Trying to Optimise the German Version of the OPTION Scale Regarding the Dyadic Aspect of Shared Decision Making

H. Keller¹; O. Hirsch¹; M. Müller-Engelmann¹; M. Heinzel-Gutenbrunner²; T. Krones¹,³; N. Donner-Banzhoff¹

¹Department of General Practice/Family Medicine, Philipps University Marburg, Marburg, Germany; ²Department of Child and Adolescent Psychiatry and Psychotherapy, Philipps University Marburg, Marburg, Germany; ³Clinical Ethics, University Hospital Zurich, Zurich, Switzerland

Keywords
Decision making, weights and measures, patient participation, cross-over studies, physician-patient relations

Summary
Objectives: The OPTION scale (“observing patient involvement in decision making”) assesses the extent to which clinicians involve patients in decisions across a range of situations in clinical practice. It so far just covers physician behavior. We intended to modify the scoring of the OPTION scale to incorporate active patient behavior in consultations.

Methods: Modification was done on scoring level, attempting a dyadic, relationship-centred approach in that high ratings can be evoked also by the behaviour of active patients. The German version of the OPTION scale was compared with a modified version by analysing video recordings of primary care consultations dealing with cardiovascular prevention. Fifteen general practitioners provided 40 videotaped consultations. Videos were analysed by two rater pairs and two experts in shared decision making (SDM).

Results: Reliability measures of the modified version were lower than those of the original scale. Significant associations of the dichotomised scale with the expert SDM rating as well as with physicians’ expertise in SDM were only found for the modified OPTION scale. Receiver Operating Characteristic (ROC) analyses confirmed a valid differentiation between the presence of SDM (yes/no) on total score level, even though the cut-off point was quite low. Standard deviations of the single items in the modified version were higher compared to the original OPTION scale, while the means of total scores were similar.

Conclusions: The original OPTION scale is physician-centered and neglects the activity and a possible self-involvement of the patient. Our modified instruction was able to capture the dyadic element partially. The development of a separate dyadic instrument might be more promising.

Correspondence to:
Dr. Oliver Hirsch, PhD
Department of General Practice/Family Medicine
Philipps-University Marburg
Karl-von-Frisch-Strasse 4
35032 Marburg
Germany
E-mail: oliver.hirsch@staff.uni-marburg.de

© Schattauer 2013

1. Introduction

Patients prefer to be involved in decisions about their medical care [1–4]. However, patient involvement still needs to be explored in order to ensure that decisions are not solely made by health professionals without involving their patients into the decision-making process [5–7]. One internationally accepted instrument that allows measuring to what degree clinicians involve patients in decision making is the OPTION (observing patient involvement in decision making) scale [8–9]. The OPTION scale was developed from a skills framework and consists of a set of competences [10] which include problem definition, explaining legitimate choices, portraying options, communicating risk, and conducting the decision process or its deferment. We examined the German version and found it to be reliable and valid at total score level [11]. However, although the developers claim the instrument depicts ‘shared decision making’ (SDM), a better description would have been ‘physician involvement in decision making’, since patient behaviour is not captured. More emphasis is put on the physicians’ side, highlighting the need for eliciting and responding to patients’ understanding and their preferences. Thus, the OPTION scale quantifies observable skills and aims to achieve an overall ‘involvement’ scale at clinician level only. As a result, a limitation of the instrument might be that in a situation in which a patient is very actively participating in decision making, the OPTION...
score may appear to be low, while a high level of shared decision making is reached [12]. We therefore modified the German version of the OPTION scale on scoring level, attempting a dyadic, relationship-centred approach to investigate active involvement of patients by analysing video recordings of primary care consultations dealing with cardiovascular prevention, e.g. prescription of statins, drugs to lower blood pressure, lifestyle interventions like diet changes, increased physical exercise.

2. Methods

In this section we describe our sample, the two used versions of the OPTION scale, the rating procedure, and statistical methodology.

2.1 Sample

Fifteen General Practitioners (GPs) provided 45 videotaped consultations with patients in whom discussion of cardiovascular risk and of preventive measures seemed indicated. Exclusion criteria were other index problems, restrictions because of language, cognitive abilities, psychiatric disorder, and severity of somatic disease. Participating GPs were 5 females and 10 males whose age ranged from 44 to 56 years. Eight physicians had participated in the preceding randomised controlled trial (RCT, Phase III) as intervention GPs, where they had received educational training in SDM. For details on the RCT please refer to Krones et al. [13] and Keller et al. [14].

