

**Reliability and validity of the Disability Assessment  
Structured Interview (DASI): a tool for assessing functional  
limitations in claimants**

*Jerry Spanjer  
Boudien Krol  
Sandra Brouwer  
Roel Popping  
Johan W. Groothoff  
Jac J.L. van der Klink*

## **ABSTRACT**

### **Objective**

The aim of this study is to investigate the reliability and validity of the Disability Assessment Structured Interview (DASI). The DASI is a semi-structured interview for assessing long-term functional limitations concerning the work disability assessment of claimants.

### **Methods**

A randomized controlled trial was conducted. Patients applying for a work-disability pension after 21 months of sick leave were independently interviewed and examined either by two physicians who had completed a DASI training period (n=32) or by two physicians from a control group (n=30) without any DASI training. Agreement percentages within both groups of physicians, eligibility for a disability benefit, and differences between the groups in terms of the scores given on the work-limitation items from the Functional Ability List (FAL) were measured to investigate reliability and concurrent validity.

To determine the content validity, the insurance physicians who completed DASI training (n=8) were asked to fill out a questionnaire concerning their opinion of the DASI. Additionally, patients filled out a questionnaire to measure their satisfaction as to the behavioral aspects of the physicians.

### **Results**

The groups showed no important differences in agreement percentages (mean percentage about 80%) and eligibility for a disability benefit. In 9 out of 21 items the physicians of the control group indicated fewer work limitations compared to physicians using the DASI. All physicians agreed on the fact that the DASI was an acceptable tool in daily practice, one that provided a realistic picture of the patient and provided sufficient information to assess functional limitations. In addition, between the two groups, no differences were found as to the satisfaction of patients concerning the behavioral aspects of the physicians.

### **Conclusion**

The DASI is a tool with a reasonable to good inter-rater reliability and content validity, and it appears to be acceptable to both patients and physicians. It did not improve inter-observer agreement beyond that of usual interview procedures used in the Netherlands. The DASI would seem to be a worthwhile tool for collecting self-reported information in order to assess functional limitations in claimants.

## INTRODUCTION

The determination of a work disability pension for patients with long-term medical impairments is of great social and financial importance. Part of the evaluation of a work disability pension is the assessment of the functional limitations of the patient. In the literature, several instruments and their psychometric properties for assessing functional limitations are described, for instance, self-report questionnaires and performance-based functional testing<sup>1-4</sup>. In most countries the actual assessment of functional limitations is carried out by a medical doctor<sup>5,6</sup>. The assessment by physicians can be based on written information (e.g., from the patient or treating physician) or can be conducted by an examination in person. In international literature a poor agreement among physicians on functional disability exists<sup>7-11</sup>. A tremendous variation in disability rating recommended by physicians given the same set of facts was found<sup>7</sup>. To decrease this variation the United States Social Security Administration (SSA) planned to “develop functional assessment instruments that are standardized, accurately measure an individual’s functional abilities and that are universally accepted by the public, the advocacy community, and health-care professionals”<sup>7,12,13</sup>.

In the Netherlands, an employer has to pay wages for two years if an employee is unable to work due to physical or mental disability. After these two years, the patient can apply for a work disability pension. Specialized insurance physicians assess the patient’s functional limitations in work as part of the application for a work disability pension. Their judgment is based on information from treating physicians, along with their own observations, physical examination and an interview with the patient. To a large extent, the assessment is based on the interview in which attention is given to activity limitations and participation, in addition to standard medical history-taking<sup>14</sup>. The assessed functional limitations are registered in a standardized list, the Functional Ability List (FAL)<sup>15</sup>.

Although in the Netherlands three semi-structured interview models for the assessment of functional limitations are available, in daily practice insurance physicians do not use a fixed model with claimants applying for disability pension<sup>14,16</sup>. One of the models available is the Disability Assessment Structured Interview (DASI)<sup>17</sup>. This is a semi-structured interview in which the three levels of functioning – impairment, activity limitation and participation – are mapped in a structured way in accordance with the International Classification of Functioning, Disability and Health (ICF)<sup>18</sup>. In Box 1 the general domains covered in the DASI are described.

