Disease oriented work ability assessment in social insurance medicine
Slebus, F.G.

Citation for published version (APA):
Slebus, F. G. (2009). Disease oriented work ability assessment in social insurance medicine.

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: http://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.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
Chapter 6

Work ability assessment of long-term sick-listed depressed employees with the use of a checklist

Slebus FG, Kuijer PP, Willems JH, Frings-Dresen MH, Sluiter JK.
(submitted for publication)
Abstract

**Purpose:** To assess the mean score and variation of work ability provided by Dutch Insurance Physicians (IPs) in five different real case history vignettes of long-term, sick-listed employees with Major Depressive Disorder (MDD) with and without the aid of a checklist.

**Method:** In a post test only randomized experiment, 25 IPs assessed work ability for five cases on a scale of 0 to 100 without the use of the checklist, while 21 IPs used the checklist. Differences between groups in mean and absolute variation of assessments were tested with independent t-tests. Intra Class Correlation (ICC) analysis was used to determine if IPs could distinguish between the vignettes.

**Results:** When using the checklist, the mean work ability score of all vignettes was 3 to 12 points higher. There was no difference in variation in work ability scores per vignette and between groups. ICC was 0.64 for both groups.

**Conclusion:** The use of the checklist increased the mean score of work ability but had no effect on its variation. The inter-rater reliability was moderate.
Introduction

The estimated annual prevalence of Major Depressive Disorder (MDD) is around 7% in the working population\(^1\), and MDD frequently has a chronic course with residual symptoms\(^2,3,4\). There is an individual and social need to keep employees with MDD in the workforce, because participation in a work environment is associated with subjective well-being and life satisfaction\(^5,6\), and because of expected labour shortages in the near future\(^7\). However, because employees with MDD have problems with functioning and work performance\(^8\), it is not surprising that MDD is associated with substantial presenteeism, absenteeism, job loss\(^9\), and ill health retirement\(^10\). The remaining work ability of sick-listed MDD employees should therefore be properly determined.

Assessing the work ability of employees with MDD involves considering aspects of work ability in relation to the symptoms of MDD presented by the employee that may be relevant to their work activities. This is a complex task because MDD is not a univocal disease but is compounded by different symptoms with varying degrees of intensity, such as fatigue or loss of energy, feelings of worthlessness, diminished ability to think or concentrate, and recurrent thoughts of death.

The assessment of work ability in the Netherlands is performed by Insurance Physicians (IPs) when employees are long-term sick-listed on the basis of an interview and examination of the client and eventually, consultation of other professionals concerned. Although the subjective interpretation of the IP who performs the assessment should be minimal\(^11\), IPs routinely interpret work ability individually\(^12\) and the inter-rater reliability of the assessment of work ability is not well studied\(^13\). More studies on the reliability of judgements of work ability are necessary to develop the professional base of work ability assessments. In relation to IPs, this means that the sources that can affect reliability are identified and reduced. Some sources may be: (a) raters may obtain different information as a result of asking different questions; (b) raters may differ in what they notice and remember when presented with the same information; (c) raters may differ in the significance they attach to what is observed; and (d) raters may use different criteria to score the same information\(^14\). In this respect, it can be hypothesized that shared starting points used to assess work ability will improve measurement outcomes. Therefore, we developed a consensus-based checklist that consists of ten aspects that were considered relevant by IPs in work ability assessments of long-term sick-listed MDD employees\(^15\). The hypothesis for the present study is that the use of checklist in work ability assessment will diminish variation in judgements because IPs will focus on the same aspects of work...
ability. To assess the effect of using the checklist on the variation in judgements of IPs, we formulated the following research question:

(i) What is the effect on work ability assessment of sick-listed MDD employees by IPs when the checklist is used compared to when the checklist is not used?
(ii) What is the effect on the variation in work ability assessment of sick-listed MDD employees by IPs when the checklist is used compared to when the checklist is not used?

Methods

We performed a post-test only randomized experiment. A group of IPs who assessed the work ability of five real case history vignettes without the help of the checklist (control group) were compared with the judgements of a group of IPs who assessed work ability of the same five real case history vignettes with the help of the checklist (intervention group).

The checklist

The checklist was developed in an earlier study\(^\text{15}\). In Table 1, the 10 aspects that have to be considered in insurance medicine when work ability is assessed for employees with MDD are presented in the first column. In the second column, examples of the aspects are presented. The participants of the intervention group had to study the items of the checklist and to take the items into account when they assessed work ability of the real case history vignettes.

The vignettes

Figure 1 shows the procedure that was followed to select the five real case history vignettes\(^{16,17,18,19,20}\) that were used in this study.

