Effectiveness of postgraduate education in occupational medicine

Smits, P.B.A.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
CHAPTER 4

Evaluation of a postgraduate educational programme for occupational physicians on work rehabilitation guidelines for patients with low-back pain

A controlled study

P.B.A. Smits1,2, J.H.A.M. Verbeek2, F.J.H. van Dijk2, J.C.M. Metz3, Th.J. ten Cate4

1 Netherlands School of Occupational Health
2 Coronel Institute, Academic Medical Centre/University of Amsterdam
3 University Medical Centre St Radboud Nijmegen
4 University Medical Centre Utrecht

Abstract

Objectives

The quality of post-graduate education for occupational physicians is an important issue in the field of occupational health. For that reason we have evaluated the effectiveness of the training programme 'guidelines for work rehabilitation of low-back pain patients'. The first question is: to what extent does knowledge of the guidelines increase. The second question is: do the participants follow the guidelines better in their work after the programme.

1 This chapter is adopted from two articles: 1. 'Evaluation of a post-graduate educational programme for occupational physicians on work rehabilitation guidelines for patients with low back pain'. Occup Environ Med 2000;57:645-6. (this article is inserted as an appendix to this thesis) and 2. 'Evaluatie van onderwijs over de concept-richtlijn 'lage rugklachten' (Evaluation of education on work rehabilitation guidelines for patients with low back pain'). TBV 2000;8:260-4.
CHAPTER 4

Methods
In a quasi-experimental design with an experimental group (n=25) and a reference group (n=20) of physicians in a post-graduate occupational health training programme knowledge was assessed with a pre-test (t1) and a post-test (t2) consisting of 45 true/false questions. The application of the guidelines in practice was evaluated with pre- and post- training recordings of performance indicators. The experimental group attended a one-and-a-half-day programme about the guidelines within a six-day module about rehabilitation. The reference group attended classes on other subjects. After six months they followed the same programme as the experimental group; after that programme the knowledge of the reference group was re-assessed (t3). All results were examined with the paired samples t-test and the independent samples t-test.

Results
Knowledge of the guidelines increased significantly in both the experimental group (t1-t2: p<.001) and the reference group (t1-t2: p<.01). The adjusted gain score is positive (independent samples t-test p<.05). The test score of the reference group increased further after the training programme (t2-t3: p<.001). The experimental group score on the performance indicators increased significantly after the training (pre- and post-recording: p<.001); the reference group’s score decreased slightly (not significant, p>.05). The adjusted gain score is positive (independent samples t-test p <.001).

Conclusions
This educational programme was effective. Knowledge and performance in practice have improved and are more in compliance with the guidelines. The knowledge test showed a possible ‘training effect’ at t2. The results and the possibilities of using performance indicators for evaluation of post-graduate education are discussed.

Keywords
Programme evaluation, post-graduate training, occupational physicians, low-back pain, guidelines
Introduction

Improvement of quality of care is regarded as a major challenge for the medical profession. The development of guidelines is seen as one way to improve quality of care. Guidelines usually aim at changing the performance of doctors who are already practising medicine, yet it is obvious that they are also important for the training and education of doctors who are new to the field. They provide a clear framework for how to act in practice, and performance indicators can be used to test whether medical performance is in compliance with the guidelines. The need for quality of care is even more important because of the increasing demand for the accountability of occupational health services. Recently an instrument for quality assessment of occupational rehabilitation for occupational physicians has been developed and evaluated. The conclusion is that it is an acceptable method which can be used for all kinds of medical audit.

In the education and training of occupational physicians there is a great need for evaluation of the effectiveness of teaching programmes. The ultimate aim of this education is to contribute to desirable outcomes such as better health, fewer diseases and less work incapacity. Evaluating whether these programmes met this aim requires longitudinal studies that are difficult to carry out. In the meantime it is worthwhile to find out if we can attain intermediate educational goals such as an increase of knowledge and improvement of performance in medical practice. To this end the same instruments can be used as in the assessment of quality of care. For this reason, part of a postgraduate training programme for occupational physicians was evaluated. The research questions were whether participating in the educational programme on occupational rehabilitation for low-back pain patients increased knowledge of the guidelines among practising physicians and whether the programme improved their practice performance. The study was carried out in 1997.

