BMJ Open

Psychometric properties of the Georgian version of Hospital Survey on Patient Safety Culture: a cross-sectional study

Nikoloz Gambashidze,[1,2] Antje Hammer,[1] Tanja Manser[3]

To cite: Gambashidze N, Hammer A, Manser T. Psychometric properties of the Georgian version of Hospital Survey on Patient Safety Culture: a cross-sectional study. *BMJ Open* 2019;**9**:e030972. doi:10.1136/bmjopen-2019-030972

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2019-030972).

Received 09 April 2019
Revised 03 June 2019
Accepted 05 July 2019

[1]Institute for Patient Safety, University Hospital Bonn, Bonn, Germany
[2]School of Health Sciences and Public Health, University of Georgia, Tbilisi, Georgia
[3]School of Applied Psychology, University of Applied Sciences and Arts Northwestern Switzerland, Olten, Switzerland

Correspondence to
Mr Nikoloz Gambashidze; nikoloz.gambashidze@ukbonn.de

ABSTRACT

Objectives To study the psychometric properties of the Georgian version of the Hospital Survey on Patient Safety Culture (HSPSC-GE).

Design Cross-sectional study.

Setting Three Georgian hospitals.

Participants Staff of participating hospitals (n=579 responses, response rate 41.6%).

Primary and secondary outcome measures Psychometric properties (Model fit, internal consistency, construct validity) of the instrument, factor structure derived from the data.

Results HSPSC-GE demonstrated acceptable construct validity but highly limited internal consistency (Cronbach's alpha 0.35–0.87). Confirmatory factor analysis with the original 12-factor model resulted in poor model fit (root mean square error of approximation (RMSEA)=0.06; standardised root mean square residuals (SRMR)=0.08; comparative fit index (CFI)=0.74; goodness of fit index (GFI)=0.81; Tucker-Lewis Index (TLI)=0.70). Accounting for reversed item bias resulted in improved fit indices. Exploratory factor analysis resulted in an alternative five-factor model including only 19 items, but with satisfactory model fit (RMSEA=0.07; SRMR=0.07; CFI=0.90; GFI=0.89; TLI=0.88).

Conclusions The HSPSC-GE as a whole demonstrated poor psychometric properties. However, a number of dimensions demonstrated acceptable internal consistency and reliability. Our results indicated presence of reversed item bias, which may be inherent to the original instrument design of the HSPSC and should be taken into account while interpreting or comparing results, as well as in analyses of psychometric properties of the instrument. Nevertheless, the HSPSC-GE provides first insights in hospital patient safety culture (PSC) in Georgia and we recommend using it in its full form to facilitate deeper analysis and further development of PSC in Georgian healthcare.

Strengths and limitations of this study

- The first study to validate the Georgian version of the Hospital Survey on Patient Safety Culture (HSPSC), an instrument to identify available strengths in local patient safety climate, and to demonstrate aspects that may require improvement.
- A comprehensive analysis of psychometric properties of the survey instrument, including analysis of the original 12-factor model and an alternative model based on the exploratory factor analysis.
- The analysis of the role of reversed item bias in psychometric evaluation of HSPSC provides additional insight into the instruments' performance.
- Study findings are limited by the study sample, which included general hospitals with n>100 hospital beds, and thus should not be directly generalised to smaller and specialised hospitals.

INTRODUCTION

Patient safety is an essential component of healthcare quality and, in order to improve patient safety, continuously developing the culture of safety is recommended.[1] Patient safety culture (PSC) represents a set of values and beliefs regarding safety, shared within the organisation, and it has been found to be associated with patient outcomes.[2 3] In hospital settings, PSC is mostly measured by means of self-administered questionnaires that typically capture a number of factors associated with PSC such as teamwork and communication, management and leadership, error reporting and organisational learning, and so on. Even though PSC is generally thought to be a multifaceted construct, there is no unified understanding of its composition.[4 5] Thus, various instruments measure slightly different factors. Moreover, studies have shown that performance of the same survey instrument in different settings may vary significantly.[6] Consequently, an increasing number of validation studies of PSC instruments are being conducted in many low/middle-income countries and developed countries, to validate the instruments for further research, and to study and report local expressions of PSC and the various elements it comprised.[5–7]

The Hospital Survey on Patient Safety Culture (HSPSC)[8] is one of the most frequently used instruments for measuring PSC in hospital settings internationally.[7 9] It has been translated into different languages and validated in many countries.[7 9] The HSPSC covers 12 different dimensions of PSC providing a wide spectrum of details useful to measure and improve PSC locally, and to analyse and understand its composition.

To date there are no data available on PSC in Georgian healthcare and no instrument has been adapted and validated for Georgian healthcare. Healthcare services in Georgia are mostly provided by private organisations, with increasing oversight of quality and safety by the state through state-funded programmes (including the Universal Health Care Program) and through regulatory agencies committed to ensuring accessible, safe and high-quality care for all citizens.[10 11] Georgian hospitals are increasingly required to have dedicated personnel and processes for ensuring patient safety and continuous quality improvement.[10 11] However, health services research in Georgia is still very limited, especially in the field of patient safety and safety culture.

