Cross-Cultural Validation of the Quality of Life in Hand Eczema Questionnaire (QOLHEQ)

Robert F. Ofenloch[1], Jart A.F. Oosterhaven[2], Päivikki Susitaival[3], Åke Svensson[4], Elke Weisshaar[1], Keiko Minamoto[5], Meltem Onder[6], Marie Louise A. Schuttelaar[2], Emel Bulbul Baskan[7], Thomas L. Diepgen[1] and Christian Apfelbacher[8]

The Quality of Life in Hand Eczema Questionnaire (QOLHEQ) is the only instrument assessing disease-specific health-related quality of life in patients with hand eczema. It is available in eight language versions. In this study we assessed if the items of different language versions of the QOLHEQ yield comparable values across countries. An international multicenter study was conducted with participating centers in Finland, Germany, Japan, The Netherlands, Sweden, and Turkey. Methods of item response theory were applied to each subscale to assess differential item functioning for items among countries. Overall, 662 hand eczema patients were recruited into the study. Single items were removed or split according to the item response theory model by country to resolve differential item functioning. After this adjustment, none of the four subscales of the QOLHEQ showed significant misfit to the item response theory model ($P < 0.01$), and a Person Separation Index of greater than 0.7 showed good internal consistency for each subscale. By adapting the scoring of the QOLHEQ using the methods of item response theory, it was possible to obtain QOLHEQ values that are comparable across countries. Cross-cultural variations in the interpretation of single items were resolved. The QOLHEQ is now ready to be used in international studies assessing the health-related quality of life impact of hand eczema.

Journal of Investigative Dermatology (2017) **137,** 1454–1460; doi:10.1016/j.jid.2017.02.969

INTRODUCTION

Hand eczema (HE) is a common and multifactorial skin disease (Coenraads, 2012). In the general population, the 1-year prevalence of HE has been estimated to be as high as 10%, with higher risk in females and in patients with contact allergy, atopy, or exposure to wet work (Thyssen et al., 2010). Often it is a chronic recurrent or persisting condition with negative socioeconomic effects, and it has been shown that about 28% of patients with HE of occupational origin are unfit to work (Diepgen et al., 2009). Health-related quality of life (HRQOL) is negatively affected in patients with HE (Apfelbacher et al., 2014; Moberg et al., 2009). HRQOL impairments in HE have been assessed by using generic HRQOL instruments like the EuroQoL-5D (Brooks, 1996) or by using skin-specific instruments like the Dermatology Life Quality Index (Finlay and Khan, 1994) or Skindex (Chren et al., 1996). The only disease-specific instrument for assessing HRQOL impairment in HE patients is the Quality of Life in Hand Eczema Questionnaire (QOLHEQ; Ofenloch et al., 2014). Although generic and skin-specific instruments enable comparability with other (skin) diseases, disease-specific instruments assess impairments caused by the disease of interest more precisely and are therefore more sensitive to change when used in clinical trials. In its validation study, the QOLHEQ was shown to be valid and reliable, and its sensitivity to change was superior compared with the EuroQoL-5D, Dermatology Life Quality Index, and Skindex-29.

Especially in chronic skin disorders, clinical severity scores alone, such as the Hand Eczema Severity Index (Held et al., 2005) or the Osnabrück Hand Eczema Index (Dulon et al., 2009), do not give enough information on the effects of treatments. This is because the clinical score is rated by a physician, and it is known that those ratings correlate only moderately with patients' perceptions of impairment (Agner et al., 2013; Ofenloch et al., 2015). Therefore, measures of HRQOL should be integrated as patient-reported outcomes in clinical trials. Cross-cultural aspects have often not been considered enough during the development of many instruments used in dermatology (Grob, 2007). When patient-reported outcome instruments are used to assess data in a cross-culturally equivalent manner, this aspect should already have been accounted for during development of the instrument.

[1]*Department of Clinical and Social Medicine, University Hospital Heidelberg, Germany;* [2]*Department of Dermatology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands;* [3]*Department of Dermatology, North Karelia Central Hospital, Joensuu, Finland;* [4]*Department of Dermatology, Skåne University Hospital, Lund University, Malmö, Sweden;* [5]*Department of Public Health, Faculty of Life Sciences, Graduate School of Kumamoto University, Kumamoto, Japan;* [6]*Department of Dermatology, Gazi University Medical Faculty, Ankara, Turkey;* [7]*Department of Dermatology, Uludag University Medical Faculty, Bursa, Turkey; and* [8]*Medical Sociology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, Germany*

Correspondence: Robert Ofenloch, Voßstraße 2, 69115 Heidelberg, Germany. E-mail: robert.ofenloch@med.uni-heidelberg.de

Abbreviations: DIF, differential item functioning; HE, hand eczema; HRQOL, health-related quality of life; IRT, item response theory; PSI, Person Separation Index; QOLHEQ, Quality of Life in Hand Eczema Questionnaire; RM, Rasch model

Received 7 November 2016; revised 24 January 2017; accepted 7 February 2017; accepted manuscript published online 1 March 2017

Table 1. Demographic characteristics of the sample

		Sex				Age			
		Male		Female					
Country	Overall n	n	%	n	%	Mean	Minimum	Maximum	SD
Germany	111	57	51.4	54	48.6	50.50[1]	18.00	78.00	12.19
Sweden	112	38	33.9	74	66.1	38.88	18.00	58.00	11.76
Finland	107	25	24.5[2]	77	75.5[2]	38.87	19.00	70.00	14.86
Turkey	112	48	42.9	64	57.1	31.97[1]	18.00	46.00	7.81
Japan	108	26	24.1[2]	82	75.9[2]	41.73	19.00	79.00	14.07
The Netherlands	112	58	51.8[2]	54	48.2[2]	43.40	18.00	77.00	14.25
Total	662	252	38.4	405	61.6	40.93	18.00	79.00	13.82

Abbreviation: SD, standard deviation.
[1]Differs significantly from the overall mean (*F* test, $P < 0.01$).
[2]Differs significantly from the overall distribution (Fisher's exact test, $P < 0.01$).

