International Journal for Quality in Health Care 2013; Volume 25, Number 2: pp. 157–166
Advance Access Publication: 18 January 2013
10.1093/intqhc/mzt001

Validity and reliability on three European language versions of the Safety Organizing Scale

DIETMAR AUSSERHOFER[1], MARIA SCHUBERT[2], MARY BLEGEN[3], SABINA DE GEEST[1] AND RENÉ SCHWENDIMANN[1]

[1]Institute of Nursing Science, University of Basel, Basel, Switzerland, [2]Centre of Clinical Nursing Science, University Hospital Zurich, Switzerland, and [3]School of Nursing–UCSF, San Francisco, 94143 CA, USA

Address reprint requests to: René Schwendimann, Institute of Nursing Science, University of Basel, Bernoullistrasse 28, CH-4056 Basel, Switzerland. Tel: +41-61-267-09-19; Fax: +41-61-267-09-55; E-mail: rene.schwendimann@unibas.ch

Accepted for publication 2 December 2012

Abstract

Background. The Safety Organizing Scale (SOS) offers a reliable snapshot of nurses' engagement in unit-level safety behaviors in hospitals. As no comparable questionnaire exists in German, French and Italian, we explored the psychometric properties of SOS translations into each of those languages.

Design and Methods. The psychometric properties of the nine-item SOS were tested according to American Educational Research Association guidelines.

Subjects and Setting. Between October 2009 and June 2010, 1633 registered medical and/or surgical nurses in 35 Swiss hospitals completed translated SOS questionnaires.

Results. For each translation, psychometric evaluation revealed evidence based on content (scale-content validity index >0.89), response patterns (e.g. average of missing values across all items = 0.80%), internal structure (e.g. comparative fit indices >0.90, root mean square error of approximation <0.08) and reliability (Cronbach's alpha >0.79). We differentiated the scale regarding one related concept (implicit rationing of nursing care). Higher SOS scores correlated with supportive leadership and lower nurse-reported medication errors, but not with nurse-reported patient falls.

Conclusions. The SOS offers a valuable measurement of engagement in safety practices that might influence patient outcomes. Initial evidence regarding the validity and reliability of the translated versions supports their use in German, French and Italian. Concurrent validity will require confirmation via further analysis using more reliable outcome measures (e.g. mortality rates). The translated versions' predictive validity needs to be established in prospective studies.

Keywords: high reliability, safety organizing, patient safety behaviors, psychometric evaluation, Swiss hospitals

Introduction

In the past years, it has become obvious that quality and patient safety are major challenges for health care [1–3]. According to patient safety experts, developing a 'safety culture' is a foundation for overcoming current safety and quality problems in health care [4–6]. Numerous conceptual definitions exist for safety culture, most of which originated in high-risk industries [7, 8]. For instance, safety culture within health-care organizations can be defined as a 'subset of organizational culture which relates specifically to the values and beliefs concerning patient safety' [9]. As values and beliefs can vary significantly across departments and across units within the same health-care organization [10, 11], a 'lived safety culture' must be visible as reflected by observable safety behaviors within each unit [12].

Safety culture and high reliability

Despite many improvements, health care is still struggling in creating a culture that sustains high levels of safety and quality performance over time [13]. Research on 'high-reliability organizations' revealed that a key feature of their safety culture that facilitates the maintenance of excellent performances is 'collective mindfulness' [14]. To stay mindful, despite hazardous environments, frontline employees consider constantly five principles: tracking small failures, resisting oversimplification, remaining sensitive to operations, maintaining capabilities for resilience and taking advantage of shifting locations of expertise [13, 14]. Studying 'collective mindfulness' in health care could help to better understand health-care professionals' behaviors underlying a patient safety culture [12] and might be a fruitful pathway leading to maintenance of safety performance over time [13].

Measuring the reliability of health-care organizations

The Safety Organizing Scale (SOS) [12], developed by Vogus and Sutcliffe, corresponds to such interrelated behavioral safety processes of 'collective mindfulness' [12]. The SOS is a 1D instrument backed by high-reliability organizational theory. The nine items assess the extent to which registered nurses (RNs) and their colleagues engage in safety behaviors and practices on their unit. Previous evidence supports the SOS's strong psychometric properties [12].

