doi: 10.1111/j.1369-7625.2009.00584.x

Confirmatory factor analysis and recommendations for improvement of the Autonomy-Preference-Index (API)

Daniela Simon PhD Dipl Psych,* Levente Kriston PhD Dipl Psych,† Andreas Loh PhD Dipl Psych,‡ Claudia Spies MD,§ Fueloep Scheibler PhD Dipl Sot,¶ Celia Wills PhD RN** and Martin Härter MD PhD Dipl Psych††,‡‡

*Research Associate, Section of Clinical Epidemiology and Health Services Research, University Medical Center, Freiburg, Germany, †Research Associate, Department of Medical Psychology, University Medical Center, Hamburg, Germany, ‡Research Associate, Department of Family Medicine, University Medical Center, Freiburg, Germany, §Professor, Department of Anaesthesia and Intensive Care, University Hospital of Berlin, Charité, Germany, ¶Research Associate, Institute for Quality and Efficiency in Health Care (IQWIG), Cologne, Germany, **Professor, College of Nursing, The Ohio State University, Columbus, OH, USA, ††Professor,Section of Clinical Epidemiology and Health Services Research, University Medical Center, Freiburg, Germany and ‡‡Professor, Department of Medical Psychology, University Medical Center, Hamburg, Germany

Correspondence
Daniela Simon PhD, Dipl. Psych.
Section of Clinical Epidemiology and Health Services Research
University Medical Center
Hauptstr. 5
79104 Freiburg
Germany
E-mail: Daniela.Simon@uniklinik-freiburg.de

Accepted for publication
4 October 2009

Keywords: confirmatory factor analysis, measurement, patient preferences, validation study

Abstract

Objective Validation of the German version of the Autonomy-Preference-Index (API), a measure of patients' preferences for decision making and information seeking.

Methods Stepwise confirmatory factor analysis was conducted on a sample of patients (n = 1592) treated in primary care for depression (n = 186), surgical and internal medicine inpatients (n = 811) and patients with minor trauma treated in an emergency department (n = 595). An initial test of the model was done on calculation and validation halves of the sample. Both local and global indexes-of-fit suggested modifications to the scale. The scale was modified and re-tested in the calculation sample and confirmed in the validation sample. Subgroup analyses for age, gender and type of treatment setting were also performed.

Results The confirmatory analysis led to a modified version of the API with better local and global indexes-of-fit for samples of German-speaking patients. Two items of the sub-scale, 'preference for decision-making', and one item of the sub-scale, 'preference for information seeking', showed very low reliability scores and were deleted. Thus, several global indexes-of-fit clearly improved significantly. The modified scale was confirmed on the validation sample with acceptable to good indices of fit. Results of subgroup analyses indicated that no adaptations were necessary.

Discussion and conclusions This first confirmatory analysis for a German-speaking population showed that the API was improved by the removal of several items. There were theoretically plausible explanations for this improvement suggesting that the modifications might also be appropriate in English and other language versions.

Introduction

Patient participation in decision making has become increasingly important especially for diseases for which more than one treatment option exists and the best choice depends on how a person values the benefits and harms of each option.[1] While most patients want detailed information and physicians often underestimate this need, far fewer patients indicate a preference to participate in decision making.[2–5] Patients' preferences for participation are not fixed and can vary depending on several factors, such as age, sex, education, experience of illness and medical care, health status, type of decision, attitude towards decision making, relationship with the physician and preference for information.[6,7] Yet, socio-demographic and disease-related variables have been shown to account for only 6.9% of the variance in participation preferences.[8] In addition, existing measures of preferences are not necessarily applicable across diverse treatment and socio-cultural contexts.[2,9] Additional research is needed on patients' preferred roles in decision making,[7] as well as measures of preferences that are appropriate for different treatment and cultural contexts.

One scale to measure patients' preferences for participation and information is the Autonomy-Preference-Index (API).[10] It was developed in a modified Delphi study involving 13 clinicians, medical sociologists and ethicists. Within the process of two iterations, key measurable dimensions of patients' preferences for autonomy were identified.[10] Patients' preference for participation and their desire for information emerged as the most important ones to discriminate between patients who prefer an active role in their care and patients who desire a more passive role. The two dimensions were operationalized by questionnaire items representing the API and were then field-tested and reviewed with patients for content validity.