First, and as is seen in Figure 1, the main office of the National Institute of Benefit Schemes randomly selected 50 reports of employees with MDD made by IPs in the period between 2006 and 2008. The medical insurance histories of those reports were scrutinized by the researchers (FS, PK, HW, MF, JS) for the presence of aspects of work ability of the checklist. Only when the medical-insurance history contained at least five of the ten items of the checklist, the medical insurance history was included as a possible vignette.
After identifying 30 possible real case histories, the cases were randomized and thereafter divided for assessment by six staff IPs of the National Institute of Benefit Schemes. In daily practice, those staff IPs coach and instruct IPs concerning work ability assessment. The staff IPs had to rate the complexity of the real case histories on a Visual Analogue Scale (VAS) (0-100 complexity points), 0 meaning ‘not at all complex’ and 100 meaning ‘very complex’. The staff IPs were also asked to comment on the case histories’ comprehensibility and usability concerning the assessment of work ability\(^\text{10}\) and to check the instructions given. Finally, they were asked to assess the work ability of the real case histories as a pre-test before the start of this study.

The range of complexity scores of the case histories as provided by the staff IPs was divided into five equal parts, each representing a complexity grade; grade 1 had the lowest complexity score and grade 5 the highest. The real case histories were graded

<table>
<thead>
<tr>
<th>Ability to:</th>
<th>Examples of ability:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Take notice</td>
<td>A truck driver should be able to notice a car accident that happens in front of him.</td>
</tr>
<tr>
<td>Sustain attention</td>
<td>A bus driver should be able to remain alert enough to drive in the correct lane even on a long, uninteresting road in the late afternoon.</td>
</tr>
<tr>
<td>Focus attention</td>
<td>A teacher should be able to concentrate on the subject of the lesson even when the students are noisy.</td>
</tr>
<tr>
<td>Complete operations</td>
<td>A baker should not only be able to put the dough in the oven but also to concentrate on, manage, and finish the whole baking process up to removing the bread from the oven.</td>
</tr>
<tr>
<td>Think in a goal-directed manner</td>
<td>An anesthetist working in an operating theater should first stabilize relevant parameters in the patient before filling in forms or performing other functions with a lower priority.</td>
</tr>
<tr>
<td>Remember</td>
<td>A hotel porter should be able to remember where he has put his guests’ luggage.</td>
</tr>
<tr>
<td>Perform routine operations</td>
<td>A school nurse should be able to vaccinate hundreds of children a day and to do this in the standard and safe way she has learned.</td>
</tr>
<tr>
<td>Undertake structured work activities</td>
<td>A bricklayer should be able to lay bricks exactly according to a given wall design.</td>
</tr>
<tr>
<td>recall</td>
<td>A medical doctor must be able to recall acquired knowledge in order to evaluate the patient’s complaints</td>
</tr>
<tr>
<td>perform autonomously</td>
<td>A general practitioner should be able to make decisions about the management of patients independently</td>
</tr>
</tbody>
</table>

Table 1  The 10 items of the checklist and examples of the items\(^\text{15}\).
Random selection of case histories (n=36) of depressed employees by National Institute of Benefit Schemes out of all assessments performed between 2006-2008

At least five aspects of the checklist are present?

no

yes

Is it possible according to staff IPs (n=6) to assess the work ability on bases of the case histories and instructions given?

no

yes

Case history not usable as vignette (n=6)

Case history not usable as vignette (n=0)

Assessment of the complexity (0 till 100 complexity points) of the case histories and the range of complexity scores

Division of the range of complexity scores in 5 equal parts to grade the case histories

Grading the case histories according to the complexity scores given by the staff IPs and randomly selecting one case history out of every five grades of complexity

Is there at least a difference of 10 complexity points between the case histories?

yes

no

Select another case history

Use case history as vignette

**Figure 1** Flow chart for the selection of the vignettes.

according to the complexity scores given by the staff IPs. Out of each grade, one real case history was randomly selected. It was then checked if the selected real case histories differed by at least 10 points in complexity of the VAS from each other. The complexity
scores for the five real case histories were: 6, 17, 34, 51 and 75, and therefore the five selected real case histories were used in this study as vignettes. The pre-test of the work ability assessment instrument was good. Fifty sets of the five vignettes in different orders were made with a randomised sequence selector (www.randomizer.com), to rule out the possibility that the order of vignettes would influence the outcomes.

**Sampling of participants**

Out of all four regions (North, East, South and West) of the Netherlands, 15 offices of the National Institute of Benefit Schemes were contacted. At those 15 offices, approximately 1000 IPs are working on a daily basis to assess work ability of work disabled employees. IPs receive four years of in-company training before they are registered as an IP. IPs who performed work ability assessments of long-term sick listed employees were asked to participate. It was estimated that two groups of around 25 IPs21 were needed to answer the research question.

