Cross-cultural adaptation into French and validation of the SCAR-Q questionnaire

Thomas Radulesco, Julien Mancini, Martin Penicaud, Jean-Jacques Grob,
Marie-Aleth Richard, Patrick Dessi, Nausicaa Malissen, Justin Michel

Title page

Title: Cross-cultural adaptation into French and validation of the SCAR-Q questionnaire.

Authors:

Thomas Radulesco [1]*

Julien Mancini [2]*

Martin Penicaud [3]†

Jean-Jacques Grob [4]†

Marie-Aleth Richard [5]‡

Patrick Dessi [3]‡

Nausicaa Malissen [4]o

Justin Michel [1]o

* joint first authors, † joint second authors, ‡ joint before last authors, ° joint senior authors

Affiliation:

1. Aix Marseille Univ, APHM, IUSTI, La Conception University Hospital, ENT-HNS, Department, Marseille, France

[2] Aix-Marseille Univ, INSERM, IRD, APHM, UMR1252, SESSTIM, Department of Public Health (BIOSTIC), Aix-Marseille University, Hôpital de la Timone, Marseille, France

[3] APHM, Department of Oto-Rhino-Laryngology and Head and Neck Surgery, La Conception University Hospital, 13385 Marseille Cedex, France

[4] INSERM, CRCM, APHM, CHU Timone, Department of Dermatology and Skin Cancer, Aix Marseille University, Marseille, France.

[5] CEReSS-EA 3279, Research Centrer in Health Services and Quality of Life Aix Marseille University , Dermatology Department, Universitary Hospital Timone, Assistance Publique Hôpitaux de Marseille, APHM, 13385, Marseille, France

Short title: SCAR-Q: French version

CORRESPONDING AUTHOR : Thomas Radulesco

Address: Department of Oto-Rhino-Laryngology and Head and Neck Surgery, La Conception University Hospital, 147 Bd Baille, 13005 Marseille, France

Tel: +33491435858   Fax: +33491435810

Email: Thomas.radulesco@ap-hm.fr

ORCID-ID : 0000-0002-5939-5372

Declarations

INTRODUCTION

Scars are natural consequences of cicatrization following surgery, trauma or burn injury. It has been shown that poor aspect scarring can result in serious mental disorders or poor self-esteem [1, 2].

Patient-reported outcome and experience measures (PROMs and PREMs) are now part of the quality assessment of the overall care of patients [3]. Klassen et al. recently developed a specific PRO instrument regarding scar evaluation [4]. The SCAR-Q self-questionnaire is composed of three scales evaluating scar appearance (12 items), symptoms (12 items) and psychological impact (5 items). The SCAR-Q was designed to evaluate all scar types in children and adults. It was considered useful in research where appearance is an important outcome. Indeed when treatments aim to specifically improve the appearance of scars, asking patients what they think about how their scar looks seems a fundamental and practical measure [4]. After rigorous development, a study among 731 patients validated the scale [5]. Like most questionnaires designed to evaluate scarring from the point of view of patients, SCAR-Q was developed in English language [6, 7]. Unfortunately, such a questionnaire cannot be used in patients with another mother language. To our knowledge, no questionnaire regarding scarring evaluation by patients is available in French.

The objective of our study was to translate the SCAR-Q instrument into French [8–11] using International guidelines issued by ISPOR (International Society for Pharmacoeconomics and Outcomes Research) and by WHO (World Health Organization) [12, 13], then to test the reliability and validity of the translated version.

MATERIALS AND METHODS

Ethical considerations

We obtained permission to use the SCAR-Q self-questionnaire from the original development team. All participating patients gave their written consent before joining this study, which was carried out in accordance with the Declaration of Helsinki (1983). We obtained an Ethical Committee Authorization (Authorization N° 2020-67, Assistance Publique des Hôpitaux de Marseille, Marseille, France).

Translation

ISPOR and WHO recommendations were used to carry out the translation process [12, 13].