The API consists of 23 items representing two sub-scales (Table 1). Eight items evaluate the preference for information, and six items measure the preference for participation. Each of the 14 items is rated using a five-point Likert scale, with response options ranging from 'strongly disagree' to 'strongly agree'. The agreement is associated with preference for information or involvement, respectively (items dm4, dm6 and is5 are reverse-scored). The participation preference is additionally assessed in relation to illness severity by

Table 1 Items of the Autonomy-Preference-Index without vignettes

	Items	
Decision-making sub-scale		
1	The important medical decisions should be made by your doctor, not by you	dm1[1]
2	You should go along with your doctor's advice even if you disagree with it	dm2
3	When hospitalized, you should not be making decisions about your own care	dm3
4	*You should feel free to make decisions about everyday medical problems*	*dm4*
5	If you were sick, as your illness became worse you would want your doctor to take greater control	dm5
6	*You should decide how frequently you need a check-up*	*dm6*
Information-seeking sub-scale		
7	As you become sicker you should be told more and more about your illness	is1[2]
8	You should understand completely what is happening inside your body as a result of your illness	is2
9	Even if the news is bad, you should be well informed	is3
10	Your doctor should explain the purpose of your laboratory tests	is4
11	*You should be given information only when you ask for it*	*is5*
12	It is important for you to know all the side effects of your medication	is6
13	Information about your illness is as important to you as treatment	is7
14	When there is more than one method to treat a problem, you should be told about each one	is8

Items in italics were removed in the modified version owing to very low reliability scores.
[1]dm, preference for participation in decision making.
[2]is, preference for information.

three clinical vignettes (three items each) on upper respiratory tract illness (representing mild disease), hypertension (moderate disease) and myocardial infarction (severe/most threatening disease). Using the vignettes patients are asked to hypothetically consider their participation preferences for different stages of disease severity. The response options for each vignette range from 'you alone', 'the doctor and you', to 'the doctor alone' should make the treatment decision.

The API was originally validated in a sample of English-speaking patients in general medicine. Factor analysis supported the clustering of items into information ($\alpha = 0.82$) and participation sub-scales ($\alpha = 0.82$).[10] The three vignettes were also analysed as separate sub-scales to investigate the effect of increasing illness severity upon patients' preferences. Scores for each scale and the vignettes are summed and linearly adjusted to a range of 0–100 for the API and 0–10 for the vignettes, respectively. Construct validity has been reported only for the participation sub-scale and was reported to have a significant positive correlation with a global item on participation ($r = 0.54$; $P < 0.0001$).[10] Convergent validity was assessed with a known group analysis: diabetic patients with high motivation for self-care and home monitoring of blood glucose had higher mean API scores compared with a sample of the general population ($P < 0.01$). No association was found between the information and participation sub-scales, supporting the orthogonality of the concepts measured by the sub-scales. Test–retest reliability after 2 weeks (Pearson r) was 0.84 for the participation sub-scale and 0.83 for the information sub-scale indicating strong stability of preferences over a 2-week time period. Internal consistency was good with Cronbach's α of 0.82 for each sub-scale.[10]

Apart from the original validation study, internal consistency has been reported previously by Spies *et al.*[11] on a German sample of chronic pain patients (participation: $\alpha = 0.61$, information: $\alpha = 0.73$) and patients in a pre-medication visit (participation: $\alpha = 0.64$; information: $\alpha = 0.66$). The authors describe that deleting the reverse-scored items increased the Cronbach's α values. Another German study reported internal consistency for the participation sub-scale ranging from 0.57 to 0.80 and recommended scale adaptations for sub-groups.[12] Furthermore, one study on a large sample of the Japanese population reported low alpha reliability coefficients (< 0.70).[13] These findings suggest additional testing of the API for groups other than native English-speaking populations. Other studies have also used modified versions of the API, in which they either added items or used selected items of the sub-scales, but did not report reasons for their modifications.[14–16] Many studies using an unmodified version of the API did not report results of psychometric testing.[17–19] In these cases it remains unclear if testing was not done, or was done but not reported owing to unfavourable outcomes or other reasons.

