The modified low back pain disability questionnaire

Reference:
Denteneer Lenie, van Daele Ulrike, Truijen Steven, De Hertogh Willem, Meirte Jill, Deckers Kristiaan, Stassijns Gaëtane.- The modified low back pain disability questionnaire
Spine - ISSN 0362-2436 - 43:5(2018), p. E292-E298
Full text (Publisher's DOI): https://doi.org/10.1097/BRS.0000000000002304

SPINE An International Journal for the study of the spine Publish Ahead of Print

DOI: 10.1097/BRS.0000000000002304

The Modified Low Back Pain Disability Questionnaire: reliability, validity and responsiveness of a Dutch language version

LenieDenteneer, Dra, MT, PT[1], GaetaneStassijns, MD, PhD[2], Steven Truijen, MSc, PhD[2], Willem De Herthogh, MT, PT, PhD[2], Jill Meirte, PT, PhD[2], KristiaanDeckers, MD[3], Ulrike Van Daele, MT, PT, PhD[2]

[1]Faculty of Medicine and Health Sciences, Rehabilitation and Physiotherapy, University of Antwerp

[2]Faculty of Medicine and Health Sciences, University of Antwerp

[3]GZA Hospitals Sint Augustinus

Corresponding Author:

LenieDenteneer, Dra, MT, PT

Faculty of Medicine and Health Sciences, rehabilitation and physiotherapy

University of Antwerp, Universiteitsplein 1, 2610 Wilrijk

Tel: 0032494884189

lenie.denteneer@uantwerpen.be

The manuscript submitted does not contain information about medical device(s)/drug(s).

No funds were received in support of this work.

No relevant financial activities outside the submitted work.

abstract

Study design: Cross sectional study

Objective: The goal of this study is totranslate the English version of the Modified Low Back Pain Disability Questionnaire (MDQ) into a Dutch version and investigate its clinimetric properties for patients with nonspecific chronic low back pain (CLBP).

Summary of Background Data: Fritz et al (2001) developed a modified version of the Oswestry Disability Questionnaire (ODI) to assess functional status and named it the MDQ. In this version, a question regarding employment and homemaking ability was substituted for the question related to sex life. Good clinimetric properties for the MDQ were identified but up until now it is not clear if the clinimetric properties of the MDQ would change if it was translated into a Dutch version.

Methods: The translation of the MDQ into Dutch was done in four steps. Test-retest reliability was investigated using the intraclass correlation coefficient (ICC) model. Validity was calculated using Pearson correlations and a 2-way analysis of variance (ANOVA) for repeated measures. Finally, responsiveness was calculated with the area under the curve (AUC), minimal detectable change (MDC) and the standardized response mean (SRM).

Results: A total of 80 completed questionnaires were collected in three different hospitals and a total of 43 patients finished a 9 weeks intervention period, completing the retest. Test-retest reliability was excellent with an ICC of 0.89 (95%CI, 0.74-0.95). To confirm the convergent validity, the MDQ answered all predefined hypothesises (r=-0.65–0.69 / p=0.01–0.00) and good results for construct validity were found (p=0.02). The MDQ had an AUC of 0.64 (95%CI, 0.47-0.81), an MDC of 8.80 points and a SRM of 0.65.

Conclusion: The Dutch version of the MDQ shows good clinimetric properties and is shown to be usable in the assessment of the functional status of Dutch speaking patients with nonspecific CLBP.

Key Words: Chronic nonspecific low back pain, Modified Low Back Pain Disability Questionnaire, Dutch, clinimetric properties

Level of Evidence:3

Introduction

Patients with low back pain (LBP) suffer from a wide range of problems. When patients are asked to report their treatment goals in advance of their rehabilitation, 96% can be classified withinthe international classification of functioning, disability, and health (ICF) category "activities and participation" with the most reported goal being "doing housework"[1]. The assessment in activity limitations and participation restrictions in patients with LBP is often performed with the use of self-report outcome measures. For the assessment of the functional status of patients with LBP, the Oswestry Disability Index (ODI) is extensively used both in the clinic and in randomized controlled trials[2-7].

The ODI is a disease-specific measure of disability with its main focus being the ICF category "activities and participation". The English and Dutch versions of the ODI are found to be reliable, valid and show medium to high levels of responsiveness[2,3,8-11].However, one question namely question number eight about sex life disability shows a poorer compliance[2] and is frequently found to be left blank[12-14]. For example, Mousavi et al reported that out of a patient sample of 100, a total of 19 patients failed to fill in the sex life question[14]. In order to minimize missing data issues in studies, Fritz et al[12] developed a modified version of the ODI and named it the Modified Low Back Pain Disability Questionnaire (MDQ).In this version, a question regarding employment and homemaking ability (which is the most reported goal for rehabilitation by patients with LBP[1])was substituted for the question related to sex life. The English version of the MDQ shows good reliability, responsiveness and validity in patients with acute LBP[12]. Following these good clinimetric properties and advantages, the literature shows an increased use of the MDQ [5,13,15-22].

To our knowledge there are no reported clinimetric properties for a Dutch version of the MDQ.Up until now it is not clear if the clinimetric properties of the MDQ would change if it

was translated into a Dutch version.The objective of this study is to translate the MDQ into Dutch and investigate its test-retest reliability, validity and responsiveness in a patient sample with nonspecific chronic LBP (CLBP).