Names of IPs from the four regions of the Netherlands who performed work ability assessments of long-term sick-listed employees were gathered from staff IPs. Next, these IPs were contacted by telephone by FS and asked whether they wanted to participate. The IPs who agreed to participate were informed of the study and signed an informed consent before the start of the study. They also completed a short form that was attached to the informed consent form to gather information about their age, gender, experience as an IP, and registration period as an IP.

IPs were randomized into two groups. To prevent ‘cross talking’ between the two groups, the offices of the National Institute of Benefit Schemes where the IPs worked were identified. According to the office where the IP worked, the IP was presented the same set of five vignettes with the checklist or without. Work ability of the vignettes was assessed on a Visual Analogue Scale (0-100); 0 meaning ‘no work ability’ and 100 meaning ‘as much work ability as before MDD’.

**Analysis**

Data of age, gender, experience as an IP, and registration period as an IP of the intake forms and, judgements of the work ability were entered in SPSS 16.0. Difference between the control and intervention group (use of the checklist) for age of IPs, years of experience, and years of registration as IP was tested with T-tests for independent samples (p<0.05), and possible differences between the two groups in gender was tested with the Chi-square test (p<0.05).
For each vignette, the mean, the range, and the standard deviation of assessed work ability score was calculated. Moreover, for the two groups of IPs for each vignette, the absolute difference of each IP with regard to their own mean group score of a vignette was calculated and used as a variation score. The distinction between the two groups for the variation score per vignette was tested for each vignette with a T-test for independent samples (p<0.05).

Intra-Class-Correlation analysis two-way random, absolute agreement, single measures was used to determine how well the IPs were able to distinguish the vignettes from each other. Good is: ICC > 0.80; moderate is: 0.50 ≤ ICC ≤ 0.80; and, poor is ICC < 0.50.

Results

Participants

Fifty-one IPs were contacted, and one IP refused to participate in the study because of involvement in another research project. Twenty-five IPs of the control group (100%) and 21 IPs of the intervention group (85%) returned their forms. The mean age of all participants was 50 years (SD 5; range 39-61); the mean years of working as an IP was 17 years (SD 6; range 5-25). The mean number of years of registration as an IP was 10 years (SD 6; range 0-23). Thirty-five percent of the IPs were female. No significant differences in age, years of working as IP, years of registration as IP, or gender were found between the intervention and the control group.

The assessment of work ability

In Table 2, the mean, the range, and standard deviation of the judgement of work ability per vignette for the control and the intervention group are presented, respectively. On a scale from 0 to 100, a wide range of mean work ability scores from 47 up to 95 per vignette were found. In the intervention group, the mean scores of work ability judgements for vignettes 1 to 5 are with 3, 12, 5, 8 and 9 points, respectively, on average higher than in the control group. There was a significant higher work ability score (p=0.04) on vignette 2 in the intervention group.

Table 3 presents, per vignette, the mean absolute variation score, the standard deviation of the variation score, and the test results between groups. No significant differences in the scores between the control and intervention group were found.
The mean assessment of work ability was higher when the checklist was used and significantly different in one out of the five used real case histories. Irrespective of the use of the checklist, there was a wide range in work ability assessments in every long-term sick-listed MDD case. No significant difference between the checklist and the

### Table 2
Mean, range (min-max), standard deviation and p-values of work ability assessment of the five vignettes for control and intervention group.

<table>
<thead>
<tr>
<th></th>
<th>Vignette 1</th>
<th>Vignette 2</th>
<th>Vignette 3</th>
<th>Vignette 4</th>
<th>Vignette 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean score</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- control</td>
<td>28</td>
<td>51</td>
<td>81</td>
<td>22</td>
<td>75</td>
</tr>
<tr>
<td>- intervention</td>
<td>31</td>
<td>63</td>
<td>86</td>
<td>30</td>
<td>84</td>
</tr>
<tr>
<td><strong>Range</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- control</td>
<td>0-77</td>
<td>10-78</td>
<td>40-10</td>
<td>0-80</td>
<td>41-100</td>
</tr>
<tr>
<td>- intervention</td>
<td>0-71</td>
<td>16-100</td>
<td>40-100</td>
<td>1-96</td>
<td>53-100</td>
</tr>
<tr>
<td><strong>Standard deviation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- control</td>
<td>23</td>
<td>20</td>
<td>15</td>
<td>24</td>
<td>18</td>
</tr>
<tr>
<td>- intervention</td>
<td>21</td>
<td>22</td>
<td>14</td>
<td>26</td>
<td>15</td>
</tr>
<tr>
<td><strong>P-value</strong></td>
<td>0.57</td>
<td>0.04</td>
<td>0.25</td>
<td>0.26</td>
<td>0.07</td>
</tr>
</tbody>
</table>

### Table 3
Mean, standard deviation, and p-values of variation of work ability assessment of the five vignettes for control and intervention group.