In addition, the psychometric properties of the API have not been more formally tested using theory-driven confirmatory factor analysis (CFA), the best means of testing the agreement between a theoretically formulated model and empirical data, neither in the original version nor in one of the adaptations. With this approach, the coherence of the sub-scales and the underlying constructs could be thoroughly investigated.

An opportunity to do further psychometric testing of the API occurred when a German version of the API was used in a research consortium on shared decision making funded by the German Ministry of Health.[20] From 2001 to 2005, the implementation of shared decision making throughout Germany was investigated in 10 projects on different medical conditions such as breast cancer, depression and hypertension.[21] The overall aim of the research consortium was to study intervention effects resulting from physician training programmes in shared decision making, the use of patient information material and patient decision aids. To create a basis for common evaluation strategies, members of the 10 projects joined a comprehensive team on methodological issues.[20] The API was one of the evaluation measures chosen by the methods team. It was translated into German by two independent researchers and back-translated by two native speakers, and the translated version was authorized by the original authors of the API.

The aim of this study was to extend the previous preliminary testing of the original factorial structure of the API for the German translation, using CFA on a mixed sample of patients from primary care, inpatient care and emergency department settings.

Methods

Study sample

The API was tested for validity and reliability on a sample of $n = 1592$ patients treated for depression in primary care ($n = 186$) or for minor traumas in an emergency department ($n = 811$). These patients participated in research projects within the research consortium. Depressed patients were subjects in a cluster randomized controlled trial on the effects of training in shared decision making for general practitioners. The data used in the present study were collected after a consultation but prior to this intervention. Patients in the emergency department sample received treatment of minor traumas at a university hospital emergency department and were screened for risky alcohol consumption. Preferences for information and participation were assessed in the context of decision making regarding behaviour changes in alcohol consumption. There was also access to data from a cooperating project in which a survey among patients receiving inpatient care on surgical and internal medicine units was conducted ($n = 595$). All studies were approved by the respective local ethic boards and are described in detail elsewhere.[11,22,23] The API vignettes on illness severity levels were not used in any of these studies, so this validation study only includes the preference for decision making and information sub-scales of the API.

Procedure

The sample was randomly divided into half to generate a calculation sample and a validation sample. The analysis was performed in four steps. First, the original two-factor model was tested on the calculation sample. Second, modifications were made as suggested by local and global indexes-of-fit for the original scale. Third, the adapted model was verified on the calculation sample. Fourth, a confirmation of the adapted model was performed on the validation sample. In addition to these steps, subgroup analyses for age, sex and setting were carried out to assess the performance of the modified API across settings and for varying age and gender groups.

Main outcome measures

To test the two-factor structure, both local and global goodness-of-fit measures were calculated. The following local *goodness-of-fit indices* (GFIs) were used to demonstrate how well the defined constructs were measured: indicator reliability, critical ratios, construct reliability and average variance extracted. Global GFIs evaluating the whole model with regard to both measurement and factorial aspects included: the *discrepancy chi-squared test, root mean square error of approximation* (RMSEA), GFIs, *adjusted goodness-of-fit-index* (AGFI), *comparative-fit index* (CFI) and *Bayesian information criterion* (BIC).[24,25] GFIs were assessed by comparing them with recommended thresholds for acceptable and good fit.[24,26,27] The exact thresholds are reported in Tables 4–6.

Statistical analysis

CFA was used to test the two-factor model defined by the original authors of the API.[10] Maximum likelihood estimation was used to analyse the covariance of items. Constructs were scaled indirectly by fixation of loadings of specific indicators. The inter-correlation of both factors was allowed. Up to three missing values were imputed using the expectation-maximization algorithm. Data imputation was needed for 0.8% of data points in outpatients, $<0.1\%$ in emergency room patients and 0.5% in inpatients. The CFA was performed with AMOS 6, a software for structural equation modelling (SEM).[28] Missing values were imputed using a programme for multiple imputation of incomplete multivariate data (NORM 2.03, Pennsylvania University, PA, USA).[29]

Table 2 Sample characteristics

	Calculation sample	Validation sample	Total sample
n	794	798	1592
Age (in years)			
Mean	47.4	48.1	47.7
SD	18.1	18.6	18.4
Sex			
Male	393 (49.5%)	433 (54.3%)	826 (51.9%)
Female	401 (50.5%)	365 (45.7%)	766 (48.1%)
Setting			
Outpatient	94 (11.8%)	92 (11.5%)	186 (11.7%)
Inpatient	313 (39.4%)	282 (35.3%)	595 (37.4%)
Emergency	387 (48.7%)	424 (53.1%)	811 (50.9%)

Results

Sample characteristics

Table 2 presents the overall sample characteristics. For the total sample, patients' mean age was 47.7 years (SD = 18.4), and 51.9% of the patients were men; 11.7 % were primary-care outpatients, 37.4 % were inpatients and 50.9 % were emergency department patients.

