



# Cross-cultural validation of the German version of the Four-Dimensional Symptom Questionnaire (4DSQ) in multimorbid elderly people

Anna Exner<sup>1</sup> · Maria Kleinstäuber<sup>2,3</sup> · Wolfgang Maier<sup>4</sup> · Angela Fuchs<sup>5</sup> · Juliana J. Petersen<sup>6</sup> · Ingmar Schäfer<sup>7</sup> · Jochen Gensichen<sup>8</sup> · Steffi G. Riedel-Heller<sup>9</sup> · Siegfried Weyerer<sup>10</sup> · Horst Bickel<sup>11</sup> · Hans-Helmut König<sup>12</sup> · Birgitt Wiese<sup>13</sup> · Gerhard Schön<sup>14</sup> · Martin Scherer<sup>7</sup> · Hendrik van den Bussche<sup>7</sup> · Berend Terluin<sup>15</sup>

Accepted: 20 June 2018 / Published online: 27 June 2018  
© Springer International Publishing AG, part of Springer Nature 2018

## Abstract

**Purpose** Psychosomatic symptoms and mental health problems are highly prevalent in multimorbid elderly people challenging general practitioners to differentiate between normal stress and psychopathological conditions. The 4DSQ is a Dutch questionnaire developed to detect anxiety, depression, somatization, and distress in primary care. This study aims to analyze measurement equivalence between a German version and the original Dutch instrument.

**Methods** A Dutch and a German sample of multimorbid elderly people, matched by gender and age, were analyzed. Equivalence of scale structures was assessed by confirmatory factor analysis (CFA). To evaluate measurement equivalence across languages, differential item functioning (DIF) was analyzed using Mantel–Haenszel method and hybrid ordinal logistic regression analysis. Differential test functioning (DTF) was assessed using Rasch analysis.

**Results** A total of 185 German and 185 Dutch participants completed the questionnaire. The CFA confirmed one-factor models for all scales of both 4DSQ versions. Nine items in three scales were flagged with DIF. The anxiety scale showed to be free of DIF. DTF analysis revealed negligible scale impact of DIF.

**Conclusions** The German 4DSQ demonstrated measurement equivalence to the original Dutch instrument. Hence, it can be considered a valid questionnaire for the screening for mental health problems in primary care.

**Keywords** Depression · Distress · Somatization · Anxiety · Measurement equivalence

## Introduction

Multimorbidity, defined as the coexistence of two or more chronic diseases in one person, has a high prevalence ranging from 55 to 98% in people aged 65+ [1] and is associated with psychological distress, depression, and other mental health problems [e.g., 2]. The health care systems of many countries are based on general practitioners (GPs) as first contact persons for all kinds of health and mental health issues [3]. Therefore, GPs play an important role

in recognizing somatic as well as psychological disorders. Despite the high prevalence of mental health problems in primary care settings, the rate of diagnosed disorders is low [e.g., 4].

The Dutch Four-Dimensional Symptom Questionnaire (4DSQ) was developed as a self-report questionnaire covering distress, depression, anxiety and somatization. It has predominantly been applied in primary care settings, with the main focus of distinguishing between patients suffering from a general non-specific stress response and persons with a pathologic psychological state [5, 6]. The questionnaire was translated into various languages and is commonly used in different countries [e.g., 7].

The aim of this study was to investigate the measurement equivalence between the German version of the 4DSQ and the original Dutch instrument in multimorbid elderly people using confirmatory factor analysis (CFA),

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s11136-018-1924-2>) contains supplementary material, which is available to authorized users.

✉ Maria Kleinstäuber  
maria.kleinstaeuber@staff.uni-marburg.de

Extended author information available on the last page of the article

and differential item and differential test functioning (DIF and DTF) analysis.

## Methods

### Study design and participants

The German 4DSQ data were collected as part of the MultiCare Cohort Study [8]. Patients aged between 65 and 85 suffering from multimorbidity and with at least one GP consultation within the preceding quarter were randomly selected from the databases of 158 GP offices at eight study centers distributed across Germany. Multimorbidity was defined as the coexistence of at least 3 chronic diseases from a list of 29 diseases. Medical conditions with a prevalence of > 25% in the aforementioned age group (e.g., hypertension) were not counted for this definition of multimorbidity. The German 4DSQ was part of the baseline interview, completed by 3189 patients between July 2008 and October 2009.