The translation process comprised 4 steps: 1/ forward translation 2/ back translation 3/ back-translation review, and 4/ patient interviews. The translation process required 3 individuals who were fluent in both English and French. Two individuals whose mother tongue was French served as forward translators (English to French). One was a surgeon and the other one professional translator specialized in medical translations. Once the two forward translations were completed, cognitive debriefing was conducted to establish consensus to merge the 2 translations and produce a single version. The last professional translator, whose mother tongue was English and was fluent in French served as back translator. The back translator did not see or review the original English version of the SCAR-Q. Once the back translation was complete, the questionnaire was returned to the original development team and analyzed by an expert review panel including various 4 medical specialists and 2 paramedics. The development team provided feedback and instructions prior to conducting the cognitive debriefing interviews. The cognitive debriefing interviews engage ten patients in the target patient population to determine the quality of the translation. The final version was not aimed to provide a literal translation, but rather a conceptually equivalent translated version worded in language patients can understand easily. As suggested by the developer of the scale SCAR-Q

scale scores were transformed into 0 (worst) to 100 (best) based on logits from Rasch measurement theory analysis [5].

Population

Inclusion criteria were: French nationality and French as their first language; age ≥18 years; having a visible scar, whatever the location, size (centimeters), date or etiology. We did our best to cover the variability of location, size of scars and the demographic variability. Exclusion criteria were: illiterate patients or those unable to understand and respond to the survey, and persons with cognitive limitation.

The patients were recruited for the consultation of dermatology or surgery. The reason for consultation was not always the scar, but the presence of a visible scar made it possible to include the patient if he agreed and if the inclusion and exclusion criteria were met. The statistical power was calculated. Computing power and minimum sample size for RMSEA: with 200 patients and a RMSEA=0.065 we had a 95.7% power (with an alpha risk=5%) to show that this RMSEA was lower than 0.08. The first 200 patients that fully answered the questionnaire were included in the study. The majority of the patients approached to complete the survey participated; however, we do not have demographic information on persons with declined to participate.

Statistical validation

We used questionnaires completed by patients and stored in a secure database. All analyses were conducted using IBM SPSS Statistics 20.0 (IBM Inc., New York, USA) and the *lavaan* package for R.3.6.3 (R Foundation for Statistical Computing, Vienna, Austria). All tests were two-sided and p-values <0.05 were considered significant.

Continuous and categorical variables were described using, respectively, means (± standard deviation) and counts (percentages). At scale level, floor and ceiling effects were considered to

be present if more than 15% of respondents achieved, respectively, the lowest or the highest possible score, [14]. At item level, these effects were considered to be present if more than 95% of respondents answered in the lowest or highest response category [15]. Items were considered redundant if the polychoric inter-item correlation was > 0.7 and irrelevant if < 0.2 [15]. Reliability was assessed using ordinal Cronbach's alpha (α) [16] and considered satisfactory if $\geq$ 0.7 [17]. Structural validity was tested using confirmatory factor analysis (CFA) with the robust weighted least squares (WLSMV) estimator and Delta parameterization, based on the original 3-factor structure of the initial version. Model fit of a correlated 3-factor structure was examined using the root mean square error of approximation (RMSEA, good fit if <0.06, poor fit if $\geq$0.10, acceptable elsewhere), the comparative fit index (CFI) and the Tucker-Lewis index (CFI and TLI, good fit if >0.95, poor fit if <0.90, acceptable elsewhere) [18].

The Spearman rank coefficient (ρ) was used to assess correlations between subscales. Three hypotheses were identified based on previous findings to determine known group validity: higher SCAR-Q scores were excepted when the scars were bigger, more recent, and localized on the face [5]. Univariate analyses were performed using ANOVAs.

To test the repeatability of this new scale, test-retest reliability was assessed in 10 patients (given the ICC estimated ($\geq$0.94) we had an 87% power to detect an ICC higher than a lowest acceptable ICC of 0.6 with a two-sided test). We considered that no clinical changes would occur in one month given the long-time scars of the patients included (75% of scar older than 12 months, scars were more than 13 years old on average). Patients answered the questionnaire again under the same conditions 1 month later. Consistency between responses was evaluated using intra-class correlation coefficient. The closer the coefficient to 1, the higher the repeatability. This test was performed for each subscale.