Methods

The relationship between assessment in theory and in practise can be illustrated using Miller's pyramid of clinical assessment. This pyra-
Figure 4.1. Miller's pyramid of clinical assessment

mid shows the relationship between assessment of knowledge ('knows'), competence ('knows how'), performance ('shows how') and action ('does'). We prefer to call the first three levels 'competence' and to call the fourth level 'performance' (in practice).\(^9\) (figure 4.1)

The educational programme was evaluated by using a test of knowledge and performance indicators to explore the four levels of the pyramid. The content of the educational programme discussed in this article is based on the Dutch guidelines for occupational physicians on work rehabilitation of employees with low-back pain. This educational programme takes one-and-a-half-days within a module of six days on work rehabilitation. It includes a lecture on the guidelines for occupational physicians and for Dutch general practitioners on low-back pain.\(^{10,11,12}\) Another presentation is about referral to a back pain clinic. Finally the students are given feedback on the low-back pain cases they submitted beforehand.

The results of this research project can be significant for the quality of the occupational physicians programme and, indirectly, for the quality of rehabilitation of employees with low-back pain. There have been earlier studies on the tasks of occupational physicians and on specific aspects of consultations.\(^{13,14}\) Along with the studies discussed above, the Coronel Institute investigated the effectiveness of occupational health care and work rehabilitation for employees with low-back pain.\(^{15,16}\) These investigations are combined in this project on the effectiveness of education.
An experimental group of 25 physicians took part in the educational programme, a reference group of 20 physicians did not (they took part in the programme six months later). The two groups included physicians of two Corvü year groups and were similar on variables as age and duration of work as a doctor. In the reference group there were less women, but there were no systematic differences in knowledge and performance scores between men and women. We assessed both groups with a pre-test and a post-test for knowledge. The performance indicators were scored in both groups in the months before and after the educational programme. This study design is known as 'quasi-experimental'.

The knowledge tests were executed directly before and after the training, in the class room, with one week intervals, in the second half of 1997. Both tests consisted of 45 true/false questions. The questions were based on the study material. The pre-test and post-test scores were analysed with the paired samples t-test. The reliability of the items in the questionnaire had been analysed with Cronbach’s alpha. Eight questions from the pre-test were not included in the analysis so the final pre-test included 37 questions. The pre-test alpha was 0.65. In the post-test all 45 questions could be included, with a Cronbach’s alpha of 0.60. The Pearson correlation between the pre-test and the post-test was not significant in either group (experimental group r = -.06; p = .83 and reference group r = .40; p = .084).

We used 'performance indicators' for the assessment of the compliance with the guidelines on work rehabilitation. A performance indicator is an essential part of the care process and is an indicator

**Table 4.1.** Research design including an untreated reference group and pre-tests and post-tests: a quasi-experimental design.

<table>
<thead>
<tr>
<th>Experimental group</th>
<th>t1 x t2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference group</td>
<td>t1 O t2 x t3</td>
</tr>
</tbody>
</table>

x = Intervention: educational programme
O = No intervention
t1 and t2: knowledge tests and performance scores
t3: knowledge test only
Table 4.2. Description of twelve performance indicators.