The Dutch reference data were selected from the CentERdata LISS (Longitudinal Internet Study in the Social Sciences) panel, a web-based register providing a representative sample of Dutch-speaking persons living in the Netherlands [9]. The 4DSQ was presented in July 2013 to all persons aged 16 and older, registered to the CentERdata LISS panel [10]. To maximize comparability to the German sample, all Dutch people aged between 65 and 85 suffering from at least three diseases according to preceding questionnaires were selected from the sample ( $N=185$ ). For every Dutch participant, one German patient was randomly chosen under consideration of matching for gender and age (5-year groups).

### Questionnaire

The 4DSQ is a Dutch self-report questionnaire consisting of 50 items covering four scales: distress, depression, anxiety and somatization [11]. Items are answered on a 5-point Likert scale. Scale values are generated by rating ‘no’ as 0, ‘sometimes’ as 1, and the remaining three categories as 2. The questionnaire was validated in various studies [e.g., 12, 13]. The original version and translations are freely available for non-commercial use at <http://www.emgo.nl/researchtools/4dsq.asp>.

First, the questionnaire was translated into German (Vierdimensionale Beschwerdenliste, 4DBL). After a critical review of this translation resulting in some minor adjustments it was back-translated into Dutch. All translations were undertaken by Dutch native speakers fluent in German by consensus.

## Analyses

For *imputing missing item scores*, the response function method was applied [14]. *Group differences* in gender, age, and 4DSQ scale mean scores were analyzed using Chi-square test, *t* test, and Mann–Whitney *U* test, respectively. As a measure of reliability, Cronbach’s alpha was calculated [15]. *Unidimensionality* of the German and Dutch 4DSQ was evaluated with a CFA. Comparative fit index (CFI) > 0.95, Tucker–Lewis index (TLI) > 0.95, and root mean square error of approximation (RMSEA) < 0.06 were used as indicators of an adequate model fit [16]. In order to analyze differential item functioning (DIF), a 2-stage method for DIF testing was used [17]: the non-parametric Mantel–Haenszel (M–H) method, detecting primarily uniform DIF [18], and the parametric hybrid ordinal logistic regression (HOLR), detecting uniform as well as non-uniform DIF [18]. The M–H method was performed with the jMetrik 3.1.2 software package [19] and HOLR with the R package ‘lordif’ [20]. DIF was identified when the absolute standardized mean item difference was > 0.1 and *p* was < .001 (M–H method) or when the increase in explained variance ( $\Delta R^2$ ) was > 2% and *p* was < .001 (HOLR method) [20]. To examine differential test functioning (DTF), the impact of DIF on scale level, Rasch analysis, as implemented in jMetrik 3.1.2 [19], was used and theta scores (estimated scores of the underlying latent trait of each scale) were calculated for each group. Raw sum scores (containing items with and without DIF) for each scale were plotted against theta values. The impact of DIF on the scale scores was defined as the vertical distance between group-specific curves at the cut-off points. Details of our statistical analysis are provided in the Online Supplementary Appendix 1.

## Results

### Initial analyses

In the German sample, 0.08% of all item scores were missing and therefore imputed, while the Dutch sample did not contain any missing values. The German group as well as the Dutch reference group consisted of 185 multimorbid patients, with 57.3% of them being female. Demographic variables, mean item scores, and Cronbach’s alpha values of the 4DSQ subscales are summarized in Table 1. In the Dutch sample, multimorbid participants aged 65 years and older show increased scores on all four dimensions of the 4DSQ compared to all LISS panel participants with an age of at least 65 years ( $N=1184$ ), (see mean values and SD in the Online Supplementary Appendix 2) [21].