The development of instruments that assess impairment in HRQOL in a valid manner across different languages and cultures is essential if one wishes to use such measures in international, multicenter studies. If clinical trials are performed in several countries, the scores obtained through a particular instrument are not necessarily comparable across these countries, as shown by Nijsten et al. (2007) for the Dermatology Life Quality Index and for Skindex in psoriasis patients. It is likely that this is true also for other HRQOL instruments, because responses to those questionnaires are often governed by social values and norms, which are likely to differ among countries (Nijsten et al., 2007).

One method to investigate if an instrument produces cross-cultural equivalent data is to test for differential item functioning (DIF) using the framework of item response theory (IRT) (Zumbo, 1999). DIF is present when the responses to a specific item from individuals with the same level of HRQOL impairment in two different countries systematically differ from each other. Differences among populations may be over- or underestimated if the items of a scale show DIF among populations (Brodersen et al., 2007). Cultural factors are likely to cause those differences in HRQOL (Scott et al., 2008). Using IRT methods, it is possible to transform the ordinal-scaled scores of an instrument into interval-scaled scores that account for DIF among subgroups, thereby making the scores comparable (Bond and Fox, 2001).

The aim of this study was to obtain a scoring for each subscale of the QOLHEQ that generates comparable values for Finland, Germany, Japan, The Netherlands, Sweden, and Turkey. All these language versions of the QOLHEQ can be found in the Supplementary Materials online.

RESULTS

Sample characteristics

Overall, 662 individuals were recruited, with a well-balanced number of 110 ± 3 individuals for each participating country. In the total sample, 61.6% of the individuals were women, with the highest proportion of women in Japan (75.9%) and the lowest proportion in The Netherlands (48.2%). Together with Finland, where the percentage of women was 75.5%, those countries differed significantly from the overall distribution ($P < 0.01$). The mean age of the study population was 40.9 years (range = 18–79 years), with the youngest subpopulation in Turkey (mean = 31.9 years) and the oldest in Germany (mean = 50.5 years). Those two countries differed significantly from the overall mean ($P < 0.01$); however, the effect size was rather small ($\eta^2 < 0.1$). The demographic characteristics of the study population are shown in total and separately for each country in detail in Table 1.

Symptoms

The first inspection of the Symptoms subscale showed a significant misfit to the Rasch model (RM) (overall $\chi^2 = 49.9$; $df = 28$; $P < 0.01$) and disordered thresholds for the item *Itch* and *Fissuring*. After adjusting those items by merging the response categories *Rarely* and *Sometimes,* the analysis on DIF was performed. Overall, the items *Pain* and *Redness* showed relevant DIF (deviation of > 0.5 logits) (Figure 1a and b), and an analysis of DIF for each country separately showed that this DIF was caused by the Swedish sample. Figure 1c and d shows that at the same level of HRQOL impairment (person location), Swedish people were more likely to report impairment (expected value) on the item *Pain* and less likely to report impairment because of *Redness* compared with the rest of the sample.

After splitting the scoring of those items for the Swedish subpopulation and retrieving a separate scoring for this group, the subscale no longer showed significant misfit to the RM (overall $\chi^2 = 53.6$; $df = 36$; $P > 0.01$). A Person Separation Index (PSI) of 0.79 indicated a good internal reliability for the adjusted subscale, which can now be used to compare scores among the participating countries.

Emotions

Although the Emotions subscale showed no initial misfit to the RM (overall $\chi^2 = 24.8$; $df = 32$; $P > 0.01$), disordered thresholds were detected for the item *Anxious*. After this item was adjusted, several items of the scale showed relevant DIF by country: (i) the Swedish subgroup was less likely to report impairment due to being *Annoyed* or *Embarrassed,* whereas (ii) the Finnish subgroup was more likely to be impaired because of being *Anxious* about the future, and (iii) the Turkish subpopulation was less likely to report being *Frustrated*.