Before initiating the study, all researchers were trained in patient interviewing. As all questionnaires were completed, there were no concerns regarding missing data.

RESULTS

Translation

Some differences were found between the two forward translations, regarding both instructions and items. For example, "How does your scar look?" was translated as "A quoi ressemble votre cicatrice?"by one translator (literally "what does your scar look like ?") and as "Quel est l'aspect de votre cicatrice?" (literally "what is the aspect of your scar?" by the other. A "reconciled" version containing both instructions and items was submitted to the back translator. No major comments were made by the development team concerning the backtranslation. Patient interviews included ten patients (mean age = 37 years old, range 25-62) with different types of scars (burns, surgical or traumatic scars). This step entailed minor modifications of item wording. For example, for item n°6 of the SCAR Appearance scale the word "irrégulière" was added to the scar description.

Statistical validation

Two hundred patients completed the questionnaire. Population characteristics are reported in **Table 1.** We included 109 females (54.5%) and 91 males (46.5%). Mean age (±SD) was 53.3 (±18) years-old (min=18; max=84). Mean scar size was 6.8 (±6.3) cm (min=1; max=35) and mean scar date was 156.7 (±169) months (min=1; max=651).

Regarding item description, no floor or ceiling effect was found (max=85%) for all items. Regarding scale description, we found a ceiling effect for all 3 subscales: in 34.0%, 33.0%, and 56.0% of patients for the appearance, symptoms, and psychological impact subscales respectively (**Figure 1**). Forty-one patients (20.5%) were not impaired at all in the three dimensions. Internal consistency was considered satisfactory with a Cronbach's alpha > 0.7 for

all subscales (0.97, 0.90 and 0.97, respectively, for SCAR-Q Appearance, Symptom and Psychological impact scales). Several items from the Appearance and Psychosocial impact scales showed redundancy, with many inter-item correlations above 0.7. The CFA of the original structure displayed reasonable fit, with RMSEA=0.065 (90% Confidence Interval: 0.057-0.072), CFI=0.974, and TLI=0.972. Subscales were positively correlated but not strongly (0.45<rho<0.65, p<0.001).

Repeatability was tested on a randomized subgroup of 10 patients who did not differ from the other patients regarding their main characteristics and initial SCAR-Q scores. Intra-class correlation coefficients were 0.95 (95% Confidence Interval: 0.83-0.99), 0.94 (0.59-0.99), and 0.99 (0.94-1.00), for Appearance, Symptom and Psychological impact scales respectively.

DISCUSSION

Modern practitioners are committed to the diagnosis and treatment of patients and to addressing the patient's objectives and assessing patient satisfaction, as part of a comprehensive approach to care [19]. Scar is a part of surgical outcome. Efforts to assess the patient's perspective will result in better therapeutic interventions to improve scar quality and acceptance [20]. For that purpose, an instrument like SCARQ must be available in other languages than English, for a much broader use. French alone is estimated to be spoken daily in Metropolitan France by 76 million native speakers and by 235 million fluent speakers worldwide.

Our translation process has already been used in other studies [21, 22]. It ensures a translation of the idea or concept, more meaningful and accurate than a mere literal translation [23]. ISPOR and WHO recommendations were chosen for the translation process, in order to obtain culturally-adapted French version of the SCAR-Q questionnaire. Combining forth and back translation methods, we reduced potential biases in the process. Two-way translation,

combined with a panel expert review and patient interviews, guarantees a culturally and socially adapted version. No translation can perfectly match the original document because of conceptual differences due to diverging languages histories. However, backtranslation into the original language ensures a conceptually valid translation[23]. By incorporating different medical specialties and paramedics in the panel of experts, we tried to make an easily understandable questionnaire. Actual patients were integrated into this process, to check whether the proposed version was also suitable for non-medical or non-paramedical users. To cover the variability of situations, we included patients with different types (keloid, contracture, hypertrophic) and locations (face and neck, chest, upper and lower members) of scars in order to be as representative as possible of target patients.