Two important characteristics of the DASI are its semi-structured way of interviewing the patient, and its method of inquiring about specific and detailed examples of limitations and concrete activities which the patient still undertakes.

Two important criteria for evaluating work-related assessments are the validity and reliability of the instruments used<sup>19,20</sup>. Validity is the extent to which an instrument measures what is intended to be measured<sup>21,22</sup>. Content validity is the degree to which the test items represent the performance domain the test is intended to measure, and it is usually determined by a panel of experts examining the relationship between the test objectives and the test items, or by detailed knowledge of the

**Box 1.** Domains covered in the DASI interview

**Introduction**

- putting the patient at ease
- short explanation of the procedure
- summarize the known data

**Work**

- type and duration of work
- content of the occupation
- perception of work by the patient

**Impairments**

- medical history and nature of current complaints
- other diseases and general anamnesis
- course of the disease
- cause of the disease (the treating physician's diagnosis; private/work stress; personality)
- treatment and medication use (now and in the past)
- if necessary: more information from treating physician, hetero-anamnesis, expert consultation

**Activity limitations**

- limitations experienced in daily life and at work, for example, lifting, walking and bending.
- concrete and detailed examples of the limitations experienced.

**Participation**

- activities of daily living (ADL), description of a usual day, hobbies, sports, housekeeping, social contacts and work. The patient is asked which activities are actually performed and for how long.

**Patient's opinion**

- Patient's view of the work limitations
- Patient's response to the provisional opinion of the physician

**Physician's opinion**

- the physician communicates his or her judgment
- explanation of the rest of the procedure
- opportunity for the patient to respond

normal practices used. Concurrent validity examines the correlation between a new measure and an accepted measure given to the same subjects<sup>22;23</sup>. Reliability involves the extent to which a test or measurement is consistent and free from error<sup>24</sup>.

In spite of the fact that assessments of functional limitations in the Netherlands are mainly based on an interview, almost no information is available on the reliability and validity of the interview as an instrument to assess functional limitations. In studies where the assessing physicians interview the patients themselves, a low inter-rater reliability was found<sup>25;26</sup>. In studies where physicians based their assessments on written reports or on video recordings of DASI interviews, reasonable to good inter-rater reliability was found<sup>27;28</sup>.

Given the immense consequences of functional assessments, it is of importance to examine the psychometric properties of such an instrument. In order to fill this gap, the aim of the present study is to evaluate in a real-life situation:

- the inter-rater reliability between physicians with and without DASI training.
- the content and concurrent validity of the DASI.
- the patient's opinion of those physicians who used and those who did not use the DASI.

The DASI method was chosen over other methods because it is a well-described method which is based on the ICF and it is the only method that has had some study done on its psychometric qualities.

## METHODS

### *Physicians*

At four out of a total of 17 branches of the Dutch Social Security Office, four insurance physicians were invited to participate in the study, resulting in 16 physicians voluntarily cooperating in this study. In each of the four locations, two insurance physicians were randomly assigned to the intervention group and two were assigned to the control group. No significant difference in the average length of time spent in professional practice between the physicians in the intervention group (15.5 years, range 7-28 years) and the control group (14.6 years, range 9-21 years) was present.

### *Training*

The intervention group was given a three-day DASI training session over a three-week period. The first week consisted of two days of instruction and practice. After demonstration of an item from the DASI by an instructor and an actress, the eight physicians practiced the items of the DASI in groups of three physicians. The role of the patient, physician and observer alternated. The next week the physicians practiced the method on their regular patients and made a video recording of the DASI interview. On the third day of the training session, in the third week, their video recordings were analyzed and assessed.

The control group did not receive any training and examined patients as usual.

### *Patients*

A total of 443 patients who applied for social disability benefit after 21 months of sick leave were asked to cooperate, of them 236 agreed (53%). Only patients with at least lower back or lower extremity problems were selected in order to obtain a homogeneous group with sufficient filling of items of the FAL (see Instruments). Of the patients who agreed to cooperate, 26% were included (n=62), 36% were diagnosed as mental complaints (n=85), and 38% had another diagnosis such as neck and upper extremity complaints, heart and lung diseases or cancer (n=89).

