Author’s response to reviews

Title: Evaluation of the theory-based Quality Improvement in Physical Therapy (QUIP) program: A one-group, pre-test post-test pilot study.

Authors:

Geert M. Rutten (g.rutten@maastrichtuniversity.nl)
Janneke Harting (j.harting@amc.uva.nl)
L.Kay Bartholomew (Leona.K.Bartholomew@uth.tmc.edu)
Angelique Schrief (a.schrief@iq.umcn.nl)
Rob A.B. Oostendorp (rob.oostendorp@planet.nl)
Nanne K. de Vries (n.devries@maastrichtuniversity.nl)

Version: 4 Date: 18 February 2013

Author’s response to reviews: see over
Maastricht, 18 February 2013

Dear editor,

Please find enclosed the re-submission to BMC Health Services Research of MS ID: 1225287207800403 entitled ‘Evaluation of the theory-based Quality Improvement in Physical Therapy (QUIP) program: A one-group, pre-test post-test pilot study’.

We thank the reviewers for their clear recommendations which have allowed us to improve the manuscript. All adjustment and explanations are presented in the subsequent pages of this covering letter. To facilitate identification, adjustments in the manuscript are written in red font.

Yours sincerely,
Geert Rutten

Maastricht University
Department of Health Promotion
Room B 1.076
P.O. Box 616
6200 MD Maastricht
Netherlands
Tel +31 (0)43 3882416
Fax +31 (0)43 3671032
Email: g.rutten@maastrichtuniversity.nl
www.gvo.unimaas.nl
Response to review: Maastricht, 18 February 2013

MS ID: 1225287207800403

Title: Evaluation of the theory-based Quality Improvement in Physical Therapy (QUIP) programme: A one-group, pre-test post-test pilot study.

Version: 2 Date: 5 November 2012

Reviewer: Annette Bishop

Reviewer's report:

Discretionary but my overall recommendation for this report.
1. Overall, this small study of a one group pre-test post-test design, using mixed methods, lends itself well to assessing the fidelity, acceptability and feasibility of the QUIP programme. The authors clearly identify the limitations of the study and suggest that effectiveness would be better tested in a randomised design. My recommendation would be to present this study as a feasibility study and thus not place too much emphasis on the tests of effectiveness, but suggest these should be tested in a larger appropriately designed future study.

We agree with the reviewer about the main purpose of the study. It is not our intention to highlight the programme’s effectiveness. In Intervention Mapping the pilot study of a newly developed programme serves the further development towards a final version that can be evaluated in an RCT. For this purpose two main questions are important: 1) is it likely that the programme is effective, in other words, should we proceed at all with the development, and 2) is the present format of the programme feasible, which is only of importance if the answer to the first question is affirmative. In addition, based on the recommendations of reviewer 2 we substantially expanded the text of the process evaluation.

In order to reduce the focus on the effectiveness we adjusted the final paragraph of the Introduction section:
This paper describes the pilot study of the QUIP programme. In IM, a pilot study serves the further development of the programme towards a final version. For this purpose two questions are important: 1) is it likely that the programme is effective, in other words, is further development worthwhile, and 2) is the present format of the programme feasible, which is only of importance if the answer to the first question is affirmative. In accordance, this study comprised an evaluation of the potential effectiveness of the programme as regards the improvement of guideline adherence and its determinants, and a process evaluation, to evaluate the programme's implementation feasibility, acceptability and fidelity. The combination of the effect and the process evaluation findings additionally served the identification of strengths and weaknesses of the programme (Steckler, 2002 #661).

We also added the following lines to the 3d par of the Discussion section
However, the main purpose of the effect evaluation was to assess whether further development of the programme would make sense at all. The results of this study indicate that the programme may have the capacity to improve guideline adherence. **Finally, to reduce the idea of a comparison of results, we downsized the 4th par. of the Discussion section:**

Most published evaluations of interventions aimed at improving guideline adherence concern randomized clinical trials (RCTs), which makes a comparison of our study findings rather complicated. Mostly, improvements found in implementation research among health care providers, including physical therapy, are 5-15% {Grimshaw, 2004 #116; Bekkering, 2005 #100; Stevenson, 2006 #775; Rebbeck, 2006 #496}.