<table>
<thead>
<tr>
<th></th>
<th>Vignette 1</th>
<th>Vignette 2</th>
<th>Vignette 3</th>
<th>Vignette 4</th>
<th>Vignette 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean variation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- control</td>
<td>19</td>
<td>15</td>
<td>12</td>
<td>19</td>
<td>15</td>
</tr>
<tr>
<td>- intervention</td>
<td>18</td>
<td>15</td>
<td>10</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td><strong>Standard deviation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- control</td>
<td>12</td>
<td>12</td>
<td>9</td>
<td>13</td>
<td>15</td>
</tr>
<tr>
<td>- intervention</td>
<td>11</td>
<td>15</td>
<td>10</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td><strong>P-value</strong></td>
<td>0.75</td>
<td>0.99</td>
<td>0.63</td>
<td>0.91</td>
<td>0.22</td>
</tr>
</tbody>
</table>

### Discussion

The mean assessment of work ability was higher when the checklist was used and significantly different in one out of the five used real case histories. Irrespective of the use of the checklist, there was a wide range in work ability assessments in every long-term sick-listed MDD case. No significant difference between the checklist and the
no-checklist group was found in the absolute variation of the judgements per vignette. Irrespective of the use of the checklist, the IPs were moderately able to distinguish between the cases described in the vignettes with regard to work ability.

**Interpretation of the results**

Contrary to our hypothesis, we were not able to demonstrate that assessment variation diminished when the checklist was used. The use of vignettes instead of real patients cannot be a sufficient explanation for this finding because the assessment of work ability is comparable with a diagnostic process\(^4\), and it was previously determined that the validity of vignettes equals standardized patients\(^{17}\). Furthermore, real case histories used as vignettes have proven reliable when diagnosis criteria were investigated\(^{16}\). Other sources of variation should, therefore, be responsible for our findings. Raters might have differed in the significance they attached to what they ‘observed’ in the text of the cases. One other possible source might be the unknown relative importance of the items in the checklist, resulting in different criteria used in the assessment of work ability. However, if the use of the checklist introduced more variation our results should have pointed in the opposite direction as we expected, which they did not.

The items of the checklist are thought to be of extra assistance in assessing work ability because all vignettes were judged higher when the checklist was used. Since both groups of IPs had assess the same real case histories, differences in obtaining information\(^{14}\) as possible sources of variance can be ruled out as source for higher work ability assessments. Therefore, the groups of IPs must have differed in what they noticed and thought was important when they assessed the work ability of the real case histories\(^{14}\).

Because the ICC of the assessments was moderate (0.64), the need for further testing of the checklist in real patients seems obvious.

**Should the checklist be used?**

In the Netherlands, the assessment of an IP is the criterion for the level of work ability, but there are no known criteria in the literature to assess work ability of long-term sick-listed MDD employees\(^{25}\). In an earlier study, we showed that IPs base their assessment for durable work ability on diagnostic aspects and think that aspects of the participation of the employee in society is important in assessing work ability\(^{26}\). When asked what aspects are important to assess whether long-term sick-listed MDD employees can participate in work, IPs provided the 10 aspects of the checklist used in
this study\textsuperscript{15}. Those aspects can be seen as criteria in work ability assessment, and, up to now, no better criteria exist in the scientific literature. Using these criteria in this study did not result in less variation among IPs’ equal work ability assessments. Therefore, the question at stake is whether these items should already be used in practice. In our opinion, the answer is yes. Besides the fact that an assessment should be reliable, an assessment should also be explicable and at least be based on the opinion of professional experts as the lowest level of evidence. Because the checklist is based on what professionals reported to find relevant in work ability assessment of long-term sick-listed MDD employees, the checklist can be considered as their professional standard. Trying to assess work ability in terms of the checklist will result in more transparency and, because of that, better quality, than explaining the judgements of work ability in terms of patients’ expressions.

It can be concluded that the use of a checklist of aspects of work ability results in higher work ability ratings but does not diminish the variation in judgements of work ability. The assessment of work ability of long-term sick-listed MDD employees by IPs contains substantial variation and is moderately reliable between raters.

It is recommended that the community of IPs establishes how to use the items of the checklist best in practice with regard to assessing work ability and instruct the users of the checklist more precisely in this respect.
References