With no validated PSC instruments available in Georgia, we aim to validate the Georgian version of the HSPSC (HSPSC-GE), more specifically to explore its psychometric properties and dimensionality. This will provide a foundation for further PSC research in Georgian healthcare. Moreover, studying the local variation of PSC in an emerging, relatively less regulated environment can provide additional insight into the composition of PSC and mechanisms of how it is developing.

METHODS

Setting

This study is based on data from a cross-sectional study *Patient Safety Culture in Georgian Healthcare (PaSCu.Ge)*. Data were gathered in three Georgian hospitals between November 2017 and March 2018. Data gathering in each hospital lasted 1 month. Two follow-up reminders were sent on the 10th and 20th days after initial invitation. Participants were offered either an electronic or a paper-based questionnaire to complete.

Patient and public involvement

Representatives of patient and public groups were not involved in the study design and implementation. Dissemination of study findings includes making the final results publicly available online (in Georgian and in English).

Sample

We included general hospitals with at least 100 hospital beds. All personnel of the participating hospitals, employed for more than 1 month, were invited to participate in the study. Participation was voluntary and anonymous, and all participants provided informed consent before completing the questionnaire.

Measure

The HSPSC[8] is a self-administered questionnaire for capturing the perceptions of hospital employees concerning PSC. The questionnaire consists of 44 items, 42 of which are grouped in 12 dimensions. On a 5-point Likert scale these 42 items measure agreement (from 'strongly disagree' to 'strongly agree') or frequency (from 'never' to 'always'). The remaining two items are the Patient Safety Grade (5-point quality scale from 'Excellent' to 'Failing') and the Number of Events Reported (6-point frequency scale from 'No event reports' to '21 or more event reports'). In addition, we collected demographic information on study participants (ie, profession, gender, tenure in the hospital and within the department).

The original US version of the HSPSC was translated from English into Georgian by a native speaker with more than 10 years of experience with the Georgian healthcare context. Next, the Georgian version was back-translated into English by a professional translator. The discrepancies between the original version and back-translation were discussed by the research team and necessary revisions were made. The revised version was pretested in a group of five local healthcare professionals (healthcare researchers, managers, physicians and nurses). The research team discussed the results of the pretest to establish a final version of the HSPSC-GE. In order to ensure better understandability and acceptability, the final version had some linguistic adaptations (eg, '*It is a pure luck that more serious errors do not happen here*' instead of '*It is just by chance that more serious mistakes don't happen around here*'), as well as minor adaptations to account for structural aspects of the Georgian healthcare system (eg, 'department' instead of 'unit'). However, in order to facilitate comparisons with results of other language versions, we maintained the overall structure and composition of the instrument intact, meaning, that all items from original US version were present in the HSPSC-GE.[8] The final version of the instrument is available on request from the corresponding author.

Analysis

Data processing and preliminary analysis

Twenty-four of the 42 items of the HSPSC-GE are positively worded (eg, '*Staff will freely speak up if they see something that may negatively affect patient care*'), with high scores corresponding to more positive PSC, while the remaining 18 are negatively worded (also called reversed coded items), with higher scores corresponding to less desirable PSC (eg, '*Staff are afraid to ask questions when something does not seem right*').[8] The negatively worded items are unequally presented in different PSC dimensions, ranging from none to all items. For consistency of interpretation, as well as for factor analysis, negatively worded items were reversed coded prior to analysis. After calculating the descriptive statistics of the sample, in order to maintain the high quality of the data, we excluded cases with more than 10% missing answers on the 42 HSPSC-GE items used in the factor analysis. The remaining missing values were imputed using multiple imputations based on the expectation maximisation (EM) algorithm.[12–14]

Before conducting exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), we evaluated Kaiser-Meyer-Olkin (KMO), measure of sampling adequacy (MSA) and Bartlett's test of sampling adequacy. The value >0.7 is desired (>0.9 perfect) for both KMO and MSA, which indicate that a sample of items, and each individual item are respectively adequate for factor analysis.[13 15] A significant p value (<0.05) of Bartlett's test indicates that it is possible to extract more than one factor.[15] We conducted all preliminary and further analyses using SAS V.9.4.

Descriptive statistics

We calculated mean scores for all 12 HSPSC-GE dimensions by averaging the corresponding items. We calculated range, mean and 95% CI for each item and dimension. We calculated the percentage of positive responses of each item and dimension by dividing the number of positive responses (4 and 5) by the total number of all non-missing responses and multiplying this value by 100%[8] and provided 95% CIs. We report percentages of positive scores only as a benchmark for comparisons, as it has been demonstrated that various scoring methods may yield different results.[16] All further analyses were conducted using the Likert scale scores.