2. If this is presented as a feasibility study I would like to see the authors recommendations for a further larger study described to show what are the key messages learnt from this pilot.

*We adjusted the final paragraph of the discussion section:*

In conclusion, although overall guideline adherence did not improve, changes on individual quality indicators suggest that a systematically developed, theory-based programme to enhance adherence to the Dutch physical therapy guidelines for low back pain has the potential to improve the quality of physical therapy care for patients with low back pain. The integrated approach of individual PTs and their PQMs, with much room for interaction, seems to benefit positive performance change. The self-regulation approach is suitable, but the programme should be re-designed in a way that it allows for a thorough self-reflection on personal performance at the beginning and at the end. Also, the steps of self-regulation should be made more explicit to enhance PT’s and PQM’s awareness of this process. Goal setting, individually and collectively, and action planning seem to be core steps in the programme, but PT’s and PQM’s need some guidance in the sound formulation of challenging yet achievable goals for improvement. In addition, more attention should be paid to translation into daily practice and to the importance of the continuity of the self regulation process. Finally, the programme should result in a thorough future practice plan for guideline implementation. In order to achieve this, the programme should allow sufficient time for attention to individual subjects and strategies and could benefit from a follow-up session to assess and support sustainability. After re-designing the programme, more sophisticated designs with larger samples are required to draw sound conclusions about its effectiveness.

3. However, if the testing of effectiveness is to be included here, the authors should add further comment on the likely impact of such a small sample size (n=24 or 25). Although overall adherence showed minimal change even descriptively some of the other determinants may have shown larger (and significant) effect sizes with a larger sample size. Inclusion of confidence intervals would be useful for the effect sizes.

*We understand the reviewers concerns as regards the presentation of the effectiveness in relation to the small sample size. We agree that a larger sample size may evoke better results. However, since this is not the main purpose of the study (see response to 2.), we prefer not to further emphasize this in our manuscript. Instead, we recommend a larger sample sizes in an effect evaluation study.*
In the limitations paragraph of the Discussion section we mention the instability of effect sizes in small samples. We adjusted these lines somewhat to make this more clear.

Third, effect sizes to express the strength of the effects can be rather unstable in small samples, resulting in large confidence intervals {Kraemer, 2006 #924}. As recommended for small samples, we therefore present them together with T-tests including significance levels {Field, 2005 #638}.

We also adjusted the final sentence of the manuscript
After re-designing the programme, more sophisticated designs with larger samples are required to draw sound conclusions about its effectiveness.

Compulsory revisions
4. More information on the purposive sampling of practices should be included including an explicit description of the inclusion criteria. Were these practices typical practices or did they only take part because they were already very aware of quality management and guidelines?

Since we intended to stay as close as possible to the real life situation, the primary inclusion criteria for practices were the presence of an (initial) quality management and at least five PTs. We decided for precisely these criteria because of the current developments in Dutch PT, in which practices are growing in their number of employees, and in which health insurance companies prefer to do business with practices with a quality certificate. Since quality certification is a recent development many practices struggle with their quality management as well as with the management of a larger organisation. Furthermore, we had some preferences for the purposive sample which have been added to the text on 1th par of the Design and recruitment section, which we added to the Design and recruitment section.