This study is part of an extensive phase IV study, investigating patient participation in the shared decision-making (SDM) process in cardiovascular prevention. The study complies with the Declaration of Helsinki. The research protocol had been approved by the local research ethics committee, University of Marburg. Informed consent had been obtained from all participating General Practitioners (GPs) and patients.

2.2 The Instruments

The OPTION scale (acronym for "observing patient involvement") assesses the extent to which clinicians involve patients in decisions across a range of situations in clinical practice. The OPTION instrument was developed from a skills framework and consists of a set of physician competences like problem definition, explaining legitimate choices, portraying options, communicating risks and benefits and conducting the decision process or its deferment [1, 8–9, 15]. In the original OPTION instrument comprising 12 items, a five-point scale is used to assess the existence and characteristics of the clinician’s communication competence. Points 0–4 correspond to the observed behaviour (0 = not observed, 1 = minimal attempt, 2 = minimal skill level, 3 = good standard, 4 = high standard). The raw total score ranges from 0 (if level 0 is rated in all items) to 48 (if level 4 is rated in all items). Consequently, it is judged whether clinicians were able to involve patients in the decision making process. Details about how each scale point should be given to differing skill levels of competences observed are provided in a manual, available from the developers at www.optioninstrument.com. Patient involvement in decision making was measured by using the original OPTION-scale and a modified version. Modification was done on category "4" by replacing ‘high standard’ with ‘active involvement of patient is observed’, assuming that rating at a high standard necessarily should imply active involvement of the patient. The videotaped consultations were rated directly.

2.3 Procedure

The OPTION ratings on the basis of the videotapes were done by four experienced raters, all psychologists with MSc and/or PhD degrees, using a crossover design. All four raters had to rate each video once, using either the original or the modified version of the OPTION scale. Thereby, two rater pairs were formed (rater1–rater3 and rater 2–rater 4), who had to assess the same videos with the same OPTION version, respectively. The cross-over design (Figure 1) was necessary to avoid rating videos twice using both scales. Rating was done in random order. All four raters are research assistants with knowledge of the principles of SDM. In order to be prepared for their task, they had to undergo an extensive OPTION training including a calibration session for each rater pair to reach a consensus about their general rating performance. In addition to the OPTION rating, they also had to decide in each case whether SDM had taken place or not [11].

In a final step, rating data were compared with a reference standard for ‘shared decision making’ (binary variable – SDM present yes/no [11]). This was evaluated independently by two experts of the shared decision-making field, who reached a consensus on each video. Assessment was done according to SDM-process steps like agreeing on task, exploring subjective and objective risk, information and discussion about options, decision making, and plan for future actions [1, 13, 16].

Figure 1 Graphical display of the cross-over design. Black: original version of the OPTION scale. Grey: modified version of the OPTION scale.
2.4 Statistical Methodology

In an observer instrument like the OPTION scale measures have to be reliable, especially between different raters, and the scale has to be valid which means that it should measure the construct it was constructed for. Consequently, reliability and validity are central aspects.

2.4.1 Reliability

Agreement between rater pairs on item level was evaluated by the Wilcoxon test for dependent data and by the marginal homogeneity test. For a measure to be reliable, ratings between raters should not differ above certain thresholds. It is recommended that a meaningful difference between two raters occurs at a p value of .25 or smaller to reduce the probability of a beta error [17, 18]. We further calculated the effect size d to assess the magnitude of the differences as significance level alone is not an adequate criterion [19, 20].

The scoring of the OPTION scale lies between an ordinal and metric scale level. We therefore applied coefficients for different scale levels. Associations between the two rater pairs on item level were assessed by Spearman correlation coefficients, weighted kappas, and intraclass correlation coefficients (ICC) [17, 21, 22]. Unadjusted intraclass correlations above .40, .60 and .80 were considered to reflect fair, moderate and substantial agreement, respectively [23]. According to Cicchetti [24] weighted kappas smaller than .40 are taken as an indicator of poor agreement, between .41 and .59 they signal moderate agreement, between .60 and .74 they are interpreted as good agreement and kappa scores larger than .74 show very good agreement.

The association of the total scores between the raters was examined by Pearson correlation coefficients. Point-biserial correlation coefficients were calculated between the sum scores of the four raters and their respective dichotomized SDM ratings [21]. The internal consistency of the whole scale was inspected with Cronbach’s α.