Materials and methods

Modified Low Back Pain Disability Questionnaire

The MDQ consists of 10 items addressing different aspects of function: pain severity, lifting, sitting, standing, walking, sleeping, personal hygiene, social life, traveling and finally employment/homemaking which replaces the sex life question reported in the ODI. Each item is scored from 0 to 5 where higher values represent greater disability. The total score (ranging from 0 to 50) is multiplied by two and expressed as a percentage.

Patients and procedure

The patient population in this study are patients with nonspecific CLBP. In-and exclusion criteria are shown in Table 1 and 2. All patients filled in an informed consent and ethical approval (B300201215600) was obtained from the local ethics committees of the University of Antwerp.

Patients were included in 3 different hospitals in Belgium and were asked to fill in following questionnaires: MDQ[12]; Visual Analogue Scale (VAS) for pain[23]; Roland Morris Disability Questionnaire (RMDQ)[24] and the 36-Item Sort Form Survey (SF-36)[25].

Patients who chose to enter a back rehabilitation program in either one of the three recruiting hospitals were asked to fill in the same series of questionnaires after an intervention period of 9 weeks. Additionally, a 7-level Global Perceived Effect scale (GPE)[26] was included at this time point. The GPE scale had 7 response options: 1= “worse than ever”, 2= “much worse”, 3= “a little worse”, 4= “about the same”, 5= “a little better”, 6= “much better”, 7=

"completely gone". The content of the intervention for these patients was exercise therapy since this can be considered as the best evidence for the rehabilitation of patients with nonspecific CLBP[27].

Translation

For the translation of a Dutch version of the MDQ, questions 1-7 and 9-10 of the already existing original Dutch version of the ODI were copied[11]. The question regarding employment/homemaking in the English version of the MDQ reported by Fritz et al[12] was translated according to 4 prescribed steps[28] which are shown in Figure 1.

Statistical analysis

Reliability

Test-retest reliability of the MDQ was investigated using the intraclass correlation coefficient model 2,1 (ICC(2,1))[29] in participants who remained "unchanged" in their LBP status between the initial and the follow-up surveys based upon the GPE. According to Davidson et al[30], we classified participants who self-reported their condition as "about the same" or "a little worse" or "a little better" as "unchanged" (GPE scores 3,4 or 5). Whatever the type of ICC that is calculated, an ICC close to 1 indicates excellent reliability. An ICC >.70 indicates good reliability, and an ICC <.70 indicates moderate to poor reliability[29].

Validity

For construct validity, we compared the changes in MDQ scores between patient groups defined as "unchanged" (GPE score 3,4 or 5) or "improved" (GPE score 6 or 7) using a 2-way analysis of variance (ANOVA) for repeated measures on the MDQ scores initially and after 9 weeks of follow-up. We hypothesized that the "improved" group would show a

progressive decrease in MDQ score at follow-up, whereas the MDQ of the "unchanged" group would not change which would be indicated by a group X time interaction[12].

Convergent validity was assessed by evaluating the correlation between the MDQ and a questionnaire that is thought to measure a similar construct[31]. In this study, Pearson correlations (chosen because of the continuous nature of the data) were calculated between the MDQ, the RMDQ and SF-36 questionnaire. Correlation values of ≤ 0.35 are generally considered to represent low correlations, 0.36 to 0.67 modest or moderate correlations, 0.68 to 0.89 strong or high and 0.90 very high correlations[32].

Responsiveness

For the anchor-based method[33], responsiveness was firstly evaluated using a receiver operating characteristic (ROC) curve. The area underthe ROC curve (AUC) reflects the ability of the test to discriminate between subjects who have "improved" from subjects who are "unchanged" based upon the GPE. A value of 1 for the AUC represents perfect (100%) accuracy, whereas a value of 0.5 represents chance alone[34]. Secondly, the minimal detectable change (MDC) which is based on the "unchanged" participants was calculated. The MDC is calculated as $1.96 \times \sqrt{2} \times SEM$[35] and the standard error of measurement (SEM) was calculated as $(sd \times [1-r]^{1/2})$, where r is the test-retest reliability coefficient and sd is the square root of the total variance[12]. The MDC can be interpreted as the magnitude of change below which there is more than a 95% chance that no real change has occurred[36].

For the distribution-based method, the standardized response mean (SRM) was calculated by dividing the mean change by the standard deviation of changed scores[37]. Values of >0.80 are large, 0.50- 0.80 moderate, and < 0.50 small [38].

Results

A total of 80 patients participated in this study. Of this sample, 70 patients chose to enter a back rehabilitation program in either one of the three recruiting hospitals. A total of 16 patients dropped out of the back rehabilitation program and a total of 11 patients did not yet complete the back rehabilitation program at the time of finishing the inclusion period for this study. Finally, a total of 43 patients were re-evaluated after an intervention period of 9 weeks.

Of the 43 patients that were re-evaluated after 9 weeks, a total of 18 patients were identified as "improved" and 23 were identified as "unchanged". Two patients reported a deterioration of their LBP after finishing the back rehabilitation program. All demographic values are presented in Table 3.

Test retest reliability

Test-retest reliability of the MDQ was excellent with an ICC of 0.89 (95% CI, 0.74-0.95).