Cronbach's alphas of the French version of SCAR-Q were high (0.97, 0.90 and 0.97 for SCAR-Q Appearance, Symptoms, and Psychological scales, respectively) thus demonstrating a high level of internal reliability. This is consistent Cronbach's alpha values of the original scale [4], ie 0.96, 0.91, and 0.95, respectively. The ceiling effect found on each subscale was probably related to the fact that scars were not the main reason for consultation in our cohort, explaining the high number of high scores (**Figure 1**). Moreover, the majority of patients had scars for more than 1 year, thus probably increasing patients' acceptance and increasing the scores. This must be confirmed by multiple patients' evaluation at different time. As expected, subscales were positively correlated, but not too strongly ($0.45 < rho < 0.65$) which supports the idea that they measure distinct dimensions. The repeatability analysis attests to the efficacy of the questionnaire although generated by translation into French.

No translation process is perfect and conceptual differences may remain. However, the combination of two methodologies aimed to minimize bias and statistical validation provided

acceptable results. Another study limit is that we do not have demographic information on persons who declined to participate.

Generalizability of study finding may be limited due to the sample characteristics. The sample included a high percentage of persons who had their scars for surgery or trauma. Our sample, smaller than in Ziolkowski et al. study [5], resulted in high ceiling effects across the 3 scales. Furthermore, our sample only included 2 patients with burn scars and our results may not be generalizable to patients with burns. While these finding demonstrate the initial validity of the French translation of the SCAR-Q, future studies are required to examine the reliability and validity in other samples, in particular with among persons with more recent and severe scars. In further studies, the French SCAR-Q could also be used simultaneously with English version in a study including both English and French-speaking patients. This would allow examination of measurement equivalence.

CONCLUSION

The SCAR-Q questionnaire is a reference in the field of patient-reported outcomes regarding scar evaluation. The 4-step translation-backtranslation process made it possible to obtain a high-quality French version in line with the original document. This translated version is now usable in France.

REFERENCES

1.   Finlay AY, Khan GK, Luscombe DK, Salek MS. (1990). Validation of Sickness Impact Profile and Psoriasis Disability Index in Psoriasis. *Br J Dermatol,* 123:751–756. https://doi.org/10.1111/j.1365-2133.1990.tb04192.x

2.   Zhang X, Liu Y, Deng X, et al. (2019). The Correlation Between Quality of Life and Acceptability of Disability in Patients With Facial Burn Scars. *Front Bioeng Biotechnol,* 7:329. https://doi.org/10.3389/fbioe.2019.00329

3.   Baumhauer JF. (2017). Patient-Reported Outcomes - Are They Living Up to Their Potential? *N Engl J Med,* 377:6–9. https://doi.org/10.1056/NEJMp1702978

4.   Klassen AF, Ziolkowski N, Mundy LR, et al. (2018). Development of a New Patient-reported Outcome Instrument to Evaluate Treatments for Scars: The SCAR-Q. *Plast Reconstr Surg Glob Open,* 6:e1672. https://doi.org/10.1097/GOX.0000000000001672

5.   Ziolkowski NI, Pusic AL, Fish JS, et al. (2020). Psychometric Findings for the SCAR-Q Patient-Reported Outcome Measure Based on 731 Children and Adults with Surgical, Traumatic, and Burn Scars from Four Countries. *Plastic & Reconstructive Surgery,* 146:331e–338e. https://doi.org/10.1097/PRS.0000000000007078

6.   Mundy LR, Miller HC, Klassen AF, et al. (2016). Patient-Reported Outcome Instruments for Surgical and Traumatic Scars: A Systematic Review of their Development, Content, and Psychometric Validation. *Aesthetic Plast Surg,* 40:792–800. https://doi.org/10.1007/s00266-016-0642-9

7.   Vercelli S, Ferriero G, Sartorio F, et al. (2003). How to assess postsurgical scars: A review of outcome measures. *Disability and Rehabilitation,* 25:2055–2063. https://doi.org/10.3109/09638280902874196

8.   Guillemin F, Bombardier C, Beaton D. (1993). Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol,* 46:1417–1432. https://doi.org/10.1016/0895-4356(93)90142-n