### *Instruments*

The Functional Ability List (FAL)<sup>15</sup> is an instrument to record functional limitations and is used in social security assessments in the Netherlands. All Dutch insurance physicians are trained and experienced in using the FAL. The FAL contains six domains in which 70 mental and physical items are addressed, and for each item the seriousness can be indicated. One example is the item “lifting or carrying”, where the insurance physician has to choose from four gradations:

#### **Lifting or carrying**

- 0 normal, can carry or lift about 15 kg (toddler)
- 1 slightly limited, can carry or lift about 10 kg (small toddler)
- 2 limited, can carry or lift about 5 kg (bag of potatoes)
- 3 severely limited, can carry or lift about 1 kg (one liter of milk)

The content validity of the DASi was assessed using a self-structured questionnaire which was filled out by the physicians who had undergone DASi training. The questionnaire contained eight questions with fixed response alternatives on a five-point ordinal rating scale. In addition, it was also possible for the physicians to make additional comments about the DASi. The questionnaire contained questions about whether the instrument was adequate for the intended purpose, whether anything essential was missing or whether any part of the instrument was irrelevant (Table 2).

In addition, the patients filled out a questionnaire that is routinely used by the Dutch Social Security Office to measure patient satisfaction with the behavioral aspects of physicians<sup>29</sup>. Lastly, the patients gave an indication of the duration of the interview.

### *Procedure*

Patients (n=62) were interviewed and examined independently by two physicians from the same group (intervention group or control group) on the same day, between June and November 2008. The patients were randomly assigned to either the intervention or the control group so as to be able to compare similar groups. The physicians recorded their assessment of those work limitations to be found in the physical items of the Functional Ability List (FAL), and provided a detailed report containing information on the interview, including their judgment and the reasons for their judgment. Furthermore, we examined whether the patients did end up qualifying for a disability benefit.

After using the DASi in daily practice, the physicians were asked to give their opinions of the DASi by filling out the questionnaire. After the interview and examination, the patients were asked their opinions as to how satisfied they were with the behavioral aspects of the physicians, also by filling out a questionnaire.

### Analysis

The “linear-weighted observed percentage agreement” on the FAL items was taken as a measurement of inter-rater reliability within each of the two groups of insurance physicians<sup>30,31</sup>. Due to the fact that the marginal distribution of the variables was skewed, the computation of an agreement index based on Cohen’s kappa could not be used. One requirement for the use of this index is that the marginals should have more or less the same frequency. If not, this will result in an overestimation of the expected agreement<sup>32</sup>. The statistical software package AGREE 7.3<sup>33</sup> was used for the calculation of these values. In general, a percentage agreement of 60 to 80% is considered reasonable to good; more than 80% is considered excellent<sup>34</sup>.

The concurrent validity was examined by comparing the mean scores on the FAL items of the intervention and the control groups. The Mann-Whitney test, a non-parametric test that is used to compare two independent groups, was used for the between-group differences in the mean scores on the FAL items.

## RESULTS

A total of 62 patients were assessed by two physicians, 32 in the intervention and 30 in the control group. There were no significant differences between the groups in terms of age, gender, terms of employment and diagnosis. The mean age of the patients in the intervention group was 49.8 years (range 30–64 years), and in the control group, 46.3 years (range 35–63 years). In the intervention group 47% of the patients were female, and in the control group, 37%. Before registering sick, the patients in the intervention group worked for an average of 31.6 hours a week (range 8–40 h), and in the control group, 33.0 hours a week (range 13–40 h). In the intervention group, nine patients had lower-extremity problems (e.g., fractured ankle, gonarthrosis or peripheral arterial disease), 15 had lower-back problems (e.g., lumbar spinal stenosis, chronic non-specific lower back pain or herniated disc) and eight patients presented more general complaints (e.g., rheumatoid arthritis, fibromyalgia or somatoform disorder). In the control group, eight patients had lower-extremity problems, 14 had lower-back problems and eight had general complaints.

Table 1 presents the “linear weighted percentage agreement” between the physicians and the “mean scores” on the items of the Functional Ability List in the control and intervention groups.