As no validated safety culture questionnaire existed in German, French and Italian, the SOS's psychometric excellence and 1D structure made it a strong candidate for cross-cultural testing. In addition, replication of studies under new conditions increases external generalizability of findings and the validity of the underlying theory [15]; if results on the SOS could be confirmed in a different cultural context, the underlying theory of high-reliability organizations in health care and the measurement of it by the SOS would be supported. To measure safety culture across a national sample of Swiss hospitals, we translated the SOS into German, French and Italian and then used the translated versions within the framework of the Swiss RN4CAST study (Nurse Forecasting: Human Resources Planning in Nursing). We surveyed RNs, as they represent the largest subgroup of health-care professionals in hospitals and play a major role in guaranteeing patient safety [16–18].

This study aims to explore the psychometric properties of the German, Italian and French versions of the SOS by translating the instrument and assessing its reliability and validity in new settings.

Methods

Design

This is a sub-study of the Swiss RN4CAST study, a multicentre cross-sectional study within the EU Seventh Framework (EU Project number: 223468).

Setting and sample

RNs working on medical, surgical and mixed medical–surgical wards of Swiss acute-care hospitals were surveyed. Hospitals were sampled according to a quota sample method based on language region and hospital type (university, cantonal and regional hospitals with ≥60 beds and ≥50 RNs). For university and cantonal hospitals, random samples of 2–6 eligible wards were selected; for smaller regional hospitals, all eligible wards were included to survey at least 50 RNs per hospital. All RNs active on the selected units were invited to participate.

Cross-cultural translation of the SOS

To adapt the SOS in a culturally relevant and comprehensible form while maintaining its original meaning and intent [19], we translated the SOS according to the adapted Brislin model [20]—an iterative process involving forward–backward translation and expert group discussion. The seven-point Likert response scale was retained. Following back translation, our expert group identified no inconsistencies in any item's meaning or wording.

Validity and reliability testing

Our general objective was to determine whether German, French and Italian versions of the SOS would reveal psychometric properties similar to those of the original. Therefore, our validation strategy was based on that of Vogus and Sutcliffe [12]. Following American Educational Research Association guidelines, [21] we developed questions and hypotheses that guided the testing of the three translations' validity and reliability (see Table 1).

Validity testing. After translating, we asked experienced RNs whether the SOS reflected a relevant content domain. For each language version, 10 experienced RNs were asked to rate the extent to which the SOS reflected nurses' safety behaviors. We established 'evidence based on content' by calculating the consensus estimates of the ratings of item scale relevance (content validity index of individual items and overall scale) [22].

'Evidence based on response processes' was compiled through assessment of distribution and skewing of the data, missing responses, multiple crosses for each item and overall scale and acceptability (number of respondents omitting no items).

To test the 'internal consistency' of the measurements and the precision of test results, we calculated Cronbach's alpha. Furthermore, to test whether the translated versions of the SOS reliably reflect a unit level construct—making aggregation of data appropriate—we computed five measures. Using *F*-statistic from a one-way variance analysis, we calculated between-group variance. We applied two types of intraclass correlations (ICCs), to calculate the proportion of variance explainable by unit membership (ICC 1) and the reliability of unit means (ICC 2). Both measures describe how strongly responses from RNs in the same unit resemble each other. [12]. We also calculated design effects to account for within-group sample size, which could have inflated ICCs. The degree to which responses of individuals within a group are interchangeable was calculated with the within-group agreement [12].

To provide 'evidence based on internal structure', we aimed to confirm the 1D structure of the original (English) SOS [12]. To test whether our model would fit the data, we conducted confirmatory factor analysis.

We assessed 'discriminant and concurrent validity' to provide evidence of the relationships of the SOS with other variables. Both higher RNs' engagement in safety behaviors and lower levels of implicit rationing of nursing care might be visible features of a safety culture and, therefore, related to each other. A recent study confirmed this assumption, suggesting that the extent to which individual RNs are rationing nursing care might also depend on the prevailing safety