<table>
<thead>
<tr>
<th>Performance indicator (P)</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>P 0 diagnosis</td>
<td>medical diagnosis; classification in aspecific back pain, radiating back pain, specific back pain, still unclear</td>
</tr>
<tr>
<td>P 1 activating approach</td>
<td>Activating approach if diagnosis is aspecific back pain</td>
</tr>
<tr>
<td>P 2-1 psycho-social judgement</td>
<td>Request for and assessment of psycho-social problems</td>
</tr>
<tr>
<td>P 2-2 psycho-social approach</td>
<td>Actions in consulting hour if necessary</td>
</tr>
<tr>
<td>P 3-1 evaluation of medical treatment</td>
<td>Is medical treatment adequate and / or impediment for rehabilitation</td>
</tr>
<tr>
<td>P 3-2 action / consultation of medical colleagues</td>
<td>Action or consultation of medical colleagues if necessary</td>
</tr>
<tr>
<td>P 4-1 evaluation of work disabilities</td>
<td>Request for and assessment of disabilities for own current work and necessary work adjustments</td>
</tr>
<tr>
<td>P 4-2 action for work adjustment</td>
<td>Actions for (temporary) work adjustment if necessary</td>
</tr>
<tr>
<td>P 5-1 evaluation of organisational obstructions</td>
<td>Request for and assessment of organisational obstructions to work rehabilitation</td>
</tr>
<tr>
<td>P 5-2 action towards the organisation / the employer</td>
<td>Actions to remove organisational obstructions if necessary</td>
</tr>
<tr>
<td>P 6 advice on return to work</td>
<td>Advice given or revision in appointment if necessary</td>
</tr>
<tr>
<td>P 7 revision policy</td>
<td>For aspecific low-back pain: revision within three weeks</td>
</tr>
</tbody>
</table>

for the quality of medical practice. Each performance indicator contains criteria which indicate the difference between good and bad quality. We used twelve well-defined performance indicators. (table 4.2)
To assess the students’ performance we asked them to randomly collect from their own practice five cases of low-back pain patients who had been absent from work in the months before the educational programme and five cases in the months after. The students were asked to fill in a form for each of these cases. By comparing these forms with the performance indicators the researcher scored to what extent the students had followed the guidelines. One researcher assessed all the cases. The possible score for each performance indicator was: ‘good’, ‘not good’, ‘not applicable’ and ‘could not be assessed’. The score was expressed per case as a percentage of the performance indicators that met the criteria versus all the assessed performance indicators that were applicable. For feedback purposes, this was done per case. For every student we calculated a mean percentage of the performance scores over all cases before and after the training. The scores and differences were calculated with a paired samples t-test. To evaluate the possible influence of differences in performance scores at base line we also used analysis of covariance to adjust for these differences. The Pearson correlation between the performance scores before and after was not significant (exp. group r=.33; p=.18, reference group r=.33; p=.29). To evaluate the educational programme we calculated a gain score and an adjusted gain score for both the knowledge tests and the performance indicators. This adjusted gain score is defined as the gain score of the experimental group minus the gain score of the reference group. The adjusted gain scores were analysed with the independent samples t-test for groups.

Results

A pre-test/post-test comparison of the knowledge test was possible for 17 physicians in the experimental group and 20 in the reference group. The knowledge test scores at t1 were low (70% and 67%). The score of the experimental group after the training at t2 was 85%, the score of the reference group, without training, 73% at t2. (table 4.3) With true/false questions a score of about 80% is supposed to be significantly different from that which can be attained by chance alone. The gain scores were significantly positive for both the experimental and the reference group. The adjusted gain score was 9%; this is signifi-
### Table 4.3. Results of pre-test and post-test on knowledge of the guidelines.

<table>
<thead>
<tr>
<th></th>
<th>Experimental group (n=17)</th>
<th>Reference Group (n=20)</th>
<th>Difference exp. and reference groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>T=1</td>
<td>70% (11.67)</td>
<td>67% (10.42)</td>
<td>3%</td>
</tr>
<tr>
<td>T=2</td>
<td>85% (6.75)</td>
<td>73% (6.68)</td>
<td>12%</td>
</tr>
<tr>
<td>T=3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gain score t1-t2</td>
<td>15%+ (7.8 - 22.0)</td>
<td>6%O (1.9 - 11.1)</td>
<td>9%# &amp; (0.4 - 16.4)</td>
</tr>
<tr>
<td>gain score t2-t3</td>
<td>11%* (7.4 - 15.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

+ paired samples t-test, T(16)=4.45; p<.001
O paired samples t-test, T(19)=2.93; p<.01
# independent samples t-test, T(35)=2.16; p<.05
& adjusted for gain score reference group
* paired samples t-test, T(18)=6.10; p<.001
(referenced group at T=3 was n=19)

The reference group participated in the educational programme six months later. At t3 we repeated the same knowledge test of t2. The knowledge score further increased to a score that was comparable with that of the experimental group after the educational programme. Notice that the standard deviation of the scores for both the experimental and the reference group decreased at t2 and even further at t3 (table 4.3).