Acceptability

We evaluated the acceptability of individual items, dimensions, as well as the complete questionnaire by means of per cent of missing answers. To further study the performance of the instrument, we calculated the floor and ceiling effects (the per cent of lowest and highest available answers, respectively). For PSC dimensions we considered 15% floor or ceiling effect as significant.[17]

Internal consistency

We evaluated the internal consistency of the instrument by calculating Cronbach's alpha for each dimension. Cronbach's alpha ≥0.6 was considered adequate[8] and alpha ≥0.7 good.[13 15] We assessed the internal consistency of the instrument using both the original 12-factor model and the alternative model resulting from the EFA.

Construct validity

We assessed construct validity by calculating Spearman's correlations between dimensions of HSPSC-GE with the single item outcome variable *Patient Safety Grade*. Because these dimensions all measure constructs related to PSC, we expected low to moderate statistically significant positive correlations. However, excessive correlation (>0.90)[13] between PSC dimensions could indicate possible collinearity.[8 13 15] To evaluate item validity, we calculated item-total correlations, expecting moderate to high positive correlations (>0.3),[15] as all the items of the instrument contribute to the common construct of PSC.

Exploratory factor analysis

To investigate the performance of the HSPSC-GE items in details, we conducted EFA and evaluated possible alternative factor structures based on our data. The study sample, stratified by hospitals, was randomly split into 'exploratory' and 'testing' subsamples. The exploratory subsample was used for EFA, and the testing subsample was later used to cross-validate EFA results in the CFA.[13]

In the EFA we used maximum likelihood for factor extraction, with varimax orthogonal prerotation, and promax oblique rotation to aid with interpretation of the factor structure.[13] Factor extraction was based on scree plot and Kaiser criterion (eigenvalues >1). Factor loadings ≥0.4 were considered significant and factor cross-loading <0.4 was considered acceptable.[13 15] We applied these criteria to achieve a satisfactory factor structure based on the exploratory subsample. Next we evaluated the fit of this model to the testing subsample.

Confirmatory factor analysis

We conducted CFA using the complete data set to evaluate the fit of the original 12-factor model with our data. The following indices and respective criteria were considered in the CFA: normed χ^2 (χ^2/df) ≤3.0; comparative fit index >0.90; goodness of fit index (GFI) >0.90; adjusted GFI >0.90; Tucker-Lewis Index/non-normed fit index >0.90; root mean square error of approximation ≤0.08; and standardised root mean square residuals ≤0.07.[13 15]

In the preliminary analysis, as well as in the EFA, we observed divergent performance of positively and negatively worded items. The use of 18 negatively worded items in the instrument may pose an additional reversed item bias,[18] meaning that participants may respond inconsistently to positively and negatively worded items. These inconsistencies in responding may affect the descriptive outcomes of the study (mean and 95% CI), and change the interitem associations (eg, correlations) and thus alter results of the CFA. To check for the presence of reversed item bias, we added separate method factors with effects on the positively or negatively worded items,[18] and tested the fit of this extended model to our data in CFA.

Lastly, we conducted CFA using the 'testing' subsample to evaluate the fit of the EFA-based model.

RESULTS

Study sample and descriptive statistics

We collected 579 questionnaires from three hospitals with an estimated total of 1391 employees, resulting in a response rate of 41.4%. Response rates in the three participating hospitals ranged from 33.7% to 50.1%. All participants chose the paper version of the questionnaire rather than using the online version. By profession, our sample was divided into three equal groups—physicians (32.5%), nurses (31.4%) and other clinical and non-clinical personnel (33.5%), all three groups being predominantly female with 61.2%, 94.5% and 85.6%, respectively. Having managerial functions was reported by 22.1% of participants, 30.5% of these were male (considerably higher compared with 18.0% in the overall sample). Descriptive characteristics of the sample are presented in table 1.

Among the 42 items included in the factor analysis, the average missing answer was 2.19%, with a maximum of

Table 1 Demographic characteristics of the study sample

Characteristics	n	%
Gender		
Male	104	18.0
Female	458	79.1
Missing	17	2.9
Profession		
Physician	188	32.5
Nurse	182	31.4
Other	194	33.5
Missing	15	2.6
Contact with patients		
Yes	459	79.3
No	101	17.4
Missing	19	3.3
Managerial functions		
Yes	128	22.1
No	412	71.2
Missing	39	6.7
Average working hours per week		
<20	20	3.5
20–39	106	18.3
40–59	336	58.0
60–79	71	12.3
80–99	18	3.1
100+	18	3.1
Missing	10	1.7
Years in the hospital		
<1	43	7.4
1–5	392	67.7
6–10	45	7.8
11–15	24	4.1
16–20	12	2.1
21+	51	8.8
Missing	12	2.1
Total sample	579	100.0

4.66% on C4 ('Staff feel free to question the decisions or actions of those with more authority'). The single item G1 (Number of Events Reported) had the highest number of missing answers with 6.56%. Most dimensions demonstrated a ceiling effect >15%, which indicates that the instrument may not be able to differentiate effectively at the high end of the construct. We did not observe the floor effect >15% in any of the dimensions. Missing answers as well as mean values and percentage of positive responses, as well as corresponding CIs for 12 dimensions, respective 42 items and 2 additional single items are presented in table 2.