9.   Beaton DE, Bombardier C, Guillemin F, Ferraz MB. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine,* 25:3186–3191. https://doi.org/10.1097/00007632-200012150-00014

10.   Chakka S, Werth VP. (2019). Cross-cultural adaptations of health-related quality of life measures. *Br J Dermatol,* 181:659–660. https://doi.org/10.1111/bjd.18272

11.   Vasquez D, Aguirre D-C, Sanclemente G. (2019). Construct validity and responsiveness of the Colombian version of Skindex-29. *Br J Dermatol,* 181:770–777. https://doi.org/10.1111/bjd.17742

12.   Wild D, Grove A, Martin M, et al (2005) Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health*, 8:94–104. https://doi.org/10.1111/j.1524-4733.2005.04054.x

13.   WHO | Process of translation and adaptation of instruments. In: WHO. https://www.who.int/substance_abuse/research_tools/translation/en/. Accessed 25 Apr 2020

14.   Terwee CB, Bot SDM, de Boer MR, et al.(2007). Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol,* 60:34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

15.   Vet HCW de, Terwee CB, Mokkink LB, Knol DL. (2011). Measurement in Medicine: A Practical Guide. In: Cambridge Core. /core/books/measurement-in-medicine/8BD913A1DA0ECCBA951AC4C1F719BCC5. Accessed 25 Apr 2020

16.   Gadermann A, Guhn M, Zumbo B. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation*, 17:1–13

17.   Bland JM, Altman DG (1997) Cronbach's alpha. *BMJ*, 314:572. https://doi.org/10.1136/bmj.314.7080.572

18.   Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives: Structural Equation Modeling: A Multidisciplinary Journal: Vol 6, No 1. https://www.tandfonline.com/doi/abs/10.1080/10705519909540118. Accessed 25 Apr 2020

19.   Tredget EE, Shupp JW, Schneider JC. (2017). Scar Management Following Burn Injury. *J Burn Care Res*, 38:146–147. https://doi.org/10.1097/BCR.0000000000000548

20.   Kim B, Sgarioto M, Hewitt D, et al. (2018). Scar outcomes in dermatological surgery. *Australas J Dermatol*, 59:48–51. https://doi.org/10.1111/ajd.12570

21.   Poulsen L, Rose M, Klassen A, et al (2017) Danish translation and linguistic validation of the BODY-Q: a description of the process. *Eur J Plast Surg*, 40:29–38. https://doi.org/10.1007/s00238-016-1247-x

22.   Radulesco T, Penicaud M, Santini L, et al. (2018). French validation of the FACE-Q Rhinoplasty module. *Clinical Otolaryngology*, https://doi.org/10.1111/coa.13267

23.   Lacasse Y, Sériès F. (2004). Health-related quality of life measurement: a readers' guide]. *Rev Mal Respir,* 21:S63-70. https://doi.org/10.1016/s0761-8425(04)71462-x

Figure legends

Figure 1. Bar graph showing scores of patients for all subscales. A ceiling effect was observed in all subscales.

Table

		Test-retest participant				
		No (n=190)		Yes (n=10)		
		Mean	SD	Mean	SD	P-value
Appearance Scale (0-100)		73.6	23.1	80.0	23.8	0.54
Symptom Scale (0-100)		81.3	16.4	81.8	13.5	0.95
Psychosocial Impact (0-100)		85.3	21.6	86.3	20.1	0.90
		N	%	N	%	
Sex	Female	104	54.7%	5	50.0%	0.99
	Male	86	45.3%	5	50.0%	
Age (y)	<=30	29	15.3%	1	10.0%	0.11
	31-60	92	48.4%	2	20.0%	
	>60	69	36.3%	7	70.0%	
Scar size (cm)	<=2	31	16.3%	1	10.0%	0.60
	3-10	130	68.4%	6	60.0%	
	>10	29	15.3%	3	30.0%	
Scar date (months)	<=3	25	13.2%	2	20.0%	0.48
	4-12	23	12.1%	0	0.0%	
	>12	142	74.7%	8	80.0%	
Scar location	Abdomen	50	26.3%	3	30.0%	0.33
	Neck	16	8.4%	0	0.0%	
	Back	7	3.7%	1	10.0%	
	Face	70	36.8%	5	50.0%	
	Lower extremity	27	14.2%	0	0.0%	

Scar etiology	Upper extremity	15	7.9%	0	0.0%	
	Breast	5	2.6%	1	10.0%	
	Burn	2	1.1%	0	0.0%	0.72
	Surgical	153	80.5%	9	90.0%	
	Traumatic	35	18.4%	1	10.0%	

Table 1. Sample characteristics.