2.4.2 Validity

We compared the dichotomized SDM expert consensus ratings with the dichotomized ratings of the four raters with kappa

---

Table 1  Means and standard deviations (mean (SD)) of the single items and the total scores of both OPTION scales listed separately for the four raters

<table>
<thead>
<tr>
<th>Item</th>
<th>rater 1 original modified</th>
<th>rater 2 original modified</th>
<th>rater 3 original modified</th>
<th>rater 4 original modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The clinician draws attention to an identified problem as one that requires a decision making process</td>
<td>3.74 (.56)</td>
<td>3.37 (1.17)</td>
<td>2.17 (1.25)</td>
<td>2.55 (1.70)</td>
</tr>
<tr>
<td>2. The clinician states that there is more than one way to deal with the identified problem (&quot;equipoise&quot;)</td>
<td>3.16 (1.21)</td>
<td>2.53 (1.54)</td>
<td>1.55 (1.36)</td>
<td>1.05 (1.28)</td>
</tr>
<tr>
<td>3. The clinician assesses the patient’s preferred approach to receiving information to assist decision making</td>
<td>.22 (.94)</td>
<td>.21 (.92)</td>
<td>.00 (.00)</td>
<td>.05 (.22)</td>
</tr>
<tr>
<td>4. The clinician lists options which can include the choice of &quot;no action&quot;</td>
<td>2.74 (1.28)</td>
<td>2.39 (1.50)</td>
<td>1.80 (1.36)</td>
<td>1.40 (1.39)</td>
</tr>
<tr>
<td>5. The clinician explains the pros and cons of options to the patient (taking &quot;no action&quot; is an option)</td>
<td>2.21 (1.44)</td>
<td>2.05 (1.51)</td>
<td>.80 (1.06)</td>
<td>.79 (1.08)</td>
</tr>
<tr>
<td>6. The clinician explores the patient’s expectations (or ideas) about how the problem(s) are to be managed</td>
<td>1.89 (1.45)</td>
<td>1.89 (1.45)</td>
<td>.15 (.37)</td>
<td>.15 (.49)</td>
</tr>
<tr>
<td>7. The clinician explores the patient’s concerns (fears) about how problem(s) are to be managed</td>
<td>1.74 (1.33)</td>
<td>1.74 (1.41)</td>
<td>.45 (.83)</td>
<td>.20 (.52)</td>
</tr>
<tr>
<td>8. The clinician checks that the patient has understood the information</td>
<td>2.29 (1.26)</td>
<td>2.32 (1.38)</td>
<td>1.84 (1.77)</td>
<td>1.11 (1.24)</td>
</tr>
<tr>
<td>9. The clinician offers the patient explicit opportunities to ask questions during the decision making process</td>
<td>3.68 (.48)</td>
<td>3.61 (1.15)</td>
<td>3.10 (1.12)</td>
<td>2.30 (1.38)</td>
</tr>
<tr>
<td>10. The clinician elicits the patient’s preferred level of involvement in decision making</td>
<td>1.06 (1.09)</td>
<td>1.47 (1.61)</td>
<td>.35 (.88)</td>
<td>.00 (.00)</td>
</tr>
<tr>
<td>11. The clinician indicates the need for a decision making (or deferring) stage</td>
<td>1.47 (1.58)</td>
<td>.84 (1.34)</td>
<td>.20 (.41)</td>
<td>.40 (.60)</td>
</tr>
<tr>
<td>12. The clinician indicates the need to review the decision (or deferment)</td>
<td>1.95 (1.68)</td>
<td>.47 (.96)</td>
<td>.35 (.81)</td>
<td>.70 (.13)</td>
</tr>
<tr>
<td>total score</td>
<td>24.64 (11.57)</td>
<td>22.41 (11.66)</td>
<td>12.50 (7.37)</td>
<td>11.00 (5.96)</td>
</tr>
</tbody>
</table>

Item scores range from 0 to 4 (0: the behaviour is not observed; 1: a minimal attempt is made to exhibit the behaviour; 2: the behaviour is observed and a minimum skill level achieved; 3: the behaviour is exhibited to a good standard; 4: active involvement of patient is observed).
coefficients. Point-biserial correlation coefficients were calculated between the total OPTION scores of both scales of the four raters and the dichotomized SDM expert consensus ratings [21].

We plotted the total OPTION scores of both scales of all four raters against the SDM expert consensus rating using Receiver Operating Characteristic (ROC) analysis [25] to search for a cut-off point that distinguishes between “SDM: yes” and “SDM: no”. Values larger than .7 are considered to show acceptable discriminability.