Physicians from the intervention group showed a mean percentage agreement of 80.6% (range 59–100%), and the control group, 83.6% (range 67–97%). Except for the item “frequent heavy lifting,” there were no differences in agreement percentages between the intervention and control groups. In 19 out of the 21 items on the FAL the physicians of the intervention group indicated more serious functional limitation scores in their assessments compared to the control group. For nine of these items, there were significant differences ( $p < 0.05$ ). Concerning the daily number of hours a patient could function, the physicians in the intervention group gave limitations in 31% of the patients; in

**Table 1.** "Linear weighted percentage agreement" between the physicians (columns 1–2) and "Mean scores" (columns 3–4) on items of the Functional Ability List in the intervention (n=32) and control (n=30) groups

Items	Agreement (%)		Mean $\pm$ SD (min-max)	
	Intervention	Controls	Intervention	Controls
<u>Body movement scale</u>				
Reaching	97	97	0.02 $\pm$ 0.13 (0-1)	0.02 $\pm$ 0.13 (0-1)
Frequent reaching	81	88	0.56 $\pm$ 0.73 (0-2)	0.55 $\pm$ 0.79 (0-3)
Bending (degrees)	78	85	0.75 $\pm$ 0.74 (0-2)*	0.48 $\pm$ 0.68 (0-2)
Frequent bending	83	86	1.50 $\pm$ 0.85 (0-3)	1.35 $\pm$ 1.01 (0-3)
Rotation¶	78	87	0.23 $\pm$ 0.43 (0-1)*	0.07 $\pm$ 0.25 (0-1)
Push or pull	75	72	1.09 $\pm$ 0.68 (0-2)*	0.62 $\pm$ 0.61 (0-2)
Lifting or carrying	85	80	1.62 $\pm$ 0.75 (0-3)*	1.07 $\pm$ 0.69 (0-2)
Frequent light lifting	78	84	0.73 $\pm$ 0.93 (0-3)	0.60 $\pm$ 0.81 (0-3)
Frequent heavy lifting¶	100*	83	1.00 $\pm$ 0.00 (1-1)*	0.88 $\pm$ 0.32 (0-1)
Walking	91	86	1.30 $\pm$ 0.79 (0-3)	1.08 $\pm$ 0.85 (0-3)
Sustained walking	90	87	1.53 $\pm$ 0.69 (0-3)*	1.20 $\pm$ 0.69 (0-2)
Climbing stairs	79	78	1.22 $\pm$ 0.93 (0-3)*	0.78 $\pm$ 0.69 (0-2)
Climbing	70	87	1.08 $\pm$ 0.78 (0-2)*	0.80 $\pm$ 0.55 (0-2)
Kneeling¶	72	67	0.16 $\pm$ 0.36 (0-1)	0.10 $\pm$ 0.30 (0-1)
<u>Body posture scale</u>				
Sitting	86	88	0.98 $\pm$ 0.85 (0-3) *	0.55 $\pm$ 0.65 (0-2)
Prolonged sitting	79	90	0.78 $\pm$ 0.70 (0-3)	0.57 $\pm$ 0.56 (0-2)
Standing	80	82	1.55 $\pm$ 0.81 (0-3)	1.40 $\pm$ 0.80 (0-3)
Prolonged standing	88	89	1.59 $\pm$ 0.71 (0-3)	1.47 $\pm$ 0.72 (0-3)
Prolonged kneeling¶	72	87	0.48 $\pm$ 0.50 (0-1)	0.63 $\pm$ 0.49 (0-1)
Prolonged bending¶	59	80	0.58 $\pm$ 0.50 (0-1)	0.43 $\pm$ 0.50 (0-1)
Working above shoulder¶	72	73	0.27 $\pm$ 0.48 (0-1)	0.36 $\pm$ 0.45 (0-1)
<b>Mean</b>	<b>80.6</b>	<b>83.6</b>		

¶ = dichotomous data, other items are ordinal; \* p < 0.05

40% of these patients the physicians were in agreement on this. In the control group, the physicians indicated a limitation in hours of daily functioning of 23% in their patients; in 29% of these patients the physicians were in agreement on this.

