative results

Table 4 presents the results of the online uploaded data on the website showing that online activities varied widely and that participation in cycle 1 was substantially lower than that in cycle 2. Perceived barriers to online activities are reported in the qualitative results section. Except for record keeping in cycle 1, peer scores were

higher than self-scores but differences were not significant. Since participants' online activities were low in cycle 1, data on the improvements made are limited. As shown in the shaded area of table 4, differences between cycles 1 and 2 were not significant for client communication, but significant for record keeping, especially regarding the lower performers at baseline. Note that these differences relate only to the limited number of participants who were active in both cycles.

DISCUSSION

This study focused on the feasibility of a quality improvement programme aiming to enhance the client-centeredness, effectiveness and transparency of physiotherapy services. The qualitative results showed that participants viewed the programme as an acceptable intervention for quality improvement purposes, allowing for stepwise self-directed quality improvement unlike the one-shot assessments of external auditors. We identified its critical success features such as training in performance appraisal and time to build a safe setting. Regarding the appropriateness of the implementation strategy to execute the programme as intended, participants reported several facilitators and barriers, allowing us to identify critical success features for broader implementation such as adequate communication of programme aims and intended outcomes at baseline, user-friendliness of the website design and competent group coaches. The weaknesses of the programme design and the barriers to programme implementation affected the impact on quality improvement and behaviour change. However, we identified meaningful self-reported results including awareness of clinical and organisational performance, improved evidence-based practice and client-centeredness and increased motivation to self-direct quality improvement. The assessed (quantitative) results showed that online activities were low in cycle 1, providing limited data on the

Table 4 Differences between self-assessment and peer assessment scores and differences between cycles 1 and 2 scores tested with non-parametric Wilcoxon signed Ranks test (Likert Scale 1–5)

	N	Mean	M/R/A	Median	Min	Max	Differences between SA1 and PA2 scores		Differences between cycle 1 and cycle 2 scores	
							Mean difference	p Value	Mean difference	p Value
SA client communication cycle 1	9	3.52	5.6%	3.60	2.33	4.75			0.10	0.674
PA client communication cycle 1	11	3.79	3.9%	3.77	3.25	4.78	0.27	0.263	−0.91	0.386
SA record keeping cycle 1	26	3.43	3.2%	3.50	2.00	5.00			0.20	0.007**
PA record keeping cycle 1	31	3.41	1.6%	3.60	1.25	4.50	−0.02	0.760	0.15	0.002**
SA client communication cycle 2	40	3.62	5.8%	3.67	2.20	5.00				
PA client communication cycle 2	45	3.70	6.4%	3.80	2.21	4.50	0.08	0.274		
SA record keeping cycle 2	48	3.62	2.8%	3.70	1.67	5.00				
PA record keeping cycle 2	63	3.75	3.3%	3.75	2.08	4.58	0.13	0.269		

**Significant at a 0.01 level.

M/NA, mean percentage of missing/perceived not relevant/perceived not applicable indicator scores; PA, peer assessment; SA, self-assessment.

improvements made in cycle 2. Despite the limited data, we observed significant improvement of self-scores and peer scores for record keeping.

When we look at programme acceptability, participants' views on the validity and the learning value of video-recordings and client records differed. We suggest that the acceptability of videos could be improved. Instead of using two single video clips, perceived as 'snapshots', several video recordings would provide more valid information as showed by a study of Ram *et al.*³⁵ However, that involves additional time and costs and might threaten long-term feasibility. Assuming that each instrument to assess professional performance has its advantages and disadvantages (standardised clients, direct observation, multisource feedback) and that there is no single best measure as shown by a systematic review of Overheem *et al.*,³⁶ the use of multiple measures is justifiable and even desirable for the purpose of gathering valid and reliable information on clinical competence.³⁷

Regarding programme implementation, we assume that the sociopolitical context—the dominant role of health insurers in quality assurance—impacted heavily on commitments to change and outcome expectancies.³⁸ Although PA aimed to provide formative feedback, emphasising learning and improvement, it was viewed as summative assessment as physiotherapists questioned the stakeholders in their efforts. Improved communication at baseline might have enhanced participant's motivation and adherence to programme guidelines. Moreover, external, top down empowerment, a trade-off between trust and control, might be critical to successful outcomes on the long term as recognition of professional accomplishment and innovation is a strong motivator of improvement.¹

Looking at the impact on quality improvement, we observed that peer scores for client communication were consistently higher than self-scores, demonstrating that participants either underestimated their own performance or overestimated their peers. In contrast to the literature on self-assessment showing that physicians generally overestimate themselves,³⁹ we assume that feelings of insecurity underlie overestimation as underestimation in this case as supported by the qualitative data. Extended exposure to critical appraisal and reinforcement of constructive feedback practices could strengthen self-efficacy beliefs according to Bandura's cognitive learning theory.⁴⁰ The results also showed that the programme was more effective in enhancing record keeping skills, than communication skills. Apparently, it took more time or effort to develop communication skills within the time span of the programme. This assumption is supported by feedback intervention theory^{31–41} explaining that the effectiveness of performance feedback is lower when the 'task novelty' and 'task complexity' is higher. Trained by audits of health insurers, participants were more familiar with assessment of record keeping. Moreover, the literature showed that clinical competency is content and context specific, meaning that competent (complex) behaviour in one case (cycle 1)

is a poor predictor for another case (cycle 2),^{42–43} and this also applies to communication skills.⁴⁴ Although this programme was not intended to produce generalisable scores, we suggest that prolonged engagement with video-assessment would yield better outcomes.