We further explored differences regarding patients’ characteristics [age, gender, education and cardiovascular risk] as well as physicians’ characteristics [age, gender and physicians’ expertise in SDM (as measured by the participation in special courses within our randomised controlled trial)] by comparing mean differences with t-tests and analysis of variance (ANOVA). Cardiovascular risk was assessed with the risk calculation part of arriba™, our decision aid for cardiovascular prevention following a global risk approach [13].

In a last step we examined the correlation of the two OPTION versions to the reference standard for SDM as well as to the physicians’ expertise in SDM as measured by the participation in special courses within our randomised controlled trial [χ²-test with effect size Cramer-V larger than .40 denoting a large effect [20], by dichotomising the two versions with regard to “patient involvement took place” [yes/no]. Thereby “no patient involvement” corresponds to ratings of “0–3” and “patient involvement corresponds to a rating of “4”. Calculations were done using PASW Statistics 18 (SPSS Inc., Chicago, IL, USA) and MedCalc 11.2.1.0 (MedCalc Software, Mariakerke, Belgium).

### 3. Results

#### 3.1 Descriptive Statistics

Of the 45 videos 5 had to be excluded since index problems different from cardiovascular risk had become central during the consultations (14 male and 26 female patients; 35 to 78 years). Table 1 depicts descriptive statistics of the four raters on item and total score level of both OPTION versions. We calculated raw total scores.

One can see in Table 1 that the mean differences on item level between the four raters are quite large. Most items are skewed towards minimal levels of SDM. The standard deviations of the single items were often higher compared to the original OPTION scale, while the means of the total scores were similar. The calibration process of rater pair 1–3 was different from the one of rater pair 2–4. This is especially obvious when looking at the differences of the total scores. Item 3 was almost never observed in all four raters.

<table>
<thead>
<tr>
<th>Item</th>
<th>rater 1 vs. rater 3 (p value)</th>
<th>rater 2 vs. rater 4 (p value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The clinician draws attention to an identified problem as one that requires a decision making process</td>
<td>&lt;.001</td>
<td>.13</td>
</tr>
<tr>
<td>2. The clinician states that there is more than one way to deal with the identified problem (“equipoise”)</td>
<td>.18</td>
<td>.004</td>
</tr>
<tr>
<td>3. The clinician assesses the patient’s preferred approach to receiving information to assist decision making</td>
<td>.32</td>
<td>1.00</td>
</tr>
<tr>
<td>4. The clinician lists options which can include the choice of “no action”</td>
<td>.49</td>
<td>.002</td>
</tr>
<tr>
<td>5. The clinician explains the pros and cons of options to the patient (taking “no action” is an option)</td>
<td>.87</td>
<td>.001</td>
</tr>
<tr>
<td>6. The clinician explores the patient’s expectations (or ideas) about how the problem(s) are to be managed</td>
<td>.65</td>
<td>.012</td>
</tr>
<tr>
<td>7. The clinician explores the patient’s concerns (fears) about how problem(s) are to be managed</td>
<td>.78</td>
<td>.006</td>
</tr>
<tr>
<td>8. The clinician checks that the patient has understood the information</td>
<td>1.00</td>
<td>.27</td>
</tr>
<tr>
<td>9. The clinician offers the patient explicit opportunities to ask questions during the decision making process</td>
<td>.01</td>
<td>.02</td>
</tr>
<tr>
<td>10. The clinician elicits the patient’s preferred level of involvement in decision making</td>
<td>.40</td>
<td>.03</td>
</tr>
<tr>
<td>11. The clinician indicates the need for a decision making (or deferring) stage</td>
<td>.001</td>
<td>.001</td>
</tr>
<tr>
<td>12. The clinician indicates the need to review the decision (or deferment)</td>
<td>.001</td>
<td>.006</td>
</tr>
</tbody>
</table>
3.2 Reliability

The reliability of the modified OPTION scale is described as differences and associations between raters and how consistent the scale is in itself.

3.2.1 Inter-rater Association

We first examined differences on the modified OPTION scale between the two rater pairs by Wilcoxon tests for dependent data and by marginal homogeneity tests as displayed in Table 2.

Table 2 shows that the distributions of the ratings especially differ between raters 2 and 4 with significant differences except for item 3 (preferred approach to receiving information). The agreement between raters 1 and 3 is higher with non-significant differences for items 3, 4 (displaying options), 5 (discussing pros and cons), 6 (exploring patient’s expectations), 7 (exploring patient’s concerns), 8 (checking understanding) and 10 (eliciting preferred level of patient’s involvement).