… at least five PTs. In order to reach an acceptable reflection of common practice, we preferred a mixture of male and female therapists of various age groups and with a difference in working experience. We also intended to include PT’s and manipulative PT’s (MPT’s), since a substantial proportion of low back pain patients visit MPT’s. For pragmatic reasons…

We also added the following information to the Response and participants section:
… work experience (SD = 9.86). Almost 33% (n = 8) were MPT’s. On the average,…

However, since this is a self selected sample, participants will probably be more motivated for delivery of high quality care. We already addressed this limitation of the study in the limitations paragraph of the Discussion section and slightly adjusted it:
A final limitation concerns the small, purposive and self-selected sample. Although age and gender of the individual participants did not substantially deviate from the national data (Kenens, 2008 #551), the results cannot be extrapolated to physical therapy practices that lack an initial quality management structure, or to PTs and PQMs that are less motivated for quality improvement. However, current developments in Dutch physical therapy, with health insurance companies monitoring the quality of care, will make an investment in the deliverance of optimal care inevitable.
5. Include a sentence to say how the clinical reasoning data was collected.
Was the factor analysis to establish the individual and organisational determinants part of the previous study (ref 31)?

*We adjusted the explanation of the clinical reasoning score and added an example:*

*Clinical reasoning* was measured by assessing the consistency of PTs’ choices over three separate quality indicators (see bottom part of Table 2) concerning the handling of psychosocial factors (Rutten, 2010 #722). Consistency in choices was operationalized as the presence (1 = present, 0 = not present) of the “conditional argument” (if-then connective) which is an important component of human reasoning. (Goel, 2007 #907) For instance, if PTs identified psychosocial factors in the case description of a vignette, did they subsequently address them in their treatment objectives? In accordance with the overall adherence score, the overall consistency measure was determined by calculating the average of the three consistency indicator scores.

*We performed a principal component analysis on the data of ref 31 retrospectively and we also performed a PCA on the data of our formative work for this project (manuscript is submitted). The main decision tools for the composition of our determinants are our theoretic framework in combination with the assessment of the determinants’ internal consistency (Cronbach’s alpha’s). The principal component analysis was mainly used as a control tool to see if our theoretic framework made enough sense. As a consequence, we do not specifically report on this in our papers.*

6. Clarify if the overall guideline adherence is a composite of the 12 quality indicators.

*We adjusted the explanation of the adherence score in the 1st par. of the Measurement and data collection section:*

The scores on the individual quality indicators (1 = meets quality indicator, 0 = does not meet quality indicator) per vignette were used to calculate an overall percentage score per indicator per therapist: (actual score on indicator / maximum score on indicator) x 100. The mean overall percentage adherence was established by calculating the average score of the 12 indicators.

7. The qualitative findings presented were all positive. Were there any negatives? Particularly were there any problems with engaging physical therapists or practices in the many hours of the programme? This should link to the discussion as it is suggested the programme was very brief and may need to more comprehensive to cover all relevant aspects, but the burden of the individuals and practices would increase if the programme was extended.

*Negative findings concerned the time investment for quality management at all, and, related to that, the financial consequences. Time investment in the course as such (4 x 3 hours for PT’s and 6 x 3 hours for PQM’s) was one and a half working day for PT’s and two working days for PQM’s. This was not judged as too many hours by PT’s as well as PQM’s. Since these were not immediately negative opinions about the programme we did not mention them before. However, we agree with the*
reviewer that these findings are of importance in relation to the Discussion. Therefore we completely revised the feasibility paragraph:

Problems with feasibility of the programme in its current form included available time, variability in completion of homework assignments, and underestimation of the needed remediation of the knowledge level for some issues. In addition, application of the self-regulation process (theoretical core) and sound clinical reasoning (basic professional skill) appeared to require more explicit instruction and guidance. Another participant’s advice was to more strictly monitor the translation to daily practice during the programme. Although all practices made progress, the plan for continuing the programme components as a normal part of practice would deserve greater attention and monitoring.

PQM’s indicated that quality management in this form was valuable, yet time consuming, and that is would help if they were in some way financially compensated for their efforts. This was especially required when this time investment was expected to become part of daily routine. PT’s realized that being involved in quality improvement activities would cost private time. Although they expressed commitment to do this, agreements had to be made with practice management about the required time investment. Concerning the number and duration of the sessions of the programme the PQM’s and the PTs indicated that 3 hour-sessions were acceptable. Since the programme would benefit from an extra session to include the issues mentioned above, some participants preferred to make the individual sessions one hour longer instead. Moreover, a six-month, follow-up session would allow monitoring and support maintenance of the quality improvement process. Two PQM’s indicated to prefer an on-site form of this programme, because it would reduce the time investment for the whole practice team. However, others indicated that this would reduce the opportunity for interaction and peer learning.