Strengths and limitations

This study evaluated what physiotherapists do in their day-to-day practice by assessing client records, video-recordings and management information. The quality improvement programme was systematically developed, and theory-based and evidence-based. Stakeholders and end-users were actively involved in programme development and implementation, and their experiences provided meaningful information on its critical success features. Participants did not fully adhere to the programme guidelines resulting in limited sample sizes threatening internal and external validity of the quantitative results. It should also be noted that we could not distinguish between 'missing' and 'not relevant or not applicable' indicator scores of active PA participants which might have biased the results. Although generalisability of the quantitative results is limited regarding the specific population of Dutch physiotherapists in primary care, we think that the qualitative results related to the acceptability and the implementation of the quality improvement programme are learning points for a broader group of healthcare professionals.

CONCLUSIONS

This study demonstrated that bottom-up quality improvement initiatives can be effective in improving healthcare quality. The results justify more rigorous evaluation to inform nationwide implementation when its critical success features are addressed. Crucial is the willingness of professionals and organisations to provide access to the confidential areas of their clinical practice. However, this information is vulnerable to summative judgement and should be protected by all stakeholders in healthcare quality. Further research is necessary to explore the sustainability of the results and the impact on client outcomes in a full-scale study.

Author affiliations

¹Institute of Health Studies, HAN University of Applied Sciences, Nijmegen, The Netherlands

²Radboud Institute for Health Sciences, IQ Healthcare, Radboud University Medical Center, Nijmegen, The Netherlands

³Dutch Institute of Allied Healthcare, Amersfoort, The Netherlands

⁴Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands

Acknowledgements The authors thank Marielle Ouwens (MO) and Menno Bouman (MB) for their input as professional auditors, the stakeholder group for their input in programme development, Frits van Trigt (FT) and Han Kingma (HK) for their efforts as knowledge brokers, all coaches and physiotherapists joined in the Network FysioGroep Haaglanden, Carol and David DeFields for proofreading the manuscript.

Collaborators Maria W G Nijhuis-van der Sanden, Femke Driehuis, Yvonne F Heerkens, Cees P M van der Vleuten, and Philip J van der Wees.

Contributors All authors read the final manuscript, gave their approval for publication and agreed to be accountable for all aspects of the work. MJMM, PJvdW and MWGNvdS contributed to study conception, design, sampling, analysis, interpretation of data, drafting and revising the manuscript. FD contributed to the conduct of the scoping review and the study. CPMvdV and YFH contributed to interpretation of the data and revision of the manuscript.

Funding This study was funded by the KNGF. The KNGF had no role in the conduct of this study, analysis or interpretation of data.

Competing interests None declared.

Ethics approval This study was approved by the Medical Ethical Committee Arnhem Nijmegen (CMO): 2015-1797.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