Statistical assumptions of structural equation modelling (SEM)

Descriptive statistics for the API are presented in Table S1. No correlations between items above 0.8 were found, suggesting non-redundancy of the included scale items. In some cases for the information sub-scale items (is1, is2, is3, is4 and is8), skewness and kurtosis values proved to be extremely high; thus, the findings regarding these items could be slightly biased (is1: 'As you become sicker, you should be told more and more about your illness', is2: 'You should understand completely what is happening inside your body as a result of your illness', is3: 'Even if the news is bad, you should be well informed', is4: 'Your doctor should explain the purpose of your laboratory tests' and is8: 'When there is more than one method to treat a problem, you should be told about each one').

Test of the original model and modifications (calculation sample)

Local goodness-of-fit

Indicator reliability exceeded 0.30 for most of the items, with the exception of certain decision-making and information-seeking sub-scale items (dm4: 'You should feel free to make decisions about everyday medical problems'; dm6: 'You should decide how frequently you need a check-up'; is5: 'You should be given information only when you ask for it'; and is6: 'It is important for you to know all the side effects of your medications'). The score of item is6 (0.23) is still rather close to the recommended threshold of 0.30. *Critical ratios* showing correlations between indicators and constructs were significant for all items but item dm6. *Construct reliability* for both sub-scales reached the necessary threshold (0.74, 0.75), but the *average variance extracted* remained below the required value of 0.5, with scores 0.41 and 0.29, respectively (Table S2). No correlation was found between the two domains (–0.02).

Global goodness-of-fit

The *discrepancy chi-squared test* exceeded the threshold and was statistically significant indicating differences between theoretical and observed relations. The RMSEA (0.085), a measure of approximate fit in the population showed only a mediocre fit. Only the GFI (0.909) reached a good fit; all other indices reached acceptable fits (AGFI: 0.875; *normed-fit index* or NFI: 0.873; *Tucker–Lewis index* or TLI: 0.867; CFI: 0.889). The BIC showed a score of 700.39 (Table 3).

Model modification

Because the decision-making sub-scale items dm4, dm6 and information sub-scale item is5 showed only poor communalities and most global indexes did not show a good fit, these items were removed and the adapted model was tested on the calculation sample (Table 3). The deletion of these items was also supported

Table 3 Measures of global fit for all models estimated

	χ^2	d.f.	*P*	χ^2/d.f.	RMSEA	GFI	AGFI	NFI	TLI	CFI	BIC
Thresholds for acceptable fit[1]			≥0.05	≤5.00	≤0.08	≥0.80	≥0.85	≥0.80	≥0.80	≥0.85	Small
Thresholds for good fit[1]			≥0.10	≤3.00	≤0.05	≥0.90	≥0.95	≥0.95	≥0.95	≥0.95	Small
Original model (calculation sample)	506.76	76	<0.001	6.67	0.085	0.909	0.875	0.873	0.867	0.889	700.39
Adapted model (calculation sample)	115.89	41	<0.001	2.83	0.048	0.974	0.958	0.969	0.973	0.980	282.82
Adapted model (validation sample)	146.74	41	<0.001	3.58	0.057	0.967	0.947	0.965	0.965	0.974	313.80

[1]Recommendations are based on Kline,[24] Byrne,[25] Kriston,[26] Hu,[27] Browne[32] and Schermelleh-Engel.[33]
RMSEA, root mean square error of approximation; GFI, goodness-of-fit index; AGFI, adjusted goodness-of-fit index; NFI, normed-fit index; TLI, Tucker–Lewis index; CFI, comparative-fit index; BIC, Bayesian information criterion.

by the fact that they were reverse-scored and may therefore lead to different response patterns compared with the non-reversed items; that is, it is possible that patients were confused by the varying wording (negative vs. positive) within the same overall scale. Because the test of the original model showed only a very small non-significant correlation between the factors representing the two API sub-scales, correlations were set to 0 in the adapted model.