After removing 21 cases with more than 10% missing answers on HSPSC-GE items 558 cases remained for imputation using multiple imputations based on the EM algorithm.

The KMO test resulted in an appropriate value of 0.84, with MSA for the items varying between 0.64 and 0.92; together with a highly significant Bartlett's test (p<0.0001), indicating that the sample was adequate for factor analysis.

Internal consistency

Only four dimensions (O2, H1, H3 and U1) demonstrated acceptable (α>0.60) to good (α>0.70) internal consistency (table 2). The remaining eight dimensions had low scores, with four dimensions (O1, U4, U5 and U7) having Cronbach's alpha scores <0.50, demonstrating extremely poor internal consistency.

Construct validity

Most dimensions demonstrated statistically significant positive correlations with other dimensions of the instrument, as well as with the single item *Patient Safety Grade*. The exception was the dimension *Staffing* (U4), which was not correlated with the single item *Patient Safety Grade*, had limited or no correlation with many other PSC dimensions and was negatively correlated with two dimensions, *Organisational Learning—Continuous Improvement* and *Communication Openness* (U2 and U7, respectively). None of the correlations were higher than 0.90, indicating that there was no collinearity between dimensions. All correlations are presented in online supplementary appendix 1. Most items had standardised item-total correlations >0.3, indicating that these items represent a common construct (ie, PSC). The three items with lowest item-total correlation were A5 ('Staff in this unit work longer hours than is best for patient care', α=−0.03), A14 ('We work in 'crisis mode' trying to do too much, too quickly', α=0.03) and B3 ('Whenever pressure builds up, my supervisor/manager wants us to work faster, even if it means taking shortcuts', α=0.08). All three were negatively worded items.

Exploring dimensions of HSPSC-GE

By conducting EFA with the exploratory sub sample (n=279) and gradually eliminating items with factor loadings <0.40 and with factor cross-loadings >0.4, 23 items were removed from the model, leading to a five factor model with 19 items (see table 3). For four original dimensions (O1, U3, U4 and U6) all items had to beremoved from the model. The negatively worded items from the three hospital-level dimensions (H1, H2 and H3) merged to form one new dimension, *Hospital-wide cooperation and support* (table 3, factor 1). Four positively worded items from three dimensions (U2, U7 and H2) formed one new dimension, *Staff's active role in promoting patient safety* (table 3, factor 2). The two negatively worded items (B3 and B4) were removed from the model leaving the dimension *Supervisor/Manager Expectations and ActionsPromoting Patient Safety* (U5) with only two items (table 3,

Table 2 HSPSC-GE dimensions and items; missing answers, mean scores and 95% CI, per cent of positive responses and corresponding 95% CI (n=579)

Dimensions/items (Cronbach's alpha)	Missing answers (%)*	Floor effect (%)†	Ceiling effect (%)‡	Mean score (±CI)	Per cent of positive responses (±CI)
Three hospital-level dimensions (H1–H3)					
H1—Management support for patient safety (α=0.65)	1.04	0.86	34.89	4.08 (±0.08)	72.74 (±2.73)
F1. Hospital management provides a work climate that promotes patient safety.	1.55	3.80	64.94	4.35 (±0.09)	82.46 (±3.13)
F8. The actions of hospital management show that patient safety is a top priority.	2.59	7.25	53.71	4.04 (±0.11)	72.52 (±3.69)
F9. Hospital management seems interested in patient safety only after an adverse event happens. (N)	3.11	7.25	47.50	3.83 (±0.11)	63.10 (±4.00)
H2—Teamwork across units (α=0.54)	1.04	0.17	22.45	3.99 (±0.07)	69.94 (±2.40)
F2. Hospital units do not coordinate well with each other. (N)	2.94	9.67	46.46	3.75 (±0.12)	63.35 (±3.99)
F4. There is good cooperation among hospital units that need to work together.	2.42	8.64	51.47	3.94 (±0.11)	68.85 (±3.82)
F6. It is often unpleasant to work with staff from other hospital units. (N)	2.59	4.32	47.15	3.96 (±0.10)	68.44 (±3.84)
F10. Hospital units work well together to provide the best care for patients.	2.07	3.45	63.56	4.32 (±0.09)	79.72 (±3.31)
H3—Handoffs and transitions (α=0.73)	1.73	0.17	25.91	3.95 (±0.08)	66.65 (±2.76)
F3. Things 'fall between the cracks' when transferring patients from one unit to another. (N)	2.25	4.15	47.50	3.91 (±0.11)	67.67 (±3.86)
F5. Important patient care information is often lost during shift changes. (N)	2.76	3.97	55.96	4.10 (±0.10)	72.65 (±3.69)
F7. Problems often occur in the exchange of information across hospital units. (N)	2.59	5.01	34.54	3.51 (±0.11)	50.18 (±4.13)
F11. Shift changes are problematic for patients in this hospital. (N)	1.90	3.28	66.32	4.27 (±0.10)	76.23 (±3.50)
Seven unit-level dimensions (U1–U7)					
U1—Teamwork within units (α=0.70)	0.17	0.17	35.92	4.37 (±0.06)	84.95 (±1.87)
A1. People support one another in this unit.	1.55	2.59	65.11	4.45 (±0.08)	88.07 (±2.66)
A3. When a lot of work needs to be done quickly, we work together as a team to get the work done.	0.86	3.11	73.75	4.57 (±0.07)	90.94 (±2.35)
A4. In this unit, people treat each other with respect.	1.90	2.59	66.32	4.44 (±0.08)	86.27 (±2.83)
A11. When one area in this unit gets really busy, others help out.	2.07	10.19	53.89	4.02 (±0.11)	74.25 (±3.60)
U2—Organisational learning—continuous improvement (α=0.58)	0.86	0.00	23.66	3.93 (±0.08)	68.14 (±2.74)
A6. We are actively doing things to improve patient safety.	1.55	1.55	73.75	4.45 (±0.09)	82.81 (±3.10)
A9. Mistakes have led to positive changes here.	2.94	12.61	33.33	3.58 (±0.11)	56.05 (±4.11)