Table I Research questions and hypotheses of this validation study

Evidence on validity and reliability	Guiding questions and hypothesis	Statistical analyses	Necessary evidence
Evidence based on content	Are the nine items relevant and appropriate in terms of safety culture in the Swiss health-care setting?	Calculation of the Content Validity Indices for individual items (I-CVI) and scale-level content validity index (S-CVI/Ave)	I-CVI of more than 0.78 and S-CVI/Ave of more than 0.90 indicate excellent content validity [22]
Evidence based on response processes	How many missing values appear in the SOS? Are there distribution abnormalities in the different items of the SOS?	Descriptive statistics (frequencies, medians, interquartile ranges, means, standard deviations, variances, graphs and cross-tabulations)	Floor effects [>50% indicating one (not at all) or two (to a very limited extent)] Ceiling effects [>50% indicating six (to a great extent) and seven (to a very great extent)].
Internal consistency (reliability)	Is the SOS instrument internally consistent and does it reflect on a unit-level construct?	Calculations of ICCs, within-group agreement and design effects based on results of one-way analysis of variance with the Safety Organizing score as the dependent variable and hospital units as independent variable	Significant between-group variance using *F*-statistic ($P < 0.05$)
		Calculation of Cronbach's alpha	ICC 1 between 0.05 and 0.30 and ICC 2 above 0.70 [12] Design effects should be ≥2 to demonstrate the nesting of the data [36] Within-group agreement values should be 0.70 or greater [12] Cronbach's alpha ≥0.70 indicate factor consistency [37]
Evidence based on internal structure	The translated German, Italian and French versions of the SOS have an 1D structure	Confirmatory factor analysis	Non-significant chi-square values Comparative fit index >0.90 Weighted root mean square residual <0.90 Upper of the root mean square error of approximation <1.00 *P* of close fit >0.05 [33]
Evidence based on relationship with other variables			
Discriminant validity	Safety culture is not strongly correlated with implicit rationing of nursing care (BERNCA).	Spearman's correlation	Less than a strong correlation ($r_s < 0.40$)
Concurrent validity	There is a positive relationship between leadership abilities of the ward nurses (measured with the PES) and safety culture (measured with the SOS)	Multilevel linear regression analysis	Significant positive regression coefficient between leadership abilities and the SOS ($P < 0.05$) after adjusting for hierarchical data structure

(*continued*)

Table 1 Continued

Evidence on validity and reliability	Guiding questions and hypothesis	Statistical analyses	Necessary evidence
Concurrent validity	There is a negative relationship between nurse-to-patient ratio and safety culture.	Multilevel linear regression analysis	Significant negative regression coefficient between leadership abilities and the SOS ($P < 0.05$) after adjusting for hierarchical data structure
Concurrent validity	There is a negative relationship between safety climate and nurse-reported medication errors and patient falls	Multilevel logistic regression analysis	Significant odds for lower frequent nurse reports for medication errors and patient falls (odds ratios <1.00, $P < 0.05$) after adjusting for hierarchical data structure

culture, i.e. collective mindfulness [23]. Yet, implicit rationing of nursing care served as a comparison concept to explore whether the SOS is discriminant from this important factor determining patient safety and quality of care [24]. Vogus and Sutcliffe provided initial evidence that better leadership abilities of nurse managers and adequate staffing levels are associated with better safety culture and fewer adverse events such as medication errors and patient falls [12, 25]. To provide concurrent validity for the translated versions of the SOS, we assessed these relationships in the Swiss cross-cultural setting.

Five variables were used for discriminant and concurrent validity testing:

Implicit rationing of nursing care, leadership abilities, staffing level, medication errors and patient falls.

Implicit rationing of nursing care, i.e. 'the withholding of or failure to carry out necessary nursing measures for patients due to a lack of nursing resources (staffing, skill mix and time)' was measured with the 32-item, 1D Basel Extent of Rationing of Nursing Care (BERNCA) instrument [24], which asks how often in their last 7 working days, nurses have been unable to carry out any of 32 listed necessary nursing tasks. Evidence is supporting the BERNCA's validity and reliability [24]. Cronbach's alphas for the German-, French- and Italian language versions were 0.94, confirming the internal consistency of this construct.

To assess leadership abilities, we used the 'Nurse Manager Ability, Leadership, and Support' subscale of the 'Practice Environment Scale' [26], which asks nurses whether specific leadership elements are present at their workplace. Psychometric strength of the practice environment scale was reported in previous studies [26, 27]. Cronbach's alphas for the German- and Italian versions were 0.76, for the French version, 0.80.

The staffing level and adverse outcomes were measured with RN self-report of the 'nurse-to-patient ratio' on the last shift, and the frequency of 'medication errors' and 'patient falls' on their units over the last year.

These three variables matched single items from the instrument battery of the RN4CAST nurse questionnaire. All instruments and items (except the SOS) used in the Swiss RN4CAST study had been used in previous studies [28, 29].