Validity

Figure 2 shows the change in MDQ scores for the participants in the "unchanged" and "improved" groups after 9 weeks of intervention. Mean MDQ scores for the total rehabilitation group, the "improved" group and the "unchanged" group at baseline and after 9 weeks evaluation are shown in Figure 3. The ANOVA analysis showed a significant interaction between both groups and the time ($p=0.02$).

The Dutch version of the MDQ showed a high correlation with the RMDQ ($r=0.69$, $p=0.00$). Also, there were moderately strong, negative correlations with the Physical Functioning and Bodily Pain domains of the SF-36, as well as weak negative correlations with the mental health and role-emotional domains of the SF-36 (Table 4).

Responsiveness

Figure 4 shows the ROC curve constructed from the changed scores for the MDQ. The AUC was 0.64 (standard error 0.09 and 95% CI, 0.47-0.81). The SEM value for the MDQ was 3.19 and based on this SEM value, the threshold for the MDC was calculated as being 8.80. Finally, the SRM was calculated as being 0.65.

Discussion

The goal of this study was to translate the English version of the MDQ into a Dutch version and investigate its clinimetric properties in a patient population with nonspecific CLBP.

Test-retest reliability

The test-retest reliability calculated in this study (ICC=0.89) was consistent with reliability coefficients found in other studies that investigated the English version of the MDQ. Hicks et al[13] reported an ICC of 0.92 in older adults with subacute and CLBP after one week follow up and Fritz et al[12] reported a ICC of 0.90 in acute LBP with 4 weeks follow up. Other studies report that a shorter follow up period may include a memory effect which may represents higher reliability[39,40]. In the current study, patients were retested after an intervention period of 9 weeks. This means that the ICC value of 0.89 in this study can be interpreted as a very good result.

validity

According to Terwee et al[41], several specific a priori hypotheses were formulated to confirm convergent validity of the MDQ in order to avoid a possible risk of bias in explanations for the associations found. Firstly, we hypothesized high positive correlations for the MDQ and RMDQ since they measure a similar construct. Secondly, we followed the predefined hypothesis suggested by Hicks et al[13] which indicate high negative correlations for the MDQ

with the Bodily Pain and Physical Function subitems of the SF-36 questionnaire. Finally, Hicks et al[13] suggest small negative correlations with the Mental Health and Role Limitations subitems of the SF-36 questionnaire since mental and emotional problems do not represent the construct of LBP-related disability. All predefined hypotheses for the SF-36 subscales were confirmed and a strong correlation for the MDQ with the RMDQ ($r=0.69$, $p<0.00$) was identified. This correlation is comparable with the original version of the ODI ($r=0.61$-0.84)[11].

To determine construct validity, we compared the changes in MDQ scores between patient groups defined as "unchanged" or "improved" based on the GPE using a 2-way ANOVA for repeated measures on the MDQ scores initially and after 9 weeks of follow-up. As hypothesized, the "improved" group showed a progressive decrease in physical impairment, whereas the impairment level of the "unchanged" group did not change (Figure 2). This finding was indicated by a significant group X time interaction and confirms a good construct validity for the MDQ. The construct validity of the English version of the MDQ was assessed by Fritz et al[12] and they found a comparable result.

Responsiveness

The AUC in this study (0.64) is slightly lower than the AUC value reported by Fritz et al[12]. They reported an AUC for the MDQ of 0.94 in patients with LBP less than 3 weeks. The European Guidelines for nonspecific CLBP[27] state that rapid improvements in functional status occur within the first month after an initial episode of LBP. After 3 months, improvement remains almost constant. This is why the detection of a clinically meaningful change will be easier in a population with acute LBP compared to a population with chronic LBP which was the target population in this study. Secondly, the 9 weeks follow up time in this study is clearly longer than the follow up time described in Fritz et al[12] who reported 4

weeks follow up. Due to recall bias, it is easier to find higher responsiveness values when a questionnaire is re-administered in a shorter timeframe. The follow up period of 9 weeks in this study was chosen to reflect a typical clinical retest period since 9 weeks is a commonly used intervention period in the clinic for comprehensive reassessment of patients with LBP[30,37]. Therefore, the slightly lower AUC value in this study can be explained by both the combination of the inclusion of a chronic LBP population and the longer follow up time.

The second anchor based method, namely the MDC is defined as the minimal variation of symptoms that is meaningful for patients and is not to be confused with Minimally Clinical Important Difference which refers to differences between patients[36]. In this study, a MDC of 8.80 points was calculated for the MDQ. Fritz et al[12] reported a MDC of 12.68 in an adult population with acute LBP and Hicks et al[13] reported a MDC of 10.66 points in an older adult population with mainly chronic LBP. In this study, a total of 41% patients rated their condition as "improved" based on the GPE scale. Based on the MDC, the MDQ was able to identify a total of 44% with a meaningful improvement in their health status. This illustrates that the MDQ was able to adequately detect a meaningful variation of symptoms in LBP over a follow up period of 9 weeks.

Finally, for the distribution-based method, the SRM of the Dutch version of the MDQ shows an identified value of 0.65 which can be considered to be moderate responsive[38]. However, Davidson et al[30] described a slightly lower SRM of 0.52. The current study calculated a higher SRM which reflects a higher responsiveness. The moderate outcome for the SRM can be again explained by the chronic nature of the patients with LBP in this study. The biggest recovery from LBP is made in the first six weeks and therefore, patients with acute LBP will show a higher change in scores on the MDQ within the first 4 weeks compared to patients with chronic LBP. For these patients, the change in MDQ scores between two assessments is

lower compared to patients with acute LBP and therefore their SRM values will also be lower.