In the intervention group, 18 out of 32 patients (56%) qualified for a work disability benefit, while in the control group 13 out of 30 patients (43%) did; this did not represent a significant difference (p = 0.31).

Table 2 presents the opinion of the eight physicians of the intervention group concerning the DASI. All physicians were in agreement that the DASI was an acceptable tool in daily practice, one which gives an objective view of the patient and enough information to assess functional ability.

**Table 2.** The physicians' opinions (n=8) on the DASI (intervention group) in percentages

The DASI ...	disagree %	neutral %	agree %
... provides an objective view of the patient			100
... attends to all relevant aspects		12.5	87.5
... attends to irrelevant aspects	87.5	12.5	
... is acceptable in daily practice			100
... provides enough information to assess functional ability			100
... differs from my own interview	25	25	50
... has added value		12.5	87.5
... allows a better founding for filling out the FAL <sup>1</sup>	12.5		87.5

<sup>1</sup>Functional Ability List

As an added value of the DASI, the physicians mentioned in particular the structuring of the interview and collecting detailed information on the functioning of the patient. One physician mentioned that the DASI mainly collected information from the patient, but that the assessment of this information into functional abilities was not addressed.

In their reports the physicians of the intervention group mentioned an average of 6.7 functional limitations as experienced by the patient (range 4–10). In the control group, an average of 4.4 functional limitations were mentioned (range 0–7) ( $p < 0.05$ ). In the case of functional limitations, 71% of the intervention group indicated the intensity of the limitations experienced, for instance, by giving an example of the limitation in daily life. In the control group, this was 40%.

The patients' satisfaction report score for physicians of both the intervention and control groups in their interviews was 7.7 on a scale from 1 to 10. Moreover, no differences between the two groups were found in terms of answers to the questions concerning behavioral aspects of the physicians (listening, empathy, meticulousness and professionalism).

According to the patients, the duration of the interview and the physical examination was on average 45–60 minutes in the intervention as well as in the control group (range < 30 min – > 60 min).

## DISCUSSION

Although accurate determination of work disability status is crucial for the health and well-being of patients and their families, the reliability and validity of bureaucratic approaches is poor and cumbersome and extraordinarily expensive. This study of the DASI demonstrates that a semi-structured interview might hold great promise as an inexpensive solution to this problem. We studied inter-rater reliability, and both content and concurrent validity, along with the patient's opinion of the DASI.

### *Reliability*

Up till now, no real life studies of inter-observer agreement among physicians in assessing functional limitations had been conducted. We hypothesized that agreement between physicians in the control group would be low and, in the DASI group, that it would be acceptable. In the end, we found an overall inter-rater reliability for the items of the FAL in the intervention group that was reasonable to good, and for some dimensions even excellent. Contrary to our expectations, the agreement in the existing practice was satisfactory too, and DASI training did not improve agreement between physicians. One explanation for this may be the fact that we used patients with relatively straightforward lower back or lower extremity problems. Possibly those patients with more complicated problems and those with mental problems might produce less satisfactory results. Because agreement between physicians in international literature is found to be very poor<sup>7-11</sup>, another explanation may be that the satisfactory agreement in existing practice is specific for the Dutch context. In the Netherlands specially trained insurance physicians assess the functional disabilities in patients. These physicians all had an interview-training, in which they were taught to ask for activity limitations and participation in addition to standard medical history-taking. This education is not always common in other countries, and may be the explanation of the relatively good agreement between the physicians.

We found a low inter-rater agreement concerning the daily number of hours a patient could function. The daily number of hours a patient can function according to the physician often has very important consequences for a work disability benefit. Therefore, the low inter-rater reliability found is undesirable. Insurance physicians in the Netherlands have a guideline for "reduced working hours"<sup>35</sup> at their disposal, but unfortunately this guideline cannot prevent the differences in outcome between the physicians. The satisfactory inter-rater reliability on the items of the FAL and the low inter-rater agreement concerning the daily number of hours a patient could function which were found in this study are comparable to Dutch studies conducted in a more controlled environment where physicians did not see patients face to face, but made an assessment based on video recordings or written reports of DASI patient interviews<sup>27,28</sup>.