We also added a paragraph to the discussion section

Some participants preferred an on-site intervention because they expected it to reduce their time investment. Although we considered this approach, we chose for meetings outside the practice. Our choice for meetings with more than one practice simultaneously was mainly driven by four reasons. First, on site interventions, such as educational outreach visits, have demonstrated small to modest effects {O’Brien, 2007 #925} on the change of professional performance. Second, we wanted to create the opportunity for interaction, which is identified as a factor that may increase the effect of educational meetings {Forsetlund, 2009 #731}. Third, we prefered to take the PTs out of their daily context, since their habitual working environment provides cues for habitual performance {Ouelette, 1998 #822}. By taking PTs out of their daily working environment these cues were avoided. Finally, we considered the costs of the programme, taking into account its implementation on a larger scale. Given the number of almost 5000 physical therapy practices in the Netherlands, an on-site programme would have been difficult to manage and very cost expensive.

8. Language corrections

The text was corrected by a native speaker and we went through it to correct further language errors.

Version: 2 Date: 16 December 2012
Reviewer: Deirdre Hurley

Reviewer’s report:

This study presents a complex, multi-level evaluation of a pilot of the quality improvement in Physiotherapy Therapy (QUIP) programme designed to increase adherence to the Dutch clinical guidelines for low back pain to evaluate its effectiveness, fidelity, acceptability and feasibility. It is topical and relevant to clinical guideline into practice implementation. The findings will be of interest to clinicians, managers and researchers.

Major Compulsory Revisions

1. Give specific details of the guideline being evaluated and how this study relates to the implementation strategy of the guideline. Also what level of implementation of the guideline was undertaken before the study commenced?

*We are concerned that a detailed description of the guidelines will add too much text to the paper. To meet the recommendation of the reviewer, we added the following lines to par 2 of the Introduction section and we also added a reference to the website of the Royal Dutch Society for Physical Therapy that contains English versions of the Dutch PT guidelines:*

The guidelines urge clinical reasoning, assessment and management of psychosocial factors, and documentation including outcome measurement. Their four main features are: applying the International Classification of Functioning, Disability and Health (ICF); identifying and applying patient profiles with duration, course, and psychosocial factors influencing recovery; limiting the number of treatment sessions in case of acute low back pain; and focusing on patient behavior to restore physical activity and social participation (an English version is available at [http://www.fysionet-evidencebased.nl/index.php/kngf-guidelines-in-english](http://www.fysionet-evidencebased.nl/index.php/kngf-guidelines-in-english)). At the time of the study, the Dutch physical therapy guidelines were disseminated by a combination of strategies. These included sending them by mail to every member of the Royal Dutch Society for Physical Therapy (KNGF), which are about 90% of Dutch physical therapists, presentation at the annual national physical therapy conference and publication in the National Journal of Physical Therapy. The guidelines also came with a competency manual in which physical therapists could test their knowledge of the guideline. At a later stage the Society for Physical Therapy developed programmes that Communities of Practice existing in Dutch physical therapy could use to improve guideline implementation. All implementation activities were voluntary and targeting individual physical therapists only.

*We also adjusted the final par. of the Introduction:*

This paper describes the pilot study of the QUIP programme. In IM, a pilot study serves the further development of the programme towards a final version. For this purpose two questions are important: 1) is it likely that the programme is effective, in other words, is further development worthwhile, and 2) is the present format of the programme feasible, which is only of importance if the answer to the first question is affirmative. In accordance, this study comprised an evaluation of the potential effectiveness of the programme as regards the improvement of guideline adherence and its determinants, and a process evaluation, to
evaluate the programme's implementation feasibility, acceptability and fidelity. The combination of the effect and the process evaluation findings additionally served the identification of strengths and weaknesses of the programme (Steckler, 2002 #661).