For the performance indicators a pre-test/post-test comparison was possible for 18 physicians in the experimental group (a total of 90 low-back pain cases before and 87 cases after) and 12 physicians in the reference group (56 cases before and 52 cases after). The gain score for the experimental group was significantly positive. For the reference group the gain score was negative, but this was not statistically significant. The adjusted gain score is 15% and this is significant (table 4.4). The standard deviation of the experimental group decreased, the standard deviation of the reference group did not (table 4.4).
Table 4.4. Pre- and post-training scores on performance indicators.

<table>
<thead>
<tr>
<th></th>
<th>Experimental group</th>
<th>Reference group</th>
<th>Difference experimental and reference groups</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n=18)</td>
<td>(n=12)</td>
<td></td>
</tr>
<tr>
<td>Before training</td>
<td>80% (8.41)</td>
<td>87% (7.24)</td>
<td>-7%</td>
</tr>
<tr>
<td>Mean % correct</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>After training</td>
<td>91% (5.79)</td>
<td>83% (7.36)</td>
<td>8%</td>
</tr>
<tr>
<td>Mean % correct</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gain score</td>
<td>11%+ (6.6-15.0)</td>
<td>-4%+ ( -9.3-1.5)</td>
<td>15%# &amp; (8.2-21.2)</td>
</tr>
<tr>
<td>(95% C.I.)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* paired samples t-test (before / after training), $T(17)=5.41$; $p<.001$
* paired samples t-test (before / after training), $T(11)=-1.60$; $p=.138$
* independent samples t-test (difference exp / reference), $T(28)=4.67$; $p<.001$
& adjusted for gain score reference group

Discussion

In this study we have shown that there was an increase in knowledge and an improvement of performance in practice after a training programme on low-back pain. These results could be biased in several ways. First, we did not use a randomised study design. For this reason the experimental and reference groups might differ in other aspects than just the training programme which we could not measure in this study. It is possible that the groups differed in learning style, which is known to influence the outcome of a training programme. Apart from this factor also other factors such as the rest of the training programme, location and time of testing differed between the two groups. However, these differences would be more likely to bias the results in the direction of not finding a difference between the two groups. The same holds for the increase in score on the knowledge test of the reference group without a training programme, which is probably due to taking the test twice. Therefore we used adjusted gain scores which take this effect into account. Even then we found a significant difference between the experimental and the reference groups. Moreover, for the performance indicators there was no such increase for the reference group. Therefore we think that the increases in knowledge and performance can be ascribed to the training programme.
Secondly, the validity of our measurement instruments can be questioned. Therefore we measured the internal consistency as a measure of the validity of our knowledge test. To increase the internal consistency we left out a couple of questions that did not seem to be related to the test as a whole. The final version of the test yielded a Cronbach's alpha of 0.65 pre-test and 0.60 post-test which is considered to be sufficient for programme evaluation purposes.\textsuperscript{21,22} The performance indicators were partly based on a physician's report of what they had themselves done in a particular case. The criteria we used to assess deviation from the guidelines for a specific performance indicator were not known explicitly by the participants. Therefore we think that the performance score based on these indicators does reflect real performance in practice and not just competence. Moreover it was shown that performing well as measured by the same indicators predicted a better outcome and greater patient satisfaction.\textsuperscript{23} However, these indicators can only reflect a part of total performance because it is impossible to cover all aspects by using one method only. Our method covers the more cognitive aspects but not the doctor-patient interaction which can better be studied by using video recordings.\textsuperscript{24} To overcome these difficulties of measurement it would be better to relate the training programme to outcome in practice. For vocational rehabilitation of low-back pain patients return to work should be measured and compared between patients treated by physicians from the reference and the intervention groups. For future research we advocate a randomised design and a more extended measurement of outcome.
References


CHAPTER 4