**Table 1** Demographic variables, mean 4DSQ scores, and Cronbach's alpha of the German ( $N=185$ ) and the Dutch ( $N=185$ ) samples

	Mean (SD)			Cronbach's alpha		
	German sample	Dutch sample	<i>p</i>	German sample	Dutch sample	<i>p</i>
Age	74.20 (5.70)	73.50 (5.85)	.263 <sup>a</sup>	–	–	–
4DSQ scale (scale range)						
Distress (0–32)	6.15 (5.42)	8.68 (7.61)	.003 <sup>b</sup>	0.85	0.93	.001 <sup>c</sup>
Depression (0–12)	0.87 (2.01)	0.92 (2.53)	.138 <sup>b</sup>	0.87	0.91	.205 <sup>c</sup>
Anxiety (0–24)	0.92 (2.53)	2.10 (3.56)	<.001 <sup>b</sup>	0.88	0.87	.849 <sup>c</sup>
Somatization (0–32)	6.97 (5.33)	8.96 (6.04)	.001 <sup>b</sup>	0.81	0.84	.268 <sup>c</sup>

<sup>a</sup>*t* test; <sup>b</sup>Mann–Whitney *U* test; <sup>c</sup>Z-score test

## Unidimensionality

For both groups, the multi-group CFA confirmed one-factor models for the 4DSQ scales (Table 2). The model fit was incrementally optimized by removing restrictions of residual covariances stepwise for item pairings sharing same content. CFI, TLI, and RMSEA indicated an adequate fit of the data to the one-factor models for all but the somatization scale, where the RMSEA showed a slight deviation from the defined threshold.

## Differential item functioning

The M–H method identified three items with DIF, while the HOLR flagged eight items (Table 3). Items 12 and 47 were detected by both methods. The anxiety scale revealed to be free of DIF, while the other scales contained three DIF items each. An analysis of the item response functions indicated a higher difficulty for German participants in four items and a lower difficulty in five items compared to the Dutch participants. Five items were identified with

(mainly) uniform, two with (mainly) non-uniform, and two with mixed uniform and non-uniform DIF.

Fig. 1 shows the category characteristic curves of two example items. Item 41 of the distress scale was detected with uniform DIF. In item 30 of the depression scale, non-uniform DIF was detected.

## Differential test functioning

Figure 2 shows the impact of DIF on scale level by comparing raw scale scores with the underlying trait estimated by Rasch analysis (theta). Additionally, DTF around validated cut-off points is depicted in Fig. 2. The highest absolute difference showed to be at the first cut-off point of the distress scale (0.31 scale points). A Dutch cut-off value of 11.00 corresponded to a German value of 11.31. As the differences in scale scores between the German and the Dutch samples due to DIF were found to be < 0.5 scale points, the impact of DIF can be considered negligible.

**Table 2** Goodness-of-fit indices of the multi-group CFA for one-factor models fitted to the data of each scale depending on residual correlations allowed

4DSQ scale	CFI Crit. > 0.95	TLI Crit. > 0.95	RMSEA Crit. < 0.06	90% CI of RMSEA	Items needing residual correlations
Distress	0.970	0.965	0.113	0.104–0.122	–
	0.986	0.984	0.077	0.067–0.088	20–39
	0.995	0.995	0.045	0.030–0.057	20–39, 47–48
Depression	1.000	1.000	0.000	0.000–0.064	–
Anxiety	1.000	1.001	0.000	0.000–0.033	–
Somatization	0.939	0.929	0.100	0.091–0.110	–
	0.950	0.942	0.091	0.081–0.100	15–16
	0.955	0.947	0.086	0.076–0.096	15–16, 12–13
	0.966	0.960	0.075	0.065–0.086	15–16, 9–12–13
	0.970	0.964	0.071	0.061–0.082	15–16, 9–12–13, 2–5
	0.976	0.970	0.065	0.054–0.076	15–16, 2–4–5, 9–12–13

CFI comparative fit index, TLI Tucker–Lewis index, RMSEA root mean square error of approximation, 90% CI 90% confidence interval of the RMSEA, Crit. critical value for model fit. Indicators of adequate model fit: Comparative fit index (CFI) > 0.95, Tucker–Lewis index (TLI) > 0.95, root mean square error of approximation (RMSEA) < 0.06