- Institute of Medicine. *Crossing the quality chasm: a new health system for the 21st century*. Washington (DC): National Academy Press, 2001.
- Eijkenaar F. Pay-for-performance for healthcare providers. Design, performance measurement, and (unintended) effects. *Health Policy (New York)* 2013;110:115–30.
- Casalino LP, Elster A, Eisenberg A, et al. Will pay-for-performance and quality reporting affect health care disparities? *Health Aff* 2007;26.
- Grol R. Quality improvement by peer review in primary care: a practical guide. *Qual Health Care* 1994;3:147–52.
- Butterfield R, McCormick B, Anderson R, et al. Quality of NHS care and external pathway peer review. 2012. <https://www.chseo.org.uk/downloads/report3-peerreview.pdf> (accessed October 2016).
- Pronovost PJ, Hudson DW. Improving healthcare quality through organisational peer-to-peer assessment: lessons from the nuclear power industry. *BMJ Qual Saf* 2012;21:872–5.
- Prochaska JO, Redding CA, Evers KE. Health behavior and health education. In: Glanz K, Rimer BK, Viswanath K, eds. *Health behavior and health education: theory, research, and practice*. Wiley & Sons, 2008:97–121.
- Ajzen I. Nature and operation of attitudes. *Annu Rev Psychol* 2001;52:27–58.
- Maas MJM, van Dulmen SA, Sagasser MH, et al. Critical features of peer assessment of clinical performance to enhance adherence to a low back pain guideline for physical therapists: a mixed methods design. *BMC Med Educ* 2015;15:203.
- Ferlie EB, Shortell SM. Improving the quality of health care in the United Kingdom and the United States: a framework for change. *Milbank Q* 2001;79:281–315.
- Australian commission on safety and quality in health care. Review by peers. A guide for professional, clinical and administrative processes. 2010. <https://www.safetyandquality.gov.au/wp-content/uploads/2012/01/37358-Review-by-Peers1.pdf> (accessed April 2016).
- Buetow SA, Roland M. Clinical governance: bridging the gap between managerial and clinical approaches to quality of care. *Qual Heal Care* 1999;8:184–90.
- Eva KW, Regehr G. Self-assessment in the health professions: a reformulation and research agenda. *Acad Med* 2005;80:S46–54.
- Norcini JJ. Peer assessment of competence. *Med Educ* 2003;37:539–43.
- Hofhuis H, van den Ende C, de Bakker D. Effects of visitation among allied health professionals. *Int J Qual Heal Care* 2006;18:397–402.
- 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;86:1120–7.
- Maas MJ, van der Wees PJ, Braam C, et al. An innovative peer assessment approach to enhance guideline adherence in physical therapy: a single-masked, cluster-randomized controlled trial. *Phys Ther* 2015;95:600–12.
- Phillips Fox D. Peer review of health care professionals: a systematic review of the literature. Melbourne. 2009. <https://www.safetyandquality.gov.au/wp-content/uploads/2012/01/25738-LitReview.pdf> (accessed April 2016).
- Van den Hombergh P, Grol R, van den Hoogen HJ, et al. Practice visits as a tool in quality improvement: acceptance and feasibility. *Qual Health Care* 1999;8:167–71.
- 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.
- Payne VL, Hysong SJ. Model depicting aspects of audit and feedback that impact physicians' acceptance of clinical performance feedback. *BMC Health Serv Res* 2016;16:260.
- Hysong SJ. Meta-analysis: audit and feedback features impact effectiveness on care quality. *Med Care* 2009;47:356–63.
- Moore GF, Audrey S, Barker M, et al. Process evaluation of complex interventions: Medical Research Council guidance. *Br Med J* 2015;350: h1258.
- Bowen DJ, Kreuter M, Spring B, et al. How we design feasibility studies. *Am J Prev Med* 2010;36:452–7.
- Hoens A, Li L. The knowledge broker's 'fit' in the world of knowledge translation. *Physiother Canada* 2014;66:223–7.
- Ward V, House A, Hamer S. Knowledge Brokering: the missing link in the evidence to action chain? *Evid Policy* 2009;5:267–79.
- 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.
- 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.
- Archer JC. State of the science in health professional education: effective feedback. *Med Educ* 2010;44:101–8.
- Sargeant JM, Mann KV, van der Vleuten CP, et al. Reflection: a link between receiving and using assessment feedback. *Adv Health Sci Educ Theory Pract* 2009;14:399–410.
- Kluger AN, Denisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bull* 1996;119:254–84.
- Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus group. *Int J Qual Heal Care* 2007;19:349–57.
- King N, Cassel C, Symon G. Using templates in the thematic analysis of texts. In: Cassell CM, Symon G, Eds. *Essential guide to qualitative methods in organisational research*. London: Sage, 2004:256–70.
- Huberman AM, Miles MB, Denzin NK, et al. Data management and analysis methods. In: Denzin NK, Lincoln YS, Eds. *Handbook of qualitative research*. Sage, 1994:428–44.
- Hobma S, Ram P, Muijtjens A, et al. Effective improvement of doctor-patient communication: a randomised controlled trial. *Br J Gen Pract* 2006;56:580–6.
- Overheem K, Faber MJ, Onyebuchi AA, et al. Doctor performance assessment development in daily practise: does it help doctors or not? A systematic review. *Med Educ* 2007;41:1039–49.
- Van Der Vleuten CPM, Schuwirth LWT, Driessen EW, et al. Twelve tips for programmatic assessment. *Med Teach* 2015;37:641–6.
- Weiner BJ. A theory of organisational readiness for change. *Implement Sci* 2009;4:67.
- Eva KW, Cunnington JPW, Reiter HI, et al. How can I know what I Don't know? Poor self assessment in a well-defined domain. *Adv Health Sci Educ Theory Pract* 2004;9:211–24.
- Bandura A. Self-efficacy: toward a unifying theory of behavioral change. *Psychol Rev* 1977;84:191–215.
- Hysong SJ, Kell HJ, Harrison J, et al. Theory-based and evidence-based design of audit and feedback programmes: examples from two clinical intervention studies. *BMJ Qual Saf* Epub ahead of print: 10 Jun 2016. doi:10.1136/bmjqs-2015-004796
- Eva KW. What every teacher needs to know about clinical reasoning. *Med Educ* 2005;39:98–106.
- Durning S, Artino AR, Pangaro L, et al. Context and clinical reasoning: understanding the perspective of the expert's voice. *Med Educ* 2011;45:927–38.
- Baig LA, Violato C, Crutcher RA. Assessing clinical communication skills in physicians: are the skills context specific or generalizable. *BMC Med Educ* 2009;9:22.