2. Provide further details of how the QUIP programme elements were developed and the level of engagement between PTs, PQMs and the research team during the process.

We understand the reviewer's request for more information about the development of the programme. A detailed description of the development of the programme is, however, reported on in another manuscript that is currently under review elsewhere. Given the extensive process of development, we think it would distract the reader from the purpose of this study if we included it in this paper. Therefore, we prefer to only present a detailed description of the final programme in this manuscript (Additional file 1 and 2)

3. Provide more details of the n values for each stage of the process, particularly the qualitative components.

Except for the vignettes (n = 24), all questionnaires were completed by all the 25 participants. We added n-values for every step of the process evaluation:

Six participant group or individual interviews (n = 21) were conducted by two members of the research team (GR and JH or AS) within 3 weeks after the programme ended. Two interviews, one after the second session and one after the final session, were conducted with the instructors who executed the programme. Guided by the evaluation questions, one of the researchers performed the interview and the other took notes. Visiting six practice locations for the interviews also provided the researchers the opportunity to observe changes in practice management and to make field notes.

Documents to be evaluated were the Personal Development Plans (n = 25) and the Practice Quality Improvement Plans (n = 7) that were written by the PTs and the PQMs as an assignment of the programme. After the last session, the participants completed a general course evaluation questionnaire (n = 25; completion time less than 15 minutes) to assess perceptions of content quality, trainers, location, organization and overall judgment of the course (1 = extremely bad to 10 = excellent).

4. Report the time involved in completing the various study measures for PTs and PQMs, the overall time frame of participation per practice and the time involved in each stage of the programme.

The exact time investment per measure was not registered. In a pretest, the time for completing the determinants questionnaire and the vignettes was 60-90 minutes. Completing the general evaluation form took less than 15 minutes after the final session. The time frame of participation in the programme per practice is presented in additional file 1.
We added the information to the Methods: Measurement and data collection section:
The clinical vignettes as well as the determinants questionnaire were completed by the PT’s and the PQM’s one week before the start of the intervention (August 2009), and within two weeks after finishing the intervention (December 2009). Completing these measures took 60 to 90 minutes.

After the last session, the participants completed a general course evaluation questionnaire (n = 25; completion time less than 15 minutes) to assess perceptions of content quality, trainers, location, organization and overall judgment of the course (1 = extremely bad to 10 = excellent).

5. Give postgraduate qualifications/additional post qualifying education of PTs in musculoskeletal area and how they relate to national profile. - Did this have any influence on their adherence?

Some of the participants were MPT’s. The reason that we did not distinguish PT’s and MPT’s is that MPT’s mostly do not perform any better in their adherence than PT’s (see Oostendorp et al, 2012, Manuelle Therapie, 16; 90-98). We added the information about additional education to the RESULTS: response and participant section:

The average age of the participants was 39 years (range 24 to 56), 55% were female (n = 15), and participants averaged 15.5 years of work experience (SD = 9.86). Almost 33% (n = 8) were MPT’s. On the average, PQMs had 0.35 FTE (range 0.1–0.5) available for quality management.

6. Give more detail of the specific results of the qualitative components of the study.

We adjusted the process evaluation results:

Programme instructors competently delivered the methods and practical applications, albeit, due to time limitations, briefly for most. This resulted in some deficits in both programme delivery and learning, including superficial reflection by PTs on their personal adherence, limited attention for goal setting, management and leadership skills and little discussion of the issue of maintenance. Furthermore, changes in the programme were required because of the unexpected low knowledge levels of the PTs on some themes, such as red flags (signs and symptoms of serious diseases), application of measurement instruments and psychosocial factors. For that purpose, PT’s advised a discussion about the content of the guideline in the first session.

Acceptability

The revised guideline was positively judged even though its recommendations were largely similar to the former version. The revision was unanimously found to be less normative, more flexible, less extensive and easier to understand and apply. The patient leaflet had only been
used by one practice, despite the judgment of all practices that the content was supportive and useful.