Analysis of the adapted model (calculation sample)

Local goodness-of-fit

Indicator reliability scores for the decision-making preference sub-scale did not change based on item deletions. Thus, it can be assumed that the deleted items did not significantly contribute to this scale. For the information sub-scale, some reliability scores slightly increased whereas others decreased. The is6 item, 'It is important for you to know all the side effects of your medications', still showed the weakest score (0.17). All *critical ratios* proved to be significant. Compared with the original model, *construct reliability* of the adapted model increased from acceptable to good scores (DM: 0.85; IS: 0.86). The overall amount of variance in the indicators accounted for by the construct (*average extracted variance*) also improved, with a good score for the decision-making sub-scale (0.60) and an acceptable score for the information-seeking sub-scale (0.47; Table S2).

Global goodness-of-fit

For the adapted version of the API both the discrepancy chi-squared test and the RMSEA decreased significantly in the desired direction (χ^2/d.f.: 2.83, RMSEA: 0.048). Moreover, GFI (0.974), AGFI (0.958), NFI (0.969), TLI (0.973) and CFI (0.980) all increased and met the threshold for good fit. Compared with the original model, the BIC (282.82) decreased clearly in favour of the adapted model (Table 3). Owing to the fact that one item (is6: 'It is important for you to know all the side effects of your medications') of the information-seeking sub-scale showed a rather mediocre association with the construct, we also tested a model without this item. This led to a slight improvement of the global fit indices (e.g. χ^2/d.f. = 2.72 and CFI: 0.984). However, as we judged this improvement to be negligible and we did not have any theoretical reason to exclude this item, we decided to keep it within the model.

Confirmation of the adapted model (validation sample)

Local goodness-of-fit

Testing the adapted version on the validation sample resulted in acceptable *indicator reliability* scores for most of the items. Only the score of item dm5 ('If you were sick, as your illness became worse you would want your doctor to take greater control') decreased slightly (0.27) but still remained close to the required threshold of 0.30. While most of the items of the sub-scale 'information seeking' increased in terms of indicator reliability, item is6, 'It is important for

you to know all the side effects of your medications', showed the lowest score (0.21). Once again all *critical ratios* were statistically significant. Scores of *construct reliability* (0.83 and 0.86) and of the *average extracted variance* (0.57 and 0.47) were almost as high as in the adapted model tested on the calculation sample (Table S2).

Global goodness-of-fit

In terms of global indexes, the discrepancy chi-squared test (χ^2/d.f.: 3.58) was slightly higher than in the calculation sample but still acceptable. The RMSEA (0.057) was clearly in the range for acceptable fit. All other tested global indexes-of-fit likewise remained close to the scores of the adapted model (GFI: 0.967; AGFI: 0.947; NFI: 0.965; TLI: 0.965; CFI: 0.974). With a score of 313.80, the BIC was almost as low as in the calculation sample and thus confirmed the good fit of the adapted model (Table 3).

Subgroup analyses (age, sex, setting)

To test the suitability of the adapted scale for subgroups of age, sex and setting global indexes-of-fit were considered. The scores were only slightly different from the validation sample and all reached the thresholds for acceptable fit (Table S3).

Discussion

A CFA of the German version of the API was performed in patients in primary care of depression, internal and surgical inpatient care and emergency care settings. Central conditions for CFA were met to perform a formal psychometric analysis of the modified API: (i) the sample size was beyond $n = 200$; (ii) skewness and kurtosis were acceptable for most items; however, items is1, is2, is3, is4 and is8 exceeded the recommended limit for skewness and item is8 was beyond the recommended limit for kurtosis; (iii) sufficient correlations between items were present to enable a structure definition.[24] Analysis of local GFIs indicated that the variance in items dm4 ('You should feel free to make decisions about everyday medical problems'), dm6 ('You should decide how frequently you need a check-up') and is5 ('You should be given information only when you ask for it') could not be sufficiently explained by the underlying constructs. Therefore, these items were deleted. Two of them belong to the preference for decision-making sub-scale and address the concepts of 'everyday medical problems' and 'medical check-up'. Further analyses are necessary to find out why items on these topics do not fit the sub-scale construct. It is possible that patients had a special medical condition on their mind when they answered the API which did not match with these rather general topics. The deleted item of the information-seeking sub-scale ('You should be given information only when you ask for it') is the only aspect of active information seeking from the patient's perspective. All other information-related items describe what patients should be told by physicians and what they should know and understand. Thus, it is possible, that the behaviour of taking active steps to receive information does not fit the construct because patients rather expect a passive role of being informed. If this is the case, the sub-scale label 'information seeking' should be revised. It is also possible that respondents had to change their mind when answering the question related to active information seeking. Further in-depth analyses of the construct would be helpful in clarifying this question. Future research might incorporate qualitative methods to explore the reasons patients report for their responses to the scale items. This type of data could illuminate people's frames of reference for responding to the questions, including if they have a particular medical condition or situation in mind when responding to the items. For example, the reason why item is6 ('It is important for you to know all the side effects of your medication') showed the poorest indicator reliability could be related to the fact that not all patients answering the API were taking medications, so it is possible that they found this item to be irrelevant or were unsure how to respond to it.

In some instances, the wording of the items may inspire a shift in frame of reference from a

specific to a general medical treatment context or *vice versa*. Interviewing patients about the reasons for their responses could also help determine if the reverse-scoring of items (that were deleted from the API) was confusing to patients or induced a different frame of reference when responding. In a future study we will investigate if the reverse-scoring has an influence on the adequacy of item fit in confirmatory model testing. For this purpose, items with reverse-scoring will be rephrased and then re-tested as part of the complete API.

In terms of construct reliability, the original model already showed acceptable scores, but we were able to improve the fit of the model by deleting certain items from the API. In contrast, the extracted variance was not satisfactory in the first model, and clearly benefited once these items were deleted from the model. This result relates to the findings of both Spies *et al.*[11] and Hamann *et al.*,[12] who reported acceptable scores of construct reliability for different medical conditions within the German Research Consortium. As long as the focus is only on internal consistency, acceptable scores indicate that the scale can be used. Furthermore, the inclusion of several items with an acceptable item-total correlation can outbalance the effect of items with low item-total correlation, thus providing good internal consistency overall. However, our analysis shows the specific ways the API can be improved by deleting certain items. This is potentially advantageous in terms of reducing respondent burden by reducing the number of items in the scale. Taking into account the extracted variance which indicates how much indicator variance is contributed by the construct can lead to a different view on the psychometric quality. Spies *et al.*[11] reported that item deletion increased the alpha score. In our study on the comprehensive sample, the alpha coefficients (factor reliabilities) increased from 0.74 to 0.85 for the decision-making sub-scale, and from 0.75 to 0.86 for the information-seeking sub-scale (see Table S2). However, the beneficial impact of item deletion was much higher on the extracted variance. Taking both indices into account thus led to a better understanding of the properties of the sub-scales. Furthermore, the scale improvement by deleting items is also supported by global indexes-of-fit. Thus, the adapted API could be confirmed and cross-validated in this study.

Another interesting result was that the underlying factors representing the sub-scales of preferences for decision making and information seeking were not correlated (−0.02). Initially, correlations were allowed in our analysis, but after this result we set them to zero. This turned out to be well supported by good indexes-of-fit. Previous findings of the original authors also reported no correlation between the two sub-scales.[10]

In addition to these results, the adapted version of the API was supported by subgroup analyses for age, sex and setting showing acceptability to good indexes-of-fit. As the score variations are only small, there is no need for further model adaptation in any of the sub-samples.

The rather small size of the outpatient sample ($n = 186$) represents one limitation of this study, as SEM should be performed on at least $n = 200$ patients. The results might also be limited by extreme scores of skewness and kurtosis that were found for some items of the information-seeking sub-scale. The results of this confirmatory analysis only apply to the German version of the API. Differences in language, health-care systems and cultures could also contribute to the bad model fit of several items, a fact that has been discussed by Scheibler *et al.*[30] with regard to the German validation of the Perceived Involvement in Care Scale (PICS). Although the number of items has been reduced, the scores can still be compared with API scores of other studies, as scoring relies on linear transformation to create a standardized sum score of 0–100, independent of the number of items contributing to the raw sum score. As the deleted reversed items did not show a relevant association with the constructs measured, the exclusion of these items did not change the meaning of the constructs, thus allowing the comparison of findings obtained with either the original or modified API sub-scales. Our analysis adds to the literature on the psychometric properties of the API by being

the first reported CFA of the API for a non English-speaking population.

Conclusion

The psychometric properties of the German API were investigated using SEM which led to an improvement by the removal of several items. As there were theoretically plausible explanations for this improvement which might also be relevant for English and other language versions, we encourage further tests of both the original and adapted versions of the API for measurement of patients' preferences for decision making and information seeking. This example of validating a translated measure indicates that a closer look at the suitability of measures in different medical conditions, settings and languages is of high value. When using the API for cross-cultural studies, it should be pre-tested in each country as psychometric properties of an instrument cannot be assumed to be cross-culturally stable on an *a priori* basis.[31] Future studies will show if reverse scoring has an impact on how well items fit the constructs. Although there is still a great need for future studies on patients' preferences for participation in decision making and information seeking, the preferences of German patients can be measured more reliably with this adaptation of the German API without impairing its validity.

Financial disclosure

This study was funded by a grant from the German Ministry of Health and Social Security (grant # 217-43794-5/6). More information can be found on the website www.shared-decision-making.org.

Acknowledgements

The authors thank all further members of the German research consortium for their continuous contribution to the work in the methods team by name Christiane Bieber, Anja Deinzer, Nadja Giersdorf, Johannes Hamann, Jürgen Kasper, Karena Leppert, Knut Müller, Tim Neumann, Hanna Rohlfing, Birgit Steuernagel, Birgitt van Oorschot and Edith Weiss-Gerlach as well as the principal investigators Reiner Anselm, Thorsten Doering, Wolfgang Eich, Christoph Heesen, Werner Kissling, Roland Schmieder, Michael Untch and Peter Zysno. Finally, they thank all the patients for their willingness to fill in the questionnaires.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table S1. Pearson product-moment correlation matrix and univariate descriptive statistics of items in the calculation sample.
Table S2. Local goodness-of-fit indexes.
Table S3. Measures of global fit for the subgroups sex, age and setting in the validation sample.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

References

1 O'Connor AM, Wennberg JE, Légaré F *et al*. Toward the 'tipping point': decision aids and informed patient choice. *Health Affairs*, 2007; **26:** 716–725.
2 Degner LF, Sloan JA. Decision-making during serious illness: what role do patients really want to play? *Journal of Clinical Epidemiology*, 1992; **45:** 941–950.
3 Guadagnoli E, Ward P. Patient participation in decision-making. *Social Science and Medicine*, 1998; **47:** 329–339.
4 Auerbach SM. Do patients want control over their own health care? A review of measures, findings, and research issues. *Journal of Health Psychology*, 2001; **6:** 191–203.
5 Kaplan RM, Frosch DL. Decision-making in medicine and health care. *Annual Review of Clinical Psychology*, 2005; **1:** 525–556.
6 Say R, Murtagh M, Thomson R. Patients' preference for involvement in medical decision-making: a narrative review. *Patient Education and Counseling*, 2006; **60:** 102–114.

7 Elwyn G, Edwards A, Britten N. ''Doing prescribing'': how might clinicians work differently for better, safer care. *Quality and Safety in Health Care*, 2003; **12** (Suppl. 1)**:** i33–i36.

8 Gillespie R, Florin D, Gillam S. *Changing Relationships: Findings from the Patient Involvement Project*. London: King's Fund, 2002.

9 Krantz DS, Baum A, von Wideman M. Assessment of preferences for self-treatment and information in health care. *Journal of Personality and Social Psychology*, 1980; **5:** 977–990.

10 Ende J, Kazis L, Ash A, Moskowitz MA. Measuring patients' desire for autonomy: decision-making and information-seeking preferences among medical patients. *Journal of General Internal Medicine*, 1989; **4:** 23–30.

11 Spies CD, Schulz CM, Weiss-Gerlach E *et al*. Preferences for shared decision-making in chronic pain patients compared with patients during a premedication visit. *Acta Anaesthesiologica Scandinavica*, 2006; **50:** 1019–1026.