Continued

Table 2 Continued

Dimensions/items (Cronbach's alpha)	Missing answers (%)*	Floor effect (%)†	Ceiling effect (%)‡	Mean score (±CI)	Per cent of positive responses (±CI)
A13. After we make changes to improve patient safety, we evaluate their effectiveness.	1.90	17.79	50.43	3.73 (±0.13)	64.79 (±3.93)
U3—Non-punitive response to error (α=0.59)	1.38	2.59	12.95	3.40 (±0.09)	49.21 (±2.86)
A8. Staff feel like their mistakes are held against them. (N)	1.90	15.54	28.84	3.14 (±0.12)	40.14 (±4.03)
A12. When an event is reported, it feels like the person is being written up, not the problem.(N)	2.42	11.74	46.46	3.71 (±0.12)	61.95 (±4.01)
A16. Staff worry that mistakes they make are kept in their personnel file. (N)	2.25	13.99	29.19	3.33 (±0.12)	45.05 (±4.10)
U4—Staffing (α=0.45)	0.69	0.00	3.63	3.34 (±0.08)	53.68 (±2.44)
A2. We have enough staff to handle the workload.	1.04	11.40	51.47	3.96 (±0.11)	75.92 (±3.50)
A5. Staff in this unit work longer hours than is best for patient care. (N)	2.94	28.15	29.88	3.01 (±0.13)	42.53 (±4.09)
A7. We use more agency/temporary staff than is best for patient care. (N)	3.28	10.36	42.31	3.61 (±0.12)	54.64 (±4.13)
A14. We work in 'crisis mode' trying to do too much, too quickly. (N)	2.07	33.85	16.58	2.72 (±0.13)	40.21 (±4.04)
U5—Supervisor/manager expectations and actions promoting patient safety (α=0.41)	0.35	0.35	17.96	4.09 (±0.06)	74.13 (±1.99)
B1. My supervisor/manager says a good word when he/she sees a job done according to established patient safety procedures.	0.86	6.04	52.33	4.18 (±0.09)	80.49 (±3.24)
B2. My supervisor/manager seriously considers staff suggestions for improving patient safety.	1.55	3.11	53.89	4.18 (±0.09)	72.11 (±3.68)
B3. Whenever pressure builds up, my supervisor/manager wants us to work faster, even if it means taking shortcuts. (N)	2.25	18.31	40.93	3.46 (±0.13)	55.65 (±4.10)
B4. My supervisor/manager overlooks patient safety problems that happen over and over. (N)	1.55	5.18	77.20	4.51 (±0.09)	87.02 (±2.76)
U6—Feedback and communication about error (α=0.57)	0.52	0.52	27.81	4.08 (±0.07)	71.72 (±2.62)
C1. We are given feedback about changes put into place based on event reports.	2.42	3.63	49.91	4.05 (±0.10)	69.91 (±3.79)
C3. We are informed about errors that happen in this unit.	3.80	3.63	44.73	3.98 (±0.10)	68.04 (±3.88)
C5. In this unit, we discuss ways to prevent errors from happening again.	2.25	3.80	57.51	4.20 (±0.09)	76.33 (±3.51)
U7—Communication openness (α=0.35)	1.04	0.52	9.33	3.51 (±0.07)	55.51 (±2.52)
C2. Staff will freely speak up if they see something that may negatively affect patient care.	2.07	7.25	46.11	3.86 (±0.11)	66.14 (±3.90)

Continued

 Gambashidze N, *et al*. *BMJ Open* 2019;**9**:e030972. doi:10.1136/bmjopen-2019-030972