Effect sizes between rater 1 and rater 3 range between $d = .04$ and 1.37. Between rater 2 and rater 4 effect sizes vary between $d = .00$ and 1.54.

Associations between the two rater pairs on the modified OPTION scale were assessed by Spearman correlation coefficients and ICCs. Agreement was measured by weighted kappas. Results are displayed in Table 3.

Spearman correlation coefficients between raters 1 and 3 reflect moderate associations for items 1 (problem definition), 2 (equipoise), 4, 5 and 9 (opportunities for asking questions). Only low associations were found for items 7 and 11 (decision-making). ICCs signal moderate associations for items 2 and fair associations for items 4, 5, 8 and 10. Weighted kappas reveal a moderate agreement for item 5. For all other items the agreement is low.

Spearman correlation coefficients between raters 2 and 4 reflect high associations on items 2, 4 and 5 and a moderate association on item 1. All other associations are low. ICCs signal an almost moderate association only on item 4. All other associations are low. Weighted kappas reveal a low agreement on all items.

The Pearson correlation between the total scores of rater 1 and rater 3 was $r = .69$ ($p = .002$) and $r = .56$ ($p = .013$) between rater 2 and rater 4. They can be considered to reflect moderate associations.

The point-biserial correlations between the sum scores of the four raters and their respective dichotomized SDM ratings were
.50 (p = .03), .53 (p = .02), .79 (p < .001) and .80 (p < .001).

3.2.2 Internal Consistency

Cronbach-α of the whole scale based on the data of all four raters is .90. The corrected item-total correlation of item 3 was .17. The other corrected item-total correlations ranged from .43 to .82.

3.3 Validity

We calculated point-biserial correlations between the total scores of the four raters and the dichotomous SDM reference consensus rating. The correlations were r = .23 (p = .33), r = .43 (p = .06), r = .62 (p = .008) and r = .83 (p < .001).

In a next step we plotted the sum scores of all four raters against the SDM expert consensus rating using Receiver Operating Characteristic (ROC) analysis. The ROC curve is shown in Figure 2.

The area under the curve (AUC) was .83 (p < .001; 95% confidence interval: .73 to .91). At a cut-off of 11 points (23 points in the scaling with a maximum of 100) the sensitivity was 81.8 % and the specificity is 76.2 %, respectively.

After a median split of patients’ age, we found a significant difference in the total OPTION score (t-test, p = .001). Consultations with patients ≤ 63 years of age were rated lower (mean 12.05; sd 9.18) than consultations with patients ≥ 63 years of age (mean 19.14; sd 8.71). This results in a medium effect size of d = 0.79.

Physicians with more expertise in SDM, as measured by the participation in special courses within our randomised controlled trial [13], received higher OPTION ratings (mean 19.81; sd 9.6) than those with less expertise (mean 12.48; sd 8.27). This results in a high effect size of d = 0.79. For all other patient and physician characteristics no differences could be found.

Considering the dichotomised scales with regard to “patient involvement took place” [yes/no], correlations to the reference SDM rating as well as to physicians’ expertise in SDM were only found for the modified OPTION. In particular, significant associations between the reference SDM rating and the OPTION ratings were found for item 1 (p = .007, Cramer-V = .30) and item 9 (p = .007, Cramer-V = .30). Items 4 and 5 also displayed an acceptable Cramer-V of .28 and .29. Physicians with more expertise in SDM were rated significantly higher (level of significance adjusted for multiple testing at p < .004) with the modified version on item 2 (p = .007, Cramer-V = .30), item 4 (p = .004, Cramer-V = .32), and item 5 (p < .001, Cramer-V = .40).

4. Discussion

As recently published, we concluded that the OPTION instrument is a useful tool to measure the concept of SDM in the consultation. The German version is reliable and valid at total score level [11]. However, the original version of the scale is physician-centred and neglects the activity and a possible self-involvement of the patient. We therefore slightly modified the original OPTION version on scoring level, attempting a dyadic, relationship-centred approach to cover active patient involvement. Hence, we took a step further, assessing not only physicians’ competences and skills but also the acquisition of behavioural aspects. However, the instruction to rate category “4” also when patient participation could be observed, did not lead to higher OPTION scores. In contrast, the associations on total score level were not as high as in the original version. One reason for this might have been that the clear instruction made it easier for the raters to come to an unambiguous assignment. Thus, a rating of “4” might have been found less appropriate compared to the original version, where selectivity between ratings of “3” and “4” is much lower. The low scores therefore reveal that many physicians failed to encourage patients’ involvement in the decision-making process.