Of the seven practices that completed the course, six were unanimously very positive about the programme (score 8 out of 10). In one practice, the opinions differed, varying from fair (6 out of 10) to very positive (8 out of 10). The main positive ideas were that it learned them something about the process of implementation of guidelines instead of treatment content and that it gave them the opportunity to compare themselves with other practices. The main critique came from one practice with a higher level management structure in that they had missed a ‘sparring partner’ on their own level. One practice dropped out. Although neither their pre-intervention adherence scores nor our observations showed better performance compared with other practices, managers explained the practice was already engaged in a quality improvement process and did not learn anything new. One of the practice’s two PQMs also indicated, however, to lack leadership skills, which may also have been a reason they dropped out. The other manager judged the programme as more suitable for practices with lower performance levels.

The PTs’ assessments of the course instructors were very positive, as were assessments of the interactive small group sessions with colleagues from the practice, the plenary discussions, presentations with peer and expert feedback, and the Meet the Expert session. Small group sessions with peers from other practices were highly appreciated by the managers, who learned from exchanging experiences, but to a lesser extent by individual PTs.

**Feasibility**

Problems with feasibility of the programme in its current form included available time, variability in completion of homework assignments, and underestimation of the needed remediation of the knowledge level for some issues. In addition, application of the self-regulation process (theoretical core) and sound clinical reasoning (basic professional skill) appeared to require more explicit instruction and guidance. Another participant’s advice was to more strictly monitor the translation to daily practice during the programme. Although all practices made progress, the plan for continuing the programme components as a normal part of practice would deserve greater attention and monitoring.

PQM’s indicated that quality management in this form was valuable, yet time consuming, and that is would help if they were in some way financially compensated for their efforts. This was especially required when this time investment was expected to become part of daily routine. PT’s realized that being involved in quality improvement activities would cost private time. Although they expressed commitment to do this, agreements had to be made with practice management about the required time investment. Concerning the number and duration of the sessions of the programme the PQM’s and the PTs indicated that 3 hour-sessions were acceptable. Since the programme would benefit from an extra session to include the issues mentioned above, some participants preferred to make the individual sessions one hour longer instead. Moreover, a six-month, follow-up session would allow monitoring and support maintenance of the quality improvement process. Two PQM’s indicated to prefer an on-site form of this programme, because it would reduce the time investment for the whole practice team. However, others indicated that this would reduce the opportunity for interaction and peer learning.
7. Did the participants provide suggestions on refinement of the programme?

See 6. We also adjusted the Discussion/Conclusion. Although it is not always phrased as an advice of participants, the conclusions are based on a synthesis of quantitative and qualitative findings, which include participant’s advices.

Participants judged the time investment of the programme as valuable but large. Some participants preferred an on-site intervention because they expected it to reduce their time investment. Although we considered this approach, we chose for meetings outside the practice. Our choice for meetings with more than one practice simultaneously was mainly driven by four reasons. First, on site interventions, such as educational outreach visits, have demonstrated small to modest effects {O’Brien, 2007 #925} on the change of professional performance. Second, we wanted to create the opportunity for interaction, which is identified as a factor that may increase the effect of educational meetings {Forsetlund, 2009 #731}. Third, we preferred to take the PTs out of their daily context, since their habitual working environment provides cues for habitual performance {Ouelette, 1998 #822}. By taking PTs out of their daily working environment these cues were avoided. Finally, we considered the costs of the programme, taking into account its implementation on a larger scale. Given the number of almost 5000 physical therapy practices in the Netherlands, an on-site programme would have been difficult to manage and very cost expensive.