In Figure 2a–d, the DIF is shown as item characteristic curves for these countries compared with the rest of the sample. By using the RM, the items were split for the corresponding countries, and country-specific,

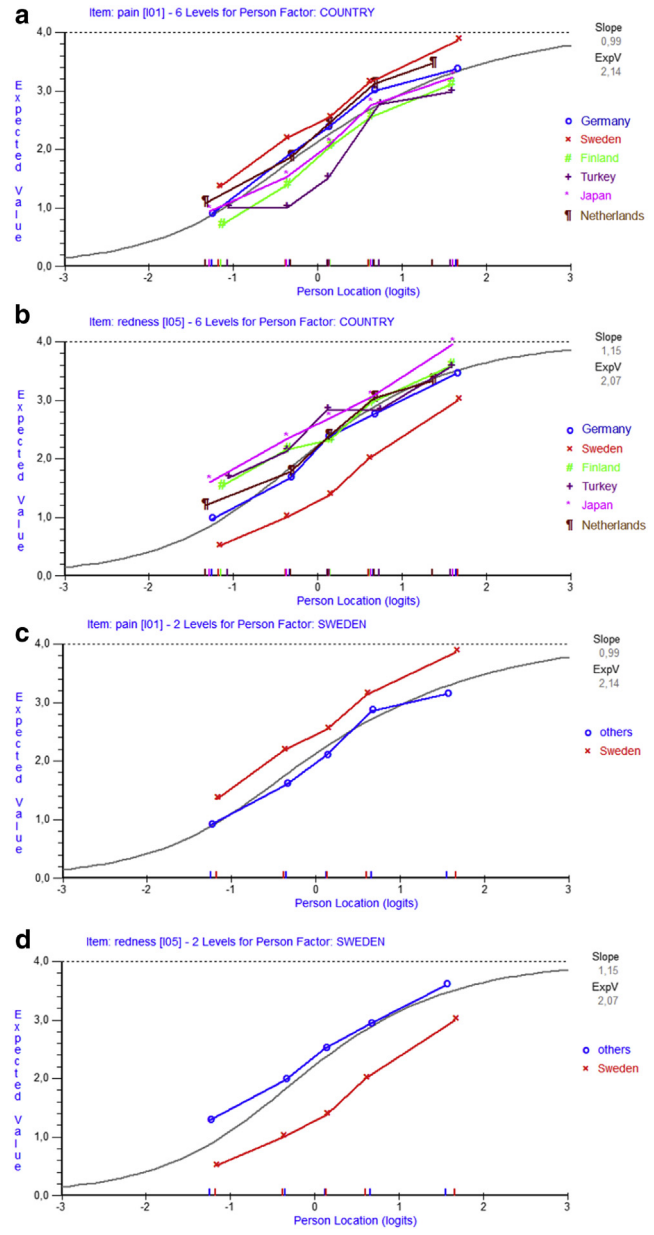

Figure 1. Item characteristic curves. Items (**a, c**) *pain* and (**b, d**) *redness* by country. ExpV, expected value.

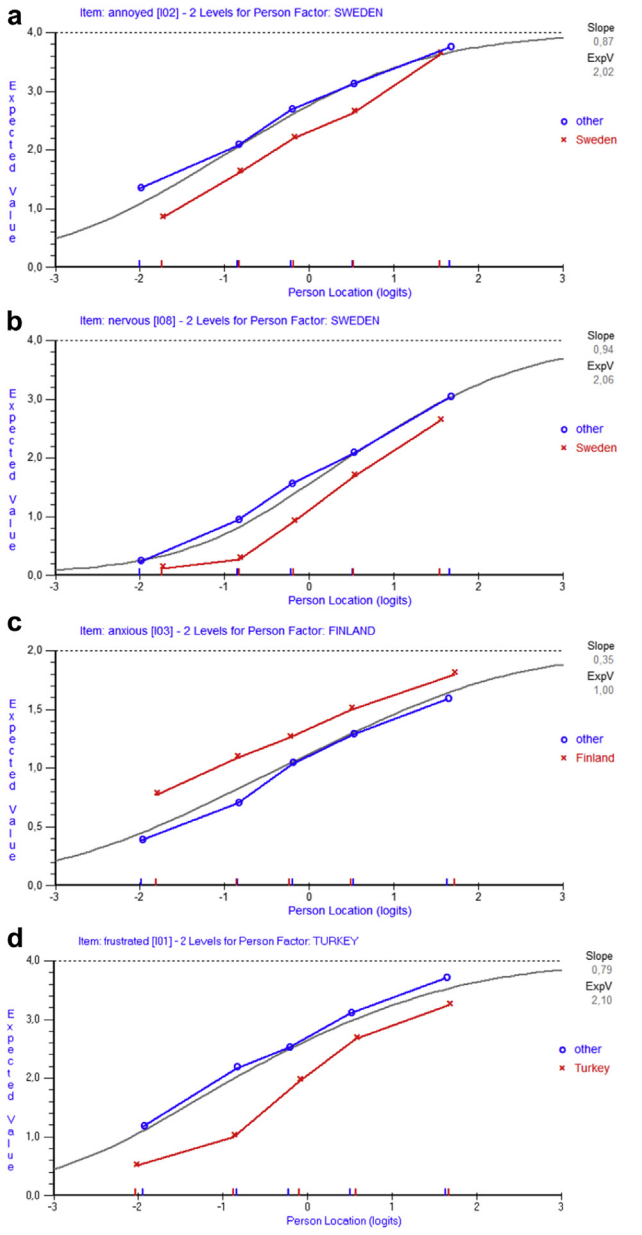

Figure 2. Item characteristic curves plotted by country. For the items of the Emotion subscale showing relevant differential item functioning by country: (**a**) annoyed by Sweden; (**b**) nervous by Sweden; (**c**) anxious by Finland; and (**d**) frustrated by Turkey. ExpV, expected value.

interval-scaled values were retrieved for Sweden, Finland, and Turkey. The final model showed no significant misfit to the RM (overall $\chi^2 = 32.2$; $df = 48$; $P > 0.01$), and a PSI of 0.88 indicated an excellent internal reliability for the adjusted subscale.