Nevertheless, the total score is still an acceptable parameter for further use. The moderate to high correlations between the sum scores of the four raters and their respective dichotomised SDM ratings highlight that the raters incorporated their OPTION ratings into their overall decision on whether SDM took place or not. The good discriminability of the OPTION total score was verified by the results of our ROC analyses. Area under the curve (AUC) values were high and confirmed that the total score is able to achieve a valid distinction between SDM present or not according to our expert raters, even though the cut-off point was quite low.

However, significant correlations of the dichotomised scales to the expert SDM rating as well as to physicians’ expertise in SDM were only found for the modified OP-
TION scale, where “4” clearly was defined in contrast to the other levels. We observed high associations between the dichotomised SDM ratings of the four raters and their respective total OPTION scores. This highlights that the raters incorporated their OPTION ratings into their overall decisions and thus serve as a quasi-validation of the whole scale.

The internal consistency of the scale measured by Cronbach-α is as high as in the original version which also confirms research results of other validation studies [8, 9, 26].

These findings, in addition with our results concerning the low inter-rater reliability, lead to the assumption that there is a necessity of a more precise explanation of all observation levels of the OPTION scale in order to achieve higher agreement between raters. In addition, SDM needs to be seen as a flexible approach for patient participation with different applicabilities and with variable patient preferences. This goes in line with the findings of Goossensen et al. [12] and Brown et al. [27]. In their systematic review on conceptual definitions of SDM, Makoul and Clayman [28] stated that there is no common definition so far and proposed a model that consists of the essential practical elements of SDM integrating concepts from the existing literature. SDM needs to be flexible because of many different potential situations in which it can be applied but determined enough in order to be operationalised. Micro aspects of behavior and cognition have to be formulated more concrete to be observable by raters or perceivable by patients or physicians in self-rating instruments.

Although the term ‘dyadic analysis’ is relatively novel in the medical encounter, there is a growing interest in the concept of interdependence in health communication research [29]. Brown et al. [27] found that patients who are successfully engaged in thinking through options and making decisions regarding their treatment have more positive outcomes than patients who simply follow orders without being actively involved. On the other hand one has to consider that patients might be more reluctant regarding negative aspects about their health [30]. Melbourne et al. [31, 32] recently described the development process of a dyadic OPTION scale. The original items of the OPTION scale were reformulated as questionnaire items so that they can be administered to patients and physicians immediately after the consultation. Consequently, the indirect assessment via observers was no longer necessary. This should be one of the topics of future research regarding the OPTION scale to further refine it. Other possible dyadic measures in shared decision making were also examined [33].

4.1 Strengths and Limitations

In contrast to the other reliability studies, we only had one index problem, namely cardiovascular prevention. Participating physicians can be considered to have a certain expertise in SDM as they had received educational training in SDM. The size and the range of our study sample might have been too small to support our conclusions. We used videotaped consultations so that the raters had additional information in the form of nonverbal behaviour (e.g. gestures, facial expressions) which might have led to increased variance between the raters. The data were partly dependent because each physician provided up to three consultations. Aspects of nonverbal communication were captured as additional information by rating videos which might have resulted in a fuller picture of the interaction. We were able to demonstrate that single items of the OPTION scale and categorical explanations need further refinement to achieve higher inter-rater agreement and higher inter-rater associations.

Expert ratings can be regarded as a quasi-validation for the scale.

5. Conclusion

We conclude that the modified German version of the OPTION instrument has more unambiguous psychometric characteristics than the original version. The inter-rater reliability is not appropriate on item level but is still reliable and valid at total score level. Our modified instruction was not able to capture the dyadic element completely and is therefore not generalizable. A possible self-involvement of the patient was still difficult to rate with the slightly modified OPTION scale. Therefore, an approach to develop a special version of a dyadic OPTION scale might be more promising.

Acknowledgments

Grateful thanks to Andreas Loh who provided training in rating the OPTION scale. We acknowledge with thanks the assistance of our study coordinator Gesa Kix. We would also like to thank all participating GPs and patients. Grateful thanks also to Christina Albohn-Kühne and Viktoria Bachmann for their OPTION ratings.

References