Data collection and data management

Data collection took place between 12 October 2009 and 30 June 2010. For each participating hospital, a pre-identified contact person (e.g. nursing expert, chief nursing officers) supported us in planning and conducting the data collection. Completed questionnaires were scanned, and data were subjected to quality control procedures.

Statistical analysis

The statistical analyses corresponding to our validity and reliability testing are listed in Table 1. To evaluate the SOS's internal structure, we performed confirmatory factor analysis using MPlus (version 6.1, 2010, Muthen and Muthen).

Descriptive, correlation, reliability and variance analyses were performed using IBM SPSS Statistics (version 18.0.2; SPSS Inc., Chicago, IL, USA); regression analysis was performed with STATA (version 11/SE; StataCorp LP), whereas ICCs, design effects and within-group agreement were calculated with Microsoft Office Excel 2011®. All hypotheses were subjected to two-sided testing, whereby the level of significance was set at $P < 0.05$.

Ethical considerations

Surveyed RN consented to participate in the study by voluntarily completing and returning the questionnaires. The study was approved by ethics committees representing all involved cantons.

Results

Questionnaires were completed and returned by 1633 RNs (overall response rate = 72%: German = 73%; French = 74%; Italian = 69%). Of these, 1630 (German = 1074; French = 401; Italian = 155) were eligible for statistical analyses (3 questionnaires were excluded from analysis as they could not be assigned to a unit). Characteristics of the participant sample are presented in Table 2.

The content validity ratings for the three language versions of the SOS revealed the relevance of both the individual items and of the overall SOS scale. Content validity indices for individual items for the German version ranged from 0.6 (items 4) to 1.0 (items 2, 3 and 8); for the French version from 0.7 (item 7) to 1.0 (items 2, 6 and 8); and for the Italian version from 0.7 (items 3 and 7) to 1.0 (items 2, 6 and 8). Scale-level content validity index was 0.91 for the German and the French versions and 0.89 for the Italian version.

Frequencies of missing values were low, ranging from 0.4 to 1.9% (average = 0.8%). In all, 1564 RNs (95.8%) submitted questionnaires with no missing values. For confirmatory factor analysis, then, we used only data from complete questionnaires.

Mean item values on the seven-point scale ranged from 4.62 (standard deviation = 1.27; item 4) to 5.62 (standard deviation = 1.07; item 5). The mean for the entire scale (SOS score) was 5.11 (standard deviation = 0.91) and the median score was 5.22 (25–75th quartile = 4.56, 5.22 and 5.78). All nine items and the SOS score were slightly left skewed, and minor ceiling effects could be observed in items 5, 6 and 9.

The alpha coefficients for this 1D construct were 0.90 (German), 0.92 (French) and 0.79 (Italian), indicating scale reliability. The reliability of the SOS as an aggregate unit measure was shown by the significant analysis of variance, within group agreement, ICC 1 and design effects (Table 3). The ICC 2 scores were slightly below recommended levels for

Table 2 Characteristics of the participating hospitals and RNs

Hospitals	Total Switzerland ($n = 35$)	German-speaking region ($n = 20$)	French-speaking region ($n = 11$)	Italian-speaking region ($n = 4$)
Hospital type, *n*				
University hospital	4	2	2	–
Cantonal hospital	15	8	5	2
Regional hospital	16	10	4	2
Hospital size (acute-care beds), *n*				
Large (>500)	6	4	2	–
Intermediate (200–500)	12	6	4	2
Small (<200)	17	10	5	2
No. units, *n*				
Medical	62	32	22	8
Surgical	59	36	17	6
Mixed medical/surgical	11	11	–	–
Unit size (beds), median (IQR)	21 (8)	20 (9)	20 (8)	22 (5)
Patients per registered nurse, median (IQR)	7 (5)	7 (5)	7 (7)	8 (2)
RNs	($n = 1630$)	($n = 1074$)	($n = 401$)	($n = 155$)
Female, %	91.7	94.0	89.1	82.1
Age (in years), %				
20–30	41.7	42.9	39.7	37.8
31–40	27.9	26.1	30.8	33.6
41–50	20.5	20.7	20.8	18.2
>51	9.9	10.3	8.7	10.5
Nurse training in Switzerland, %	78.0	84.0	63.0	72.0
Employment, %				
>90%	48.5	47.5	12.8	9.0
51–90%	32.6	29.9	37.6	38.2
10–50%	18.9	22.6	49.6	52.8
Professional experience (in years), median (IQR)				
As a nurse	8 (15)	8 (16)	7 (13)	8 (14)
In this hospital	5 (10)	5 (10)	5 (10)	6 (10)

IQR, interquartile range.