Tables

Characteristics	N (total = 200)	(%)
Sex		
Female	109	54.5
Male	91	46.6
Age (years)	*Min 18; max = 84; mean = 53.3 (±18)*	
≤30	30	15
31-60	94	47
> 60	76	38
Scar size (cm)	*Min = 1; Max = 35, mean = 6.8 (±6.3)*	
≤ 2 cm	32	16
3-10 cm	136	68
> 10 cm	32	16
Scar date (months)	*Min = 1; max = 651, mean 156.7 (±169)*	
≤ 3 months	27	13.5
4 months – 1 year	23	11.5
> 1 year	150	75
Scar etiology		
Surgical	162	81
Traumatic	36	18
Burn	2	1
Scar location		
Face	75	37.5
Abdomen	53	26.5
Neck	16	8
Back	8	4
Lower extremity	27	13.5
Upper extremity	15	7.5
Breast	6	3

Table 1. Sample characteristics.

SCAR-Q Appearance scale

	App	App2	App3	App4	App5	App6	App7	App8	App9	App10	App11
Appr1	1.00										
App2	0.76	1.00									
App3	0.77	0.80	1.00								
App4	0.67	0.74	0.75	1.00							
App5	0.75	0.81	0.83	0.76	1.00						
App6	0.62	0.77	0.77	0.76	0.79	1.00					
App7	0.57	0.70	0.73	0.68	0.72	0.72	1.00				
App8	0.63	0.73	0.70	0.69	0.76	0.77	0.84	1.00			
App9	0.67	0.77	0.80	0.74	0.80	0.87	0.72	0.80	1.00		
App10	0.68	0.82	0.83	0.72	0.90	0.86	0.78	0.80	0.87	1.00	
App11	0.70	0.80	0.83	0.81	0.86	0.84	0.74	0.78	0.84	0.88	1.00
App12	0.60	0.71	0.65	0.61	0.76	0.66	0.64	0.59	0.66	0.72	0.74

SCAR-Q Symptom scale

	Sym1	Sym2	Sym3	Sym4	Sym5	Sym6	Sym7	Sym8	Sym9	Sym10	Sym11
Sym1	1.00										
Sym2	0.49	1.00									
Sym3	0.51	0.59	1.00								
Sym4	0.60	0.34	0.40	1.00							
Sym5	0.37	0.51	0.62	0.48	1.00						
Sym6	0.47	0.54	0.53	0.35	0.40	1.00					
Sym7	0.53	0.63	0.48	0.43	0.54	0.58	1.00				
Sym8	0.44	0.72	0.48	0.30	0.53	0.41	0.43	1.00			
Sym9	0.52	0.59	0.45	0.44	0.66	0.35	0.52	0.61	1.00		
Sym10	0.40	0.43	0.68	0.35	0.46	0.54	0.44	0.34	0.39	1.00	
Sym11	0.46	0.38	0.52	0.45	0.35	0.46	0.25	0.21	0.27	0.50	1.00
Sym12	0.36	0.47	0.40	0.46	0.38	0.33	0.58	0.21	0.37	0.47	0.43

SCAR -Q Psychological impact

	Psy1	Psy2	Psy3	Psy4	Psy5
Psy1	1.00				
Psy2	0.83	1.00			
Psy3	0.81	0.87	1.00		
Psy4	0.69	0.82	0.80	1.00	
Psy5	0.79	0.83	0.85	0.76	1.00

Additional Table 1. Polychoric inter-item correlations for the 3 subscales. App = Appearance;

Sym = symptom.