Functioning

In the primary analysis with the RM, the Functioning subscale showed no significant misfit to the RM (overall $\chi^2 = 34.2$; $df = 32$; $P > 0.01$) and no disordered thresholds. Five items of the subscale showed relevant DIF. Individual analysis by country showed that although the Swedish subpopulation showed a slightly higher likelihood for scoring higher on the item of being impaired in *Washing*, the DIF in the four other items was caused by the Turkish subpopulation (Figure 3). Individuals of the Turkish subpopulation were, compared with those from other countries at the same level of HRQOL impairment, less likely to report problems in doing *Home duties* or *Hobbies* but, on the other hand, more likely to experience impairment because of *Avoiding contact* with others or while *Touching family*.

Nevertheless, all five items could be split using the RM, and country-specific, interval-scaled values were retrieved for Sweden and Turkey. The final model showed no significant misfit to the RM (overall $\chi^2 = 71.4$; $df = 52$; $P > 0.01$) and a PSI equal to 0.83 also indicated good internal reliability for the adjusted subscale.

Treatment and prevention

The Treatment and Prevention subscale showed no significant misfit to the RM initially (overall $\chi^2 = 21.6$; $df = 28$;

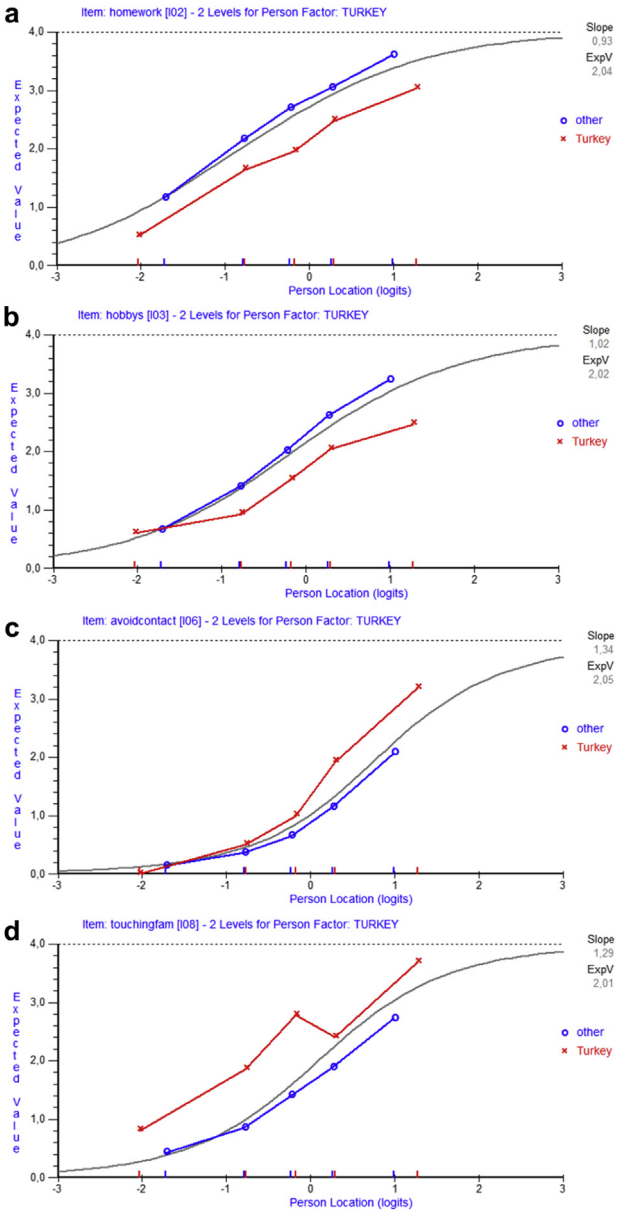

Figure 3. Item characteristic curves. For the items of the Functioning subscale showing relevant differential item functioning for Turkey: (**a**) doing homework; (**b**) hobbys; (**c**) avoiding contact with other people; and (**d**) touching family or partner. ExpV, expected value.