### Validity

Preferably, validity is assessed by comparing the measurement studied to a gold standard. For assessing functional limitations, however, no gold standard is available. Different methods for assessment, for instance, self-assessment questionnaires, clinical examination and performance tests, lead to different outcomes<sup>36</sup>. From the reports made by the physicians in this study, it appeared that the same information could lead to different outcomes. One example was the assessment of a 56-year-old patient with depression and lower-back problems as a result of a somatoform disorder. One physician assessed no functional limitations when considering the diagnosis and an absence of objective functional defects. The other physician assessed the same patient and concluded the patient was limited in lifting ability (10 kg maximum), sitting (1 h maximum) and walking (half an hour maximum) because the patient made a genuine impression, and offered a plausible and consistent story. The question might be raised as to whether consistency in a patient's behavior together with the functional limitations experienced should in fact be leading factors in the assessment, this despite the fact that there might be no actual objective medical findings present. In this light, part of the assessment of functional limitations would seem to lie in the realm of a social rather than a medical concept.

This study showed a satisfactory content validity for the DASI. Without a single exception, the physicians agreed on the fact that the DASI was an acceptable tool in daily practice and one which gave an objective view of the patient and enough information in order to assess functional abilities. Seven out of eight physicians found the DASI to be an even better basis for the assessment of functional limitations than the interview they usually applied.

For assessing concurrent validity, the outcomes of two measurements administered to the same patients were compared. In this study, we compared the outcomes of the intervention group and the control group in different groups of patients. Because patients were randomly allocated to the intervention group and the control group, however, the groups were comparable. This is supported by the fact that there were no significant differences for both groups in terms of age, gender, terms of employment and diagnosis. We found that in almost half of the items of the FAL, the physicians using the DASI gave substantially more severe functional limitations in their assessments than did the control group. But this did not lead to an increased number of patients who qualified for a disability benefit. One explanation for the more severe functional limitations may be that the DASI focuses more attention on problems concerning activities and functional limitations as compared to "care as usual." That this did not lead to more disability benefits can be explained by the Dutch system for determining the benefit. An occupational expert investigates what jobs the patient is theoretically still able to perform in light of the limitations. The earning capacity will determine the disability benefit. Apparently the more severe functional limitations did not lead to a greater loss in earning capacity. The literature describes the fact that insurance physicians show limited attention to the detailed information regarding the functional limitations the patients experience<sup>14</sup>. This is in line with the findings in this study where the physicians in the intervention group reported

significantly more severe functional limitations.

Even though no differences between the intervention and the control groups were found in disability benefit outcome, the difference in outcome for the functional limitations was important because functional limitations are needed for reintegration into appropriate work. Because of the lack of a gold standard, it is unknown whether the more severe ratings in the DASI group are more valid than those of the “usual care” physicians. The physicians in the intervention group found the DASI to be a better basis for the assessment than the interview they usually applied. Therefore, we think the DASI contributes to a more thorough assessment of the functional limitations.

#### *Patients' opinions*

The patients' mean report score for satisfaction with the DASI was 7.7 on a scale from 1 to 10. The same score was found for the interviews of the physicians in the control group. Apparently the DASI did not improve or worsen patient satisfaction.

#### *Study limitations*

Some limitations of this study should be noted. The physicians knew they were being monitored; this might have influenced their assessments. However, it was practically impossible to conduct a study which would not have had this disadvantage. Furthermore, the assessments were aimed at physically based functional limitations; mentally based functional limitations might well present a rather different outcome. Finally, although the physicians in the intervention group received DASI training, it is possible they did not implement this in daily practice. We studied the reports of the assessments to check whether the physicians who received DASI training actually performed the interview as it was taught. One important characteristic of the DASI is the presence of concrete and detailed information on functional limitations as experienced by the patient. The reports on those physicians who received the training contained more functional limitations and more detailed information on this point, indicating that the intervention group actually performed what they had been trained to do.