Study limitations

To translate a questionnaire, the guidelines for the process of cross-cultural adaptation of self-report measures which were reported by Beaton et al[42] suggest the use of an informed and uniformed forward translator to avoid information bias and to elicit unexpected meanings of the items in the translated questionnaire. In the current study, the forward translation of the English version of the MDQ into a Dutch version was completed by two physical therapists who both have a M.Sc. degree in manual therapy. Both were aware of the concepts that were being measured by the MDQ. If an uninformed translator is implemented, he or she is more likely to detect different meaning of the original questionnaire than the informed translator. This translator could be less influenced by an academic goal and could offer a translation that reflects the language used by that population, often highlighting ambiguous meanings in the original questionnaire. It is important to mention that this could have had a possible bias on the translation of the MDQ.

In this study there was a drop-out ratio of 27% which could have affected the study representativeness. However, if we compare this percentage to some other studies that have investigated patients with CLBP, drop-out ratios ranging from 14% to 33% can be noted [21,43,44]. This means that the drop-out ratio within this study can be considered as representative. A possible reason for these relative high percentages is that patients with chronic complaints might be more likely to lose their internal motivation if no immediate results are experienced when starting the treatment.

Conclusion

The results of this study show an excellent test-retest reliability. To confirm the convergent validity, the MDQ answered all predefined hypotheses and good results for construct validity were found. The MDQ has an AUC of 0.64, a MDC of 8.80 points and a SRM of 0.65.

In conclusion, the Dutch version of the MDQ shows good clinimetric properties and is shown to be usable in the assessment of the functional status of Dutch speaking chronic nonspecific LBP patients.

References:

1. Bagraith KS, Hayes J, Strong J. Mapping patient goals to the International Classification of Functioning, Disability and Health (ICF): examining the content validity of the low back pain core sets. *Journal of rehabilitation medicine* 2013;45:481-7.
2. Beurskens AJ, de Vet HC, Koke AJ, et al. Measuring the functional status of patients with low back pain. Assessment of the quality of four disease-specific questionnaires. *Spine* 1995;20:1017-28.
3. Chapman JR, Norvell DC, Hermsmeyer JT, et al. Evaluating common outcomes for measuring treatment success for chronic low back pain. *Spine* 2011;36:S54-68.
4. Deyo RA, Battie M, Beurskens AJ, et al. Outcome measures for low back pain research. A proposal for standardized use. *Spine* 1998;23:2003-13.
5. Brennan GP, Fritz JM, Hunter SJ, et al. Identifying subgroups of patients with acute/subacute "nonspecific" low back pain: results of a randomized clinical trial. *Spine* 2006;31:623-31.
6. Goldby LJ, Moore AP, Doust J, et al. A randomized controlled trial investigating the efficiency of musculoskeletal physiotherapy on chronic low back disorder. *Spine* 2006;31:1083-93.
7. Rossignol M, Abenhaim L, Seguin P, et al. Coordination of primary health care for back pain. A randomized controlled trial. *Spine* 2000;25:251-8; discussion 8-9.
8. Fairbank JC, Pynsent PB. The Oswestry Disability Index. *Spine* 2000;25:2940-52; discussion 52.
9. Frost H, Lamb SE, Stewart-Brown S. Responsiveness of a patient specific outcome measure compared with the Oswestry Disability Index v2.1 and Roland and Morris Disability Questionnaire for patients with subacute and chronic low back pain. *Spine* 2008;33:2450-7; discussion 8.
10. Vianin M. Psychometric properties and clinical usefulness of the Oswestry Disability Index. *Journal of chiropractic medicine* 2008;7:161-3.
11. van Hooff ML, Spruit M, Fairbank JC, et al. The Oswestry Disability Index (version 2.1a): validation of a Dutch language version. *Spine* 2015;40:E83-90.
12. Fritz JM, Irrgang JJ. A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. *Physical therapy* 2001;81:776-88.
13. Hicks GE, Manal TJ. Psychometric properties of commonly used low back disability questionnaires: are they useful for older adults with low back pain? *Pain medicine (Malden, Mass.)* 2009;10:85-94.
14. Mousavi SJ, Parnianpour M, Mehdian H, et al. The Oswestry Disability Index, the Roland-Morris Disability Questionnaire, and the Quebec Back Pain Disability Scale: translation and validation studies of the Iranian versions. *Spine* 2006;31:E454-9.
15. Brandt Y, Currier L, Plante TW, et al. A Randomized Controlled Trial of Core Strengthening Exercises in Helicopter Crewmembers with Low Back Pain. *Aerospace medicine and human performance* 2015;86:889-94.
16. Hicks GE, Fritz JM, Delitto A, et al. Preliminary development of a clinical prediction rule for determining which patients with low back pain will respond to a stabilization exercise program. *Archives of physical medicine and rehabilitation* 2005;86:1753-62.
17. Childs JD, Fritz JM, Flynn TW, et al. A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: a validation study. *Annals of internal medicine* 2004;141:920-8.
18. Fritz JM, Delitto A, Erhard RE. Comparison of classification-based physical therapy with therapy based on clinical practice guidelines for patients with acute low back pain: a randomized clinical trial. *Spine* 2003;28:1363-71; discussion 72.