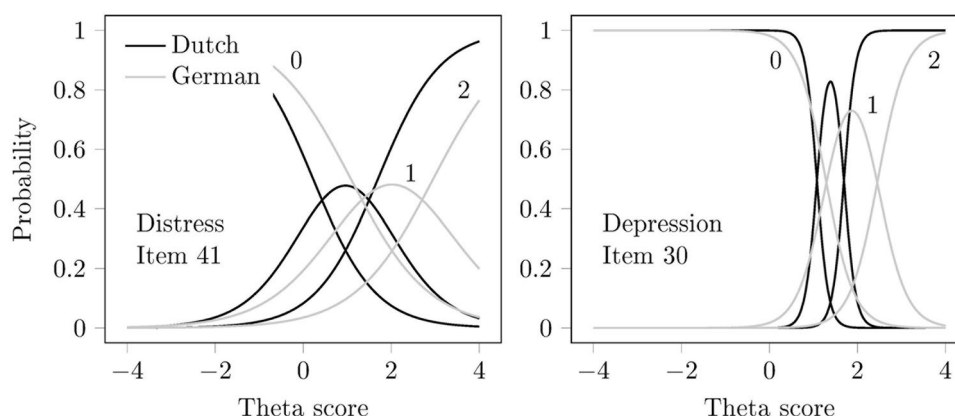
**Table 3** Items identified with significant DIF

4DSQ scale	Item no.	Short description	M–H method <sup>a</sup>	HOLR method <sup>b</sup>	DIF type	Direction <sup>c</sup>
Distress	17	Feeling down or depressed	0.27	2.69	Mixed	+
	41	Easily become emotional		2.46	Uniform	–
	47	Fleeting images of any upsetting event(s)		2.72	Uniform	+
Depression	30	Life is not worth while	0.24	5.00	Non-uniform	–
	34	Can't enjoy anything anymore		3.08	Uniform	–
	46	Think 'I wish I was dead'		2.08	Mixed	+
Somatization	3	Fainting	–0.19	3.64	Non-uniform	+
	9	Bloated feeling in the abdomen			Uniform	+
	12	Nausea or an upset stomach		3.72	Uniform	–

<sup>a</sup>M–H method: standardized mean differences (SMD) calculated by Mantel–Haenszel method

<sup>b</sup>HOLR method: difference in  $R^2$  ( $\Delta R^2$ ) calculated by hybrid ordinal logistic regression ( $\times 100$ )

<sup>c</sup>Direction of DIF: the item is globally more difficult (–)/easier (+) for German participants



**Fig. 1** Category characteristic curves of DIF items. The curves display the probability of responding to an item by choosing the option ‘no’ (category ‘0’), ‘sometimes’ (category ‘1’) or ‘regularly,’ ‘often’ or ‘very often or constantly’ (category ‘2’) in dependence on the latent trait estimated by HOLR analysis and language group. The response functions of each category of the German item 41 are uniformly shifted on the x-axis towards higher values of the latent trait of distress (German participants needed a higher level of distress to

reach higher item scores). At low levels of item 30 (depression scale), there was an equal probability of reaching category ‘0’ for both groups. The probability of category ‘1’ rose slightly earlier for German than for Dutch participants. However globally at higher levels of depression, reaching higher item scores was ‘easier’ for Dutch participants, shown by the steep rise of the Dutch category ‘2’ curve at lower depression values than the corresponding German curve