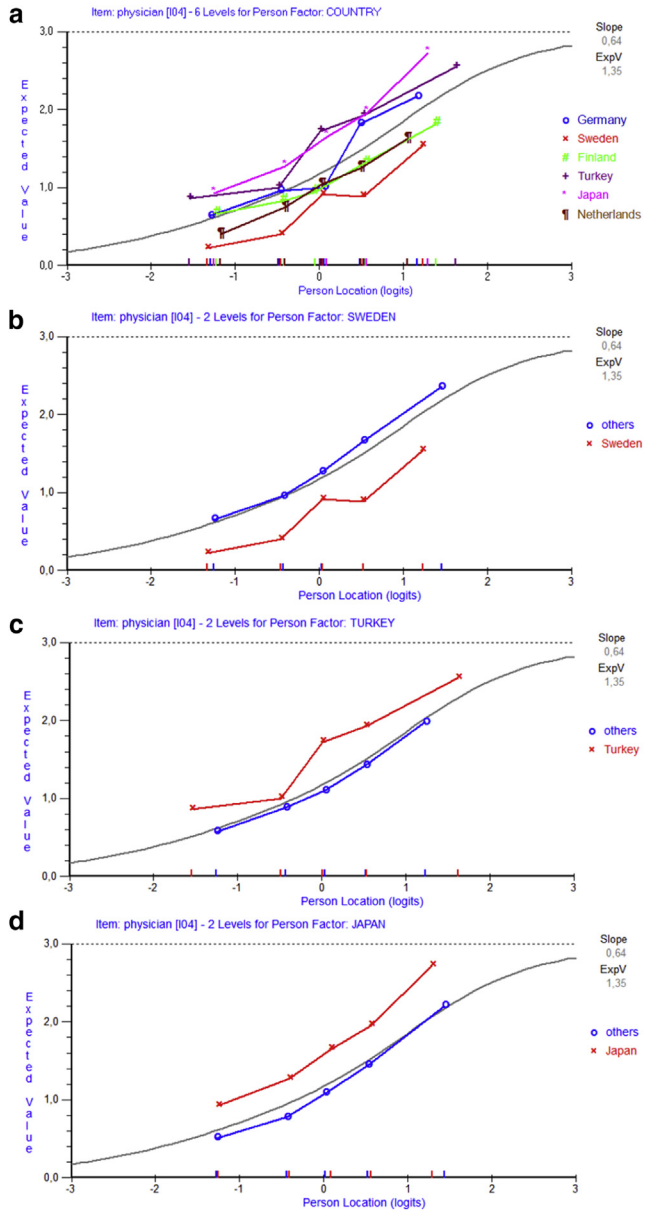

Figure 4. Item characteristic curves. For the item *visiting a physician* for (**a**) each country, (**b**) Sweden, (**c**) Turkey, and (**d**) Japan.

$P > 0.01$), but disordered thresholds were found for the item for feeling impaired because of *Visiting physicians*. After adjusting the response categories for this item, significant DIF was found in the analysis where all countries were tested in parallel (Figure 4a). Individual analysis for each country showed that this DIF was relevant for Sweden, Turkey, and Japan (Figure 4b–d). The Swedish subgroup was less likely to be impaired because of *Visiting physicians,* whereas the likelihood for the Turkish and Japanese subgroups to be impaired on this item was higher (Figure 4).

The DIF for this item was intensely discussed at an international meeting of the developers, and it was agreed that this DIF may be caused not only by cross-cultural factors but also by differences in the health care systems of the different countries. The Swedish population, living in a country with a national health system, may in general be less likely to visit a dermatologist often, even if HE is severe. Therefore, at the same level of HRQOL impairment compared with individuals from other countries, Swedish people were less likely to be impaired in this area because clinical visits are so scarce. If interpreted this way, the DIF found for *Visiting physicians* would not be an anomaly but would reflect true differences among countries. Nevertheless, because this fact is unlikely to explain the DIF for other countries, it was decided to conservatively remove this item from the scale to gain comparable HRQOL results across countries. Although one item was removed from the subscale, the final model showed no significant misfit to the RM (overall $\chi^2 = 14.5$; $df = 24$; $P > 0.01$), and the internal reliability remained good (PSI = 0.74).

Structural equation model and scoring of the QOLHEQ

The raw scores of the recoded subscales were introduced into a structural equation model representing the multidimensional structure of the QOLHEQ (see Supplementary Figure S1

online). According to the recommendations of Schermelleh-Engel et al. (2003), the fit indices received in this international sample indicated a good fit of the adjusted QOLHEQ scoring (standardized root mean square residual = 0.06; goodness-of-fit index =0.98; adjusted goodness-of-fit index = 0.98; normed fit index = 0.98). Removing the item *Visiting physicians* from the domain treatment/prevention lead to a higher correlation of that domain with the higher-order factor HRQOL compared with the original scoring (Ofenloch et al., 2014) ($\beta = 0.94$ vs. $\beta = 0.87$).

The Rasch-transformed interval-scaled values for each subscale are given overall and separately for Sweden, Finland, and Turkey (wherever a subscale was adjusted for DIF by country) in Supplementary Table S1 online. However, to use an HRQOL measure with this high precision for international comparison, great effort is needed to perform the scoring of the instrument: first, the raw scores need to be created (giving values from 0–4 [see Supplementary Table S2 online] for each answer on the QOLHEQ and summing them up by subscale), then each of those scores needs to be translated into country-specific values (see Supplementary Table S1), which leads to a rescoring of 278 values overall. This virtually cannot be performed without using modern statistical software. To enhance the use of the QOLHEQ with precise values for international comparison, an SPSS-Syntax, performing the QOLHEQ scoring by considering all those aspects and additionally transforming each subscale to a score with a range from 0–100, can be downloaded, together with the different language versions of the QOLHEQ, at www.QOLHEQ.dermis.net or found in the Supplementary Materials.

Effects of the cross-cultural adjustments

To visualize the effects that a cross-culturally inequivalent measurement can have on international comparisons, the mean values of the QOLHEQ before and after the adaption are given by country in Table 2. We assessed whether a QOLHEQ mean of a given country differed significantly from the QOLHEQ mean of the remaining countries. Before rescoring, the German and Dutch populations showed a significantly decreased mean in the Emotion subscale; after adapting for DIF, those effects disappeared—in case of the Dutch population, the value was even slightly increased (although not significant). On the other hand, the Japanese and Finnish populations showed no significant deviation before adapting for DIF; afterward, the values were significantly higher for the Japanese and significantly lower for the Finnish population.