Table 2 Continued

Dimensions/items (Cronbach's alpha)	Missing answers (%)*	Floor effect (%)†	Ceiling effect (%)‡	Mean score (±CI)	Per cent of positive responses (±CI)
C4. Staff feel free to question the decisions or actions of those with more authority.	4.66	25.22	14.51	2.70 (±0.12)	31.88 (±3.89)
C6. Staff are afraid to ask questions when something does not seem right. (N)	1.55	6.74	45.94	3.92 (±0.10)	66.67 (±3.87)
Two outcome dimensions (O1–O2)					
O1—Overall perceptions of patient safety (α=0.40)	1.04	0.17	21.24	3.94 (±0.07)	69.34 (±2.25)
A10. It is just by chance that more serious mistakes do not happen around here. (N)	2.59	9.84	54.92	3.95 (±0.12)	68.79 (±3.83)
A15. Patient safety is never sacrificed to get more work done.	2.76	10.02	57.17	4.15 (±0.11)	79.40 (±3.34)
A17. We have patient safety problems in this unit. (N)	1.90	12.95	50.95	3.77 (±0.12)	62.50 (±3.98)
A18. Our procedures and systems are good at preventing errors from happening.	1.38	8.46	47.84	3.88 (±0.11)	67.08 (±3.86)
O2—Frequency of events reported (α=0.87)	0.35	4.66	21.07	3.39 (±0.10)	47.21 (±3.54)
D1. When a mistake is made, but is caught and corrected before affecting the patient, how often is this reported?	0.69	9.84	30.05	3.34 (±0.11)	46.26 (±4.08)
D2. When a mistake is made, but has no potential to harm the patient, how often is this reported?	1.90	10.36	24.70	3.23 (±0.11)	40.49 (±4.04)
D3. When a mistake is made that could harm the patient, but does not, how often is this reported?	2.42	9.84	40.07	3.61 (±0.12)	55.04 (±4.11)
Two single item outcomes (E1, G1)					
E1. Patient safety grade	0.52	0.17	11.74	3.64 (±0.06)	54.69 (±4.07)
G1. Number of events reported	6.56	78.07	1.04	NA	16.45 (±3.13)§

(N) denotes negatively worded items; total sample n=579.

*Percentage of missing answers before imputation.
†Percentage of participants indicating lowest answer category.
‡Percentage of participants indicating highest answer category.
§Percentage of participants reporting one or more errors in the past 12 months.
HSPSC-GE, Georgian version of the Hospital Survey on Patient Safety Culture; NA, not applicable.

factor 5). Two dimensions, *Frequency of Events Reported* and *Teamwork within Units* (O2 and U1), were independently present in the model (table 3, factors 3 and 4).

Fit of the data with different factor models

CFA of the EFA-based five-factor model using the testing subsample (n=279) resulted in acceptable fit indices. In contrast, CFA of the 12-factor model with the complete sample (n=558) resulted in poor model fit. Next, to account for the item wording, we extended the model with additional method factors for negatively worded and positively worded items, which improved the model fit. The results of the three CFAs are presented in table 4.

DISCUSSION

In this study, we evaluated the psychometric properties of the HSPSC-GE. The original 12-factor model demonstrated poor fit with our data and internal consistency of many dimensions was not satisfactory. We were also able to show that parts of the instrument are relatively stable and demonstrate acceptable psychometric properties.

In our study, 8 out of 12 dimensions of PSC showed poor internal consistency. Four of these, namely *Overall Perceptions of Patient Safety*, *Staffing*, *Non-punitive Response to Error* and *Feedback and Communication about Error*, were completely eliminated during EFA. Other validation

Table 3 Rotated factor structure of the five-factor model resulting from the EFA

Factor (α)/item	Factor loadings
Factor 1: Hospital-wide cooperation and support (α=0.79)	
F2. Hospital units do not coordinate well with each other. (N)	0.55
F3. Things 'fall between the cracks' when transferring patients from one unit to another. (N)	0.64
F5. Important patient care information is often lost during shift changes. (N)	0.67
F6. It is often unpleasant to work with staff from other hospital units. (N)	0.57
F7. Problems often occur in the exchange of information across hospital units. (N)	0.57
F9. Hospital management seems interested in patient safety only after an adverse event happens. (N)	0.52
F11. Shift changes are problematic for patients in this hospital. (N)	0.58
Factor 2: Staff's active role in promoting patient safety (α=0.77)	
A6. We are actively doing things to improve patient safety.	0.62
A13. After we make changes to improve patient safety, we evaluate their effectiveness.	0.66
C2. Staff will freely speak up if they see something that may negatively affect patient care.	0.73
F4. There is good cooperation among hospital units that need to work together.	0.69
Factor 3: Frequency of events reported (α=0.87)	
D1. When a mistake is made, but is caught and corrected before affecting the patient, how often is this reported?	0.86
D2. When a mistake is made, but has no potential to harm the patient, how often is this reported?	0.89
D3. When a mistake is made that could harm the patient, but does not, how often is this reported?	0.70
Factor 4: Teamwork within units (α=0.71)	
A1. People support one another in this unit.	0.86
A3. When a lot of work needs to be done quickly, we work together as a team to get the work done.	0.51
A4. In this unit, people treat each other with respect.	0.74
Factor 5: Supervisor/manager expectations and actions promoting patient safety (α=0.65)	
B1. My supervisor/manager says a good word when he/she sees a job done according to established patient safety procedures.	0.52
B2. My supervisor/manager seriously considers staff suggestions for improving patient safety.	0.94