## Discussion

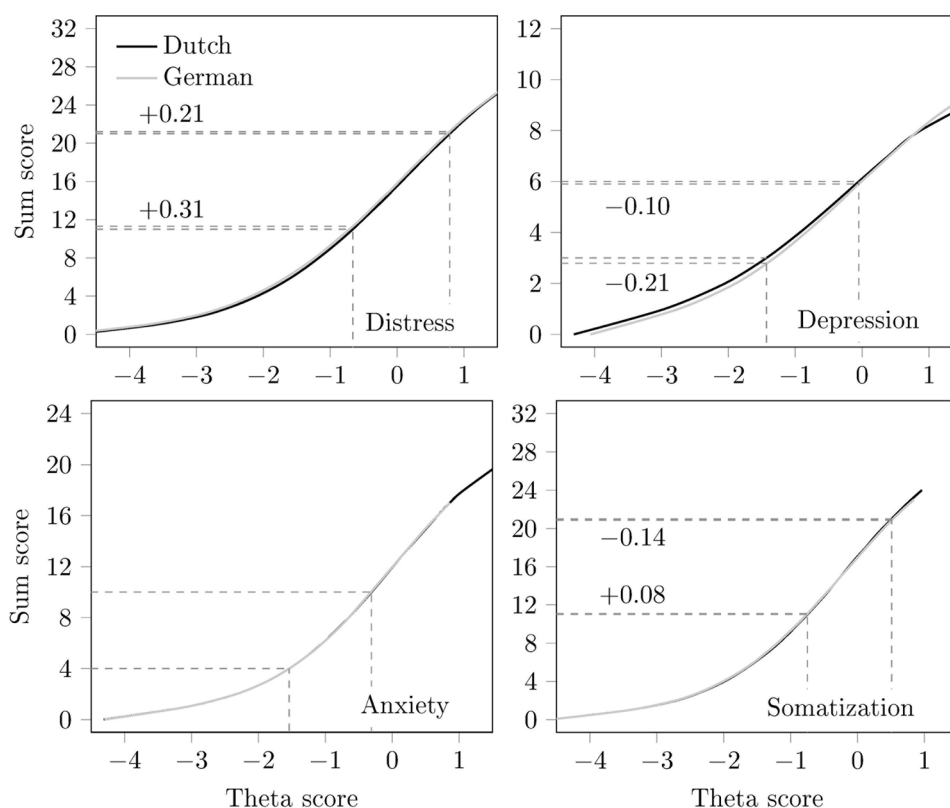
The aim of this study was to analyze the measurement equivalence between the German version of the 4DSQ and the original Dutch questionnaire in multimorbid elderly people. DIF was detected in 9 out of 50 items. DIF analyses of other translations into languages with a closer linguistic relation to Dutch and samples with a comparable cultural background showed a similar number of DIF laden items [e.g., 22].

Even though DIF was detected on item level, the effects on scale level were negligible. Impact values of +0.31 (distress scale) or lower on raw scale scores corresponding

to the estimated latent trait at the cut-off points imply the same meaning of cut-off points in both language groups.

While the Dutch sample was a general population sample recruited from households nationwide in the Netherlands and paid for their participation, in the German study only GP patients with a consultation during the last quarter were included. We assume that in the Dutch sample primary health care utilization varied more than in the German sample. A regular usage of primary care is associated with various positive outcomes on individual disease management [23]. Taking this under consideration, the higher mean 4DSQ scores (and also the higher variances) in the Dutch sample seem to be plausible.

**Fig. 2** Impact of DIF on scales scores. For each scale, an estimation of the underlying latent trait calculated by Rasch analysis is plotted against the raw sum score (calculated from DIF free and DIF laden items). The impact of DIF can be defined as the vertical distance between group-specific curves



The current study suggests that the German version of the 4DSQ measures the dimensions depression, distress, anxiety and somatization in the same way as the original Dutch instrument in multimorbid elderly people and that German scores can be interpreted in the same way as Dutch scores. Although the Dutch 4DSQ demonstrates measurement equivalence across age [21], this does not necessarily imply that the German 4DSQ is equivalent to the Dutch questionnaire in other non-multimorbid or non-elderly samples. This should be tested in other samples.

**Acknowledgements** In this paper, we make use of data of the LISS (Longitudinal Internet Studies for the Social Sciences) panel administered by CentERdata (Tilburg University, The Netherlands). We are grateful to the general practitioners in Bonn, Düsseldorf, Frankfurt/Main, Hamburg, Jena, Leipzig, Mannheim, and Munich who supplied the clinical information on their patients.

**Funding** The MultiCare Cohort Study was supported by the German Federal Ministry of Education and Research (Grant Numbers 01ET0725-31, 01ET1006A-K). The LISS panel was supported by the Netherlands Organization for Scientific Research (NOW): Measurement and Experimentation in the Social Sciences (MESS) Project (Project Number 176.010.2005.017).

### Compliance with ethical standards

**Conflict of interest** BT is the copyright owner of the 4DSQ and receives copyright fees from companies that use the 4DSQ on a commercial basis (the 4DSQ is freely available for non-commercial use in

health care and research). BT received fees from various institutions for workshops on the application of the 4DSQ in primary care settings. All other authors declare that they have no conflict of interests.