DISCUSSION

The QOLHEQ is the only instrument to assess disease-specific HRQOL in patients with HE, and it can now be used for the comparison of HRQOL impairment in international clinical trials or epidemiological studies using the German, Dutch, Finnish, Swedish, Japanese, and Turkish versions of the QOLHEQ. In this validation study, we applied the methods of modern test theory, which are now widely accepted as the new standard in the dermatological community for assessing patient-reported outcomes (Liu et al., 2016; Nijsten et al., 2006, 2007; Tennant et al., 2004; Twiss et al., 2012). We were able to show that a

Table 2. Mean values for each subscale of the Quality of Life in Hand Eczema Questionnaire by country before (raw score) and after (value) cross-cultural adaptation[1,2]

	Symptoms		Emotions		Functions		Treatment	
Country	**Raw Score**	**Value**	**Raw Score**	**Value**	**Raw Score**	**Value**	**Raw Score**	**Value**
Germany								
Mean	**45.65**	**46.50**	**36.12**	30.60	**33.02**	**42.78**	**43.43**	**44.23**
SD	21.38	14.57	22.54	22.08	22.17	18.92	20.86	16.84
Sweden								
Mean	52.88	52.32	42.80	32.38	43.28	49.80	48.18	48.67
SD	19.43	14.28	23.00	24.20	22.64	17.95	20.43	16.80
Finland								
Mean	56.73	53.70	43.26	**29.83**	**35.43**	45.28	49.29	49.14
SD	20.51	14.04	21.52	20.72	21.70	18.13	19.69	14.60
Turkey								
Mean	**64.29**	**59.61**	**57.11**	36.57	**50.67**	**55.38**	**67.46**	**65.18**
SD	20.82	15.77	26.77	26.99	26.34	18.31	21.96	18.20
Japan								
Mean	55.93	53.39	41.82	**38.28**	39.76	48.55	48.56	48.94
SD	19.17	13.30	21.78	22.24	21.79	16.57	19.53	15.00
Netherlands								
Mean	**49.87**	**48.77**	**37.85**	37.55	38.42	47.57	**42.13**	**45.35**
SD	20.78	14.13	20.06	20.76	20.34	15.95	18.48	15.46
Total								
Mean	54.21	52.38	43.18	34.22	40.14	48.26	49.87	50.29
SD	21.12	14.91	23.64	23.12	23.22	18.04	21.80	17.60

Abbreviation: SD, standard deviation.
[1]Scores transformed to a range from 0–100.
[2]Boldface indicates values that differ significantly ($P < 0.05$) from the mean value of the remaining countries.

cross-cultural inequivalent measurement can lead to false conclusions about differences among populations. This highlights the importance of applying methods of modern test theory and testing for DIF before drawing international comparisons with a given measure. Still, the differences presented in Table 2 should not be interpreted as representative for the whole countries investigated, because we obtained only convenience samples and did not collect information on reasons for nonparticipation or clinical characteristics of the patients included.

Other dermatology-specific HRQOL instruments have shown cross-cultural inequivalence; however, no adaption was performed to those scales to obtain comparable values (Nijsten et al., 2007). According to the definition of Nijsten et al., the QOLHEQ can now be considered as a third-generation instrument for assessing HRQOL, because detailed information is given about dimensionality and response categories and an adaption for DIF was performed (Nijsten, 2012).

To our knowledge, this is the first study assessing cross-cultural aspects of HRQOL in a sample of dermatological patients from six countries. The strength of this study is that those aspects have been investigated in a sample of 662 patients with HE, who were equally distributed across countries so as not to overweight the impact of a single culture in the analysis. However, sampling within each country did not occur at random. As described, patients were sampled in a consecutive manner in the different centers. Still, random sampling would have been impossible, largely because the totality of HE patients is unknown; therefore, drawing a random sample did not seem to be possible. Further, it would have been beneficial to include other language versions of the QOLHEQ in this study. We acknowledge that we have no representation from Africa, Latin America, and Oceania. These languages need to be investigated in future studies.

Although we found some variations in the demographic characteristics of the participating centers, it was not expected that this affected the following DIF analyses on cross-cultural equivalence, because an assessment of DIF by age groups and sex in the German validation study showed that there was no significant DIF for the QOHEQ in those categories (Ofenloch et al., 2014). With the results of this international/cross-cultural validation study, the QOLHEQ is the first HRQOL instrument in dermatology with country-specific values that account for DIF among countries. However, further investigations of the QOLHEQ are needed to enhance the interpretability of national and international HRQOL impairment in HE patients. A banding study, like the one performed by Hongbo et al. (2005) for the Dermatology Life Quality Index, could provide a meaningful categorization of the QOLHEQ scores. In addition, further psychometric properties should also be assessed in samples at the country level, like in the German validation study (Ofenloch et al, 2014), because the cross-cultural validity reported here is only a part of the whole construct validity.

The QOLHEQ has now been rescored on an international level (and one item was removed from scoring), and the scoring of the QOLHEQ at the country level may deviate from its international scoring (e.g., the most precise scoring of the QOLHEQ in a purely German population is still the one presented in the German validation study [Ofenloch et al., 2014]). Therefore, national validation studies are needed to achieve the best psychometric properties for the instrument and the highest precision in measuring HRQOL in HE patients at country level. Ideally, in the future, studies using the QOLHEQ should report both national and international values.