Table 3 Measures to determine the effect of data nesting of the SOS within hospital units

Measures	German version ($n = 1038$)	French version ($n = 385$)	Italian version ($n = 141$)
F-statistics	$F(85\,952) = 4.19$***	$F(35\,349) = 2.48$***	$F(13\,127) = 2.48$**
Within-group agreement ($r_{wg(j)}$)	0.96	0.93	0.94
ICC (1)	0.19	0.11	0.11
ICC (2)	0.76	0.59	0.59
Design effect	3.09	2.21	2.21

**$P < 0.01$, ***$P < 0.001$.

the French and Italian versions, which might be an effect of the lower average group size (RNs per unit) when compared with units of German-speaking hospitals. However, design effects and the other measures support the assumption that all translated SOS versions are meaningful at the unit level and justify the aggregation of individual responses to the unit level.

For each language version, confirmatory factor analyses described the closest possible data fit. As Table 4 shows, items had highly significant factor loadings for all three versions. For the German and French versions, our model demonstrated excellent fit across all fit indices; for the Italian version, we failed only to demonstrate an acceptable root mean square error of approximation (Table 4). In summary, these results provide evidence of the 1D structure of all three translated versions of the SOS and justify the aggregation of the nine individual items to a single SOS score.

To show discriminant validity, correlation analysis between the SOS and the BERNCA revealed a weak, but statistically significant negative relationship for all three translations (German $r_s = -0.24$ $P < 0.01$; French $r_s = -0.24$, $P < 0.01$; Italian $r_s = -0.28$, $P < 0.01$). These results distinguish safety culture, as measured with the SOS, from implicit rationing of nursing care.

As reported in Table 5, concurrent validity was examined via several regression models. The first shows the association of nurse-reported leadership and patient-to-RN ratios on the SOS Score. For all three translations, leadership was positively related to the SOS, whereas no significant relationships were found between the patient-to-RN ratios and the SOS (Table 5). Model 2 shows that higher SOS scores were associated with lower nurse-reported medication errors for all three translations (German: odds ratio = 0.941; French: odds ratio = 0.959; Italian: odds ratio = 0.878), whereas in model 3, no SOS score was significantly related to patient falls in any language version (Table 5).

Discussion

This study provides evidence supporting the validity and reliability of all three translated versions of the SOS, which we tested in the cultural context of the Swiss health-care system, pursuing procedures conforming to rigorous international standards [21].

The SOS measures health-care professionals' engagement in crucial safety behaviors at the unit level, reflecting the safety culture of health-care organizations. Most instruments measuring the patient safety culture are based on multidimensional conceptualizations [30, 31] but capture few patient safety behaviors. In our understanding, safety culture is characterized by visible features of a safety culture [32], such as safety behaviors of professionals, that directly influence patient outcomes. The SOS's quick diagnostic sampling of crucial patient safety behaviors reflecting on 'collective mindfulness' makes it a valuable tool for monitoring the reliability of health-care organizations.

Overall, our results indicate similar psychometric properties to those for the original SOS. 'Evidence based on content' confirmed our accurate and rigorous translation process. This might have resulted in a low proportion of missing values and indicated the instrument's practicability and acceptability ('evidence based on response processes'). The observed slightly positive data skewing and ceiling effects for three out of nine items give little reason for concern (e.g. systematic measurement error), because all nine items still reflected variability across units. However, the SOS's responsiveness to change needs to be explored in future safety and quality improvement research.

The structure of the tool as 1D was supported for all versions. Item loadings on one factor were strong for all three versions. The one weak value was for the Italian SOS version; we failed to demonstrate adequate root mean square error of approximation. The root mean square error of approximation depends on the sample size, represented in the denominator of its equation [33]. Although only a few parameters required estimation, the sample size for the Italian version was likely not sufficient for an adequate estimation of this fit index and its confidence interval. Future studies should re-confirm the internal structure of the Italian version with a larger sample size.

Whereas we confirmed the hypothesis that supportive leadership was positively related with the SOS (H3), we found no association between the patient-to-RN ratio and the SOS (H4). The recall of the workload of each RN for the last shift might not be accurate, and there was little variability in the data across the hospitals, making it difficult to show a relationship with the SOS. Similar to Vogus and Sutcliffe [12], SOS scores were inversely associated with nurse-reported medication errors, but, in contrast to their findings, we found no relationship with nurse-reported patient falls. We assume that the SOS might be less sensitive to this type of adverse events, as there are many patient-related risk factors triggering patient falls [34], which are difficult to influence by RNs' general safety behaviors. Our initial inconsistent findings on the relationship between the translated versions of the SOS and patient outcomes require confirmation via further analysis using more reliable outcome measures (e.g. mortality rates), as our nurse-reported patient outcomes may be subject to bias (recall bias) [35]. The RN4CAST study used a cross-sectional design, which does

Table 4 Standardized factor loadings, standard errors and fit indices of the confirmatory factor analysis for the three language versions of the SOS

Items of the SOS	German version ($n = 1038$)		French version ($n = 385$)		Italian version ($n = 141$)	
	Factor loadings	Standard error	Factor loadings	Standard error	Factor loadings	Standard error
We have a good 'map' of each other's talents and skills	0.708***	0.018	0.760***	0.026	0.402***	0.068
We talk about mistakes and ways to learn from them	0.823***	0.014	0.814***	0.021	0.783***	0.042
We discuss our unique skills with each other so we know who on the unit has relevant specialized skills and knowledge	0.828***	0.012	0.859***	0.016	0.762***	0.044
We discuss alternatives as to how to go about our normal work activities	0.760***	0.014	0.879***	0.016	0.826***	0.036
When giving report to an oncoming nurse, we usually discuss what to look out for	0.616***	0.023	0.746***	0.028	0.369***	0.075
When attempting to resolve a problem, we take advantage of the unique skills of our colleagues	0.804***	0.013	0.771***	0.021	0.665***	0.051
We spend time identifying activities we do not want to go wrong	0.630***	0.021	0.653***	0.029	0.742***	0.038
When errors happen, we discuss how we could have prevented them	0.736***	0.019	0.799***	0.021	0.843***	0.040
When a patient crisis occurs, we rapidly pool our collective expertise to attempt to resolve it	0.611***	0.022	0.708***	0.027	0.489***	0.060
Chi-square	50.710***		26.003*		24.811*	
Degrees of freedom	13		13		13	
Comparative fit index	0.997		0.998		0.993	
Root mean square error of approximation (90% confidence interval)	0.053 (0.038, 0.069)		0.051 (0.021, 0.079)		0.080 (0.028, 0.128)	
P of close fit	0.352		0.437		0.139	
Weighted root mean square residual	0.462		0.319		0.375	

*$P < 0.05$, ***$P < 0.001$.

Table 5 Multilevel regression analysis results for relationship of the SOS with other variables (concurrent validity)

Language versions	Variables	SOS (model 1)	Medication errors (model 2)	Patient falls (model 3)			
		Coefficient (*P*)	95% CI	Coefficient (*P*)	95% CI	Coefficient (*P*)	95% CI
German version (*n* = 1038)	Safety Organizing Score	–	–	−0.060 (<0.001)	−0.092 to −0.028	−0.013 (0.279)	−0.036–0.010
	Leadership	0.612 (<0.001)	0.524–0.700	−0.016 (0.604)	−0.064–0.031	0.007 (0.684)	−0.027–0.042
	Patient-to-RN ratio	−0.006 (0.142)	−0.014–0.002	0.001 (0.336)	−0.002 to 0.006	0.001 (0.362)	−0.001–0.004
	Intercept (86 units)	2.977 (0.001)	2.488–3.466	0.545 (<0.001)	0.281–0.809	0.079 (0.414)	−0.111–0.270
French version (*n* = 385)	Safety Organizing Score	–	–	−0.042 (0.031)	−0.081 to −0.004	−0.017 (0.340)	−0.051–0.018
	Leadership	0.603 (<0.001)	0.443–0.762	−0.013 (0.682)	−0.075–0.049	−0.035 (0.224)	−0.091–0.021
	Patient-to-RN ratio	0.003 (0.773)	−0.016–0.022	0.007 (0.048)	0.00007–0.014	0.002 (0.490)	−0.004–0.009
	Intercept (36 units)	2.415 (0.001)	1.502–3.328	0.517 (0.004)	0.166–0.867	0.432 (0.007)	0.119–0.744
Italian version (*n* = 141)	Safety Organizing Score	–	–	−0.130 (0.017)	−0.236 to −0.023	−0.092 (0.086)	−0.196–0.013
	Leadership	0.434 (<0.001	0.201–0.667	0.026 (0.703)	−0.108–0.160	−0.013 (0.847)	−0.143–0.117
	Patient-to-RN ratio	0.008 (0.663)	−0.029–0.046	0.005 (0.650)	−0.016–0.026	0.001 (0.895)	−0.021–0.024
	Intercept (14 units)	4.483 (0.001)	3.126–5.839	0.211 (0.640)	−0.673–1.095	0.601 (0.189)	−0.296–1.498

Multilevel linear (model 1) and logistic (models 2 and 3) regression analyses were performed with STATA 11 with unit-level random effects. Including the random effects accounts for the hierarchical structure of the data (RNs nested within units). Coefficient in model 1 refers to the unstandardized regression coefficients (B), in model 2 and 3 to logit coefficients. All models were adjusted for sociodemographic characteristics of RNs (age, education, employment grade and professional experience).
P, *P*-value; CI, confidence interval.

not allow deriving causal inferences between the SOS and patient outcomes. Predictive validity of the translated versions of the SOS for patient safety outcomes needs to be established in future prospective studies.

Conclusions

In summary, the SOS is a valuable tool to measure RNs' engagement in safety behaviors and processes with possible direct relationships to patient outcomes. Initial evidence on the validity and reliability of the translated SOS versions supports their use in German-, French- and Italian-speaking health-care settings. In clinical practice, it can be used to describe and monitor both the extent of health-care professionals' engagement in the tested behaviors/processes and the prevailing 'collective mindfulness' at unit level. As the individual items entail information on changeable behaviors/processes, the SOS allows unit and hospital leaders to plan, implement and evaluate interventions to improve the safety culture [7].

With further validity testing of the German, French and Italian SOS versions, this measure can be used in outcome research to explain its interaction with other known outcome influencing factors such as rationing of nursing care or the nurse practice environment. This would augment our knowledge of the most significant factors of clinical practice related to patient safety and quality of care.

Acknowledgements

The authors thank all hospitals for participating in the Swiss RN4CAST study. Special thanks go to Tracy Glass, PhD (Biostatistician) for her statistical support and to Chris Shultis for his work of editing the article.

Funding

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 223468. For more information on the RN4CAST project, please visit www.rn4cast.eu.

References

1. Soop M, Fryksmark U, Koster M *et al*. The incidence of adverse events in Swedish hospitals: a retrospective medical record review study. *Int J Qual Health Care* 2009;**21**:285–91.
2. Aranaz-Andres JM, Aibar-Remon C, Vitaller-Burillo J *et al*. Impact and preventability of adverse events in Spanish public hospitals: results of the Spanish National Study of Adverse Events (ENEAS). *Int J Qual Health Care* 2009;**21**:408–14.
3. de Vries EN, Ramrattan MA, Smorenburg SM *et al*. The incidence and nature of in-hospital adverse events: a systematic review. *Qual Saf Health Care* 2008;**17**:216–23.
4. Conen D. Maßnahmen zur Verbesserung der Patientensicherheit. *Bundesgesundheitsblatt–Gesundheitsforschung–Gesundheitsschutz* 2011;**54**:171–5.
5. Vincent C, Aylin P, Franklin BD *et al*. Is health care getting safer? *BMJ* 2008;**337**:a2426.
6. Institute of Medicine. *Patient Safety. Achieving a New Standard for Care*. Washington, DC: The National Academies Press, 2004.
7. Halligan M, Zecevic A. Safety culture in healthcare: a review of concepts, dimensions, measures and progress. *BMJ Qual Saf* 2011;**20**:338–43.
8. Guldenmund FW. The nature of safety culture: a review of theory and research. *Saf Sci* 2000;**34**:215–57.
9. Feng X, Bobay K, Weiss M. Patient safety culture in nursing: a dimensional concept analysis. *J Adv Nurs* 2008;**63**:310–9.
10. Deilkas E, Hofoss D. Patient safety culture lives in departments and wards: multilevel partitioning of variance in patient safety culture. *BMC Health Serv Res* 2010;**10**:85.
11. Singer SJ, Gaba DM, Falwell A *et al*. Patient safety climate in 92 US hospitals: differences by work area and discipline. *Med Care* 2009;**47**:23–31.
12. Vogus TJ, Sutcliffe KM. The Safety Organizing Scale: development and validation of a behavioral measure of safety culture in hospital nursing units. *Med Care* 2007;**45**:46–54.
13. Chassin MR, Loeb JM. The ongoing quality improvement journey: next stop, high reliability. *Health Aff (Millwood)* 2011;**30**:559–68.
14. Weick KE, Sutcliffe KM. *Managing the Unexpected: Resilient Performance in an Age of Uncertainty*. 2nd edn. San Francisco, CA: John Wiley, 2009.
15. McKinley W. Organizational theory development: displacement of ends? *Organ Stud* 2010;**31**:47–68.
16. Kendall-Gallagher D, Blegen MA. Competence and certification of registered nurses and safety of patients in intensive care units. *Am J Crit Care* 2009;**18**:106–13.
17. Teng CI, Dai YT, Shyu YI *et al*. Professional commitment, patient safety, and patient-perceived care quality. *J Nurs Scholarsh* 2009;**41**:301–9.
18. Aiken LH, Clarke SP, Cheung RB *et al*. Educational levels of hospital nurses and surgical patient mortality. *JAMA* 2003;**290**:1617–23.
19. Sperber AD. Translation and validation of study instruments for cross-cultural research. *Gastroenterology* 2004;**126**:S124–S8.
20. Jones PS, Lee JW, Phillips LR *et al*. An adaptation of Brislin's translation model for cross-cultural research. *Nurs Res* 2001;**50**:300–4.
21. American Educational Research Association. *Standards for Educational and Psychological Testing*. Washington DC: American Educational Research Association, 1999.
22. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Res Nurs Health* 2007;**30**:459–67.
23. Schubert M, Ausserhofer D, Desmedt M *et al*. Levels and correlates of implicit rationing of nursing care in Swiss acute care hospitals–a cross sectional study. *Int J Nurs Stud* in-press.

24. Schubert M, Glass TR, Clarke SP *et al*. Validation of the basel extent of rationing of nursing care instrument. *Nurs Res* 2007;**56**:416–24.

25. Vogus TJ, Sutcliffe KM. The impact of safety organizing, trusted leadership, and care pathways on reported medication errors in hospital nursing units. *Med Care* 2007;**45**:997–1002.

26. Lake ET. Development of the practice environment scale of the Nursing Work Index. *Res Nurs Health* 2002;**25**:176–88.

27. Gajewski BJ, Boyle DK, Miller PA *et al*. A multilevel confirmatory factor analysis of the Practice Environment Scale: a case study. *Nurs Res* 2010;**59**:147–53.

28. Schubert M, Glass TR, Clarke SP *et al*. Rationing of nursing care and its relationship to patient outcomes: the Swiss extension of the International Hospital Outcomes Study. *Int J Qual Health Care* 2008;**20**:227–37.

29. Aiken LH, Clarke SP, Sloane DM. Hospital staffing, organization, and quality of care: cross-national findings. *Nurs Outlook* 2002;**50**:187–94.

30. Sexton JB, Helmreich RL, Neilands TB *et al*. The Safety Attitudes Questionnaire: psychometric properties, benchmarking data, and emerging research. *BMC Health Serv Res* 2006;**6**:44.

31. Nieva VF, Sorra J. Safety culture assessment: a tool for improving patient safety in healthcare organizations. *Qual Saf Health Care* 2003;**12**:ii17–23.

32. Flin R, Mearns K, O'Connor P *et al*. Measuring safety climate: identifying the common features. *Saf Sci* 2000;**34**: 177–92.

33. Kline RB. *Principles and Practice of Structural Equation Modeling*. 3rd edn. New York: The Guilford Press, 2011.

34. Cameron ID, Murray GR, Gillespie LD *et al*. Interventions for preventing falls in older people in nursing care facilities and hospitals. *Cochrane Database Syst Rev* 2010;**20**: CD005465.

35. Singer S, Lin S, Falwell A *et al*. Relationship of safety climate and safety performance in hospitals. *Health Serv Res* 2009;**44**:399–421.

36. Sorra JS, Dyer N. Multilevel psychometric properties of the AHRQ hospital survey on patient safety culture. *BMC Health Serv Res* 2010;**10**:199.

37. Pett M, Lackey N, Sullivan J. *Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research*. Thousand Oaks: Sage Publications, 2003.