The table demonstrates standardised regression coefficients for items remaining in the model.
Underlined denotes new dimensions that were not part of original 12-factor model.
(N) denotes negatively worded items.
EFA, exploratory factor analysis.

studies have found similar problems with the dimensions *Overall Perceptions of Safety*[6 9 19–21] and *Hospital Management Support for Patient Safety* and *Staffing*.[19] Also, the dimensions *Communication Openness—Continuous Learning* and *Feedback and Communication about Error* often merge together into one common factor.[6 9 21–26] These dimensions may be particularly unstable in translated versions, indicating the need for improvement in the item set and/or wording to support international use of the instrument.

Our results demonstrate that study participants responded differently to positively and negatively worded items. In general, negatively worded items had lower mean values and percentages of positive responses, compared with positively worded items. In the alternative five-factor model, disproportionately more negatively worded items were eliminated. Moreover, in our EFA-based model all five dimensions consisted either entirely of positively or negatively worded items. Finally, our extended model that accounted for the reversed item bias resulted in better fit indices, demonstrating that at least part of the variance in our data can be explained by direction of item wording. These results may suggest that study participants perceive and interpret positively and negatively worded items differently. HSPSC-GE has the same item composition and wording as the original US version,[8] and so it may be reasonable to suggest that the reversed item bias is an inherent part to the original instrument design, rather than a feature of the local version. As such, it may affect other language versions of the instruments as well. Similarly, significant effect of item wording on per cent of positive scores was reported by the experimental study using HSPSC,[27] where control group was asked to fill in the 19 items from HSPSC with original wording, while the wording of the same items was reversed for the study group. The authors concluded that the wording may affect the outcomes, and, to facilitate

Table 4 Indices of confirmatory factor analyses using the original 12-factor model, the EFA-based five-factor model and additional method factors

Model fit indices in CFA	Criteria for good model fit*	Original 12-factor model †	Original model ‡ extended with method factors	EFA-based five-factor model §
Sample size	NA	558	558	279¶
Number of factors	NA	12	12	¶
χ^2/df	<3.00	3.3	2.8	2.2
Root mean square error of approximation (RMSEA)	<0.08	0.065	0.057	0.065
Standardised root mean square residuals (SRMR)	<0.07	0.081	0.070	0.068
Goodness of fit index (GFI)	>0.90	0.81	0.85	0.89
Adjusted GFI	>0.90	0.77	0.82	0.86
Normed fit index	≥0.95	0.67	0.73	0.84
Comparative fit index ≥0.90	≥0.90	0.74	0.80	0.90
Tucker-Lewis Index/non-normed fit index non-normed fit index	≥0.90	0.70	0.77	0.88

*Model fits in accordance with Hair *et al*.[13]
†All 12 dimensions of the original model (H1–H3, U1–U7, O1–O2).
‡Original 12-factor model, extended with method factors for positively and negatively worded items.
§EFA-based five-factor model (19 items from dimensions O2, H1, H2, H3, U1, U2, U5 and U7).
¶Testing subsample.
CFA, confirmatory factor analysis; EFA, exploratory factor analysis; NA, not applicable.

reliable measurement of various components of PSC, they argue for balancing out the number of positively and negatively worded items within all dimensions. Studies using the HSPSC frequently report less positive results for dimensions with predominantly negatively worded items (*Non-punitive Response to Error, Staffing* and *Hospital Handoffs and Transitions*).[7 28 29] Although these dimensions may represent truly problematic aspects of hospital safety culture, rather lower scores may be at least partially explained by the reversed item bias (reduced scores on negatively worded items) coupled with unequal presence of negatively worded items in PSC dimensions. Therefore, this method bias should be taken into account when using the HSPSC, while interpreting and comparing the results, as well as in factor analyses.

Relatively limited internal consistency and construct validity in our results may be partially due to characteristics of the study population, and not just by properties of the instrument. Specifically, because the concept of PSC is relatively new for Georgian healthcare, participants might find it difficult to associate certain ideas or behaviours with common constructs. This can be addressed with targeted educational activities and trainings, familiarising healthcare personnel with relevant concepts. Additionally, we observed considerable ceiling effect in most PSC dimensions, indicating a grouping of the results on the highest response category. One could speculate on different factors 'pushing' the results towards the positive end. This might be factors associated with study method, like social desirability bias,[30] but also factors associated with participants, like, for example, fear of retribution or possibly lower expectations regarding patient safety-related issues. The factors associated with the sample might be mitigated through education and training. The same analysis using a sample of participants with a more structured and somewhat shared understanding of concepts of PSC could result in better properties of the instrument. This should be considered in further investigations on safety culture in Georgian hospitals.