MATERIALS AND METHODS

The QOLHEQ

The QOLHEQ was developed by an international expert group consisting of health scientists and dermatologists with special expertise in HE from Australia, Denmark, Finland, Germany, Japan, and Sweden. The development process was performed by this international group to build items that assess HRQOL in a cross-culturally equivalent manner, enabling the comparison of HRQOL impairment across countries. To receive a valid instrument covering all relevant aspects of HRQOL, patients suffering from HE were also involved in the development process through standardized questionnaires and focus groups. The translation process was then performed according to international guidelines, which are described in detail elsewhere (Oosterhaven et al., in press). The QOLHEQ consists of 30 items and assesses disease-specific HRQOL in HE patients using four scales covering impairment because of (i) symptoms, (ii) emotions, (iii) functioning, and (iv) treatment/prevention. A large validation study carried out in German HE patients showed the QOLHEQ to be a valid, reliable, and sensitive measure for assessing HRQOL in that population (Ofenloch et al., 2014).

Sampling

The HE patients participating in this international study were recruited consecutively at the North Karelia Central Hospital in Joensuu (Finland), University Hospital Heidelberg (Germany), the hospital and private clinics in Kumamoto (Japan), the University Medical Center Groningen (The Netherlands), Skåne University Hospital in Lund (Sweden), and Uludag University Medical Faculty Bursa and Sakarya University Medical Faculty Adapazar (Turkey). The study was approved by the local institutions, and written informed consent was received from all subjects included. Because an unbalanced sample size among groups might affect analyses of variance (Shaw and Mitchell-Olds, 1993) we aimed at recruiting a balanced sample of about 110 subjects per group. Each center consecutively recruited all patients with active HE and a history of HE within the last week into the study. An exclusion criterion was age younger than 18 years. According to the rules of the developers (Ofenloch et al., 2014), data were excluded from analysis if data for more than three items of the QOLHEQ were missing. This lead to an exclusion of 26 participants, who were equally distributed across the countries. It was only in the Japanese dataset that there were no missing data overall.

Statistical analysis

Basic statistical calculations were performed using SPSS 23 (IBM, Armonk, NY). As a method of IRT, a Rasch analysis with the partial credit model (Masters, 1982) was performed for each subscale separately using RUMM2030 (Rumm Laboratory Pty. Ltd., Duncraig, Western Australia, Australia). The initial scoring of the QOLHEQ in this analysis was performed according to the results of the primary validation study (Ofenloch et al., 2014). In a first step, the overall fit to the RM was assessed by (i) using a chi-square test for the item-trait interaction, (ii) checking for disordered thresholds of the item categories, and (iii) assessing the fit residuals for item mean interaction. To receive results comparable with the analysis of the primary validation study (Ofenloch et al., 2014), the chi-square test was performed with an adjusted sample size of n = 350 using the

chi-square test adjustment function in RUMM2030. If a disordered threshold was detected, a rescoring of single items was performed, to gain fit of the subscale to the RM. The fit of the final model to the RM was again assessed using a chi-square test over the item-trait interaction. The internal reliability of each subscale was assessed using the PSI. A value of PSI greater than 0.7 was considered to be evidence for good internal reliability.

After adjusting the subscales to resolve disordered thresholds, an analysis of variance was performed to assess cross-cultural equivalence by testing for DIF among countries. We ran this analysis in two steps: (i) we tested DIF for each item among all countries in parallel and (ii) we tested DIF for each country compared with the rest of the sample separately to identify the language version that actually caused the DIF for a specific item. The second analysis step was done to enhance interpretation of DIF for items showing DIF in the first step of the analysis. At an international meeting of the developers of the QOLHEQ, it was decided to assess, in addition to the significance of DIF, the magnitude of the deviation in terms of the fit residuals by country for each item showing DIF. The fit residual is the mean deviation of the response pattern for an item by country on a logit scale. In the case of uniform DIF, a deviation of +0.5 logits indicates that an individual in one country is about 20% more likely to score one response category higher on a specific item compared with an individual with the same degree of impairment from another country in the sample. It was decided that a mean deviation of greater than 0.5 logits for a subscale is defined as clinically relevant DIF, which was adjusted for in the ongoing analysis.

The adjustment for DIF was performed by splitting items for the calculation of the Rasch estimates, which means that those items are rendered unique for the groups showing DIF (Tennant et al., 2004). If, for example, an item shows DIF for Sweden, it is split into one separate item for Sweden containing missing values for all other countries and one item for all the other countries (which contains missing values for Sweden). This way, separate location and threshold values can be calculated for this item by country.

Before and after the rescoring of the subscales, the QOLHEQ was introduced into a structural equation model using AMOS 23 (IBM) representing all four domains and the higher-order factor HRQOL in one model. This was done to assess if the raw scores of the QOLHEQ in a sample of all countries combined still represented a valid multidimensional construct of HRQOL, as shown in the German validation study (Ofenloch et al., 2014).

CONFLICT OF INTEREST

The authors state no conflict of interest.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at www.jidonline.org, and at http://dx.doi.org/10.1016/j.jid.2017.02.969.

REFERENCES

Agner T, Jungersted JM, Coenraads PJ, Diepgen T. Comparison of four methods for assessment of severity of hand eczema. Contact Dermatitis 2013;69:107–11.