**Ethical approval** The study protocol of this trial was approved by the Ethics Committee of the Medical Association of Hamburg in February 2008 and amended in November 2008 (Approval-No. 2881). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

**Informed consent** Informed consent was obtained from all individual participants included in the study.


### References

1. Marengoni, A., Angleman, S., Melis, R., Mangialasche, F., Karp, A., Garmen, A., et al. (2011). Aging with multimorbidity: A systematic review of the literature. *Ageing Research Reviews*, 10(4), 430–439.
2. Swartz, J. A., & Jantz, I. (2014). Association between nonspecific severe psychological distress as an indicator of serious mental illness and increasing levels of medical multimorbidity. *American Journal of Public Health*, 104(12), 2350–2358.
3. Rattay, P., Butschalowsky, H., Rommel, A., Prütz, F., Jordan, S., Nowossadeck, E., et al. (2013). Utilization of outpatient and inpatient health services in Germany. Results of the German Health Interview and Examination Survey for Adults (DEGS1) [Inanspruchnahme der ambulanten und stationären medizinischen Versorgung in Deutschland: Ergebnisse der Studie zur Gesundheit



- Erwachsener in Deutschland (DEGS1)]. Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz, 56, 832–844.
4. Olario, E., Forero, C. G., Castro-Rodriguez, J. I., Rodrigo-Calvo, M. T., Álvarez, P., Martín-López, L. M., et al. (2015). Detection of anxiety disorders in primary care: A meta-analysis of assisted and unassisted diagnosis. *Depression and Anxiety*, 32(7), 471–484.
  5. Terluin, B. (1996). De Vierdimensionale Klachtenlijst (4DKL). Een vragenlijst voor het meten van distress, depressie, angst en somatisatie. *Huisarts Wet*, 39(12), 538–547.
  6. Terluin, B., van Marwijk, H. W. J., Ader, H. J., de Vet, H. C. W., Penninx, B. W. J. H., Hermens, M. L. M., et al. (2006). The Four-Dimensional Symptom Questionnaire (4DSQ): A validation study of a multidimensional self-report questionnaire to assess distress, depression, anxiety and somatization. *BMC Psychiatry*, 6, 34.
  7. Terluin, B., Unalan, P. C., Turfaner, N., Özkul, S. A., & Marwijk, H. W. J. V. (2016). Cross-cultural validation of the Turkish Four-Dimensional Symptom Questionnaire (4DSQ) using differential item and test functioning (DIF and DTF) analysis. *BMC Family Practice*, 17(1), 53.
  8. Schäfer, I., Hansen, H., Schön, G., Maier, W., Höfels, S., Altiner, A., et al. (2009). The German MultiCare-study: Patterns of multimorbidity in primary health care - protocol of a prospective cohort study. *BMC Health Services Research*, 9, 145.
  9. Scherpenzeel, A. (2011). Data collection in a probability-based Internet panel: How the LISS panel was built and how it can be used. *Bulletin de Méthodologie Sociologique*, 109(1), 56–61.
  10. Terluin, B., Marchand, M. A. G., De Vet, H. C. W., & Brouwers, E. P. M. (2015). Impact of verb tense on response to the Four-Dimensional Symptom Questionnaire (4DSQ). *European Journal of Psychological Assessment*, 31(3), 202–210.
  11. Terluin, B., Brouwers, E. P. M., Van Marwijk, H. W. J., Verhaak, P. F. M., & Van Der Horst, H. E. (2009). Detecting depressive and anxiety disorders in distressed patients in primary care; comparative diagnostic accuracy of the Four-Dimensional Symptom Questionnaire (4DSQ) and the Hospital Anxiety and Depression Scale (HADS). *BMC Family Practice*, 10, 58.
  12. De Vroege, L., Emons, W. H. M., Sijtsma, K., Hoedeman, R., & Van Der Feltz-Cornelis, C. M. (2015). Validation of the 4DSQ somatization subscale in the occupational health care setting as a screener. *Journal of Occupational Rehabilitation*, 25(1), 105–115.
  13. Terluin, B., Oosterbaan, D. B., Brouwers, E. P. M., Van Straten, A., Van De Ven, P. M., Langerak, W., et al. (2014). To what extent does the anxiety scale of the Four-Dimensional Symptom Questionnaire (4DSQ) detect specific types of anxiety disorder in primary care? A psychometric study. *BMC Psychiatry*, 14, 121.
  14. Van Ginkel, J. R., & Van Der Ark, L. A. (2005). SPSS syntax for missing value imputation in test and questionnaire data. *Applied Psychological Measurement*, 29(2), 152–153.
  15. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
  16. Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.
  17. Hambleton, R. K. (2006). Good practices for identifying differential item functioning. *Medical Care*, 44(11), 182–188.
  18. Teresi, J. A. (2006). Different approaches to differential item functioning in health applications: Advantages, disadvantages and some neglected topics. *Medical Care*, 44(11 Suppl 3), 152–170.
  19. Meyer, J. P. (2014). *Applied measurement with jMetrik*. London: Routledge.
  20. Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software*, 39(8), 1–30.
  21. Terluin, B., Smits, N., Brouwers, E. P., & de Vet, H. C. (2016). The Four-Dimensional Symptom Questionnaire (4DSQ) in the general population: Scale structure, reliability, measurement invariance and normative data: A cross-sectional survey. *Health and Quality of Life Outcomes*, 14(1), 130.
  22. Terluin, B., Smits, N., & Miedema, B. (2014). The English version of the four-dimensional symptom questionnaire (4DSQ) measures the same as the original Dutch questionnaire: A validation study. *European Journal of General Practice*, 20(4), 320–326.
  23. Smith, J. J., Berman, M. D., Hiratsuka, V. Y., & Frazier, R. R. (2015). The effect of regular primary care utilization on long-term glycemic and blood pressure control in adults with diabetes. *Journal of the American Board of Family Medicine*, 28(1), 28–37.