12 Hamann J, Neuner B, Kasper J *et al*. Participation preferences of patients with acute and chronic conditions. *Health Expectations*, 2007; **10:** 358–863.

13 Hashimoto H, Fukuhara S. The influence of locus of control on preferences for information and decision-making. *Patient Education and Counseling*, 2003; **55:** 236–240.

14 Cyran EM, Crane LA, Palmer L. Physician sex and other factors associated with type of breast cancer surgery in older women. *Archives of Surgery*, 2001; **136:** 185–191.

15 Schouten BC, Eijkman MA, Hoogstraten J. Information and participation preferences of dental patients. *Journal of Dental Research*, 2004; **83:** 961–965.

16 Ishikawa H, Hashimoto H, Yano E. Patients' preferences for decision-making and the feeling of being understood in the medical encounter among patients with rheumatoid arthritis. *Arthritis and Rheumatism*, 2006; **55:** 878–883.

17 Neuner B, Dizner-Golab A, Gentilello LM *et al*. Trauma patients' desire for autonomy in medical decision-making is impaired by smoking and hazardous alcohol consumption – a bi-national study. *Journal of International Medical Research*, 2007; **35:** 609–614.

18 Nomura K, Ohno M, Fujinuma Y, Ishikawa H. Patient autonomy preferences among hypertensive outpatients in a primary care setting in Japan. *Internal Medicine*, 2007; **46:** 1403–1408.

19 Neuner B, Spies C, Miller P *et al*. Unobserved heterogeneity in trauma patients' desire for autonomy in medical decision-making in an emergency department. *Medical Science Monitor*, 2008; **14:** 366–371.

20 Giersdorf N, Loh A, Bieber C *et al*. Development and validation of assessment instruments for shared decision-making. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*, 2004; **47:** 969–976.

21 Härter M, Loh A, Spies C. *Gemeinsam entscheiden – erfolgreich behandeln*. Köln: Deutscher Ärzteverlag, 2005.

22 Steffen P. Patientenzufriedenheit als Zielgröße für das strategische Krankenhausmanagement. In: Pfaff H, Lütticke J, Badura B, Piekarski C, Richter P (eds) *''Weiche'' Kennzahlen für das strategische Krankenhausmanagement. Stakeholder-Interessen zielgerichtet erkennen und einbeziehen*. Göttingen: Verlag Hans Huber, 2004: 85–108.

23 Loh A, Kremer N, Giersdorf N *et al*. Information and participation interests of patients with depression in clinical decision-making in primary care. *Zeitschrift für ärztliche Fortbildung und Qualitätssicherung*, 2004; **98:** 101–107.

24 Kline RB. *Principles and Practice of Structural Equation Modeling*. New York: Guilford Press, 1998.

25 Byrne BM. *Structural Equation Modeling with AMOS. Basic Concepts, Applications, and Programming*. Hillsdale: Lawrence Erlbaum, 2001.

26 Kriston L, Gunzler C, Harms A, Berner M. Confirmatory factor analysis of the German version of the international index of erectile function (IIEF): a comparison of four models. *Journal of Sexual Medicine*, 2008; **5:** 92–99.

27 Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modelling*, 1999; **6:** 1–55.

28 SPSS Inc., Chicago, Illinois, USA.

29 Schafer JL. *NORM: Multiple Imputation of Incomplete Multivariate Data under a Normal Model [Computer Software]*. University Park: Department of Statistics, Pennsylvania State University, 1999.

30 Scheibler F, Freise D, Pfaff H. Die Einbeziehung von Patienten in die Behandlung. Validierung der deutschen PICS-Skalen. *Journal of Public Health*, 2004; **12:** 199–209.

31 Simon D, Loh A, Härter M. Die Entwicklung und Evaluation von Interventionen zur Förderung Partizipativer Entscheidungsfindung – Rahmenkonzept und Messinstrumente *Zeitschrift für Medizinische. Psychologie*, 2008; **17:** 149–159.

32 Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS (eds), *Testing Structural Equation Models*. Newbury Park: Sage, 1993: 445–455.

33 Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. *MPR-Online*, 2003; **8:** 23–74.