Apfelbacher CJ, Molin S, Weisshaar E, BauEr A, Elsner P, Mahler V, et al. Characteristics and provision of care in patients with chronic hand eczema: updated data from the CARPE registry. Acta Derm Venereol 2014;94:163–7.

Bond TG, Fox CM. Applying the Rasch model: fundamental measurement in the human sciences. Mahwah, NJ: Lawrence Erlbaum; 2001.

Brodersen J, Meads D, Kreiner S, Thorsen H, Doward L, McKenna S. Methodological aspects of differential item functioning in the Rasch model. J Med Econ 2007;10:309–24.

Brooks R. EuroQol: the current state of play. Health Policy 1996;37:53–72.

Chren MM, Lasek RJ, Quinn LM, Mostow EN, Zyzanski SJ. Skindex, a quality-of-life measure for patients with skin disease: reliability, validity, and responsiveness. J Invest Dermatol 1996;107:707–13.

Coenraads P-J. Hand eczema. New Engl J Med 2012;367(19):1829–37.

Diepgen T, Andersen KE, Brandao F, Bruze M, Bruynzeel D, Frosch P, et al. Hand eczema classification: a cross-sectional, multicentre study of the aetiology and morphology of hand eczema. Br J Dermatol 2009;160:353–8.

Dulon M, Skudlik C, Nübling M, John SM, Nienhaus A. Validity and responsiveness of the Osnabrück Hand Eczema Severity Index (OHSI): a methodological study. Br J Dermatol 2009;160:137–42.

Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)—a simple practical measure for routine clinical use. Clin Exp Dermatol 1994;19:210–6.

Grob JJ. Why are quality of life instruments not recognized as reference measures in therapeutic trials of chronic skin disorders? J Invest Dermatol 2007;127:2299–301.

Held E, Skoet R, Johansen JD, Agner T. The hand eczema severity index (HECSI): a scoring system for clinical assessment of hand eczema. A study of inter-and intraobserver reliability. Br J Dermatol 2005;152:302–7.

Hongbo Y, Thomas CL, Harrison MA, Salek MS, Finlay AY. Translating the science of quality of life into practice: what do dermatology life quality index scores mean. J Invest Dermatol 2005;125:659–64.

Liu Y, Li T, An J, Zeng W, Xiao S. Rasch analysis holds no brief for the use of the Dermatology Life Quality Index (DLQI) in Chinese neurodermatitis patients. Health Qual Life Outcomes 2016;14(1):1.

Masters GN. A Rasch model for partial credit scoring. Psychometrika 1982;47:149–74.

Moberg C, Alderling M, Meding B. Hand eczema and quality of life: a population-based study. Br J Dermatol 2009;161:397–403.

Nijsten T. Dermatology life quality index: time to move forward. J Invest Dermatol 2012;132:11–3.

Nijsten T, Meads DM, de Korte J, Sampogna F, Gelfand JM, Ongenae K, et al. Cross-cultural inequivalence of dermatology-specific health-related quality of life instruments in psoriasis patients. J Invest Dermatol 2007;127:2315–22.

Nijsten TE, Sampogna F, Chren MM, Abeni DD. Testing and reducing skindex-29 using Rasch analysis: Skindex-17. J Invest Dermatol 2006;126:1244–50.

Ofenloch RF, Diepgen TL, Popielnicki A, Weisshaar E, Molin S, Bauer A, et al. Severity and functional disability of patients with occupational contact dermatitis: validation of the German version of the Occupational Contact Dermatitis Disease Severity Index. Contact Dermatitis 2015;72(2):84–9.

Ofenloch RF, Weisshaar E, Dumke AK, Molin S, Diepgen TL, Apfelbacher C. The Quality of Life in Hand Eczema Questionnaire (QOLHEQ): validation of the German version of a new disease-specific measure of quality of life for patients with hand eczema. Br J Dermatol 2014;171:304–12.

Oosterhaven JAF, Schuttelaar ML, Apfelbacher C, Diepgen TL, Ofenloch RF. Guideline for the translation and national validation of the Quality of life in hand eczema Questionnaire (QOLHEQ). Contact Dermatitis, in press.

Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. Methods Psychol Res Online 2003;8(2):23–74.

Scott NW, Fayers PM, Aaronson NK, Bottomley A, de Graeff A, Groenvold M, et al. The relationship between overall quality of life and its subdimensions was influenced by culture: analysis of an international database. J Clin Epidemiol 2008;61:788–95.

Shaw RG, Mitchell-Olds T. ANOVA for unbalanced data: an overview. Ecology 1993;74:1638–45.

Tennant A, Penta M, Tesio L, Grimby G, Thonnard JL, Slade A, et al. Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. Med Care 2004;42(1 Suppl.): I37–48.

Thyssen JP, Johansen JD, Linneberg A, Menne T. The epidemiology of hand eczema in the general population—prevalence and main findings. Contact Derm 2010;62(2):75–87.

Twiss J, Meads DM, Preston EP, Crawford SR, McKenna SP. Can we rely on the Dermatology Life Quality Index as a measure of the impact of psoriasis or atopic dermatitis? J Invest Dermatol 2012;132:76–84.

Zumbo BD. A handbook on the theory and methods of differential item functioning (DIF). Ottawa: National Defense Headquarters; 1999.