19. Hsieh LL, Kuo CH, Lee LH, et al. Treatment of low back pain by acupressure and physical therapy: randomised controlled trial. *BMJ (Clinical research ed.)* 2006;332:696-700.
20. Browder DA, Childs JD, Cleland JA, et al. Effectiveness of an extension-oriented treatment approach in a subgroup of subjects with low back pain: a randomized clinical trial. *Physical therapy* 2007;87:1608-18; discussion 577-9.
21. Rabin A, Shashua A, Pizem K, et al. A clinical prediction rule to identify patients with low back pain who are likely to experience short-term success following lumbar stabilization exercises: a randomized controlled validation study. *The Journal of orthopaedic and sports physical therapy* 2014;44:6-b13.
22. Denteneer L, Stassijns G, De Hertogh W, et al. Derivation and validation phase for the development of clinical prediction rules for rehabilitation in chronic nonspecific low back pain patients: study protocol for a randomized controlled trial. *Trials* 2015;16:4.
23. Boonstra AM, Schiphorst Preuper HR, Reneman MF, et al. Reliability and validity of the visual analogue scale for disability in patients with chronic musculoskeletal pain. *International journal of rehabilitation research. Internationale Zeitschrift fur Rehabilitationsforschung. Revue internationale de recherches de readaptation* 2008;31:165-9.
24. Brouwer S, Kuijer W, Dijkstra PU, et al. Reliability and stability of the Roland Morris Disability Questionnaire: intra class correlation and limits of agreement. *Disability and rehabilitation* 2004;26:162-5.
25. Aaronson NK, Muller M, Cohen PD, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. *Journal of clinical epidemiology* 1998;51:1055-68.
26. Kamper SJ, Ostelo RW, Knol DL, et al. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *Journal of clinical epidemiology* 2010;63:760-6.e1.
27. Airaksinen O, Brox JI, Cedraschi C, et al. Chapter 4. European guidelines for the management of chronic nonspecific low back pain. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society* 2006;15 Suppl 2:S192-300.
28. Hoben M, Bar M, Mahler C, et al. Linguistic validation of the Alberta Context Tool and two measures of research use, for German residential long term care. *BMC research notes* 2014;7:67.
29. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin* 1979;86:420-8.
30. Davidson M, Keating JL. A comparison of five low back disability questionnaires: reliability and responsiveness. *Physical therapy* 2002;82:8-24.
31. Smith BH, Penny KI, Purves AM, et al. The Chronic Pain Grade questionnaire: validation and reliability in postal research. *Pain* 1997;71:141-7.
32. Taylor R. Interpretation of the Correlation Coefficient: A Basic Review. *Journal of Diagnostic Medical Sonography* 1990;6:35-9.
33. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29-36.
34. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *Journal of chronic diseases* 1986;39:897-906.
35. Ostelo RW, de Vet HC, Knol DL, et al. 24-item Roland-Morris Disability Questionnaire was preferred out of six functional status questionnaires for post-lumbar disc surgery. *Journal of clinical epidemiology* 2004;57:268-76.

36. Kovacs FM, Abraira V, Royuela A, et al. Minimal clinically important change for pain intensity and disability in patients with nonspecific low back pain. *Spine* 2007;32:2915-20.
37. Stratford PW, Binkley JM, Riddle DL. Health status measures: strategies and analytic methods for assessing change scores. *Physical therapy* 1996;76:1109-23.
38. Sivan M. Interpreting effect size to estimate responsiveness of outcome measures. *Stroke; a journal of cerebral circulation* 2009;40:e709; author reply e10-1.
39. Kopec JA, Esdaile JM, Abrahamowicz M, et al. The Quebec Back Pain Disability Scale: conceptualization and development. *Journal of clinical epidemiology* 1996;49:151-61.
40. Gronblad M, Hupli M, Wennerstrand P, et al. Intercorrelation and test-retest reliability of the Pain Disability Index (PDI) and the Oswestry Disability Questionnaire (ODQ) and their correlation with pain intensity in low back pain patients. *The Clinical journal of pain* 1993;9:189-95.
41. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology* 2007;60:34-42.
42. Beaton DE, Bombardier C, Guillemin F, et al. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000;25:3186-91.
43. Rasmussen-Barr E, Ang B, Arvidsson I, et al. Graded exercise for recurrent low-back pain: a randomized, controlled trial with 6-, 12-, and 36-month follow-ups. *Spine* 2009;34:221-8.
44. Cairns MC, Foster NE, Wright C. Randomized controlled trial of specific spinal stabilization exercises and conventional physiotherapy for recurrent low back pain. *Spine* 2006;31:E670-81.

Figure 1: Steps for the translation of the English MDQ into a Dutch version

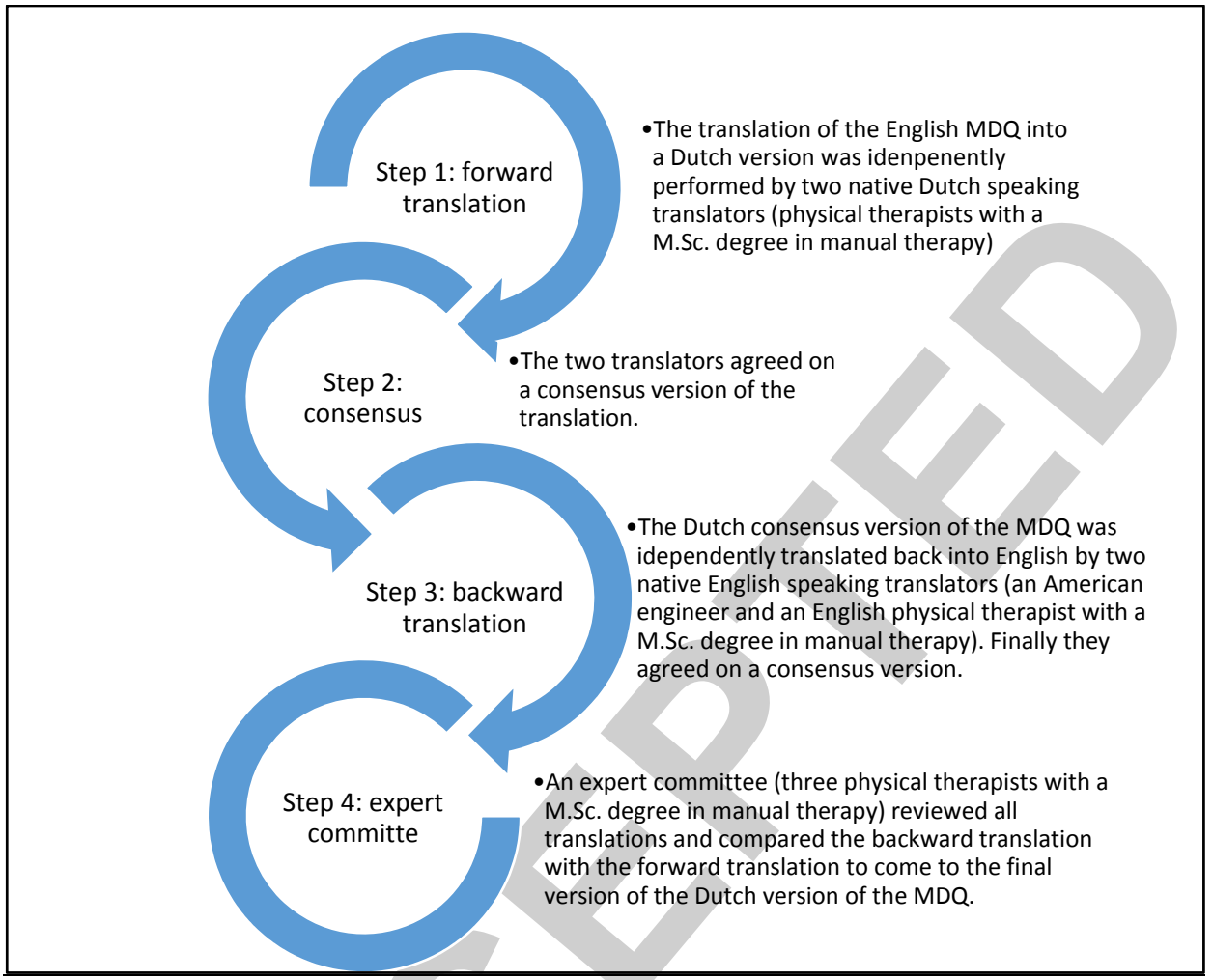

Figure 2 : Graph of the MDQ scores for the groups defined as“improved” or “stable” based on the Global Perceived Effect rating. The interaction between time and group was significant (p=0.02)

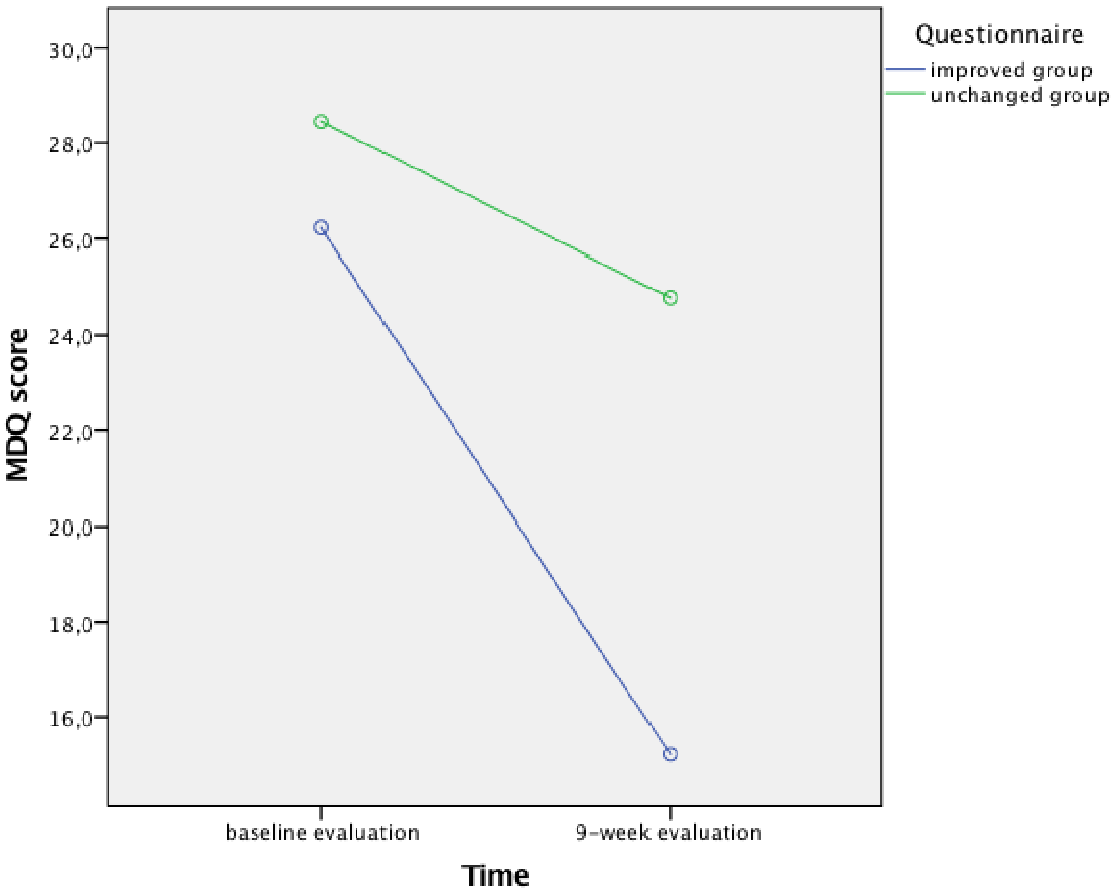

Figure 3: Mean MDQ scores for the “total rehabilitation” group, the “improved” group and the “unchanged” group at baseline and after 9 weeks evaluation.

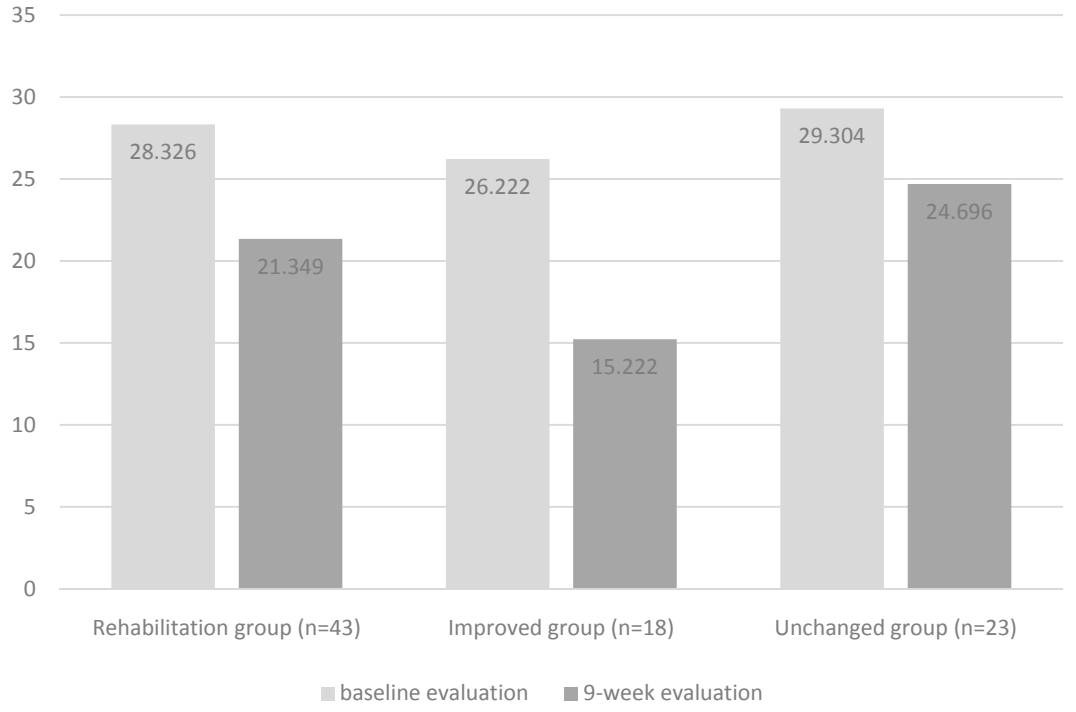

Figure 4: Receiver operating characteristic curve forthe MDQ.

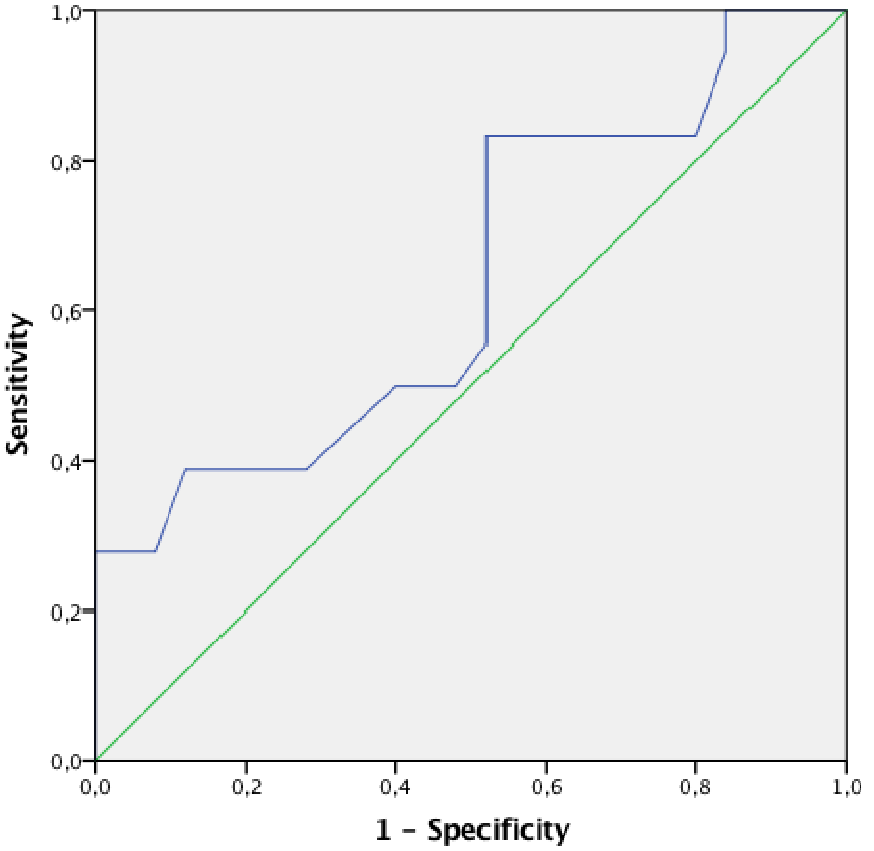

Table 1: Inclusion criteria

Inclusion criteria	**Rationale**
18 to 65 years old	Chronic low back pain in older adults is more likely to have specific causes (e.g., spinal canal stenosis)
Current nonspecific low back pain persisting ≥ 3 months	Condition studied is specifically chronic
Dutch fluency sufficient to follow treatment instructions and answer survey questions	Fully informed consent and data collection

Table 2: Exclusion criteria

Exclusion criteria	**Rationale**
Spinal canal stenosis Spondylolisthesis spondylitis Large herniated disc sciatica radiating pain below the knee Previous back surgery History of vertebral fracture malignancy muscle-, nerve-, skin- or joint diseases	Back pain possibly due to, specific disease
Known pregnancy	Pregnancy-related low back pain is different in etiology and time course than the target condition for the study (nonspecific chronic low back pain)
Lack of consent	Research policy

Table 3: Baseline characteristics of "total sample", " total rehabilitation", "unchanged" and "improved" groups

	Total sample (n=80)	Total rehabilitation group (n=43)	Unchanged group (n=23)	Improved group (n=18)
Age (years)	37.8 (± 12.7)	41.3 (± 12.4)	43.4 (± 11.9)	39.1 (± 12.8)
Gender (% men)	46.3	62.8	52.2	77.8
VAS now (0-10)	3.7 (± 2.8)	3.6 (± 2.6)	3.5 (± 2.6)	3.5 (± 2.8)
Vas min (0-10)	2.2 (± 2.3)	2.2 (± 2.2)	2.8 (± 2.3)	1.6 (± 2.0)
Vas max (0-10)	7.2 (± 2.1)	7.3 (± 2.0)	6.9 (± 2.1)	7.8 (± 1.9)
MDQ (0-100)	27.2 (± 15.0)	28.3 (± 13.6)	29.3 (± 14.9)	26.2 (± 12.5)
RMDQ (0-24)	8.7 (± 4.9)	8.5 (± 4.4)	9.0 (± 4.7)	7.4 (± 3.8)
TAMPA (17-68)	36.5 (± 8.5)	36.2 (± 9.0)	38.6 (± 9.1)	32.7 (± 7.1)
SF 36 Vitality (0-100)	62.6 (± 23.2)	65.0 (± 25.3)	65.5 (± 25.8)	65.3 (± 21.9)
SF36 Physical Functioning (0-100)	65.3 (± 20.7)	63.3 (± 19.4)	62.6 (± 20.6)	66.7 (± 17.1)
SF36 Bodily pain (0-100)	41.7 (± 17.5)	40.9 (± 15.2)	42.0 (± 13.9)	42.3 (± 15.8)
SF36 General Health Perceptions (0-100)	64.2 (± 22.9)	64.9 (± 24.0)	62.9 (± 24.6)	68.3 (± 21.3)
SF36 Physical Role Functioning (0-100)	13.9 (± 20.2)	17.9 (± 25.2)	16.0 (± 16.9)	22.2 (± 33.7)
SF36 Emotional role functioning (0-100)	25.2 (± 28.8)	29.8 (± 33.0)	27.2 (± 30.7)	35.2 (± 37.2)
SF36 Social Role Functioning (0-100)	72.2 (± 24.5)	73.3 (± 22.4)	72.3 (± 24.1)	74.3 (± 20.3)
SF36 Mental Health (0-100)	78.8 (± 21.6)	80.8 (± 21.7)	80.2 (± 20.8)	81.1 (± 22.8)

Table 4: Pearson correlations between the MDQ and the RMDQ/SF36 domains

	MDQ
RMDQ	0.69 (p= 0.00)
SF 36 Vitality	-0.34 (p= 0.00)
SF36 Physical Functioning	-0.59 (p= 0.00)
SF36 Bodily pain	-0.65 (p= 0.00)
SF36 General Health Perceptions	-0.38 (p= 0.00)
SF36 Physical Role Functioning	-0.29 (p= 0.01)
SF36 Emotional role functioning	-0.32 (p= 0.00)
SF36 Social Role Functioning	-0.56 (p= 0.00)
SF36 Mental Health	-0.46 (p= 0.00)