In conclusion, although overall guideline adherence did not improve, changes on individual quality indicators suggest that a systematically developed, theory-based programme to enhance adherence to the Dutch physical therapy guidelines for low back pain has the potential to improve the quality of physical therapy care for patients with low back pain. The integrated approach of individual PTs and their PQMs, with much room for interaction, seems to benefit positive performance change. The self-regulation approach is suitable, but the programme should be re-designed in a way that it allows for a thorough self-reflection on personal performance at the beginning and at the end. Also, the steps of self-regulation should be made more explicit to enhance PT’s and PQM’s awareness of this process. Goal setting, individually and collectively, and action planning seem to be core steps in the programme, but PT’s and PQM’s need some guidance in the sound formulation of challenging yet achievable goals for improvement. In addition, more attention should be paid to translation into daily practice and to the importance of the continuity of the self regulation process. Finally, the programme should result in a thorough future practice plan for guideline implementation. In order to achieve this, the programme should allow sufficient time for attention to individual subjects and strategies and could benefit from a follow-up session to assess and support sustainability. After re-designing the programme, more sophisticated designs with larger samples are required to draw sound conclusions about its effectiveness.

8. Justify the study design in more detail given the limitations which are extensively discussed.

See also our response to recommendation 1 of reviewer 1.

We furthermore adjusted the first lines of the METHODS - Design and recruitment section: Since this pilot study served the further development of the programme, the potential effects of the QUIP programme were evaluated in a single group, pre-test post-test design. This
design is not appropriate to thoroughly assess the effectiveness of an intervention, but it is productive when pre-test data are obtained shortly before and post-test data shortly after the intervention {Polit, 2004 #558}.

9. Discuss the relevance of guideline adherence in the context of clinical outcome.

*We briefly mentioned this in the Introduction section, but we explained this a little more:*

In addition, adherence of Dutch physical therapists to these guidelines had repeatedly been shown to be limited (42%-67%) {Bekkering, 2005 #100; Swinkels, 2005 #115; Rutten, 2010 #722}, while previous studies, including a study on the Dutch guidelines for low back pain, indicated that higher adherence rates were related to better treatment results and lower utilization of care {Fritz, 2007 #427; Rutten, 2010 #722}.

10. Expand on the discussion of the feasibility of the programme in terms of training required for its delivery, and time involved by clinicians and managers in participation.

*We added the following paragraphs to the Discussion section:*

Participants judged the time investment of the programme as valuable but large. Some of them preferred an on-site intervention because they expected it to reduce their time investment. Although we have considered this approach, we chose for meetings outside the practice. Our choice for meetings with more than one practice simultaneously was mainly driven by four reasons. First, on site interventions, such as educational outreach visits, have demonstrated small to modest effects {O'Brien, 2007 #925} on the change of professional performance. Second, we wanted to create the opportunity for interaction, which is identified as a factor that may increase the effect of educational meetings {Forsetlund, 2009 #731}. Third, we preferred to take the PTs out of their daily context, since their habitual working environment provides cues for habitual performance {Ouelette, 1998 #822}. By taking PTs out of their working environment these cues were avoided. Finally, we considered the costs of the programme, taking into account its implementation on a larger scale. Given the number of almost 5000 physical therapy practices in the Netherlands, an on-site programme would have been difficult to manage and very cost expensive.

Implementation of the programme on a larger scale requires trainers with knowledge and skills in physical therapy, education and management. In order to facilitate the dissemination we involved two trainers with the required knowledge and skills in physical therapy practice, physical therapy education and management education in the development of the programme. These trainers provide nationwide additional post qualifying education. The programme has now been included in their curriculum.

Minor Essential Revision

Some missing words ie p5 first sentence insert 'a' before bridge; p8 paragraph 2,
second sentence missing a fullstop after 'physical therapy'.

*We adjusted the text as recommended.*

Discretionary Revisions
1. p10 - paragraph 3 first sentence - reduce number of abbreviations

*We assume the reviewer means the following sentence, which we have adjusted.*

Documents to be evaluated were the Personal Development Plans (n = 25) and the Practice Quality Improvement Plans (n = 7) that were written by the PTs and the PQMs as an assignment of the programme.

2. Suggest a flow chart with time line of each stage of the process

*We hope the adjustments made in the text clarify the time line sufficiently. We are not sure if the further clarification of the process by the addition of a flow chart outweighs the disadvantage of the further expansion of the paper due to the addition of an extra figure. Therefore, we prefer not to add this flow chart.*