#### *The DASI in daily practice*

Several tools are used to assess the functional abilities of people with medical impairments, but no single currently existing test provides a valid measurement of functional limitations<sup>12,37</sup>. Functional capacity tests and questionnaires alone cannot properly assess functional limitations without an appraisal of the outcome of these tests. A combination of specialized physicians and instruments such as functional capacity tests and questionnaires looks the most promising. It might be useful to provide self-report questionnaires about function to the patient before the DASI in order to increase the efficiency and specificity of the interview. Then, clinical examination and a semi-structured interview, like the DASI, could be conducted by a physician or, in part, even by a trained nurse<sup>28</sup>. Based on this information, individually selected functional capacity tests could be conducted to

confirm or disconfirm the initial results of the interview. Guidelines and protocols might narrow down further the differences in assessment among physicians<sup>25</sup>.

#### *Future research*

Further research into the value of guidelines and protocols – especially where the assessment of limitations as to the number of hours a patient can function daily is a factor – as well as additional studies concerning the use of the DASI in mental-function limitations may be useful. Concurrent validity can be assessed by comparing outcomes of the DASI with self-report questionnaires and functional capacity tests. For research into the validity of instruments to assess functional limitations, a gold standard is needed. A gold standard might be approached by looking for consensus among a number of physicians after medical examination, an interview protocol, questionnaires and performance tests.

#### **CONCLUSION**

In conclusion, we would state that the DASI is a tool with a reasonable to good inter-rater reliability and content validity, and that it appears to be acceptable to both patients and physicians. The DASI did not improve inter-observer agreement beyond that of usual interview procedures used in the Netherlands. The DASI would seem to be a worthwhile tool for collecting self-reported information in order to assess functional limitations in claimants. Because the physicians who used the DASI assessed more functional limitations as compared to usual practice, further research into the interpretation of the self-reported information is needed.

**Reference List**

1. Bot SDM, Terwee CB, van der Windt DAWM, Bouter LM, Dekker J, de Vet HCW. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. *Annals Of The Rheumatic Diseases* 2004;63:335-341.
2. Grotle M, Brox JI, Vollestad NK. Functional status and disability questionnaires: what do they assess? A systematic review of back-specific outcome questionnaires. *Spine* 2005;3:130-140.
3. Wind H, Gouttebarga V, Kuijer PP, Frings-Dresen MH. Assessment of functional capacity of the musculoskeletal system in the context of work, daily living, and sport: a systematic review. *J Occup Rehabil* 2005;15:253-272.
4. Matheson L, Mayer J, Mooney V, Sarkin A, Dreisinger T, Verna J et al. A method to provide a more efficient and reliable measure of self-report physical work capacity for patients with spinal pain. *J Occup Rehabil* 2008;18:46-57.
5. De Boer WEL, Besseling JJM, Willems JHBM. Organisation of disability evaluation in 15 countries. *Pratiques et Organisations de Soins* 2007;38:205-217.
6. Wahlstrom R, Alexanderson K. Chapter 11. Physicians' sick-listing practices. *Scandinavian Journal of Public Health* 2004;32:222-255.
7. Clark W, Haldeman S. The development of guideline factors for the evaluation of disability in neck and back injuries. *Spine* 1993;18:1736-1745.
8. Söderberg E, Alexanderson K. Sickness certification practices of physicians: a review of the literature. *Scandinavian Journal of Public Health* 2003;31:460-474.
9. Lax MB, Manetti FA, Klein RA. Medical evaluation of work-related illness: evaluations by a treating occupational medicine specialist and by independent medical examiners compared. *International Journal of Occupational & Environmental Health* 2004;10:1-12.
10. Wolfson AM, Doctor JN, Burns SP. Clinician judgments of functional outcomes: how bias and perceived accuracy affect rating. *Archives of Physical Medicine & Rehabilitation* 2000;81:1567-1574.
11. Rainville J, Pransky G, Indahl A, Mayer EK. The physician as disability advisor for patients with musculoskeletal complaints. *Spine* 2005;30:2579-2584.
12. Matheson LN, Kane M, Rodbard D. Development of new methods to determine work disability in the United States. *J Occup Rehabil* 2001;11:143-154.
13. Social Security Administration. Plan for a new disability claim process. Washington, DC: Social Security Administration, 1994.
14. Bont AA de, Berendsen L, Boonk MPA, Brink JC van den. In de spreekkamer van de verzekeringsarts. Een onderzoek naar het verzekeringsgeneeskundige deel van de WAO-claimbeoordeling. [In the consulting room of the insurance physician. A study of the role of the insurance physician in disability benefit]. Zoetermeer, the Netherlands: CTSV, 2000.
15. Functional Ability List. ClaimBeoordelings- en BorgingsSysteem (CBBS). Amsterdam: Lisv, 2002.
16. De Boer WEL, Wijers JHL, Spanjer J, Beijl I van der, Zuidam W, Venema A. Gespreksmodellen in de verzekeringsgeneeskunde. [Interview models in insurance medicine]. Hoofddorp, the Netherlands: TNO, 2004.
17. Spanjer J. Disability Assessment Structured Interview (DASI), description of an assessment method. Zuidlaren, the Netherlands, 2007. A copy can be acquired from the author.
18. International classification of functioning, disability and health. Geneva: World Health Organisation, 2001.
19. Innes E, Straker L. Attributes of excellence in work-related assessments. *Work* 2003;20:63-76.
20. Matheson LN, Gaudino EA, Mael F, Hesse BW. Improving the validity of the impairment evaluation process: A proposed theoretical framework. *J Occup Rehabil* 2000;10:311-320.

21. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice. 2 ed. Upper Saddle River, New Jersey, Prentice Hall, Inc, 2000.
22. Innes E, Straker L. Validity of work-related assessments. *Work* 1999;13:125-152.
23. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. 2nd edn. Oxford: Oxford University Press, 1995.
24. Innes E, Straker L. Reliability of work-related assessments. *Work* 1999;13:107-124.
25. Spanjer J. De reproduceerbaarheid van WAO beoordelingen, een literatuuronderzoek. [The reproducibility in disability assessment: a literature search]. *Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde* 2001;9:195-198.
26. Brouwer S, Dijkstra PU, Gerrits EHJ, Schellekens JMH, Groothoff JW, Geertzen JHB et al. Intra- en inter-beoordelaarsbetrouwbaarheid 'FIS-belastbaarheidspatroon' en 'Functionele Mogelijkheden Lijst'. [Intra- and inter-rater reliability 'Functional Information System' and 'Functional Ability List']. *Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde* 2002;11:360-367.
27. Spanjer J. De betrouwbaarheid van WAO beoordelingen. [The inter- and intra- rater reliability of disability pension assessments]. *Tijdschrift voor Bedrijfs- en Verzekeringsgeneeskunde* 2001;9:234-241.
28. Spanjer J, Krol B, Brouwer S, Groothoff JW. Inter-rater reliability in disability assessment based on a semi-structured interview report. *Disability & Rehabilitation* 2008;30:1885-1890.
29. Burg CL van der. Clientmonitor. Leiden, the Netherlands: AStri, 2007.
30. Schouten HJA. Measuring pairwise agreement among many observers. *Biometrical Journal* 1980;22:497-504.
31. Schouten HJA. Measuring pairwise agreement among many observers. II. Some improvements and additions. *Biometrical Journal* 1982;24:431-435.
32. Popping R. On agreement indices for nominal data. In: Saris W.E., Gallhofer I.N. *Sociometric Research. Volume 1. Data collection and scaling.* New York: St. Martin's Press, 1988, pp. 90-105.
33. Popping R. AGREE, A package for computing nominal scale agreement. *Comput Stat Data Anal* 1984;2:182-185.
34. Landis JR, Koch GG. The measurements of observer agreement for categorical data. *Biometrics* 1977;33:159-174.
35. Standaard verminderde arbeidsduur. [Guideline reduced working hours]. Amsterdam: Livs, 2007.
36. Brouwer S, Dijkstra PU, Stewart RE, Göeken LNH, Groothoff JW, Geertzen JHB. Comparing self-report, clinical examination and functional testing in the assessment of work-related limitations in patients with chronic low back pain. *Disability & Rehabilitation* 2005;27:999-1005.
37. Gaudino EA, Matheson LN, Mael FA. Development of the Functional Assessment Taxonomy. *J Occup Rehabil* 2001;11:155-175.