## Affiliations

Anna Exner<sup>1</sup> · Maria Kleinstäuber<sup>2,3</sup>  · Wolfgang Maier<sup>4</sup> · Angela Fuchs<sup>5</sup> · Juliana J. Petersen<sup>6</sup> · Ingmar Schäfer<sup>7</sup> · Jochen Gensichen<sup>8</sup> · Steffi G. Riedel-Heller<sup>9</sup> · Siegfried Weyerer<sup>10</sup> · Horst Bickel<sup>11</sup> · Hans-Helmut König<sup>12</sup> · Birgitt Wiese<sup>13</sup> · Gerhard Schön<sup>14</sup> · Martin Scherer<sup>7</sup> · Hendrik van den Bussche<sup>7</sup> · Berend Terluin<sup>15</sup>

<sup>1</sup> Department of Education Studies and Psychology, University of Siegen, Adolf-Reichwein-Str. 2a, 57076 Siegen, Germany

<sup>2</sup> Department of Clinical Psychology and Psychotherapy, Philipps-University of Marburg, Gutenbergstraße 18, 35037 Marburg, Germany

<sup>3</sup> Department of Medical Psychology, School of Medicine, University of Auckland, 2 Park Road, Grafton, Auckland 1023, New Zealand

<sup>4</sup> Clinic of Psychiatry and Psychotherapy, University Hospital Bonn, Sigmund-Freud-Str. 25, 53127 Bonn, Germany

<sup>5</sup> Institute of General Practice (ifam), Medical Faculty, Heinrich Heine-University Düsseldorf, Moorenstr. 5, 40225 Düsseldorf, Germany

<sup>6</sup> Institute of General Practice, Goethe-University Frankfurt am Main, Theodor Stern Kai 7, 60590 Frankfurt/main, Germany

<sup>7</sup> Department of Primary Medical Care, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany

<sup>8</sup> Institute of General Practice and Family Medicine, University Hospital of Ludwig-Maximilians-University Munich, Pettenkoferstr. 10, 80336 Munich, Germany

<sup>9</sup> Institute for Social Medicine, Occupational Health and Public Health, University of Leipzig, Semmelweisstr. 10, 04103 Leipzig, Germany

<sup>10</sup> Central Institute of Mental Health, Medical Faculty Mannheim/Heidelberg University, J5, 68159 Mannheim, Germany

- <sup>11</sup> Department of Epidemiological Psychiatry, Technical University of Munich, Ismaninger Straße 22, 81675 Munich, Germany
- <sup>12</sup> Department of Health Economics and Health Services Research, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany
- <sup>13</sup> Institute of General Practice, Hannover Medical School, Carl-Neuberg-Str.1, 30625 Hannover, Germany
- <sup>14</sup> Department of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany
- <sup>15</sup> Department of General Practice and Elderly Care Medicine, Amsterdam Public Health Research Institute, VU University Medical Center, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands