BMJ Open

Psychometric properties of the German Hospital Survey on Patient Safety Culture and cross-national comparison of the instrument's dimensionality

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-018366
Article Type:	Research
Date Submitted by the Author:	26-Jun-2017
Complete List of Authors:	Gambashidze, Nikoloz; University hospital Bonn, Institute for patient safety Hammer, Antje; University hospital Bonn, Institute for patient safety Broesterhaus, Mareen; University hospital Bonn, Institute for patient safety Manser, Tanja; University hospital Bonn, Institute for patient safety
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Research methods
Keywords:	Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Health & safety < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, International health services < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, STATISTICS & RESEARCH METHODS

Title: Psychometric properties of the German Hospital Survey on Patient Safety Culture and cross-national comparison of the instrument's dimensionality

Authors: Nikoloz Gambashidze*[1], Antje Hammer*[1], Mareen Brösterhaus[1], Tanja Manser[1], On behalf of the WorkSafeMed Consortium**

[1] University Hospital Bonn, Institute for Patient Safety, Bonn, Germany

* Co-first author.

** The WorkSafeMed Project Consortium comprises: Luntz E, Rieger MA (project lead), Sturm H, Wagner A(Institute of Occupational and Social Medicine and Health Services Research, University Hospital of Tuebingen), Hammer A, Manser T (Institute for Patient Safety, University Hospital Bonn), Martus P (Institute for Clinical Epidemiology and Applied Biometry, University Hospital of Tuebingen), Holderied M (University Hospital Tuebingen).

Corresponding author:

Nikoloz Gambashidze,

Institute for Patient Safety,

University Hospital Bonn,

Sigmund-Freud-Str. 25,

53127 Bonn, Germany

Phone: +49 228 287 13948

E-Mail: Nikoloz.Gambashidze@ukbonn.de

Running title: Psychometrics of German HSPSC

Word count, abstract: 248.

Word count, manuscript: 2805.

Abstract

Objective: To study the psychometric characteristics of German version of the Hospital Survey on Patient Safety Culture, and to compare its dimensionality to other language versions in order to understand the instrument's potential for cross-national studies.

Design: Cross-sectional multicenter study to establish psychometric properties of German version of the survey instrument.

Setting: 73 units from 37 departments of two German university hospitals.

Participants: Clinical personnel (n=995 responses, response rate 39.6%).

Primary and secondary outcome measures: Psychometric properties (e.g. Model fit, internal consistency) of the instrument, and comparison of dimensionality across different language translations.

Results: The instrument demonstrated acceptable to good internal consistency (Cronbach's alpha 0.64-0.88). Confirmatory factor analysis of the original 12-factor model resulted in marginally satisfactory model fit (RMSEA= 0.05; SRMR; CFI=0.90; GFI=0.88; TLI=0.88). Exploratory factor analysis resulted in an alternative 8-factor model with good model fit (RMSEA= 0.05; SRMR; CFI=0.95; GFI=0.91; TLI=0.94) and internal consistency (Cronbach's alpha 0.73-0.87). Analysis of the dimensionality compared to models from 10 other language versions revealed eight dimensions with relatively stable composition and appearance across different versions and four requiring further improvement.

Conclusions: The German version of Hospital survey on Patient Safety Culture demonstrated satisfactory psychometric properties for use in German hospitals. However, our comparison of instrument dimensionality across different language versions indicates limitations concerning cross-national studies. Results of this study can be considered in interpreting findings across national contexts, in further refinement of the instrument for cross-national studies, and to better understand various facets and dimensions of patient safety culture.

Keywords: Quality in health care; Health & safety; International health services; statistics & research methods.

Article Summary

Strengths and limitations of this study

- Our study supports the development of a more uniform factor structure for the Hospital Survey on Patient Safety Culture across language versions in order to facilitate its use in cross-national research.
- By evaluating commonalities and variations in different language versions of the Hospital Survey on Patient Safety Culture we identify relatively stable factors, as well as those in need for improvement.
- First study to validate the German version of the Hospital Survey on Patient Safety Culture for clinical personnel.
- The considerable diversity in study methodology and reporting of studies with different language versions of the Hospital Survey on Patient Safety Culture presents an obstacle for cross-national use of the instrument that has yet to be overcome.

Introduction

All healthcare organizations face specific sets of risks and challenges regarding patient safety. These challenges change dynamically over time, reflecting developments within the organization as well as in its operating environment such as changes in demographics and epidemiology or in patient behavior. To effectively manage these challenges, it is recommended for healthcare organizations to develop a culture of safety that prioritizes safety and organizational learning among other organizational goals [1]. Safety culture is generally considered to be a relatively stable construct, rooted in organizational culture [2].

A number of instruments for measuring safety culture in healthcare organizations have been developed. These instruments enable researchers and decision-makers to evaluate and compare results on different levels of the healthcare system [3]. Comparing results across units and hospitals and establishing benchmarks can drive continuous patient safety improvement. One of the most widely used instruments for evaluating healthcare providers' perception of safety culture in hospital setting is the Hospital Survey on Patient Safety Culture (HSPSC) [4]. The instrument has been translated into many languages and used in different countries around the world [5–16].

There are two gaps that this study aims to address. Firstly, so far no German version of HSPSC has been validated for healthcare personnel in Germany. Secondly, despite some attempts at comparing safety culture at the international level [17, 18], the comparability of the different language versions of the instrument has not been studied systematically. While satisfactory psychometric properties were reported for the original North-American version [4] with 12 dimensions of patient safety culture, alternative factor structures have been reported for other language versions, with the number of dimensions ranging from 8 to 12 [6–15]. Because an instrument's dimensionality determines the interpretation of results, similarities and differences in dimensionality across different language versions should be considered for cross-national studies of patient safety culture.

Therefore, the aim of this study is two-fold: (1) validation of German version of HSPSC (HSPSC-D) by evaluation of its psychometric properties and (2) evaluation of the instrument's potential for cross-national studies, by comparative analysis of instrument's dimensionality as reported for different language versions.

Methods

Setting

This study was based on data from the cross-sectional, multicenter, mixed methods study *Working conditions, safety culture and patient safety in hospitals – what predicts the safety of the medication process* (WorkSafeMed), conducted between 2014 and 2017. In this article, we focus on HSPSC-D data to evaluate its psychometric properties. The WorkSafeMed study with all its components has been approved by the responsible ethics committees of the medical faculties of the project partners in Bonn (#350/14) and Tubingen (#547/2014BO1). Each partner complied with confidentiality requirements according to German law.

Sample

Safety culture data were collected in two German university hospitals April to July 2015. We included staff from inpatient units with ≥500 patients a year. Intensive care and psychiatric units were excluded. Across the two hospitals a total of 73 units from 37 departments participated in the study. The HSPSC-D questionnaire was distributed to 2512 healthcare professionals.

Measure

In order to develop a version of the HSPSC for German healthcare professionals (HSPSC-D), we used two previous German language versions as a starting point. A first translation of the HSPSC for hospital staff in the German speaking part of Switzerland [11] had been culturally and linguistically adapted for use in Swiss hospitals. Hammer et al. [19] used the Swiss version as a starting point for developing a management version of HSPSC to study perceptions of safety culture among medical directors in German hospitals. In our study, the instrument was adapted to be used with healthcare personnel in German hospitals.

The resulting HSPSC-D questionnaire follows the structure of the original North-American version [4] and includes 44 items, 42 of which compose 12 dimensions (10 safety culture dimensions and 2 outcome dimensions). These 42 items use a five-point Likert scale to measure agreement ranging from "strongly disagree" (1) to "strongly agree" (5), or frequency ranging from "never" (1) to "always" (5). The remaining two single item measures are *Number of events reported* (measured on six frequency groups from "No event reports" to "21 event reports or more") and *Patient safety grade* (measured on five-point scale from "Failing" to "Excellent").

<u>Analysis</u>

Data processing and preliminary analysis. After excluding responses with more than 30% missing values in HSPSC-D items, we conducted multiple imputations based on the expectation maximization (EM) algorithm using the statistical software NORM 2.03 [20, 21] to replace remaining missing values. Negatively worded items were reverse coded before further analysis.

Several indices were taken into account to ensure that our study sample, as well as every subset used in further analysis was appropriate for factor analysis. Kaiser-Meyer-Olkin (KMO) indicates if the sample of items is adequate for factor analysis, while Measure of Sampling Adequacy (MSA) indicates if an individual item is adequate for factor analysis. For both indices the value >0.7 is desired, and the value of >0.9 is considered perfect [22]. A significant p-value (<0.05) of Bartlett's test of sampling adequacy indicates that it is possible to extract more than one factor [22]. The analyses were performed using SAS 9.4.

Descriptive statistics. We calculated composite scores for each dimension suggested by Sorra and Nieva [4] by calculating the average of corresponding items. We also calculated percentages of positive responses for each dimension by dividing the number of positive responses on corresponding items by the number of non-missing answers in the dimension. Descriptive statistics for each item and dimension were evaluated, including range, mean, and standard deviation.

Exploratory Factor Analysis (EFA). We used EFA to evaluate the factor structure emerging from the study data. In general, EFA and Confirmatory Factor Analysis (CFA) should be performed using different subsets [23]. Thus, we performed the split-half cross validation, by randomly splitting our sample in two: "Exploring" (for EFA) and "Testing" subsets (for subsequent CFA). EFA using maximum likelihood was conducted using the "Exploring" subset. We used Varimax orthogonal pre-rotation, and Promax oblique rotation to aid with interpretation of factor model [23]. We used scree plot and Kaiser Criterion (Eigenvalues >1) for factor extraction. Factor loadings ≥0.4 were considered significant, and factor cross loading <0.4 was considered acceptable [22, 23]. Applying these criteria, we gradually eliminated problematic items until EFA resulted in a satisfactory factor structure.

Confirmatory Factor Analysis (CFA). We evaluated the model fit of the factor structure resulting from the EFA by conducting CFA using the "Testing" subset. By conducting a series of CFA using the complete dataset, we evaluated model fit of original 12-factor model [4], as well as other factor models reported by studies of different language versions of HSPSC. From the official website of the Agency for Healthcare Research and Quality (AHRQ) [24] we retrieved a list of studies including psychometric evaluation of the instrument and identified those reporting a different factor structure. Indices evaluated in CFA are presented in table 4.

Internal consistency. Internal consistency was evaluated by calculating Cronbach's alpha as an indicator of correlation between each item and the factor. As described by Sorra and Nieva [4], we considered Cronbach's alpha ≥0.6 as acceptable. Cronbach's alpha >0.7 was considered to be good [22]. Cronbach's alphas were calculated for all factor models considered in the CFA, including the factor model that emerged from EFA.

Evaluation of common dimensionality. In order to evaluate the potential of the instrument for cross-national studies, we evaluated its dimensionality as reported for different language versions. We evaluated appearance and composition of each of the 12 dimensions proposed by Sorra and Nieva [4] and of the 42 corresponding items in all factor models identified from AHRQ web-page [24].

Results

Study sample and descriptive statistics.

Out of 2512 distributed questionnaires 995 were completed, resulting in a response rate of 39.6%. Sample characteristics are presented in table 1.

Table 1: Characteristics of study sample.

Variables	N	%
Study site	**995**	**100.0%**
Hospital A	575	57.8%
Hospital B	420	42.2%
Gender	**995**	**100.0%**
Female	656	65.9%
Male	291	29.2%
missing	48	4.8%
Professional Groups	**995**	**100.0%**
Physician	183	18.4%
Physicians' assistant	198	19.9%
Nurse	552	55.5%
other	34	3.4%
missing	28	2.8%
Managerial functions	**995**	**100.0%**
Yes	195	19.6%
No	759	76.3%
missing	41	4.1%
Contact with patients	**995**	**100.0%**
Yes	965	97.0%

No	7	0.7%
missing	23	2.3%
Age (years)	**995**	**100.0%**
<25	61	6,1%
25-34	360	36,2%
35-44	230	23,1%
45-54	170	17,1%
>54	84	8,4%
missing	90	9,0%

21 responses contained more than 30% missing values and were not included in the analysis. Descriptive statistics of HSPSC-D items and dimensions after imputing remaining missing answers and reverse coding of the negatively worded items are presented in the Table 2.

Table 2: Descriptive statistics of HSPSC-D items and dimensions.

Dimension / Item [1,2,3]	Percent of positive responses [4]	Mean	St.d.
01. Teamwork Within Hospital Units	**42,3%**	**3,32**	**0,61**
A1. People support one another in this unit.	58,3%	3,65	0,78
A3. When a lot of work needs to be done quickly, we work together as a team to get the work done.	51,2%	3,50	0,84
A4. In this unit, people treat each other with respect.	40,9%	3,36	0,78
A11. When one area in this unit gets really busy, others help out.	18,8%	2,79	0,91
02. Organizational Learning—Continuous improvement	**32,7%**	**3,06**	**0,70**
A6. We are actively doing things to improve patient safety.	50,1%	3,40	0,91
A9. Mistakes have led to positive changes here.	23,5%	2,88	0,89
A13. After we make changes to improve patient safety, we evaluate their effectiveness.	24,4%	2,90	0,89
03. Non-punitive Response To Error	**50,2%**	**3,38**	**0,80**
A8. (R)[b] Staff feel like their mistakes are held against them.	40,1%	3,19	0,96
A12. (R) When an event is reported, it feels like the person is being written up, not the problem.	48,3%	3,33	0,99
A16. (R) Staff worry that mistakes they make are kept in their personnel file.	62,1%	3,62	0,99
04. Staffing	**24,9%**	**2,57**	**0,79**
A2. We have enough staff to handle the workload.	7,5%	2,01	0,97
A5. (R) Staff in this unit work longer hours than is best for patient care	23,1%	2,57	1,18
A7. (R) We use more agency/temporary staff than is best for patient care.	58,2%	3,57	1,20
A14. (R) We work in "crisis mode," trying to do too much, too quickly.	10,9%	2,13	1,02
05. Overall Perceptions of Safety	**34,4%**	**3,03**	**0,79**
A10. (R) It is just by chance that more serious mistakes don't happen around here.	41,1%	3,08	1,20
A15. Patient safety is never sacrificed to get more work done.	25,4%	2,75	1,04
A17. (R) We have patient safety problems in this unit.	43,9%	3,29	0,97
A18. Our procedures and systems are good at preventing errors from happening.	27,2%	3,00	0,89
06. Supervisor/manager expectations & actions promoting safety	**48,5%**	**3,34**	**0,71**
B1. My supervisor/manager says a good word when he/she sees a job done according to established patient safety procedures.	33,7%	3,03	1,02
B2. My supervisor/manager seriously considers staff suggestions for improving patient safety.	55,9%	3,51	0,87
B3. (R) Whenever pressure builds up, my supervisor/manager wants us to work faster, even if it means taking shortcuts.	42,8%	3,19	0,98
B4. (R) My supervisor/manager overlooks patient safety problems that happen over and over.	61,7%	3,61	0,89
07. Frequency of Event Reporting	**38,0%**	**3,00**	**1,03**
D1. When a mistake is made, but is caught and corrected before affecting the patient, how often is this reported?	39,0%	3,03	1,17
D2. When a mistake is made, but has no potential to harm the patient, how often is this reported?	30,1%	2,77	1,14
D3. When a mistake is made that could harm the patient, but does not, how often is this reported?	45,0%	3,19	1,13
08. Feedback and Communication About Error	**48,0%**	**3,36**	**0,85**
C1. We are given feedback about changes put into place based on event reports.	40,0%	3,18	1,04

C3. We are informed about errors that happen in this unit.	50,1%	3,41	0,99
C5. In this unit, we discuss ways to prevent errors from happening again.	53,9%	3,50	0,95
09. Communication Openness	**58,6%**	**3,60**	**0,68**
C2. Staff will freely speak up if they see something that may negatively affect patient care.	66,2%	3,74	0,87
C4. Staff feel free to question the decisions or actions of those with more authority.	45,4%	3,35	0,89
C6. (R) Staff are afraid to ask questions, when something does not seem right.	64,1%	3,71	0,91
10. Hospital Management Support for Patient Safety	**23,4%**	**2,79**	**0,86**
F1. Hospital management provides a work climate that promotes patient safety.	22,4%	2,83	0,94
F8. The actions of hospital management show that patient safety is a top priority.	21,1%	2,74	0,97
F9. (R) Hospital management seems interested in patient safety only after an adverse event happens.	26,8%	2,79	1,04
11. Teamwork Across Hospital Units	**29,0%**	**3,03**	**0,61**
F2. (R) Hospital units do not coordinate well with each other.	14,7%	2,57	0,91
F4. There is good cooperation among hospital units that need to work together.	22,6%	3,03	0,73
F6. (R) It is often unpleasant to work with staff from other hospital units.	49,1%	3,39	0,82
F10. Hospital units work well together to provide the best care for patients.	29,7%	3,14	0,77
12. Hospital Handoffs & Transitions	**35,3%**	**3,07**	**0,64**
F3. (R) Things "fall between the cracks" when transferring patients from one unit to another.	13,2%	2,50	0,88
F5. (R) Important patient care information is often lost during shift changes.	37,1%	3,16	0,89
F7. (R) Problems often occur in the exchange of information across hospital units.	29,3%	3,04	0,81
F11. (R) Shift changes are problematic for patients in this hospital.	61,5%	3,59	0,82
E1. Please give your work area/unit in this hospital an overall grade on patient safety.	**14,2%**	**2,78**	**0,76**
E2. In the past 12 months, how many event reports have you filled out and submitted?	**17,4%**	**2,90**	**0,76**

Note: Answers 4 and 5 ("A*gree*" and "S*trongly agree*" or "*Most of the time*" and "A*lways*") were considered as positive. Prior to analysis, negatively worded items were reverse coded.
[1]01-12 – corresponding dimension according to original North-American 12-factor model.
[2]A1-A18; B1-B4; C1-C6; D1-D3; E1-E2; F1-F11: Codes of questionnaire items.
[3](R) – negatively worded items, which were reverse coded prior to the analysis.
[4]N=974.

KMO for the complete sample was 0.93, and MSA for individual items ranged between 0.87 and 0.96. For "Exploring" and "Testing" subsets, KMO was 0.91 and 0.92 respectively, and MSA of individual items in both subsets ranged between 0.84 and 0.96. Bartlett's test was highly significant ($p<0.001$) for the dataset, as well as for both subsets. Preliminary analyses indicated that our sample and the subsets were adequate for factor analysis.

Exploratory Factor Analysis (EFA). We conducted EFA using the "Exploring" subset. 14 items not meeting the criteria (factor loading >0.4, cross loading <0.4) were excluded from the model, resulting in an 8-factor model with 28 items. The dimension *Organizational learning – continuous improvement* was completely removed. The dimensions *Staffing* and *Overall perceptions of safety* were merged together, as were the dimensions *Feedback and communication about error* with *Communication openness*, and *Teamwork across hospital units* with *Handoffs and transitions*. The resulting 8-factor model is presented in table 3.

Table 3: Appearance of HSPSC items in 12 analyzed factor models (8-factor EFA model, original 12-factor model, and 10 different versions).

HSPSC items[1,2]	Germany (Exploratory Factor Analysis)	**USA (Sorra & Nieva, 2004)**	England (UK) (Waterson et al., 2010)	Scotland (UK) (Sarac et al., 2011)	France (Occelli et al. 2013)	Switzerland (Perneger, 2013)	Switzerland (Pfeiffer and Manser, 2010)	Netherlands (Smits et al. 2008)	Sweden (Hedsköld et al., 2013)	Slovenia (Robida, 2013)	Turkey (Bodur, 2010)	Palestine (Najjar et al., 2013)
01. Teamwork within units												
A1	1	**1**	1	1	1	1	1	1	1	1	1	1
A3	1	**1**	1	1	1	1	1	1	1	1	1	1
A4	1	**1**	1	1	1	1	1	1	1	1	1	1
A11	(N)	**1**	(N)	1	(N)	1	(N)	1	1	1	1	1
02. Organizational learning												
A6	(N)	**2**	(N)	2	8	2	1	8	8	2	2	2
A9	(N)	**2**	(N)	2	8	2	(N)	8	8	2	3	2
A13	(N)	**2**	(N)	2	8	2	6	8	8	(N)	2	2
03. Nonpunitive response to error												
A8	3	**3**	3	3	3	3	3	3	3	3	3	3
A12	3	**3**	(N)	3	3	3	3	3	3	3	3	3
A16	3	**3**	3	3	3	3	3	3	3	3	3	3
04. Staffing												
A2	4	**4**	4	4	4	4	4	4	4	4	4	4
A5	4	**4**	(N)	4	4	4	4	4	4	4	2	4
A7	(N)	**4**	(N)	4	(N)	4	(N)	4	4	4	4	(N)
A14	4	**4**	4	4	4	4	4	5	4	4	4	4
05. Overall perceptions of safety												
A10	4	**5**	4	4	5	5	4	5	4	4	5	(N)
A15	(N)	**5**	(N)	4	5	5	(N)	(N)	4	4	5	5
A17	(N)	**5**	4	4	5	5	4	5	4	4	5	5
A18	(N)	**5**	(N)	(N)	5	5	7	5	8	4	5	5
06. Supervisor/manager expectations/actions												
B1	6	**6**	6	6	6	6	6	6	6	6	6	6
B2	6	**6**	6	6	6	6	6	6	6	6	6	6
B3	(N)	**6**	(N)	6	6	6	6	6	6	6	6	6
B4	6	**6**	(N)	6	6	6	6	6	6	6	6	6
07. Frequency of event reporting												
D1	7	**7**	7	7	7	7	7	7	7	7	7	7
D2	7	**7**	7	7	7	7	7	7	7	7	7	7
D3	7	**7**	7	7	7	7	7	7	7	7	7	7
08. Feedback and communication about error												
C1	8	**8**	8	8	8	8	8	8	8	8	8	(N)
C3	8	**8**	8	8	8	8	8	8	8	8	8	8
C5	8	**8**	8	8	8	8	8	8	8	8	8	8
09. Communication openness												
C2	8	**9**	9	8	9	8	8	9	8	8	8	8
C4	(N)	**9**	9	8	9	8	8	9	8	8	8	8
C6	(N)	**9**	9	8	9	8	(N)	9	3	(N)	8	(N)
10. Hospital management support for patient safety												
F1	10	**10**	(N)	10	10	10	10	10	10	10	10	10
F8	10	**10**	(N)	10	10	10	10	10	10	10	10	10
F9	10	**10**	(N)	10	10	10	10	10	10	10	10	10

11. Teamwork across hospital units												
F2	11	**11**	11	11	11	11	11	11	10	11	10	11
F4	11	**11**	11	11	11	11	11	11	10	11	10	11
F6	11	**11**	(N)	(N)	11	11	11	(N)	12	11	12	11
F10	11	**11**	11	11	10	11	11	11	10	11	10	11
12. Hospital handoffs and transitions												
F3	(N)	**12**	12	(N)	11	11	11	(N)	12	11	12	12
F5	(N)	**12**	12	12	11	11	(N)	12	12	11	12	12
F7	11	**12**	12	12	11	11	11	(N)	12	11	12	12
F11	(N)	**12**	12	12	(N)	11	(N)	12	12	(N)	12	12

Note:
The uncolored cells represent "No change" compared to original 12-factor model.
Colored boxes indicate items that were deleted (N) or moved to different dimension (Dimension number 1-12);
(N): items removed form factor model.
[1]01-12 – corresponding dimension according to original North-American 12-factor model.
[2]A1-A18; B1-B4; C1-C6; D1-D3; E1-E2; F1-F11: Codes of the questionnaire items.

Confirmatory factor analysis (CFA). CFA using the „Testing" subset demonstrated a satisfactory model fit of the factor structure that emerged from EFA (see table 4). The model satisfied desired thresholds of most analyzed indices (RMSEA=0.05; SRMR=0.05; GFI=0.90; CFI=0.93; TLI/NNFT=0.91).

Table 4. Results of Confirmatory Factor Analysis (CFA) of all 12 factor models analyzed.

Variables / Indices analyzed in CFA	Criteria	Germany (EFA)	USA (Sorra & Nieva, 2004)	England (UK) (Waterson et al., 2010)	Scotland (UK) Sarac et al., 2011)	France (Occelli et al. 2013)	Switzerland (Perneger, 2013)	Switzerland (Pfeiffer and Manser, 2010)	Netherlands (Smits et al. 2008)	Sweden (Hedsköld et al., 2013)	Slovenia (Robida, 2013)	Turkey (Bodur, 2010)	Palestine (Najjar et al., 2013)
Number of Observations	NA	487	974	974	974	974	974	974	974	974	974	974	974
Number of Variables	NA	28	42	27	39	39	42	35	38	42	39	42	38
Number of Factors	NA	8	12	9	10	10	10	8	11	8	9	10	11
Root Mean Square Error of Approximation (RMSEA)	<0.07	0,05	0,05	0,05	0,05	0,06	0,05	0,06	0,05	0,06	0,05	0,06	0,05
Standardized Root Mean Residual (SRMR)	<0.08	0,05	0,05	0,05	0,06	0,06	0,05	0,08	0,06	0,06	0,06	0,06	0,05
Root Mean Square Residual (RMSR / RMR)	NA	0,04	0,05	0,04	0,05	0,05	0,05	0,07	0,05	0,06	0,05	0,06	0,04
Goodness of Fit Index (GFI)	>0.90	0,91	0,88	0,92	0,88	0,87	0,86	0,86	0,89	0,83	0,87	0,84	0,90
Adjusted GFI (AGFI)	>0.90	0,90	0,86	0,90	0,86	0,85	0,84	0,84	0,86	0,81	0,85	0,81	0,87
Normed Fit Index (NFI)	>0.95	0,90	0,86	0,90	0,86	0,85	0,84	0,84	0,86	0,80	0,85	0,81	0,88
Comparative Fit Index (CFI)	≥0.90	0,95	0,90	0,93	0,89	0,88	0,88	0,87	0,90	0,84	0,89	0,85	0,91
Tucker-Lewis Index / Non-normed Fit Index (TLI / NNFI)	≥0.90	0,94	0,88	0,91	0,88	0,87	0,87	0,85	0,88	0,83	0,87	0,83	0,90

Note: Colored cells contain values that do not meet requirements.

From the official website of AHRQ [24] we retrieved the list of 23 articles reporting psychometric analyses on international level. From these articles we extracted 10 factor models that differed from the original North-American version. These factor models were from the following countries: England (UK) [15], Scotland (UK) [13], France [9], Switzerland (French [10] and German [11]), Netherland [14], Sweden [7], Slovenia [12], Turkey [6] and Palestine [8]. The eleventh factor model considered in the analysis was the original 12-factor model [4].

Subsequent series of CFA revealed satisfactory fit of the models from England (UK) [15] (RMSEA=0.05; SRMR=0.05; GFI=0.92; CFI=0.93; TLI/NNFT=0.91) and Palestine [8] (RMSEA=0.05; SRMR=0.05; GFI=0.90; CFI=0.91; TLI/NNFT=0.90) to our data. The original 12-factor model resulted in marginally satisfactory model fit (RMSEA=0.05; SRMR=0.05; GFI=0.88; CFI=0.90; TLI/NNFT=0.88). The models from Scotland (UK), France, Switzerland, Netherland and Slovenia, resulted in suboptimal values of CFA indices (table 4). Models from Sweden and Turkey demonstrated unsatisfactory model fit in CFA.

Internal consistency. Cronbach's alpha was acceptable (0.6-0.7) or good (>0.7) for all dimensions of most models proposed for the various language versions, with the exception of *Organizational learning* in the Slovenian (0.51) and Turkish (0.53) models, and *Staffing* in the Dutch model (0.53) (see table 5). Five dimensions, namely *Teamwork within units*, *Non-punitive response to error*, *Supervisor expectations & actions promoting patient safety*, *Frequency of events reported*, and *Feedback & communication about error*, were present in all models and demonstrated acceptable (>0.6) or good (>0.7) Cronbach's alphas.

Table 5: Internal consistency (Cronbach's alpha) of all 12 factor models analyzed.

Dimensions (from original 12-factor model)	**Germany (EFA)**	USA (Sorra & Nieva, 2004)	England (UK) (Waterson et al., 2010)	Scotland (UK) Sarac et al., 2011)	France (Occelli et al. 2013)	Switzerland (Perneger, 2013)	Switzerland (Pfeiffer and Manser, 2010)	Netherlands (Smits et al. 2008)	Sweden (Hedsköld et al., 2013)	Slovenia (Robida, 2013)	Turkey (Bodur, 2010)	Palestine (Najjar et al., 2013)
01. Teamwork Within Units	0,78	0,74	0,79	0,74	0,79	0,74	0,75	0,74	0,74	0,74	0,74	0,74
02. Organizational Learning - Continuous Improvement		0,68		0,68		0,68				0,51	0,53	0,68
03. Nonpunitive Response to Error	0,73	0,74	0,61	0,74	0,74	0,74	0,74	0,74	0,72	0,74	0,72	0,74
04. Staffing	0,79	0,70	0,80	0,80	0,73	0,70	0,80	0,53	0,80	0,82	0,65	0,73
05. Overall Perceptions of Patient Safety		0,77			0,77	0,77		0,79			0,77	0,71
06. Supervisor Expectations & Actions Promoting Patient Safety	0,75	0,75	0,72	0,75	0,75	0,75	0,74	0,75	0,75	0,75	0,75	0,75
07. Frequency of Events Reported	0,87	0,88	0,88	0,88	0,88	0,88	0,80	0,88	0,88	0,88	0,88	0,88
08. Feedback & Communication About Error	0,83	0,81	0,81	0,82	0,83	0,82	0,83	0,83	0,86	0,83	0,82	0,80
09. Communication Openness		0,64	0,64		0,64			0,64				
10. Management Support for Patient Safety	0,83	0,84		0,84	0,82	0,84	0,84	0,84	0,84	0,84	0,84	0,84
11. Teamwork Across Units	0,79	0,75	0,75	0,75	0,79	0,83	0,82	0,75		0,82		0,75
12. Handoffs & Transitions		0,75	0,75	0,68				0,66	0,76		0,76	0,75

<0,6 – Not satisfactory (cells colored in dark grey)
0,6-0,7 – Acceptable
>0,7 – Good
Empty cell (colored in light grey) – Dimension is not present in the model.

Evaluation of common dimensionality. We analyzed the appearance and role of each individual item and dimension from the original 12-factor model in factor model from EFA, and in 10 models reported by studies from different language versions. Table 3 presents 42 items of the original 12-factor model and their appearance in all 12 analyzed models. The uncolored cells represent no change, where the item retains its original role in the factor model. Changes are represented by colored boxes, which indicate elimination of the questionnaire item (N), or moving it to a different dimension (labeled from 1 to 12).

14 items were eliminated from analysis in EFA. 11 of these 14 items demonstrated significant inconsistency since in at least half of 10 analyzed factor models they were also eliminated, moved or merged with another dimension. All of the remaining 28 items of our 8-factor factor model demonstrated relative stability by retaining a similar role in at least 50% of the 10 analyzed factor models; 23 items maintained their role in 80% or more of the models.

Eight dimensions, including *Teamwork within units*, *Non-punitive response to error*, *Supervisor expectations & actions promoting patient safety*, *Frequency of events reported*, *Staffing*, *Feedback & communication about error*, *Management support for patient safety* and *Teamwork across hospital units* demonstrated relative stability over the different language models, appearing in 80% or more of the 10 analyzed models. The dimension *Communication openness* was merged with the dimension *Feedback and communication about error* in 7 models [6–8, 10–13]. Similarly, the dimension *Hospital handoffs and transitions* was merged with the dimension *Teamwork across hospital units* in 5 models [9–12, 14], and the dimension *Overall perceptions of safety* with the dimension *Staffing* in 5 models [7, 11–13, 15]. The items from the dimension *Organizational learning - continuous improvement* were shown to be highly inconsistent across various models. In five models, the items from this dimension were either removed from the model [15] or merged with other dimensions [7, 9, 11, 14] (e.g. with *Feedback and communication about error*).

Discussion

The aim of this study was to evaluate the psychometric properties of the HSPSC-D and compare its dimensionality with factor structures derived for different language versions of the HSPSC. Our split-half validation resulted in an alternative 8-factor model with good psychometric properties. Most parts of the instrument demonstrate relative stability over different language versions and appear suitable for cross-national studies. However, items of four safety culture dimensions require further improvement to support a common structure for comparison across language versions.

In our study HSPSC-D demonstrated marginally satisfactory psychometric properties, allowing for its use in German hospitals. HSPSC-D demonstrated a somewhat unsatisfactory model fit in CFA with the original 12-factor model. EFA resulted in an alternative 8-factor model, with good model fit. Nevertheless, the instrument demonstrated satisfactory to good internal consistency in both models. Studies with other language versions of the HSPSC have repeatedly reported similar results – good model fit of different factor structure and mostly good internal consistency [6, 7, 9–13, 15]. These findings indicate that the HSPSC is a useful instrument for measuring and comparing patient safety culture within a healthcare system for which the particular HSPSC version has previously been validated.

Our analysis of instrument dimensionality across language versions revealed that whilst some dimensions maintain relative stability of appearance and composition across language versions, others vary significantly. When analyzing 12 different factor models, including the original North American 12-factor model and the 8-factor model resulting from our EFA, we found that items from eight dimensions can be considered in international studies. The remaining four dimensions, namely *Organizational Learning—Continuous improvement Overall Perceptions of Safety*, *Communication Openness* and *Hospital Handoffs & Transitions*, appeared in only ≤50% of analyzed models since corresponding items were either removed, migrated to or merged with other dimensions. These dimensions require further refinement before use in cross-national studies.

Evaluation of psychometric properties of a translated version of the instrument is important, as results of validated instruments can be interpreted and used for comparison in local contexts. A number of studies reported that the original 12-factor model did not fit the data well, and alternative factor models were suggested. For example, Hedskoeld [7] revealed a 9-factor model but argues against removing items and dimensions from the instrument, stating that they can still be used to understand and improve patient safety.

Concerning the international use of the instrument, several articles highlight the importance of a common factor structure. For example, Occelli [9] underlines the need to adapt the tool to each country's environment while stating that "for international comparison purposes, a core set of dimensions consistently assessed as valid should be defined and measured in all countries." Perneger [10] further argues that local improvements to a translated version can be ineffective, due to several unresolved issues inherent in the instrument, such as limited internal consistency of some dimensions, different dimensionality found in various language versions and the lack of external validation of study results.

Limitations

The data analysis and results in the study were limited to two German university hospitals. Also, our findings shouldn't be generalized to all hospital employees, as the study sample mainly consists of nurses and physicians. However, our findings regarding psychometric properties of the instrument, as well as its dimensionality are in line with those of similar studies from other countries. While exploring the common dimensionality of various language versions our analysis was limited to research articles retrieved from the official web-page of AHRQ [24]. Taking into account more studies that report a different factor structure based on a systematic review could improve the analysis. Lastly, the diversity of study methodology and reporting of studies with different language versions of HSPSC, may be considered an additional obstacle for cross-national use of the instrument.

Conclusions

Overall the German version of the HSPSC demonstrated acceptable psychometric properties for surveying clinical personnel in German hospitals. We found that most safety culture dimensions were relatively stable across different language models. However, other dimensions demonstrate high variability and inconsistency. Such inconsistencies need to be refined in order to support a more uniform factor structure across language versions in order to facilitate the use of HSPSC at the cross-national level.

Contributors.

Data were collected by the WorkSafeMed Consortium. Data analysis was carried out by NG under the supervision of TM and AH. NG and AH wrote the manuscript that was then revised by MB, TM. The final version of the manuscript has been approved by all authors.

Competing interests: None.

Acknowledgments

We thank the members of the advisory board for their valuable advice at various stages of the project; especially for their critical assessment of instrument development: (The names were removed for the review purposes, and will be added before publishing).

Last but not least we would like to thank all study participants. We acknowledge the support of the hospital management, the efforts of study coordinators in participating departments and units to facilitate data collection and the respondents for their effort and time to fill in the surveys.

Funding

The „WorkSafeMed“ study was funded by the Federal Ministry of Education and Research (FKZ 01GY1325A and 01GY1325B). We acknowledge additional financial support by the German Research Foundation and the administrative support by the DLR Project Management Agency.

Data Sharing Statement

Because of data security aspects, data from the WorkSafeMed study will not be made available in the public domain. However, data will be used by students of both project

partners for their theses. Data will be stored in accordance with national and regional data security standards.

REFERENCES

1 Waterson P, ed. Patient safety culture: Theory, methods, and application. Farnham, Surrey, UK: Ashgate 2014.
2 Guldenmund FW. The nature of safety culture: a review of theory and research. Saf Sci 2000;34:215–57.
3 Sorra J, Dyer N. Multilevel psychometric properties of the AHRQ hospital survey on patient safety culture. BMC Health Serv Res 2010;10:199.
4 Sorra J, Nieva V. Hospital Survey on Patient Safety Culture. (Prepared by Westat, under Contract No. 290-96-0004). AHRQ Publication No. 04-0041. Rockville, MD 2004.
5 Elmontsri M, Almashrafi A, Banarsee R, et al. Status of patient safety culture in Arab countries: a systematic review. BMJ Open 2017;7(2):e013487.
6 Bodur S, Filiz E. Validity and reliability of Turkish version of “Hospital Survey on Patient Safety Culture” and perception of patient safety in public hospitals in Turkey. BMC Health Serv Res 2010;10:28.
7 Hedskold M, Pukk-Harenstam K, Berg E, et al. Psychometric properties of the Hospital Survey on Patient Safety Culture, HSOPSC, applied on a large Swedish health care sample. BMC Health Serv Res 2013;13:332.
8 Najjar S, Hamdan M, Baillien E, et al. The Arabic version of the hospital survey on patient safety culture: a psychometric evaluation in a Palestinian sample. BMC Health Serv Res 2013;13:193.
9 Occelli P, Quenon J-L, Kret M, et al. Validation of the French version of the Hospital Survey on Patient Safety Culture questionnaire. Int J Qual Health Care 2013;25:459–68.
10 Perneger TV, Staines A, Kundig F. Internal consistency, factor structure and construct validity of the French version of the Hospital Survey on Patient Safety Culture. BMJ Qual Saf 2014;23:389–97.
11 Pfeiffer Y, Manser T. Development of the German version of the Hospital Survey on Patient Safety Culture: Dimensionality and psychometric properties. Saf Sci 2010;48:1452–62.
12 Robida A. Hospital Survey on Patient Safety Culture in Slovenia: a psychometric evaluation. Int J Qual Health Care 2013;25:469–75.
13 Sarac C, Flin R, Mearns K, et al. Hospital survey on patient safety culture: psychometric analysis on a Scottish sample. BMJ Qual Saf 2011;20:842–48.
14 Smits M, Christiaans-Dingelhoff I, Wagner C, et al. The psychometric properties of the ‘Hospital Survey on Patient Safety Culture’ in Dutch hospitals. BMC Health Serv Res 2008;8:230.
15 Waterson P, Griffiths P, Stride C, et al. Psychometric properties of the Hospital Survey on Patient Safety Culture: findings from the UK. Qual Saf Health Care 2010;19(5):2e.
16 Hammer A, Manser T. The Use of the Hospital Survey on Patient Safety Culture in Europe. In: Waterson P, ed. Patient safety culture: Theory, methods, and application. Farnham, Surrey, UK: Ashgate 2014:229–62.
17 Wagner C, Smits M, Sorra J, et al. Assessing patient safety culture in hospitals across countries. Int J Qual Health Care 2013;25(3):213–21.
18 Haugen AS, Softeland E, Eide GE, et al. Patient safety in surgical environments: cross-countries comparison of psychometric properties and results of the Norwegian version of the Hospital Survey on Patient Safety. BMC Health Serv Res 2010;10:279.
19 Hammer A, Ernstmann N, Ommen O, et al. Psychometric properties of the Hospital Survey on Patient Safety Culture for hospital management (HSOPS_M). BMC Health Serv Res 2011;11:165.
20 Schafer JL, Graham JW. Missing data: Our view of the state of the art. Psychol Methods 2002;7:147–77.

21 Wirtz M. On the problem of missing data: How to identify and reduce the impact of missing data on findings of data analysis. Rehabilitation (Stuttg) 2004;43(2):109–15.
22 Field A. Discovering statistics using SPSS: (and sex and drugs and rock’n’roll), 3rd edn. Los Angeles: SAGE 2012.
23 Hair JF, Black WC, Babin BJ, et al. Multivariate data analysis. Upper Saddle River, N.J.: Pearson Education 2014.
24 the Agency for Healthcare Research and Quality (AHRQ). Surveys on Patient Safety Culture: Psychometric Analyses—International 2017. Available at: www.ahrq.gov/professionals/quality-patient-safety/patientsafetyculture/resources Accessed May 02, 2017.

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1,2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			**5-6**
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			**6-9**
Study design	4	Present key elements of study design early in the paper	6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6, 7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6,7
Bias	9	Describe any efforts to address potential sources of bias	*NA*
Study size	10	Explain how the study size was arrived at	NA
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	8,9
		(*b*) Describe any methods used to examine subgroups and interactions	NA
		(*c*) Explain how missing data were addressed	7
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	NA
		(*e*) Describe any sensitivity analyses	NA
Results			9-19

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	9
		(b) Give reasons for non-participation at each stage	-
		(c) Consider use of a flow diagram	-
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	9-10
		(b) Indicate number of participants with missing data for each variable of interest	9-10
Outcome data	15*	Report numbers of outcome events or summary measures	9-12
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	NA
		(*b*) Report category boundaries when continuous variables were categorized	10
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	13-19
Discussion			20-21
Key results	18	Summarise key results with reference to study objectives	20
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	21
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	20
Generalisability	21	Discuss the generalisability (external validity) of the study results	21-22
Other information			23
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	23

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Evaluation of psychometric properties of the German Hospital Survey on Patient Safety Culture and its potential for cross-cultural comparisons: A cross-sectional study.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-018366.R1
Article Type:	Research
Date Submitted by the Author:	12-Aug-2017
Complete List of Authors:	Gambashidze, Nikoloz; University hospital Bonn, Institute for patient safety Hammer, Antje; University hospital Bonn, Institute for patient safety Broesterhaus, Mareen; University hospital Bonn, Institute for patient safety Manser, Tanja; University hospital Bonn, Institute for patient safety
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Research methods
Keywords:	Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Health & safety < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, International health services < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, STATISTICS & RESEARCH METHODS

Title: Evaluation of psychometric properties of the German Hospital Survey on Patient Safety Culture and its potential for cross-cultural comparisons: A cross-sectional study.

Authors: Nikoloz Gambashidze*[1], Antje Hammer*[1], Mareen Brösterhaus[1], Tanja Manser[1], On behalf of the WorkSafeMed Consortium**

[1] University Hospital Bonn, Institute for Patient Safety, Bonn, Germany

* Co-first author.

** The WorkSafeMed Project Consortium comprises: Luntz E, Rieger MA (project lead), Sturm H, Wagner A(Institute of Occupational and Social Medicine and Health Services Research, University Hospital of Tuebingen), Hammer A, Manser T (Institute for Patient Safety, University Hospital Bonn), Martus P (Institute for Clinical Epidemiology and Applied Biometry, University Hospital of Tuebingen), Holderied M (University Hospital Tuebingen).

Corresponding author:

Nikoloz Gambashidze,

Institute for Patient Safety,

University Hospital Bonn,

Sigmund-Freud-Str. 25,

53127 Bonn, Germany

Phone: +49 228 287 13948

E-Mail: Nikoloz.Gambashidze@ukbonn.de

Running title: Psychometrics of German HSPSC

Word count, abstract: 255.

Word count, manuscript: 3266.

Abstract

Objective: To study the psychometric characteristics of German version of the Hospital Survey on Patient Safety Culture, and to compare its dimensionality to other language versions in order to understand the instrument's potential for cross-national studies.

Design: Cross-sectional multicenter study to establish psychometric properties of German version of the survey instrument.

Setting: 73 units from 37 departments of two German university hospitals.

Participants: Clinical personnel (n=995 responses, response rate 39.6%).

Primary and secondary outcome measures: Psychometric properties (e.g. Model fit, internal consistency, construct validity) of the instrument, and comparison of dimensionality across different language translations.

Results: The instrument demonstrated acceptable to good internal consistency (Cronbach's alpha 0.64-0.88). Confirmatory factor analysis of the original 12-factor model resulted in marginally satisfactory model fit (RMSEA= 0.05; SRMR; CFI=0.90; GFI=0.88; TLI=0.88). Exploratory factor analysis resulted in an alternative 8-factor model with good model fit (RMSEA= 0.05; SRMR; CFI=0.95; GFI=0.91; TLI=0.94) and good internal consistency (Cronbach's alpha 0.73-0.87) and construct validity. Analysis of the dimensionality compared to models from 10 other language versions revealed eight dimensions with relatively stable composition and appearance across different versions and four dimensions requiring further improvement.

Conclusions: The German version of Hospital survey on Patient Safety Culture demonstrated satisfactory psychometric properties for use in German hospitals. However, our comparison of instrument dimensionality across different language versions indicates limitations concerning cross-national studies. Results of this study can be considered in interpreting

findings across national contexts, in further refinement of the instrument for cross-national studies, and to better understand various facets and dimensions of patient safety culture.

Keywords: Quality in health care; Health & safety; International health services; statistics & research methods.

Article Summary

Strengths and limitations of this study

- Our study supports the development of a more uniform factor structure for the Hospital Survey on Patient Safety Culture across language versions in order to facilitate its use in cross-national research.
- By evaluating commonalities and variations in different language versions of the Hospital Survey on Patient Safety Culture we identify relatively stable factors, as well as those in need for improvement.
- First study to validate the German version of the Hospital Survey on Patient Safety Culture for clinical personnel.
- The considerable diversity in study methodology and reporting of studies with different language versions of the Hospital Survey on Patient Safety Culture presents an obstacle for cross-national use of the instrument that has yet to be overcome.

Introduction

All healthcare organizations face specific sets of risks and challenges regarding patient safety. These challenges change dynamically over time, reflecting developments within the organization as well as in its operating environment such as changes in demographics and epidemiology or in patient behavior. To effectively manage these challenges, it is recommended for healthcare organizations to develop a culture of safety that prioritizes safety and organizational learning among other organizational goals [1]. Safety culture is generally considered to be a relatively stable construct, rooted in organizational culture [2].

A number of instruments for measuring safety culture in healthcare organizations have been developed. These instruments enable researchers and decision-makers to evaluate and compare results on different levels of the healthcare system [3]. Comparing results across units and hospitals and establishing benchmarks can drive continuous patient safety improvement. One of the most widely used instruments for evaluating healthcare providers' perception of safety culture in hospital setting is the Hospital Survey on Patient Safety Culture (HSPSC) [4]. The instrument has been translated into many languages and used in different countries around the world [5–16].

There are two gaps that this study aims to address. Firstly, so far no German version of HSPSC has been validated for healthcare personnel in Germany. Secondly, despite some attempts at comparing safety culture at the international level [17, 18], the comparability of the different language versions of the instrument has not been studied systematically. While satisfactory psychometric properties were reported for the original North-American version [4] with 12 dimensions of patient safety culture, alternative factor structures have been reported for other language versions, with the number of dimensions ranging from 8 to 12 [5–7, 9–12, 14–16]. Because an instrument's dimensionality determines the interpretation of results, similarities and differences in dimensionality across different language versions should be considered for cross-national studies of patient safety culture.

Therefore, the aim of this study is two-fold: (1) validation of German version of HSPSC (HSPSC-D) by evaluation of its psychometric properties and (2) evaluation of the instrument's potential for cross-national studies, by comparative analysis of instrument's dimensionality as reported for different language versions.

Methods

Setting

This study was based on data from the cross-sectional, multicenter, mixed methods study *Working conditions, safety culture and patient safety in hospitals – what predicts the safety of the medication process* (WorkSafeMed), conducted between 2014 and 2017. In this article, we focus on HSPSC-D data to evaluate its psychometric properties. The WorkSafeMed study with all its components has been approved by the responsible ethics committees of the medical faculties of the project partners in Bonn (#350/14) and Tubingen (#547/2014BO1). Each partner complied with confidentiality requirements according to German law.

Sample

Safety culture data were collected in two German university hospitals April to July 2015. We included staff from inpatient units with ≥500 patients a year. Intensive care and psychiatric units were excluded. Across the two hospitals a total of 73 units from 37 departments participated in the study. The HSPSC-D questionnaire was distributed to 2512 healthcare professionals. All participants received an initial invitation to participate in the study, followed by two reminders. Study material included all required information regarding the study and data handling. Participation in the study was anonymous and participants' consent was implied by returning completed questionnaires. Non-responder analysis was not performed.

Measure

In order to develop a version of the HSPSC for German healthcare professionals (HSPSC-D), we used two previous German language versions as a starting point. A first translation of the HSPSC for hospital staff in the German speaking part of Switzerland [7] had been culturally and linguistically adapted for use in Swiss hospitals. Hammer et al. [19] used the Swiss version as a starting point for developing a management version of HSPSC to study perceptions of safety culture among medical directors in German hospitals. In our study, the instrument was adapted to be used with healthcare personnel in German hospitals.

The resulting HSPSC-D questionnaire follows the structure of the original North-American version [4] and includes 44 items, 42 of which compose 12 dimensions (10 safety culture dimensions and 2 outcome dimensions). These 42 items use a five-point Likert scale to measure agreement ranging from “strongly disagree” (1) to “strongly agree” (5), or frequency ranging from “never” (1) to “always” (5). The remaining two single item measures are *Number of events reported* (measured on six frequency groups from “No event reports” to “21 event reports or more”) and *Patient safety grade* (measured on five-point scale from “Failing” to “Excellent”).

<u>Analysis</u>

Data processing and preliminary analysis. After excluding responses with more than 30% missing values in HSPSC-D items, we conducted multiple imputations based on the expectation maximization (EM) algorithm using the statistical software NORM 2.03 [20, 21] to replace remaining missing values. Negatively worded items were reverse coded before further analysis.

Several indices were taken into account to ensure that our study sample, as well as every subset used in further analysis was appropriate for factor analysis. Kaiser-Meyer-Olkin (KMO) indicates if the sample of items is adequate for factor analysis, while Measure of Sampling Adequacy (MSA) indicates if an individual item is adequate for factor analysis. For both indices the value >0.7 is desired, and the value of >0.9 is considered perfect [22]. A significant p-value (<0.05) of Bartlett's test of sampling adequacy indicates that it is possible to extract more than one factor [22]. The analyses were performed using SAS 9.4.

Descriptive statistics. We calculated composite scores for each dimension suggested by Sorra and Nieva [4] by calculating the average of corresponding items. We also calculated percentages of positive responses for each dimension by dividing the number of positive responses on corresponding items by the number of non-missing answers in the dimension. Descriptive statistics for each item and dimension were evaluated, including range, mean, and standard deviation.

Exploratory Factor Analysis (EFA). We used EFA to evaluate the factor structure emerging from the study data. In general, EFA and Confirmatory Factor Analysis (CFA) should be performed using different subsets [23]. Thus, we performed the split-half cross validation, by randomly splitting our sample in two: "Exploring" (for EFA) and "Testing" subsets (for subsequent CFA). EFA using maximum likelihood was conducted using the "Exploring" subset. We used Varimax orthogonal pre-rotation, and Promax oblique rotation to aid with interpretation of factor model [23]. We used scree plot and Kaiser Criterion (Eigenvalues >1) for factor extraction. Factor loadings ≥0.4 were considered significant, and factor cross loading <0.4 was considered acceptable [22, 23]. Applying these criteria, we gradually eliminated problematic items until EFA resulted in a satisfactory factor structure.

Confirmatory Factor Analysis (CFA). We evaluated the model fit of the factor structure resulting from the EFA by conducting CFA using the "Testing" subset. By conducting a series of CFA using the complete dataset, we evaluated model fit of original 12-factor model [4], as well as other factor models reported by studies of different language versions of HSPSC. From the official website of the Agency for Healthcare Research and Quality (AHRQ) [24] we retrieved a list of studies including psychometric evaluation of the instrument and identified those reporting a different factor structure.

Internal consistency. Internal consistency was evaluated by calculating Cronbach's alpha as an indicator of correlation between each item and the factor. In their exploratory study Sorra and Nieva [4] considered Cronbach's alpha ≥0.6 as acceptable. We used Cronbach's alpha ≥0.7, as it is typically used in later studies using the HSPSC [5, 6, 9, 11, 14, 15, 17, 19], and is well supported by the literature [22, 23]. Cronbach's alphas were calculated for all factor models considered in the CFA, including the factor model that emerged from EFA.

Construct validity. By calculating average of corresponding non-missing items we calculated mean values for each dimension for the original 12-factor model and for the new model that emerged from EFA. Pearson's correlations were evaluated between dimensions in each model. We expected low to moderate correlations between dimensions. However, correlations >0.85 would indicate possible multicollinearity [4, 22]. We also evaluated the correlations between dimensions of both models with two single item outcome variables – *Patient safety grade* and *Number of incidents reported*.

Evaluation of common dimensionality. In order to evaluate the potential of the instrument for cross-national studies, we evaluated its dimensionality as reported for different language versions. We evaluated appearance and composition of each of the 12 dimensions proposed by Sorra and Nieva [4] and of the 42 corresponding items in all factor models identified from AHRQ web-page [24].

Results

Study sample and descriptive statistics.

Out of 2512 distributed questionnaires 995 were completed, resulting in a response rate of 39.6%. Sample characteristics are presented in table 1.

Table 1: Characteristics of study sample.

Variables	N	%
Study site	**995**	**100.0%**
Hospital A	575	57.8%
Hospital B	420	42.2%
Gender	**995**	**100.0%**
Female	656	65.9%
Male	291	29.2%
missing	48	4.8%
Professional Groups	**995**	**100.0%**
Physician	183	18.4%
Physicians' assistant	198	19.9%
Nurse	552	55.5%
other	34	3.4%
missing	28	2.8%
Managerial functions	**995**	**100.0%**
Yes	195	19.6%
No	759	76.3%
missing	41	4.1%
Contact with patients	**995**	**100.0%**
Yes	965	97.0%
No	7	0.7%
missing	23	2.3%
Age (years)	**995**	**100.0%**
<25	61	6.1%
25-34	360	36.2%
35-44	230	23.1%
45-54	170	17.1%
>54	84	8.4%
missing	90	9.0%

21 responses contained more than 30% missing values and were not included in the analysis. Descriptive statistics of HSPSC-D items and dimensions after imputing remaining missing answers and reverse coding of the negatively worded items are presented in the Table 2.

Table 2: Descriptive statistics of HSPSC-D items and dimensions.

Dimension / Item [1,2,3]	Percent of positive responses [4]	Mean	SD[5]
01. Teamwork Within Hospital Units	**42.3%**	**3.32**	**0.61**
A1. People support one another in this unit.	58.3%	3.65	0.78
A3. When a lot of work needs to be done quickly, we work together as a team to get the work done.	51.2%	3.50	0.84
A4. In this unit, people treat each other with respect.	40.9%	3.36	0.78
A11. When one area in this unit gets really busy, others help out.	18.8%	2.79	0.91
02. Organizational Learning—Continuous improvement	**32.7%**	**3.06**	**0.70**
A6. We are actively doing things to improve patient safety.	50.1%	3.40	0.91
A9. Mistakes have led to positive changes here.	23.5%	2.88	0.89
A13. After we make changes to improve patient safety, we evaluate their effectiveness.	24.4%	2.90	0.89
03. Non-punitive Response To Error	**50.2%**	**3.38**	**0.80**
A8. (R) Staff feel like their mistakes are held against them.	40.1%	3.19	0.96
A12. (R) When an event is reported, it feels like the person is being written up, not the problem.	48.3%	3.33	0.99
A16. (R) Staff worry that mistakes they make are kept in their personnel file.	62.1%	3.62	0.99
04. Staffing	**24.9%**	**2.57**	**0.79**
A2. We have enough staff to handle the workload.	7.5%	2.01	0.97
A5. (R) Staff in this unit work longer hours than is best for patient care	23.1%	2.57	1.18
A7. (R) We use more agency/temporary staff than is best for patient care.	58.2%	3.57	1.20
A14. (R) We work in "crisis mode," trying to do too much, too quickly.	10.9%	2.13	1.02
05. Overall Perceptions of Safety	**34.4%**	**3.03**	**0.79**
A10. (R) It is just by chance that more serious mistakes don't happen around here.	41.1%	3.08	1.20
A15. Patient safety is never sacrificed to get more work done.	25.4%	2.75	1.04
A17. (R) We have patient safety problems in this unit.	43.9%	3.29	0.97
A18. Our procedures and systems are good at preventing errors from happening.	27.2%	3.00	0.89
06. Supervisor/manager expectations & actions promoting safety	**48.5%**	**3.34**	**0.71**
B1. My supervisor/manager says a good word when he/she sees a job done according to established patient safety procedures.	33.7%	3.03	1.02
B2. My supervisor/manager seriously considers staff suggestions for improving patient safety.	55.9%	3.51	0.87
B3. (R) Whenever pressure builds up, my supervisor/manager wants us to work faster, even if it means taking shortcuts.	42.8%	3.19	0.98
B4. (R) My supervisor/manager overlooks patient safety problems that happen over and over.	61.7%	3.61	0.89
07. Frequency of Event Reporting	**38.0%**	**3.00**	**1.03**
D1. When a mistake is made, but is caught and corrected before affecting the patient, how often is this reported?	39.0%	3.03	1.17
D2. When a mistake is made, but has no potential to harm the patient, how often is this reported?	30.1%	2.77	1.14
D3. When a mistake is made that could harm the patient, but does not, how often is this reported?	45.0%	3.19	1.13
08. Feedback and Communication About Error	**48.0%**	**3.36**	**0.85**
C1. We are given feedback about changes put into place based on event reports.	40.0%	3.18	1.04

C3. We are informed about errors that happen in this unit.	50.1%	3.41	0.99
C5. In this unit, we discuss ways to prevent errors from happening again.	53.9%	3.50	0.95
09. Communication Openness	**58.6%**	**3.60**	**0.68**
C2. Staff will freely speak up if they see something that may negatively affect patient care.	66.2%	3.74	0.87
C4. Staff feel free to question the decisions or actions of those with more authority.	45.4%	3.35	0.89
C6. (R) Staff are afraid to ask questions, when something does not seem right.	64.1%	3.71	0.91
10. Hospital Management Support for Patient Safety	**23.4%**	**2.79**	**0.86**
F1. Hospital management provides a work climate that promotes patient safety.	22.4%	2.83	0.94
F8. The actions of hospital management show that patient safety is a top priority.	21.1%	2.74	0.97
F9. (R) Hospital management seems interested in patient safety only after an adverse event happens.	26.8%	2.79	1.04
11. Teamwork Across Hospital Units	**29.0%**	**3.03**	**0.61**
F2. (R) Hospital units do not coordinate well with each other.	14.7%	2.57	0.91
F4. There is good cooperation among hospital units that need to work together.	22.6%	3.03	0.73
F6. (R) It is often unpleasant to work with staff from other hospital units.	49.1%	3.39	0.82
F10. Hospital units work well together to provide the best care for patients.	29.7%	3.14	0.77
12. Hospital Handoffs & Transitions	**35.3%**	**3.07**	**0.64**
F3. (R) Things "fall between the cracks" when transferring patients from one unit to another.	13.2%	2.50	0.88
F5. (R) Important patient care information is often lost during shift changes.	37.1%	3.16	0.89
F7. (R) Problems often occur in the exchange of information across hospital units.	29.3%	3.04	0.81
F11. (R) Shift changes are problematic for patients in this hospital.	61.5%	3.59	0.82
E1. Please give your work area/unit in this hospital an overall grade on patient safety.	**32.9%**	**3.22**	**0.76**

Note: Answers 4 and 5 ("A*gree*" and "S*trongly agree*" or "*Most of the time*" and "A*lways*") were considered as positive. Prior to analysis, negatively worded items were reverse coded.
[1]01-12 – corresponding dimension according to original North-American 12-factor model.
[2]A1-A18; B1-B4; C1-C6; D1-D3; E1; F1-F11: Codes of questionnaire items.
[3](R) – negatively worded items, which were reverse coded prior to the analysis.
[4]n=974.
[5] SD – standard deviation.

KMO for the complete sample was 0.93, and MSA for individual items ranged between 0.87 and 0.96. For “Exploring” and “Testing” subsets, KMO was 0.91 and 0.92 respectively, and MSA of individual items in both subsets ranged between 0.84 and 0.96. Bartlett’s test was highly significant ($p<0.001$) for the dataset, as well as for both subsets. Preliminary analyses indicated that our sample and the subsets were adequate for factor analysis.

Exploratory Factor Analysis (EFA). We conducted EFA using the “Exploring” subset. We considered factor loadings ≥0.4 as significant, as this cut-off value was typically used in similar studies [4–6, 10–12, 14–16] and was supported by the literature [22, 23]. 14 items not meeting the criteria (factor loading ≥0.4, cross loading <0.4) were excluded from the model, resulting in an 8-factor model with 28 items. The dimension *Organizational learning – continuous improvement* was completely removed. The dimensions *Staffing* and *Overall perceptions of safety* were merged together, as were the dimensions *Feedback and communication about error* with *Communication openness*, and *Teamwork across hospital units* with *Handoffs and transitions*. The resulting 8-factor model is presented in table 3.

Table 3: Appearance of HSPSC items in 12 analyzed factor models (8-factor EFA model, original 12-factor model, and 10 different versions).

HSPSC items[1,2]	Germany (Exploratory Factor Analysis)	**USA (Sorra & Nieva, 2004)**	England (UK) (Waterson et al., 2010)	Scotland (UK) (Sarac et al., 2011)	France (Occelli et al. 2013)	Switzerland (Perneger, 2013)	Switzerland (Pfeiffer and Manser, 2010)	Netherlands (Smits et al. 2008)	Sweden (Hedsköld et al., 2013)	Slovenia (Robida, 2013)	Turkey (Bodur, 2010)	Palestine (Najjar et al., 2013)
01. Teamwork within units												
A1	1	**1**	1	1	1	1	1	1	1	1	1	1
A3	1	**1**	1	1	1	1	1	1	1	1	1	1
A4	1	**1**	1	1	1	1	1	1	1	1	1	1
A11	(N)	**1**	(N)	1	(N)	1	(N)	1	1	1	1	1
02. Organizational learning												
A6	(N)	**2**	(N)	2	8	2	1	8	8	2	2	2
A9	(N)	**2**	(N)	2	8	2	(N)	8	8	2	3	2
A13	(N)	**2**	(N)	2	8	2	6	8	8	(N)	2	2
03. Nonpunitive response to error												
A8	3	**3**	3	3	3	3	3	3	3	3	3	3
A12	3	**3**	(N)	3	3	3	3	3	3	3	3	3
A16	3	**3**	3	3	3	3	3	3	3	3	3	3
04. Staffing												
A2	4	**4**	4	4	4	4	4	4	4	4	4	4
A5	4	**4**	(N)	4	4	4	4	4	4	4	2	4
A7	(N)	**4**	(N)	4	(N)	4	(N)	4	4	4	4	(N)
A14	4	**4**	4	4	4	4	4	5	4	4	4	4
05. Overall perceptions of safety												
A10	4	**5**	4	4	5	5	4	5	4	4	5	(N)
A15	(N)	**5**	(N)	4	5	5	(N)	(N)	4	4	5	5
A17	(N)	**5**	4	4	5	5	4	5	4	4	5	5
A18	(N)	**5**	(N)	(N)	5	5	7	5	8	4	5	5
06. Supervisor/manager expectations/actions												
B1	6	**6**	6	6	6	6	6	6	6	6	6	6
B2	6	**6**	6	6	6	6	6	6	6	6	6	6
B3	(N)	**6**	(N)	6	6	6	6	6	6	6	6	6
B4	6	**6**	(N)	6	6	6	6	6	6	6	6	6
07. Frequency of event reporting												
D1	7	**7**	7	7	7	7	7	7	7	7	7	7
D2	7	**7**	7	7	7	7	7	7	7	7	7	7
D3	7	**7**	7	7	7	7	7	7	7	7	7	7
08. Feedback and communication about error												
C1	8	**8**	8	8	8	8	8	8	8	8	8	(N)
C3	8	**8**	8	8	8	8	8	8	8	8	8	8
C5	8	**8**	8	8	8	8	8	8	8	8	8	8
09. Communication openness												
C2	8	**9**	9	8	9	8	8	9	8	8	8	8
C4	(N)	**9**	9	8	9	8	8	9	8	8	8	8
C6	(N)	**9**	9	8	9	8	(N)	9	3	(N)	8	(N)
10. Hospital management support for patient safety												
F1	10	**10**	(N)	10	10	10	10	10	10	10	10	10
F8	10	**10**	(N)	10	10	10	10	10	10	10	10	10
F9	10	**10**	(N)	10	10	10	10	10	10	10	10	10

11. Teamwork across hospital units												
F2	11	**11**	11	11	11	11	11	11	10	11	10	11
F4	11	**11**	11	11	11	11	11	11	10	11	10	11
F6	11	**11**	(N)	(N)	11	11	11	(N)	12	11	12	11
F10	11	**11**	11	11	10	11	11	11	10	11	10	11
12. Hospital handoffs and transitions												
F3	(N)	**12**	12	(N)	11	11	11	(N)	12	11	12	12
F5	(N)	**12**	12	12	11	11	(N)	12	12	11	12	12
F7	11	**12**	12	12	11	11	11	(N)	12	11	12	12
F11	(N)	**12**	12	12	(N)	11	(N)	12	12	(N)	12	12

Note:
The uncolored cells represent "No change" compared to original 12-factor model.
Colored boxes indicate items that were deleted (N) or moved to different dimension (Dimension number 1-12);
(N): items removed form factor model.
[1]01-12 – corresponding dimension according to original North-American 12-factor model.
[2]A1-A18; B1-B4; C1-C6; D1-D3; F1-F11: Codes of the questionnaire items.

Confirmatory factor analysis (CFA). CFA using the „Testing" subset demonstrated a satisfactory model fit of the factor structure that emerged from EFA (see table 4). The model satisfied desired thresholds of most analyzed indices (RMSEA=0.05; SRMR=0.05; GFI=0.90; CFI=0.93; TLI/NNFT=0.91).

Table 4. Results of Confirmatory Factor Analysis (CFA) of all 12 factor models analyzed.

Variables / Indices analyzed in CFA	Criteria [8, 23]	Germany (EFA)	USA (Sorra & Nieva, 2004)	England (UK) (Waterson et al., 2010)	Scotland (UK) Sarac et al., 2011)	France (Occelli et al. 2013)	Switzerland (Perneger, 2013)	Switzerland (Pfeiffer and Manser, 2010)	Netherlands (Smits et al. 2008)	Sweden (Hedsköld et al., 2013)	Slovenia (Robida, 2013)	Turkey (Bodur, 2010)	Palestine (Najjar et al., 2013)
Number of Observations	NA	487	974	974	974	974	974	974	974	974	974	974	974
Number of Variables	NA	28	42	27	39	39	42	35	38	42	39	42	38
Number of Factors	NA	8	12	9	10	10	10	8	11	8	9	10	11
Root Mean Square Error of Approximation (RMSEA)	<0.07	0.05	0.05	0.05	0.05	0.06	0.05	0.06	0.05	0.06	0.05	0.06	0.05
Standardized Root Mean Residual (SRMR)	<0.08	0.05	0.05	0.05	0.06	0.06	0.05	0.08	0.06	0.06	0.06	0.06	0.05
Root Mean Square Residual (RMSR / RMR)	NA	0.04	0.05	0.04	0.05	0.05	0.05	0.07	0.05	0.06	0.05	0.06	0.04
Goodness of Fit Index (GFI)	>0.90	0.91	0.88	0.92	0.88	0.87	0.86	0.86	0.89	0.83	0.87	0.84	0.90
Adjusted GFI (AGFI)	>0.90	0.90	0.86	0.90	0.86	0.85	0.84	0.84	0.86	0.81	0.85	0.81	0.87
Normed Fit Index (NFI)	>0.95	0.90	0.86	0.90	0.86	0.85	0.84	0.84	0.86	0.80	0.85	0.81	0.88
Comparative Fit Index (CFI)	≥0.90	0.95	0.90	0.93	0.89	0.88	0.88	0.87	0.90	0.84	0.89	0.85	0.91
Tucker-Lewis Index / Non-normed Fit Index (TLI / NNFI)	≥0.90	0.94	0.88	0.91	0.88	0.87	0.87	0.85	0.88	0.83	0.87	0.83	0.90

Note: Colored cells contain values that do not meet requirements.

From the official website of AHRQ [24] we retrieved the list of 23 articles reporting psychometric analyses on international level. From these articles we extracted 10 factor models that differed from the original North-American version. These factor models were from the following countries: England (UK) [9], Scotland (UK) [5], France [15], Switzerland (French [14] and German [7]), Netherland [10], Sweden [11], Slovenia [6], Turkey [12] and Palestine [16]. The eleventh factor model considered in the analysis was the original 12-factor model [4].

Subsequent series of CFA revealed satisfactory fit of the models from England (UK) [9] (RMSEA=0.05; SRMR=0.05; GFI=0.92; CFI=0.93; TLI/NNFT=0.91) and Palestine [16] (RMSEA=0.05; SRMR=0.05; GFI=0.90; CFI=0.91; TLI/NNFT=0.90) to our data. The original 12-factor model resulted in marginally satisfactory model fit (RMSEA=0.05; SRMR=0.05; GFI=0.88; CFI=0.90; TLI/NNFT=0.88). The models from Scotland (UK), France, Switzerland, Netherland and Slovenia, resulted in suboptimal values of CFA indices (table 4). Models from Sweden and Turkey demonstrated unsatisfactory model fit in CFA.

Internal consistency. The original 12-factor model demonstrated good Cronbach's alpha for all dimensions except *Organizational learning – Continuous improvement* (0.68) and *Communication openness* (0.64). Cronbach's alpha for dimensions of 8-factor model were between 0.73 and 0.87. Two dimensions, *Teamwork within units* and *Communication openness*, demonstrated consistently low alphas in other factor models analyzed. Three dimensions, *Nonpunitive response to error*, *Staffing* and *Handoffs & transitions*, had lower than 0.7 values only in one or two of analyzed models. Cronbach's alpha for the remaining seven dimensions in all analyzed models was ≥0.7, if present in the model (table 5).

Table 5: Internal consistency (Cronbach's alpha) of all 12 factor models analyzed.

Dimensions (from original 12-factor model)	**Germany (EFA)**	USA (Sorra & Nieva, 2004)	England (UK) (Waterson et al., 2010)	Scotland (UK) (Sarac et al., 2011)	France (Occelli et al. 2013)	Switzerland (Perneger, 2013)	Switzerland (Pfeiffer and Manser, 2010)	Netherlands (Smits et al. 2008)	Sweden (Hedsköld et al., 2013)	Slovenia (Robida, 2013)	Turkey (Bodur, 2010)	Palestine (Najjar et al., 2013)
01. Teamwork Within Units	0.78	0.74	0.79	0.74	0.79	0.74	0.75	0.74	0.74	0.74	0.74	0.74
02. Organizational Learning - Continuous Improvement		0.68		0.68		0.68				0.51	0.53	0.68
03. Nonpunitive Response to Error	0.73	0.74	0.61	0.74	0.74	0.74	0.74	0.74	0.72	0.74	0.72	0.74
04. Staffing	0.79	0.70	0.80	0.80	0.73	0.70	0.80	0.53	0.80	0.82	0.65	0.73
05. Overall Perceptions of Patient Safety		0.77			0.77	0.77		0.79			0.77	0.71
06. Supervisor Expectations & Actions Promoting Patient Safety	0.75	0.75	0.72	0.75	0.75	0.75	0.74	0.75	0.75	0.75	0.75	0.75
07. Frequency of Events Reported	0.87	0.88	0.88	0.88	0.88	0.88	0.80	0.88	0.88	0.88	0.88	0.88
08. Feedback & Communication About Error	0.83	0.81	0.81	0.82	0.83	0.82	0.83	0.83	0.86	0.83	0.82	0.80
09. Communication Openness		0.64	0.64		0.64			0.64				
10. Management Support for Patient Safety	0.83	0.84		0.84	0.82	0.84	0.84	0.84	0.84	0.84	0.84	0.84
11. Teamwork Across Units	0.79	0.75	0.75	0.75	0.79	0.83	0.82	0.75		0.82		0.75
12. Handoffs & Transitions		0.75	0.75	0.68				0.66	0.76		0.76	0.75

<0.7 – Not satisfactory (cells colored in dark grey)
≥0.7 – Good [23]
Empty cell (colored in light grey) – Dimension is not present in the model.

Construct validity. Correlation between dimensions of original 12-factor model were between 0.10 and 0.61 ($p<0.01$). All 12 dimensions were positively correlated with the outcome variable *Patient safety grade* (correlations between 0.26 and 0.70, $p<0.01$). Dimensions of 8-factor model from EFA were also positively inter-correlated (0.18-0.54, $p<0.01$) and positively correlated with the outcome variable *Patient safety grade* (0.29-0.58, $p<0.01$). All dimensions in both factor models resulted in no or week correlation (<0.2) with outcome variable *Number of events reported*. All correlations are presented in the online Appendix 1."

Evaluation of common dimensionality. We analyzed the appearance and role of each individual item and dimension from the original 12-factor model in factor model from EFA, and in 10 models reported by studies from different language versions. Table 3 presents 42 items of the original 12-factor model and their appearance in all 12 analyzed models. The uncolored cells represent no change, where the item retains its original role in the factor model. Changes are represented by colored boxes, which indicate elimination of the questionnaire item (N), or moving it to a different dimension (labeled from 1 to 12).

14 items were eliminated from analysis in EFA. 11 of these 14 items demonstrated significant inconsistency since in at least half of 10 analyzed factor models they were also eliminated, moved or merged with another dimension. All of the remaining 28 items of our 8-factor factor model demonstrated relative stability by retaining a similar role in at least 50% of the 10 analyzed factor models; 23 items maintained their role in 80% or more of the models.

Eight dimensions, including *Teamwork within units*, *Non-punitive response to error*, *Supervisor expectations & actions promoting patient safety*, *Frequency of events reported*, *Staffing*, *Feedback & communication about error*, *Management support for patient safety* and *Teamwork across hospital units* demonstrated relative stability over the different language models, appearing in 80% or more of the 10 analyzed models. The dimension *Communication openness* was merged with the dimension *Feedback and communication about error* in 7 models [5–7, 11, 12, 14, 16]. Similarly, the dimension *Hospital handoffs and transitions* was merged with the dimension *Teamwork across hospital units* in 4 models [6, 7, 14, 15], and the dimension *Overall perceptions of safety* with the dimension *Staffing* in 5 models [5–7, 9, 11]. The items from the dimension *Organizational learning - continuous improvement* were shown to be highly inconsistent across various models. In five models, the items from this dimension were either removed from the model [9] or merged with other dimensions [7, 10, 11, 15] (e.g. with *Feedback and communication about error*).

Discussion

The aim of this study was to evaluate the psychometric properties of the HSPSC-D and compare its dimensionality with factor structures derived from different language versions of the HSPSC. Our split-half validation resulted in an alternative 8-factor model with good psychometric properties. Most parts of the instrument demonstrate relative stability over different language versions and appear suitable for cross-national studies. However, items of four safety culture dimensions require further improvement to support a common structure for comparison across language versions.

In our study HSPSC-D demonstrated marginally satisfactory psychometric properties, allowing for its use in German hospitals. HSPSC-D demonstrated a somewhat unsatisfactory model fit in CFA with the original 12-factor model. EFA resulted in an alternative 8-factor model, with good model fit. Nevertheless, the instrument demonstrated satisfactory to good internal consistency in both models. Studies with other language versions of the HSPSC have repeatedly reported similar results – good model fit of different factor structure and mostly good internal consistency [5–7, 9, 11, 12, 14, 15]. These findings indicate that the HSPSC is a useful instrument for measuring and comparing patient safety culture within a healthcare system for which the particular HSPSC version has previously been validated.

Our analysis of instrument dimensionality across language versions revealed that whilst some dimensions maintain relative stability of appearance and composition across language versions, others vary significantly. When analyzing 12 different factor models, including the original North American 12-factor model and the 8-factor model resulting from our EFA, we found that items from eight dimensions maintain relative stability in appearance and composition over different cultural adaptations. These dimensions were *Teamwork within units*, *Nonpunitive response to error*, *Staffing*, *Supervisor/manager expectations/actions*, *Frequency of event reporting*, *Feedback and communication about error*, *Hospital management support for patient safety* and *Teamwork across hospital units*. The items from these dimensions seem to maintain their coherence and measure one common factor in different language adaptations and different healthcare systems. In contrast the remaining four dimensions, namely *Organizational Learning—Continuous improvement, Overall Perceptions of Safety*, *Communication Openness* and *Hospital Handoffs & Transitions*, appeared in only ≤60% of analyzed models since corresponding items were either removed, or migrated to or merged with other dimensions. Similarly, Hedskoeld [7] revealed a 9-factor model but argues against removing items and dimensions from the instrument, stating that they can still be used to understand and improve patient safety. Even though these dimensions and corresponding items may be very important in studies of patient safety culture, they need to be refined in order to support their stability over different cultural adaptations.

Evaluation of psychometric properties of a translated version of the instrument is important, as only the results of validated instruments can be properly interpreted and used for comparison in local contexts. A number of studies reported that the original 12-factor model did not fit the data well, and alternative factor models were suggested [5–7, 9–12, 14–16]. Variation in the factor structure may be partially attributed to the differences between study samples and study populations. These studies differ by setting, sample size, representation of different professional groups and other characteristics, which can have influence on the performance of the instrument, hence should be considered in analysis. Finally the specific characteristics of study population's culture, as well as of local healthcare system influences how the respondents perceive, understand and respond to each individual item in the questionnaire, ultimately altering the factor structure and interpretation of the results. Concerning the international use of the instrument, several articles highlight the importance of a common factor structure. For example, Occelli [15] underlines the need to adapt the tool to each country's environment while stating that "for international comparison purposes, a core set of dimensions consistently assessed as valid should be defined and measured in all countries." Perneger [14] further argues that local improvements to a translated version can be ineffective, due to several unresolved issues inherent in the instrument, such as limited internal consistency of some dimensions, different dimensionality found in various language versions and the lack of external validation of study results.

Limitations

The data analysis and results in the study were limited to two German university hospitals. Also, our findings shouldn't be generalized to all hospital employees, as the study sample mainly consists of nurses and physicians. However, our findings regarding psychometric properties of the instrument, as well as its dimensionality are in line with those of similar studies from other countries. While exploring the common dimensionality of various language versions our analysis was limited to research articles retrieved from the official web-page of AHRQ [24]. Taking into account more studies that report a different factor structure based on a systematic review could improve the analysis. Lastly, the diversity of study methodology and reporting of studies with different language versions of HSPSC, may be considered an additional obstacle for cross-national use of the instrument.

Conclusions

Overall the German version of the HSPSC demonstrated acceptable psychometric properties for surveying clinical personnel in German hospitals. We found that most safety culture dimensions were relatively stable across different language models. However, other dimensions demonstrate high variability and inconsistency. Such inconsistencies need to be refined in order to support a more uniform factor structure across language versions in order to facilitate the use of HSPSC at the cross-national level.

Contributors.

Data were collected by the WorkSafeMed Consortium. Data analysis was carried out by NG under the supervision of TM and AH. NG and AH wrote the manuscript that was then revised by MB, TM. The final version of the manuscript has been approved by all authors.

Competing interests: None.

Acknowledgments

We thank the members of the advisory board for their valuable advice at various stages of the project; especially for their critical assessment of instrument development: (The names were removed for the review purposes, and will be added before publishing).

Last but not least we would like to thank all study participants. We acknowledge the support of the hospital management, the efforts of study coordinators in participating departments and units to facilitate data collection and the respondents for their effort and time to fill in the surveys.

Funding

The „WorkSafeMed" study was funded by the Federal Ministry of Education and Research (FKZ 01GY1325A and 01GY1325B). We acknowledge additional financial support by the German Research Foundation and the administrative support by the DLR Project Management Agency.

Data Sharing Statement

Because of data security aspects, data from the WorkSafeMed study will not be made available in the public domain. However, data will be used by students of both project

partners for their theses. Data will be stored in accordance with national and regional data security standards.

REFERENCES

1 Waterson P, ed. Patient safety culture: Theory, methods, and application. Farnham, Surrey, UK: Ashgate 2014.
2 Guldenmund FW. The nature of safety culture: a review of theory and research. *Saf Sci* 2000;34:215–57.
3 Sorra J, Dyer N. Multilevel psychometric properties of the AHRQ hospital survey on patient safety culture. *BMC Health Serv Res* 2010;10:199.
4 Sorra J, Nieva V. Hospital Survey on Patient Safety Culture. (Prepared by Westat, under Contract No. 290-96-0004). AHRQ Publication No. 04-0041. Rockville, MD 2004.
5 Sarac C, Flin R, Mearns K, et al. Hospital survey on patient safety culture: psychometric analysis on a Scottish sample. *BMJ Qual Saf* 2011;20:842–48.
6 Robida A. Hospital Survey on Patient Safety Culture in Slovenia: a psychometric evaluation. *Int J Qual Health Care* 2013;25:469–75.
7 Pfeiffer Y, Manser T. Development of the German version of the Hospital Survey on Patient Safety Culture: Dimensionality and psychometric properties. *Saf Sci* 2010;48:1452–62.
8 Hammer A, Manser T. The Use of the Hospital Survey on Patient Safety Culture in Europe. In: Waterson P, ed. Patient safety culture: Theory, methods, and application. Farnham, Surrey, UK: Ashgate 2014:229–62.
9 Waterson P, Griffiths P, Stride C, et al. Psychometric properties of the Hospital Survey on Patient Safety Culture: findings from the UK. *Qual Saf Health Care* 2010;19(5):2e.
10 Smits M, Christiaans-Dingelhoff I, Wagner C, et al. The psychometric properties of the ‚Hospital Survey on Patient Safety Culture' in Dutch hospitals. *BMC Health Serv Res* 2008;8:230.
11 Hedskold M, Pukk-Harenstam K, Berg E, et al. Psychometric properties of the Hospital Survey on Patient Safety Culture, HSOPSC, applied on a large Swedish health care sample. *BMC Health Serv Res* 2013;13:332.
12 Bodur S, Filiz E. Validity and reliability of Turkish version of „Hospital Survey on Patient Safety Culture" and perception of patient safety in public hospitals in Turkey. *BMC Health Serv Res* 2010;10:28.
13 Elmontsri M, Almashrafi A, Banarsee R, et al. Status of patient safety culture in Arab countries: a systematic review. *BMJ Open* 2017;7(2):e013487.
14 Perneger TV, Staines A, Kundig F. Internal consistency, factor structure and construct validity of the French version of the Hospital Survey on Patient Safety Culture. *BMJ Qual Saf* 2014;23:389–97.
15 Occelli P, Quenon J-L, Kret M, et al. Validation of the French version of the Hospital Survey on Patient Safety Culture questionnaire. *Int J Qual Health Care* 2013;25:459–68.
16 Najjar S, Hamdan M, Baillien E, et al. The Arabic version of the hospital survey on patient safety culture: a psychometric evaluation in a Palestinian sample. *BMC Health Serv Res* 2013;13:193.
17 Haugen AS, Softeland E, Eide GE, et al. Patient safety in surgical environments: cross-countries comparison of psychometric properties and results of the Norwegian version of the Hospital Survey on Patient Safety. *BMC Health Serv Res* 2010;10:279.
18 Wagner C, Smits M, Sorra J, et al. Assessing patient safety culture in hospitals across countries. *Int J Qual Health Care* 2013;25(3):213–21.

19 Hammer A, Ernstmann N, Ommen O, et al. Psychometric properties of the Hospital Survey on Patient Safety Culture for hospital management (HSOPS_M). *BMC Health Serv Res* 2011;11:165.

20 Wirtz M. On the problem of missing data: How to identify and reduce the impact of missing data on findings of data analysis. *Rehabilitation (Stuttg)* 2004;43(2):109–15.

21 Schafer JL, Graham JW. Missing data: Our view of the state of the art. *Psychol Methods* 2002;7:147–77.

22 Field A. Discovering statistics using SPSS: (and sex and drugs and rock'n'roll), 3rd edn. Los Angeles: SAGE 2012.

23 Hair JF, Black WC, Babin BJ, et al. Multivariate data analysis. Upper Saddle River, N.J.: Pearson Education 2014.

24 the Agency for Healthcare Research and Quality (AHRQ). Surveys on Patient Safety Culture: Psychometric Analyses—International 2017. Available at: www.ahrq.gov/professionals/quality-patient-safety/patientsafetyculture/resources Accessed May 02, 2017.

Appendix 1

Pearson's correlations between single item outcome variables (*Patient safety grade* (E1) and *Number of events reported* (G1)) and HSPSC dimensions in two factor models.

A. Original 12-factor model	Number of items	**E1**	**G1**[+]	01	02	03	04	05	06	07	08	09	10	11
E1. Patient Safety Grade	1													
G1: Number of Events Reported	1	**-0,04**												
01. Teamwork within units	4	**0,32****	**0,08***											
02. Organizational learning	3	**0,48****	**0,02**	0,45**										
03. Nonpunitive response to error	3	**0,41****	**-0,05**	0,32**	0,38**									
04. Staffing	4	**0,46****	**-0,08***	0,26**	0,30**	0,42**								
05. Overall perceptions of safety	4	**0,70****	**-0,06***	0,37**	0,57**	0,46**	0,61**							
06. Supervisor/manager expectations/actions	4	**0,41****	**0,00**	0,35**	0,47**	0,38**	0,32**	0,44**						
07. Frequency of event reporting	3	**0,34****	**0,14****	0,22**	0,39**	0,18**	0,14**	0,34**	0,28**					
08. Feedback and communication about error	3	**0,43****	**0,01**	0,40**	0,61**	0,32**	0,22**	0,44**	0,50**	0,50**				
09. Communication openness	3	**0,26****	**0,06**	0,38**	0,43**	0,35**	0,10**	0,28**	0,38**	0,36**	0,59**			
10. Hospital management support for patient safety	3	**0,58****	**-0,03**	0,38**	0,55**	0,40**	0,47**	0,61**	0,39**	0,36**	0,49**	0,29**		
11. Teamwork across hospital units	4	**0,43****	**0,00**	0,40**	0,43**	0,38**	0,32**	0,46**	0,33**	0,32**	0,42**	0,31**	0,54**	
12. Hospital handoffs and transitions	4	**0,37****	**0,03**	0,32**	0,33**	0,31**	0,25**	0,36**	0,29**	0,25**	0,36**	0,35**	0,41**	0,60**

B. new 8-factor model	Number of items	**E1**	**G1**[+]	01	02	03	04	05	06	07	08	09	10	11
E1. Patient safety grade	1													
G1: Number of events reported	1	**-0,04**												
01. Teamwork within units	3	**0,28****	**0,09****											
02. Organizational learning [D]	-	-	-	-										
03. Nonpunitive response to error	3	**0,41****	**-0,05**	0,30**	-									
04. Staffing	4	**0,54****	**-0,09****	0,24**	-	0,42**								
05. Overall perceptions of safety [D]	-	-	-	-	-	-	-							
06. Supervisor/manager expectations/actions	3	**0,40****	**0,01**	0,30**	-	0,34**	0,31**	-						
07. Frequency of event reporting	3	**0,34****	**0,14****	0,20**	-	0,18**	0,19**	-	0,29**					
08. Feedback and communication about error	4	**0,41****	**0,03**	0,38**	-	0,33**	0,24**	-	0,52**	0,50**				
09. Communication openness [D]	-	-	-	-	-	-	-	-	-	-	-			
10. Hospital management support for patient safety	3	**0,58****	**-0,03**	0,32**	-	0,40**	0,51**	-	0,38**	0,36**	0,47**	-		
11. Teamwork across hospital units	5	**0,42****	**0,00**	0,36**	-	0,40**	0,37**	-	0,31**	0,31**	0,40**	-	0,54**	
12. Hospital handoffs and transitions [D]	-	-	-	-	-	-	-	-	-	-	-	-	-	-

n=974, except for G1.
[+] n=952 for the variable G1: Number of Events Reported.
*P < 0.05; **P < 0.01.
[D] Dimension is removed from the model (collored in dark gray).

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1,2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2,3
Introduction			**5-6**
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			**6-9**
Study design	4	Present key elements of study design early in the paper	6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-7
Bias	9	Describe any efforts to address potential sources of bias	*NA*
Study size	10	Explain how the study size was arrived at	6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7-9
		(*b*) Describe any methods used to examine subgroups and interactions	NA
		(*c*) Explain how missing data were addressed	7
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	NA
		(*e*) Describe any sensitivity analyses	NA
Results			**9-20**

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	9-10
		(b) Give reasons for non-participation at each stage	-
		(c) Consider use of a flow diagram	-
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	9-10
		(b) Indicate number of participants with missing data for each variable of interest	9-10
Outcome data	15*	Report numbers of outcome events or summary measures	9-12
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	NA
		(*b*) Report category boundaries when continuous variables were categorized	10
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	13-20
Discussion			**20-23**
Key results	18	Summarise key results with reference to study objectives	20-21
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	22
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	21-22
Generalisability	21	Discuss the generalisability (external validity) of the study results	21-22
Other information			**24-25**
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	24-25

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

Evaluation of psychometric properties of the German Hospital Survey on Patient Safety Culture and its potential for cross-cultural comparisons: A cross-sectional study.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-018366.R2
Article Type:	Research
Date Submitted by the Author:	08-Sep-2017
Complete List of Authors:	Gambashidze, Nikoloz; University hospital Bonn, Institute for patient safety Hammer, Antje; University hospital Bonn, Institute for patient safety Broesterhaus, Mareen; University hospital Bonn, Institute for patient safety Manser, Tanja; University hospital Bonn, Institute for patient safety
Primary Subject Heading:	Health services research
Secondary Subject Heading:	Research methods
Keywords:	Quality in health care < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, Health & safety < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, International health services < HEALTH SERVICES ADMINISTRATION & MANAGEMENT, STATISTICS & RESEARCH METHODS

Title: Evaluation of psychometric properties of the German Hospital Survey on Patient Safety Culture and its potential for cross-cultural comparisons: A cross-sectional study.

Authors: Nikoloz Gambashidze*[1], Antje Hammer*[1], Mareen Brösterhaus[1], Tanja Manser[1], On behalf of the WorkSafeMed Consortium**

[1] University Hospital Bonn, Institute for Patient Safety, Bonn, Germany

* Co-first author.

** The WorkSafeMed Project Consortium comprises: Luntz E, Rieger MA (project lead), Sturm H, Wagner A(Institute of Occupational and Social Medicine and Health Services Research, University Hospital of Tuebingen), Hammer A, Manser T (Institute for Patient Safety, University Hospital Bonn), Martus P (Institute for Clinical Epidemiology and Applied Biometry, University Hospital of Tuebingen), Holderied M (University Hospital Tuebingen).

Corresponding author:

Nikoloz Gambashidze,

Institute for Patient Safety,

University Hospital Bonn,

Sigmund-Freud-Str. 25,

53127 Bonn, Germany

Phone: +49 228 287 13948

E-Mail: Nikoloz.Gambashidze@ukbonn.de

Running title: Psychometrics of German HSPSC

Word count, abstract: 255.

Word count, manuscript: 3312.

Abstract

Objective: To study the psychometric characteristics of German version of the Hospital Survey on Patient Safety Culture, and to compare its dimensionality to other language versions in order to understand the instrument's potential for cross-national studies.

Design: Cross-sectional multicenter study to establish psychometric properties of German version of the survey instrument.

Setting: 73 units from 37 departments of two German university hospitals.

Participants: Clinical personnel (n=995 responses, response rate 39.6%).

Primary and secondary outcome measures: Psychometric properties (e.g. Model fit, internal consistency, construct validity) of the instrument, and comparison of dimensionality across different language translations.

Results: The instrument demonstrated acceptable to good internal consistency (Cronbach's alpha 0.64-0.88). Confirmatory factor analysis of the original 12-factor model resulted in marginally satisfactory model fit (RMSEA= 0.05; SRMR; CFI=0.90; GFI=0.88; TLI=0.88). Exploratory factor analysis resulted in an alternative 8-factor model with good model fit (RMSEA= 0.05; SRMR; CFI=0.95; GFI=0.91; TLI=0.94) and good internal consistency (Cronbach's alpha 0.73-0.87) and construct validity. Analysis of the dimensionality compared to models from 10 other language versions revealed eight dimensions with relatively stable composition and appearance across different versions and four dimensions requiring further improvement.

Conclusions: The German version of Hospital survey on Patient Safety Culture demonstrated satisfactory psychometric properties for use in German hospitals. However, our comparison of instrument dimensionality across different language versions indicates limitations concerning cross-national studies. Results of this study can be considered in interpreting findings across national contexts, in further refinement of the instrument for cross-national studies, and to better understand various facets and dimensions of patient safety culture.

Keywords: Quality in health care; Health & safety; International health services; statistics & research methods.

Article Summary

Strengths and limitations of this study

- Our study supports the development of a more uniform factor structure for the Hospital Survey on Patient Safety Culture across language versions in order to facilitate its use in cross-national research.
- By evaluating commonalities and variations in different language versions of the Hospital Survey on Patient Safety Culture we identify relatively stable factors, as well as those in need for improvement.
- First study to validate the German version of the Hospital Survey on Patient Safety Culture for clinical personnel.
- The considerable diversity in study methodology and reporting of studies with different language versions of the Hospital Survey on Patient Safety Culture presents an obstacle for cross-national use of the instrument that has yet to be overcome.

Introduction

All healthcare organizations face specific sets of risks and challenges regarding patient safety. These challenges change dynamically over time, reflecting developments within the organization as well as in its operating environment such as changes in demographics and epidemiology or in patient behavior. To effectively manage these challenges, it is recommended for healthcare organizations to develop a culture of safety that prioritizes safety and organizational learning among other organizational goals [1]. Safety culture is generally considered to be a relatively stable construct, rooted in organizational culture [2].

A number of instruments for measuring safety culture in healthcare organizations have been developed. These instruments enable researchers and decision-makers to evaluate and compare results on different levels of the healthcare system [3]. Comparing results across units and hospitals and establishing benchmarks can drive continuous patient safety improvement. One of the most widely used instruments for evaluating healthcare providers' perception of safety culture in hospital setting is the Hospital Survey on Patient Safety Culture (HSPSC) [4]. The instrument has been translated into many languages and used in different countries around the world [5–16].

There are two gaps that this study aims to address. Firstly, so far no German version of HSPSC has been validated for healthcare personnel in Germany. Secondly, despite some attempts at comparing safety culture at the international level [17, 18], the comparability of the different language versions of the instrument has not been studied systematically. While satisfactory psychometric properties were reported for the original North-American version [4] with 12 dimensions of patient safety culture, alternative factor structures have been reported for other language versions, with the number of dimensions ranging from 8 to 12 [5–7, 9–12, 14–16]. Because an instrument's dimensionality determines the interpretation of results, similarities and differences in dimensionality across different language versions should be considered for cross-national studies of patient safety culture.

Therefore, the aim of this study is two-fold: (1) validation of German version of HSPSC (HSPSC-D) by evaluation of its psychometric properties and (2) evaluation of the instrument's potential for cross-national studies, by comparative analysis of instrument's dimensionality as reported for different language versions.

Methods

Setting

This study was based on data from the cross-sectional, multicenter, mixed methods study *Working conditions, safety culture and patient safety in hospitals – what predicts the safety of the medication process* (WorkSafeMed), conducted between 2014 and 2017. In this article, we focus on HSPSC-D data to evaluate its psychometric properties. The WorkSafeMed study with all its components has been approved by the responsible ethics committees of the medical faculties of the project partners in Bonn (#350/14) and Tubingen (#547/2014BO1). Each partner complied with confidentiality requirements according to German law.

Sample

Safety culture data were collected in two German university hospitals April to July 2015. We included staff from inpatient units with ≥500 patients a year. Intensive care and psychiatric units were excluded. Across the two hospitals a total of 73 units from 37 departments participated in the study. The HSPSC-D questionnaire was distributed to 2512 healthcare professionals. All participants received an initial invitation to participate in the study, followed by two reminders. Study material included all required information regarding the study and data handling. Participation in the study was anonymous and participants' consent was implied by returning completed questionnaires. Non-responder analysis was not performed.

Measure

In order to develop a version of the HSPSC for German healthcare professionals (HSPSC-D), we used two previous German language versions as a starting point. A first translation of the HSPSC for hospital staff in the German speaking part of Switzerland [7] had been culturally and linguistically adapted for use in Swiss hospitals. Hammer et al. [19] used the Swiss version as a starting point for developing a management version of HSPSC to study perceptions of safety culture among medical directors in German hospitals. In our study, the instrument was adapted to be used with healthcare personnel in German hospitals.

The resulting HSPSC-D questionnaire follows the structure of the original North-American version [4] and includes 44 items, 42 of which compose 12 dimensions (10 safety culture dimensions and 2 outcome dimensions). These 42 items use a five-point Likert scale to measure agreement ranging from “strongly disagree” (1) to “strongly agree” (5), or frequency ranging from “never” (1) to “always” (5). The remaining two single item measures are *Number of events reported* (measured on six frequency groups from “No event reports” to “21 event reports or more”) and *Patient safety grade* (measured on five-point scale from “Failing” to “Excellent”).

Analysis

Data processing and preliminary analysis. After excluding responses with more than 30% missing values in HSPSC-D items, we conducted multiple imputations based on the expectation maximization (EM) algorithm using the statistical software NORM 2.03 [20, 21] to replace remaining missing values. Negatively worded items were reverse coded before further analysis.

Several indices were taken into account to ensure that our study sample, as well as every subset used in further analysis was appropriate for factor analysis. Kaiser-Meyer-Olkin (KMO) indicates if the sample of items is adequate for factor analysis, while Measure of Sampling Adequacy (MSA) indicates if an individual item is adequate for factor analysis. For both indices the value >0.7 is desired, and the value of >0.9 is considered perfect [22]. A significant p-value (<0.05) of Bartlett's test of sampling adequacy indicates that it is possible to extract more than one factor [22]. The analyses were performed using SAS 9.4.

Descriptive statistics. We calculated composite scores for each dimension suggested by Sorra and Nieva [4] by calculating the average of corresponding items. We also calculated percentages of positive responses for each dimension by dividing the number of positive responses on corresponding items by the number of non-missing answers in the dimension. Descriptive statistics for each item and dimension were evaluated, including range, mean, and standard deviation.

Exploratory Factor Analysis (EFA). We used EFA to evaluate the factor structure emerging from the study data. In general, EFA and Confirmatory Factor Analysis (CFA) should be performed using different subsets [23]. Thus, we performed the split-half cross validation, by randomly splitting our sample in two: "Exploring" (for EFA) and "Testing" subsets (for subsequent CFA). EFA using maximum likelihood was conducted using the "Exploring" subset. We used Varimax orthogonal pre-rotation, and Promax oblique rotation to aid with interpretation of factor model [23]. We used scree plot and Kaiser Criterion (Eigenvalues >1) for factor extraction. Factor loadings ≥0.4 were considered significant, and factor cross loading <0.4 was considered acceptable [22, 23]. Applying these criteria, we gradually eliminated problematic items until EFA resulted in a satisfactory factor structure.

Confirmatory Factor Analysis (CFA). We evaluated the model fit of the factor structure resulting from the EFA by conducting CFA using the "Testing" subset. By conducting a series of CFA using the complete dataset, we evaluated model fit of original 12-factor model [4], as well as other factor models reported by studies of different language versions of HSPSC. From the official website of the Agency for Healthcare Research and Quality (AHRQ) [24] we retrieved a list of studies including psychometric evaluation of the instrument and identified those reporting a different factor structure.

Internal consistency. Internal consistency was evaluated by calculating Cronbach's alpha as an indicator of correlation between each item and the factor. In their exploratory study Sorra and Nieva [4] considered Cronbach's alpha ≥0.6 as acceptable. We used Cronbach's alpha ≥0.7, as it is typically used in later studies using the HSPSC [5, 6, 9, 11, 14, 15, 17, 19], and is well supported by the literature [22, 23]. Cronbach's alphas were calculated for all factor models considered in the CFA, including the factor model that emerged from EFA.

Construct validity. By calculating average of corresponding non-missing items we calculated mean values for each dimension for the original 12-factor model and for the new model that emerged from EFA. Pearson's correlations were evaluated between dimensions in each model. We expected low to moderate correlations between dimensions. However, correlations >0.85 would indicate possible multicollinearity [4, 22]. We also evaluated the correlations between dimensions of both models with two single item outcome variables – *Patient safety grade* and *Number of incidents reported*.

Evaluation of common dimensionality. In order to evaluate the potential of the instrument for cross-national studies, we evaluated its dimensionality as reported for different language versions. We evaluated appearance and composition of each of the 12 dimensions proposed by Sorra and Nieva [4] and of the 42 corresponding items in all factor models identified from AHRQ web-page [24].

Results

Study sample and descriptive statistics.

Out of 2512 distributed questionnaires 995 were completed, resulting in a response rate of 39.6%. Sample characteristics are presented in table 1.

Table 1: Characteristics of study sample.

Variables	N	%
Study site	**995**	**100.0%**
Hospital A	575	57.8%
Hospital B	420	42.2%
Gender	**995**	**100.0%**
Female	656	65.9%
Male	291	29.2%
missing	48	4.8%
Professional Groups	**995**	**100.0%**
Physician	183	18.4%
Physicians' assistant	198	19.9%
Nurse	552	55.5%
other	34	3.4%
missing	28	2.8%
Managerial functions	**995**	**100.0%**
Yes	195	19.6%
No	759	76.3%
missing	41	4.1%
Contact with patients	**995**	**100.0%**
Yes	965	97.0%
No	7	0.7%
missing	23	2.3%
Age (years)	**995**	**100.0%**
<25	61	6.1%
25-34	360	36.2%
35-44	230	23.1%
45-54	170	17.1%
>54	84	8.4%
missing	90	9.0%

Out of our sample of n=995, 766 responses (76.98%) had no missing values on HSPSC items. 21 responses (2.1%) contained more than 30% missing values on HSPSC items and were thus not included in the analysis. Remaining missing values were imputed using multiple imputations based on the expectation maximization (EM) algorithm. As a result, n=974 cases were available for further analysis. Descriptive statistics of HSPSC-D items and dimensions after imputing remaining missing answers and reverse coding of the negatively worded items are presented in the Table 2.

Table 2: Descriptive statistics of HSPSC-D items and dimensions.

Dimension / Item [1,2,3]	Percent of positive responses [4]	Mean	SD[5]
01. Teamwork Within Hospital Units	**42.3%**	**3.32**	**0.61**
A1. People support one another in this unit.	58.3%	3.65	0.78
A3. When a lot of work needs to be done quickly, we work together as a team to get the work done.	51.2%	3.50	0.84
A4. In this unit, people treat each other with respect.	40.9%	3.36	0.78
A11. When one area in this unit gets really busy, others help out.	18.8%	2.79	0.91
02. Organizational Learning—Continuous improvement	**32.7%**	**3.06**	**0.70**
A6. We are actively doing things to improve patient safety.	50.1%	3.40	0.91
A9. Mistakes have led to positive changes here.	23.5%	2.88	0.89
A13. After we make changes to improve patient safety, we evaluate their effectiveness.	24.4%	2.90	0.89
03. Non-punitive Response To Error	**50.2%**	**3.38**	**0.80**
A8. (R) Staff feel like their mistakes are held against them.	40.1%	3.19	0.96
A12. (R) When an event is reported, it feels like the person is being written up, not the problem.	48.3%	3.33	0.99
A16. (R) Staff worry that mistakes they make are kept in their personnel file.	62.1%	3.62	0.99
04. Staffing	**24.9%**	**2.57**	**0.79**
A2. We have enough staff to handle the workload.	7.5%	2.01	0.97
A5. (R) Staff in this unit work longer hours than is best for patient care	23.1%	2.57	1.18
A7. (R) We use more agency/temporary staff than is best for patient care.	58.2%	3.57	1.20
A14. (R) We work in "crisis mode," trying to do too much, too quickly.	10.9%	2.13	1.02
05. Overall Perceptions of Safety	**34.4%**	**3.03**	**0.79**
A10. (R) It is just by chance that more serious mistakes don't happen around here.	41.1%	3.08	1.20
A15. Patient safety is never sacrificed to get more work done.	25.4%	2.75	1.04
A17. (R) We have patient safety problems in this unit.	43.9%	3.29	0.97
A18. Our procedures and systems are good at preventing errors from happening.	27.2%	3.00	0.89
06. Supervisor/manager expectations & actions promoting safety	**48.5%**	**3.34**	**0.71**
B1. My supervisor/manager says a good word when he/she sees a job done according to established patient safety procedures.	33.7%	3.03	1.02
B2. My supervisor/manager seriously considers staff suggestions for improving patient safety.	55.9%	3.51	0.87
B3. (R) Whenever pressure builds up, my supervisor/manager wants us to work faster, even if it means taking shortcuts.	42.8%	3.19	0.98
B4. (R) My supervisor/manager overlooks patient safety problems that happen over and over.	61.7%	3.61	0.89
07. Frequency of Event Reporting	**38.0%**	**3.00**	**1.03**
D1. When a mistake is made, but is caught and corrected before affecting the patient, how often is this reported?	39.0%	3.03	1.17
D2. When a mistake is made, but has no potential to harm the patient, how often is this reported?	30.1%	2.77	1.14
D3. When a mistake is made that could harm the patient, but does	45.0%	3.19	1.13

not, how often is this reported?			
08. Feedback and Communication About Error	**48.0%**	**3.36**	**0.85**
C1. We are given feedback about changes put into place based on event reports.	40.0%	3.18	1.04
C3. We are informed about errors that happen in this unit.	50.1%	3.41	0.99
C5. In this unit, we discuss ways to prevent errors from happening again.	53.9%	3.50	0.95
09. Communication Openness	**58.6%**	**3.60**	**0.68**
C2. Staff will freely speak up if they see something that may negatively affect patient care.	66.2%	3.74	0.87
C4. Staff feel free to question the decisions or actions of those with more authority.	45.4%	3.35	0.89
C6. (R) Staff are afraid to ask questions, when something does not seem right.	64.1%	3.71	0.91
10. Hospital Management Support for Patient Safety	**23.4%**	**2.79**	**0.86**
F1. Hospital management provides a work climate that promotes patient safety.	22.4%	2.83	0.94
F8. The actions of hospital management show that patient safety is a top priority.	21.1%	2.74	0.97
F9. (R) Hospital management seems interested in patient safety only after an adverse event happens.	26.8%	2.79	1.04
11. Teamwork Across Hospital Units	**29.0%**	**3.03**	**0.61**
F2. (R) Hospital units do not coordinate well with each other.	14.7%	2.57	0.91
F4. There is good cooperation among hospital units that need to work together.	22.6%	3.03	0.73
F6. (R) It is often unpleasant to work with staff from other hospital units.	49.1%	3.39	0.82
F10. Hospital units work well together to provide the best care for patients.	29.7%	3.14	0.77
12. Hospital Handoffs & Transitions	**35.3%**	**3.07**	**0.64**
F3. (R) Things "fall between the cracks" when transferring patients from one unit to another.	13.2%	2.50	0.88
F5. (R) Important patient care information is often lost during shift changes.	37.1%	3.16	0.89
F7. (R) Problems often occur in the exchange of information across hospital units.	29.3%	3.04	0.81
F11. (R) Shift changes are problematic for patients in this hospital.	61.5%	3.59	0.82
E1. Please give your work area/unit in this hospital an overall grade on patient safety.	**35.5%**	**3.22**	**0.76**

Note: Answers 4 and 5 ("A*gree*" and "S*trongly agree*" or "*Most of the time*" and "A*lways*") were considered as positive. Prior to analysis, negatively worded items were reverse coded.
[1]01-12 – corresponding dimension according to original North-American 12-factor model.
[2]A1-A18; B1-B4; C1-C6; D1-D3; E1; F1-F11: Codes of questionnaire items.
[3](R) – negatively worded items, which were reverse coded prior to the analysis.
[4]n=974.
[5] SD – standard deviation.

KMO for the complete sample was 0.93, and MSA for individual items ranged between 0.87 and 0.96. For "Exploring" and "Testing" subsets, KMO was 0.91 and 0.92 respectively, and MSA of individual items in both subsets ranged between 0.84 and 0.96. Bartlett's test was highly significant ($p<0.001$) for the dataset, as well as for both subsets. Preliminary analyses indicated that our sample and the subsets were adequate for factor analysis.

Exploratory Factor Analysis (EFA). We conducted EFA using the "Exploring" subset. We considered factor loadings ≥0.4 as significant, as this cut-off value was typically used in similar studies [4–6, 10–12, 14–16] and was supported by the literature [22, 23]. 14 items not meeting the criteria (factor loading ≥0.4, cross loading <0.4) were excluded from the model, resulting in an 8-factor model with 28 items. The dimension *Organizational learning – continuous improvement* was completely removed. The dimensions *Staffing* and *Overall perceptions of safety* were merged together, as were the dimensions *Feedback and communication about error* with *Communication openness*, and *Teamwork across hospital units* with *Handoffs and transitions*. The resulting 8-factor model is presented in table 3.

Table 3: Appearance of HSPSC items in 12 analyzed factor models (8-factor EFA model, original 12-factor model, and 10 different versions).

HSPSC items[1,2]	Germany (Exploratory Factor Analysis)	**USA (Sorra & Nieva, 2004)**	England (UK) (Waterson et al., 2010)	Scotland (UK) (Sarac et al., 2011)	France (Occelli et al. 2013)	Switzerland (Perneger, 2013)	Switzerland (Pfeiffer and Manser, 2010)	Netherlands (Smits et al. 2008)	Sweden (Hedsköld et al., 2013)	Slovenia (Robida, 2013)	Turkey (Bodur, 2010)	Palestine (Najjar et al., 2013)
01. Teamwork within units												
A1	1	**1**	1	1	1	1	1	1	1	1	1	1
A3	1	**1**	1	1	1	1	1	1	1	1	1	1
A4	1	**1**	1	1	1	1	1	1	1	1	1	1
A11	(N)	**1**	(N)	1	(N)	1	(N)	1	1	1	1	1
02. Organizational learning												
A6	(N)	**2**	(N)	2	8	2	1	8	8	2	2	2
A9	(N)	**2**	(N)	2	8	2	(N)	8	8	2	3	2
A13	(N)	**2**	(N)	2	8	2	6	8	8	(N)	2	2
03. Nonpunitive response to error												
A8	3	**3**	3	3	3	3	3	3	3	3	3	3
A12	3	**3**	(N)	3	3	3	3	3	3	3	3	3
A16	3	**3**	3	3	3	3	3	3	3	3	3	3
04. Staffing												
A2	4	**4**	4	4	4	4	4	4	4	4	4	4
A5	4	**4**	(N)	4	4	4	4	4	4	4	2	4
A7	(N)	**4**	(N)	4	(N)	4	(N)	4	4	4	4	(N)
A14	4	**4**	4	4	4	4	4	5	4	4	4	4
05. Overall perceptions of safety												
A10	4	**5**	4	4	5	5	4	5	4	4	5	(N)
A15	(N)	**5**	(N)	4	5	5	(N)	(N)	4	4	5	5
A17	(N)	**5**	4	4	5	5	4	5	4	4	5	5
A18	(N)	**5**	(N)	(N)	5	5	7	5	8	4	5	5
06. Supervisor/manager expectations/actions												
B1	6	**6**	6	6	6	6	6	6	6	6	6	6
B2	6	**6**	6	6	6	6	6	6	6	6	6	6
B3	(N)	**6**	(N)	6	6	6	6	6	6	6	6	6
B4	6	**6**	(N)	6	6	6	6	6	6	6	6	6
07. Frequency of event reporting												
D1	7	**7**	7	7	7	7	7	7	7	7	7	7
D2	7	**7**	7	7	7	7	7	7	7	7	7	7
D3	7	**7**	7	7	7	7	7	7	7	7	7	7
08. Feedback and communication about error												
C1	8	**8**	8	8	8	8	8	8	8	8	8	(N)
C3	8	**8**	8	8	8	8	8	8	8	8	8	8
C5	8	**8**	8	8	8	8	8	8	8	8	8	8
09. Communication openness												
C2	8	**9**	9	8	9	8	8	9	8	8	8	8
C4	(N)	**9**	9	8	9	8	8	9	8	8	8	8
C6	(N)	**9**	9	8	9	8	(N)	9	3	(N)	8	(N)
10. Hospital management support for patient safety												
F1	10	**10**	(N)	10	10	10	10	10	10	10	10	10
F8	10	**10**	(N)	10	10	10	10	10	10	10	10	10
F9	10	**10**	(N)	10	10	10	10	10	10	10	10	10

11. Teamwork across hospital units												
F2	11	**11**	11	11	11	11	11	11	10	11	10	11
F4	11	**11**	11	11	11	11	11	11	10	11	10	11
F6	11	**11**	(N)	(N)	11	11	11	(N)	12	11	12	11
F10	11	**11**	11	11	10	11	11	11	10	11	10	11
12. Hospital handoffs and transitions												
F3	(N)	**12**	12	(N)	11	11	11	(N)	12	11	12	12
F5	(N)	**12**	12	12	11	11	(N)	12	12	11	12	12
F7	11	**12**	12	12	11	11	11	(N)	12	11	12	12
F11	(N)	**12**	12	12	(N)	11	(N)	12	12	(N)	12	12

Note:
The uncolored cells represent "No change" compared to original 12-factor model.
Colored boxes indicate items that were deleted (N) or moved to different dimension (Dimension number 1-12);
(N): items removed form factor model.
[1]01-12 – corresponding dimension according to original North-American 12-factor model.
[2]A1-A18; B1-B4; C1-C6; D1-D3; F1-F11: Codes of the questionnaire items.

Confirmatory factor analysis (CFA). CFA using the „Testing" subset demonstrated a satisfactory model fit of the factor structure that emerged from EFA (see table 4). The model satisfied desired thresholds of most analyzed indices (RMSEA=0.05; SRMR=0.05; GFI=0.90; CFI=0.93; TLI/NNFT=0.91).

Table 4. Results of Confirmatory Factor Analysis (CFA) of all 12 factor models analyzed.

Variables / Indices analyzed in CFA	Criteria	Germany (EFA)	USA (Sorra & Nieva, 2004)	England (UK) (Waterson et al., 2010)	Scotland (UK) Sarac et al., 2011)	France (Occelli et al. 2013)	Switzerland (Perneger, 2013)	Switzerland (Pfeiffer and Manser, 2010)	Netherlands (Smits et al. 2008)	Sweden (Hedsköld et al., 2013)	Slovenia (Robida, 2013)	Turkey (Bodur, 2010)	Palestine (Najjar et al., 2013)
Number of Observations	NA	487	974	974	974	974	974	974	974	974	974	974	974
Number of Variables	NA	28	42	27	39	39	42	35	38	42	39	42	38
Number of Factors	NA	8	12	9	10	10	10	8	11	8	9	10	11
Root Mean Square Error of Approximation (RMSEA)	<0.07	0.05	0.05	0.05	0.05	0.06	0.05	0.06	0.05	0.06	0.05	0.06	0.05
Standardized Root Mean Residual (SRMR)	<0.08	0.05	0.05	0.05	0.06	0.06	0.05	0.08	0.06	0.06	0.06	0.06	0.05
Root Mean Square Residual (RMSR / RMR)	NA	0.04	0.05	0.04	0.05	0.05	0.05	0.07	0.05	0.06	0.05	0.06	0.04
Goodness of Fit Index (GFI)	>0.90	0.91	0.88	0.92	0.88	0.87	0.86	0.86	0.89	0.83	0.87	0.84	0.90
Adjusted GFI (AGFI)	>0.90	0.90	0.86	0.90	0.86	0.85	0.84	0.84	0.86	0.81	0.85	0.81	0.87
Normed Fit Index (NFI)	>0.95	0.90	0.86	0.90	0.86	0.85	0.84	0.84	0.86	0.80	0.85	0.81	0.88
Comparative Fit Index (CFI)	≥0.90	0.95	0.90	0.93	0.89	0.88	0.88	0.87	0.90	0.84	0.89	0.85	0.91
Tucker-Lewis Index / Non-normed Fit Index (TLI / NNFI)	≥0.90	0.94	0.88	0.91	0.88	0.87	0.87	0.85	0.88	0.83	0.87	0.83	0.90

Note: Colored cells contain values that do not meet requirements.

From the official website of AHRQ [24] we retrieved the list of 23 articles reporting psychometric analyses on international level. From these articles we extracted 10 factor models that differed from the original North-American version. These factor models were from the following countries: England (UK) [9], Scotland (UK) [5], France [15], Switzerland (French [14] and German [7]), Netherland [10], Sweden [11], Slovenia [6], Turkey [12] and Palestine [16]. The eleventh factor model considered in the analysis was the original 12-factor model [4].

Subsequent series of CFA revealed satisfactory fit of the models from England (UK) [9] (RMSEA=0.05; SRMR=0.05; GFI=0.92; CFI=0.93; TLI/NNFT=0.91) and Palestine [16] (RMSEA=0.05; SRMR=0.05; GFI=0.90; CFI=0.91; TLI/NNFT=0.90) to our data. The original 12-factor model resulted in marginally satisfactory model fit (RMSEA=0.05; SRMR=0.05; GFI=0.88; CFI=0.90; TLI/NNFT=0.88). The models from Scotland (UK), France, Switzerland, Netherland and Slovenia, resulted in suboptimal values of CFA indices (table 4). Models from Sweden and Turkey demonstrated unsatisfactory model fit in CFA.

Internal consistency. The original 12-factor model demonstrated good Cronbach's alpha for all dimensions except *Organizational learning – Continuous improvement* (0.68) and *Communication openness* (0.64). Cronbach's alpha for dimensions of 8-factor model were between 0.73 and 0.87. Two dimensions, *Teamwork within units* and *Communication openness*, demonstrated consistently low alphas in other factor models analyzed. Three dimensions, *Nonpunitive response to error*, *Staffing* and *Handoffs & transitions*, had lower than 0.7 values only in one or two of analyzed models. Cronbach's alpha for the remaining seven dimensions in all analyzed models was ≥0.7, if present in the model (table 5).

Table 5: Internal consistency (Cronbach's alpha) of all 12 factor models analyzed.

Dimensions (from original 12-factor model)	**Germany (EFA)**	USA (Sorra & Nieva, 2004)	England (UK) (Waterson et al., 2010)	Scotland (UK) Sarac et al., 2011)	France (Occelli et al. 2013)	Switzerland (Perneger, 2013)	Switzerland (Pfeiffer and Manser, 2010)	Netherlands (Smits et al. 2008)	Sweden (Hedsköld et al., 2013)	Slovenia (Robida, 2013)	Turkey (Bodur, 2010)	Palestine (Najjar et al., 2013)
01. Teamwork Within Units	0.78	0.74	0.79	0.74	0.79	0.74	0.75	0.74	0.74	0.74	0.74	0.74
02. Organizational Learning - Continuous Improvement		0.68		0.68		0.68				0.51	0.53	0.68
03. Nonpunitive Response to Error	0.73	0.74	0.61	0.74	0.74	0.74	0.74	0.74	0.72	0.74	0.72	0.74
04. Staffing	0.79	0.70	0.80	0.80	0.73	0.70	0.80	0.53	0.80	0.82	0.65	0.73
05. Overall Perceptions of Patient Safety		0.77			0.77	0.77		0.79			0.77	0.71
06. Supervisor Expectations & Actions Promoting Patient Safety	0.75	0.75	0.72	0.75	0.75	0.75	0.74	0.75	0.75	0.75	0.75	0.75
07. Frequency of Events Reported	0.87	0.88	0.88	0.88	0.88	0.88	0.80	0.88	0.88	0.88	0.88	0.88
08. Feedback & Communication About Error	0.83	0.81	0.81	0.82	0.83	0.82	0.83	0.83	0.86	0.83	0.82	0.80
09. Communication Openness		0.64	0.64		0.64			0.64				
10. Management Support for Patient Safety	0.83	0.84		0.84	0.82	0.84	0.84	0.84	0.84	0.84	0.84	0.84
11. Teamwork Across Units	0.79	0.75	0.75	0.75	0.79	0.83	0.82	0.75		0.82		0.75
12. Handoffs & Transitions		0.75	0.75	0.68				0.66	0.76		0.76	0.75

<0.7 – Not satisfactory (cells colored in dark grey)
≥0.7 – Good [23]
Empty cell (colored in light grey) – Dimension is not present in the model.

Construct validity. Correlation between dimensions of original 12-factor model were between 0.10 and 0.61 (p<0.01). All 12 dimensions were positively correlated with the outcome variable *Patient safety grade* (correlations between 0.26 and 0.70, p<0.01). Dimensions of 8-factor model from EFA were also positively inter-correlated (0.18-0.54, p<0.01) and positively correlated with the outcome variable *Patient safety grade* (0.29-0.58, p<0.01). All dimensions in both factor models resulted in no or week correlation (<0.2) with outcome variable *Number of events reported*. All correlations are presented in the online Appendix 1."

Evaluation of common dimensionality. We analyzed the appearance and role of each individual item and dimension from the original 12-factor model in factor model from EFA, and in 10 models reported by studies from different language versions. Table 3 presents 42 items of the original 12-factor model and their appearance in all 12 analyzed models. The uncolored cells represent no change, where the item retains its original role in the factor model. Changes are represented by colored boxes, which indicate elimination of the questionnaire item (N), or moving it to a different dimension (labeled from 1 to 12).

14 items were eliminated from analysis in EFA. 11 of these 14 items demonstrated significant inconsistency since in at least half of 10 analyzed factor models they were also eliminated, moved or merged with another dimension. All of the remaining 28 items of our 8-factor factor model demonstrated relative stability by retaining a similar role in at least 50% of the 10 analyzed factor models; 23 items maintained their role in 80% or more of the models.

Eight dimensions, including *Teamwork within units*, *Non-punitive response to error*, *Supervisor expectations & actions promoting patient safety*, *Frequency of events reported*, *Staffing*, *Feedback & communication about error*, *Management support for patient safety* and *Teamwork across hospital units* demonstrated relative stability over the different language models, appearing in 80% or more of the 10 analyzed models. The dimension *Communication openness* was merged with the dimension *Feedback and communication about error* in 7 models [5–7, 11, 12, 14, 16]. Similarly, the dimension *Hospital handoffs and transitions* was merged with the dimension *Teamwork across hospital units* in 4 models [6, 7, 14, 15], and the dimension *Overall perceptions of safety* with the dimension *Staffing* in 5 models [5–7, 9, 11]. The items from the dimension *Organizational learning - continuous improvement* were shown to be highly inconsistent across various models. In five models, the items from this dimension were either removed from the model [9] or merged with other dimensions [7, 10, 11, 15] (e.g. with *Feedback and communication about error*).

Discussion

The aim of this study was to evaluate the psychometric properties of the HSPSC-D and compare its dimensionality with factor structures derived from different language versions of the HSPSC. Our split-half validation resulted in an alternative 8-factor model with good psychometric properties. Most parts of the instrument demonstrate relative stability over different language versions and appear suitable for cross-national studies. However, items of four safety culture dimensions require further improvement to support a common structure for comparison across language versions.

In our study HSPSC-D demonstrated marginally satisfactory psychometric properties, allowing for its use in German hospitals. HSPSC-D demonstrated a somewhat unsatisfactory model fit in CFA with the original 12-factor model. EFA resulted in an alternative 8-factor model, with good model fit. Nevertheless, the instrument demonstrated satisfactory to good internal consistency in both models. Studies with other language versions of the HSPSC have repeatedly reported similar results – good model fit of different factor structure and mostly good internal consistency [5–7, 9, 11, 12, 14, 15]. These findings indicate that the HSPSC is a useful instrument for measuring and comparing patient safety culture within a healthcare system for which the particular HSPSC version has previously been validated.

Our analysis of instrument dimensionality across language versions revealed that whilst some dimensions maintain relative stability of appearance and composition across language versions, others vary significantly. When analyzing 12 different factor models, including the original North American 12-factor model and the 8-factor model resulting from our EFA, we found that items from eight dimensions maintain relative stability in appearance and composition over different cultural adaptations. These dimensions were *Teamwork within units*, *Nonpunitive response to error*, *Staffing*, *Supervisor/manager expectations/actions*, *Frequency of event reporting*, *Feedback and communication about error*, *Hospital management support for patient safety* and *Teamwork across hospital units*. The items from these dimensions seem to maintain their coherence and measure one common factor in different language adaptations and different healthcare systems. In contrast the remaining four dimensions, namely *Organizational Learning—Continuous improvement, Overall Perceptions of Safety*, *Communication Openness* and *Hospital Handoffs & Transitions*, appeared in only ≤60% of analyzed models since corresponding items were either removed, or migrated to or merged with other dimensions. Similarly, Hedskoeld [7] revealed a 9-factor model but argues against removing items and dimensions from the instrument, stating that they can still be used to understand and improve patient safety. Even though these dimensions and corresponding items may be very important in studies of patient safety culture, they need to be refined in order to support their stability over different cultural adaptations.

Evaluation of psychometric properties of a translated version of the instrument is important, as only the results of validated instruments can be properly interpreted and used for comparison in local contexts. A number of studies reported that the original 12-factor model did not fit the data well, and alternative factor models were suggested [5–7, 9–12, 14–16]. Variation in the factor structure may be partially attributed to the differences between study samples and study populations. These studies differ by setting, sample size, representation of different professional groups and other characteristics, which can have influence on the performance of the instrument, hence should be considered in analysis. Finally the specific characteristics of study population’s culture, as well as of local healthcare system influences how the respondents perceive, understand and respond to each individual item in the questionnaire, ultimately altering the factor structure and interpretation of the results. Concerning the international use of the instrument, several articles highlight the importance of a common factor structure. For example, Occelli [15] underlines the need to adapt the tool to each country's environment while stating that "for international comparison purposes, a core set of dimensions consistently assessed as valid should be defined and measured in all countries." Perneger [14] further argues that local improvements to a translated version can be ineffective, due to several unresolved issues inherent in the instrument, such as limited internal consistency of some dimensions, different dimensionality found in various language versions and the lack of external validation of study results.

Limitations

The data analysis and results in the study were limited to two German university hospitals. Also, our findings shouldn't be generalized to all hospital employees, as the study sample mainly consists of nurses and physicians. However, our findings regarding psychometric properties of the instrument, as well as its dimensionality are in line with those of similar studies from other countries. While exploring the common dimensionality of various language versions our analysis was limited to research articles retrieved from the official web-page of AHRQ [24]. Taking into account more studies that report a different factor structure based on a systematic review could improve the analysis. Lastly, the diversity of study methodology and reporting of studies with different language versions of HSPSC, may be considered an additional obstacle for cross-national use of the instrument.

Conclusions

Overall the German version of the HSPSC demonstrated acceptable psychometric properties for surveying clinical personnel in German hospitals. We found that most safety culture dimensions were relatively stable across different language models. However, other dimensions demonstrate high variability and inconsistency. Such inconsistencies need to be refined in order to support a more uniform factor structure across language versions in order to facilitate the use of HSPSC at the cross-national level.

Contributors.

Data were collected by the WorkSafeMed Consortium. Data analysis was carried out by NG under the supervision of TM and AH. NG and AH wrote the manuscript that was then revised by MB, TM. The final version of the manuscript has been approved by all authors.

Competing interests: None.

Acknowledgments

We thank the members of the advisory board for their valuable advice at various stages of the project; especially for their critical assessment of instrument development: (The names were removed for the review purposes, and will be added before publishing).

Last but not least we would like to thank all study participants. We acknowledge the support of the hospital management, the efforts of study coordinators in participating departments and units to facilitate data collection and the respondents for their effort and time to fill in the surveys.

Funding

The „WorkSafeMed“ study was funded by the Federal Ministry of Education and Research (FKZ 01GY1325A and 01GY1325B). We acknowledge additional financial support by the German Research Foundation and the administrative support by the DLR Project Management Agency.

Data Sharing Statement

Because of data security aspects, data from the WorkSafeMed study will not be made available in the public domain. However, data will be used by students of both project

partners for their theses. Data will be stored in accordance with national and regional data security standards.

REFERENCES

1 Waterson P, ed. Patient safety culture: Theory, methods, and application. Farnham, Surrey, UK: Ashgate 2014.
2 Guldenmund FW. The nature of safety culture: a review of theory and research. *Saf Sci* 2000;34:215–57.
3 Sorra J, Dyer N. Multilevel psychometric properties of the AHRQ hospital survey on patient safety culture. *BMC Health Serv Res* 2010;10:199.
4 Sorra J, Nieva V. Hospital Survey on Patient Safety Culture. (Prepared by Westat, under Contract No. 290-96-0004). AHRQ Publication No. 04-0041. Rockville, MD 2004.
5 Sarac C, Flin R, Mearns K, et al. Hospital survey on patient safety culture: psychometric analysis on a Scottish sample. *BMJ Qual Saf* 2011;20:842–48.
6 Robida A. Hospital Survey on Patient Safety Culture in Slovenia: a psychometric evaluation. *Int J Qual Health Care* 2013;25:469–75.
7 Pfeiffer Y, Manser T. Development of the German version of the Hospital Survey on Patient Safety Culture: Dimensionality and psychometric properties. *Saf Sci* 2010;48:1452–62.
8 Hammer A, Manser T. The Use of the Hospital Survey on Patient Safety Culture in Europe. In: Waterson P, ed. Patient safety culture: Theory, methods, and application. Farnham, Surrey, UK: Ashgate 2014:229–62.
9 Waterson P, Griffiths P, Stride C, et al. Psychometric properties of the Hospital Survey on Patient Safety Culture: findings from the UK. *Qual Saf Health Care* 2010;19(5):2e.
10 Smits M, Christiaans-Dingelhoff I, Wagner C, et al. The psychometric properties of the 'Hospital Survey on Patient Safety Culture' in Dutch hospitals. *BMC Health Serv Res* 2008;8:230.
11 Hedskold M, Pukk-Harenstam K, Berg E, et al. Psychometric properties of the Hospital Survey on Patient Safety Culture, HSOPSC, applied on a large Swedish health care sample. *BMC Health Serv Res* 2013;13:332.
12 Bodur S, Filiz E. Validity and reliability of Turkish version of "Hospital Survey on Patient Safety Culture" and perception of patient safety in public hospitals in Turkey. *BMC Health Serv Res* 2010;10:28.
13 Elmontsri M, Almashrafi A, Banarsee R, et al. Status of patient safety culture in Arab countries: a systematic review. *BMJ Open* 2017;7(2):e013487.
14 Perneger TV, Staines A, Kundig F. Internal consistency, factor structure and construct validity of the French version of the Hospital Survey on Patient Safety Culture. *BMJ Qual Saf* 2014;23:389–97.
15 Occelli P, Quenon J-L, Kret M, et al. Validation of the French version of the Hospital Survey on Patient Safety Culture questionnaire. *Int J Qual Health Care* 2013;25:459–68.
16 Najjar S, Hamdan M, Baillien E, et al. The Arabic version of the hospital survey on patient safety culture: a psychometric evaluation in a Palestinian sample. *BMC Health Serv Res* 2013;13:193.
17 Haugen AS, Softeland E, Eide GE, et al. Patient safety in surgical environments: cross-countries comparison of psychometric properties and results of the Norwegian version of the Hospital Survey on Patient Safety. *BMC Health Serv Res* 2010;10:279.
18 Wagner C, Smits M, Sorra J, et al. Assessing patient safety culture in hospitals across countries. *Int J Qual Health Care* 2013;25(3):213–21.

19 Hammer A, Ernstmann N, Ommen O, et al. Psychometric properties of the Hospital Survey on Patient Safety Culture for hospital management (HSOPS_M). *BMC Health Serv Res* 2011;11:165.

20 Wirtz M. On the problem of missing data: How to identify and reduce the impact of missing data on findings of data analysis. *Rehabilitation (Stuttg)* 2004;43(2):109–15.

21 Schafer JL, Graham JW. Missing data: Our view of the state of the art. *Psychol Methods* 2002;7:147–77.

22 Field A. Discovering statistics using SPSS: (and sex and drugs and rock'n'roll), 3rd edn. Los Angeles: SAGE 2012.

23 Hair JF, Black WC, Babin BJ, et al. Multivariate data analysis. Upper Saddle River, N.J.: Pearson Education 2014.

24 the Agency for Healthcare Research and Quality (AHRQ). Surveys on Patient Safety Culture: Psychometric Analyses—International 2017. Available at: www.ahrq.gov/professionals/quality-patient-safety/patientsafetyculture/resources Accessed May 02, 2017.

Appendix 1

Pearson's correlations between single item outcome variables (*Patient safety grade* (E1) and *Number of events reported* (G1)) and HSPSC dimensions in two factor models.

A. Original 12-factor model	Number of items	**E1**	**G1**[+]	01	02	03	04	05	06	07	08	09	10	11
E1. Patient Safety Grade	1													
G1: Number of Events Reported	1	**-0,04**												
01. Teamwork within units	4	**0,32****	**0,08***											
02. Organizational learning	3	**0,48****	**0,02**	0,45**										
03. Nonpunitive response to error	3	**0,41****	**-0,05**	0,32**	0,38**									
04. Staffing	4	**0,46****	**-0,08***	0,26**	0,30**	0,42**								
05. Overall perceptions of safety	4	**0,70****	**-0,06***	0,37**	0,57**	0,46**	0,61**							
06. Supervisor/manager expectations/actions	4	**0,41****	**0,00**	0,35**	0,47**	0,38**	0,32**	0,44**						
07. Frequency of event reporting	3	**0,34****	**0,14****	0,22**	0,39**	0,18**	0,14**	0,34**	0,28**					
08. Feedback and communication about error	3	**0,43****	**0,01**	0,40**	0,61**	0,32**	0,22**	0,44**	0,50**	0,50**				
09. Communication openness	3	**0,26****	**0,06**	0,38**	0,43**	0,35**	0,10**	0,28**	0,38**	0,36**	0,59**			
10. Hospital management support for patient safety	3	**0,58****	**-0,03**	0,38**	0,55**	0,40**	0,47**	0,61**	0,39**	0,36**	0,49**	0,29**		
11. Teamwork across hospital units	4	**0,43****	**0,00**	0,40**	0,43**	0,38**	0,32**	0,46**	0,33**	0,32**	0,42**	0,31**	0,54**	
12. Hospital handoffs and transitions	4	**0,37****	**0,03**	0,32**	0,33**	0,31**	0,25**	0,36**	0,29**	0,25**	0,36**	0,35**	0,41**	0,60**
B. new 8-factor model	Number of items	**E1**	**G1**[+]	01	02	03	04	05	06	07	08	09	10	11
E1. Patient safety grade	1													
G1: Number of events reported	1	**-0,04**												
01. Teamwork within units	3	**0,28****	**0,09****											
02. Organizational learning [D]	-	-	-	-										
03. Nonpunitive response to error	3	**0,41****	**-0,05**	0,30**	-									
04. Staffing	4	**0,54****	**-0,09****	0,24**	-	0,42**								
05. Overall perceptions of safety [D]	-	-	-	-	-	-	-							
06. Supervisor/manager expectations/actions	3	**0,40****	**0,01**	0,30**	-	0,34**	0,31**	-						
07. Frequency of event reporting	3	**0,34****	**0,14****	0,20**	-	0,18**	0,19**	-	0,29**					
08. Feedback and communication about error	4	**0,41****	**0,03**	0,38**	-	0,33**	0,24**	-	0,52**	0,50**				
09. Communication openness [D]	-	-	-	-	-	-	-	-	-	-	-			
10. Hospital management support for patient safety	3	**0,58****	**-0,03**	0,32**	-	0,40**	0,51**	-	0,38**	0,36**	0,47**	-		
11. Teamwork across hospital units	5	**0,42****	**0,00**	0,36**	-	0,40**	0,37**	-	0,31**	0,31**	0,40**	-	0,54**	
12. Hospital handoffs and transitions [D]	-	-	-	-	-	-	-	-	-	-	-	-	-	-

n=974, except for G1.
[+] n=952 for the variable G1: Number of Events Reported.
*P < 0.05; **P < 0.01.
[D] Dimension is removed from the model (collored in dark gray).

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1,2
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2,3
Introduction			**5-6**
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			**6-9**
Study design	4	Present key elements of study design early in the paper	6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-7
Bias	9	Describe any efforts to address potential sources of bias	*NA*
Study size	10	Explain how the study size was arrived at	6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7-9
		(*b*) Describe any methods used to examine subgroups and interactions	NA
		(*c*) Explain how missing data were addressed	7
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	NA
		(*e*) Describe any sensitivity analyses	NA
Results			**9-20**

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	9-10
		(b) Give reasons for non-participation at each stage	-
		(c) Consider use of a flow diagram	-
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	9-10
		(b) Indicate number of participants with missing data for each variable of interest	9-10
Outcome data	15*	Report numbers of outcome events or summary measures	9-12
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	NA
		(*b*) Report category boundaries when continuous variables were categorized	10
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	13-20
Discussion			**20-23**
Key results	18	Summarise key results with reference to study objectives	20-21
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	22
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	21-22
Generalisability	21	Discuss the generalisability (external validity) of the study results	21-22
Other information			**24-25**
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	24-25

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.