One of the purposes of PSC assessment is to compare the results between different settings (unit/team, hospitals, healthcare systems) or time periods (monitoring the change over time). In order to support such comparisons, a common measurement instrument should be used, which has adequate psychometric properties for all settings. Although our results demonstrate considerably limited psychometric properties of the HSPSC-GE and that some dimensions with extremely limited internal consistency should be interpreted with caution, we still argue against significant changes in the factor structure and item composition. Several arguments can support this claim: (1) First, the psychometric properties, including the dimensionality of the HSPSC-GE may change in time with the evolution of the field of PSC in Georgian healthcare, as the study participants will have increasingly shared understanding and perception of PSC in their organisations. Exposure of study participants to the internationally shared concepts may also facilitate this process. (2) Using the common item set will ensure continuous collection of local data on a common spectrum of relevant items for future analysis, and the ability to compare results with studies from other developed countries and low/middle-income countries. (3) Because problems with some dimensions and items are not unique to our study, but reported rather frequently in validation studies in different languages, the instrument needs to be improved on a larger, international level. (4) And finally, even though the dimensionality of the instrument, as well as its understanding by the participants may vary, the individual items of the instrument are still relevant for the field of PSC and thus should be monitored further.

Limitations

This is the first study validating an established PSC instrument in Georgian healthcare. While we used a large data set from three hospitals, our findings are limited by the study sample which included only general hospitals with more than 100 hospital beds and should not be generalised to smaller or specialised hospitals in Georgia. The generalisability of our results may be also limited by the modest response rate, which however is comparable to similar studies. In 2018, the Agency for Healthcare Research and Quality comparative database report[29] reported average response rate of 56%, ranging from 12% to 100%. According to a recent review, response rate in comparable validation studies from other countries may go as low as 23%.[7] Moreover, the poor performance of the instrument in our sample might be bound to the language version we developed for this study, and this should be taken into consideration in further investigations of safety culture in Georgia. Lastly, we were not able to include the association of PSC with objective patient outcomes of the hospitals (external validity). However, there is growing evidence supporting the positive correlation between PSC and various outcome variables.[2 3]

CONCLUSIONS

HSPSC-GE demonstrated poor psychometric properties, and many items and dimensions may need to be further developed. However, parts of the instrument demonstrated sufficient internal consistency and acceptable reliability to be used in studies of PSC in Georgian hospitals. We were able to demonstrate that negatively worded items may be prone to reversed item bias, which may have an effect on the mean values, as well as on dimensionality of the instrument. It is likely that this effect is inherent in the HSPSC design, and so should be accounted for when interpreting and comparing results, and when analysing the psychometric properties of any language version. Since the problems we found with various dimensions are not unique for our sample, improvement of the instrument should be done on the global, not a local scale. Meanwhile, HSPSC-GE extracts necessary information for developing PSC in healthcare organisations, and we recommend using it in its full form to facilitate further analysis of results and development of the field.

Acknowledgements We thank all the study participants. We acknowledge the support of the hospital management, the efforts of study coordinators and cooperation of the hospital employees to find time to fill in the surveys.

Contributors Data were collected by the local study coordinators in the participating hospitals under the supervision of NG. Data analysis was carried out by NG. The manuscript was prepared by NG in collaboration with TM and AH. The final version of the manuscript has been approved by all the authors.

Funding This work was supported by the German Academic Exchange Service (Funding programme: 57214224 to NG); and the BONFOR programme of the University Hospital Bonn (grant number: 2017-5-01).

Competing interests None declared.

Ethics approval The study was approved by the ethics committees of the University Hospital Bonn (No 117/17) and the University of Georgia (No 05-13779).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data from the PaSCu.Ge study will not be made available in the public domain to preserve data security. Data will be stored in accordance with national and regional data security standards.

REFERENCES

1. Groves PS, Meisenbach RJ, Scott-Cawiezell J. Keeping patients safe in healthcare organizations: a structuration theory of safety culture. *J Adv Nurs* 2011;67:1846–55.
2. Singer S, Lin S, Falwell A, *et al*. Relationship of safety climate and safety performance in hospitals. *Health Serv Res* 2009;44:399–421.
3. DiCuccio MH. The relationship between patient safety culture and patient outcomes. *J Patient Saf* 2015;11:135–42.
4. Pumar-Méndez MJ, Attree M, Wakefield A. Methodological aspects in the assessment of safety culture in the hospital setting: a review of the literature. *Nurse Educ Today* 2014;34:162–70.
5. Alsalem G, Bowie P, Morrison J. Assessing safety climate in acute hospital settings: a systematic review of the adequacy of the psychometric properties of survey measurement tools. *BMC Health Serv Res* 2018;18:353.
6. Gambashidze N, Hammer A, Brösterhaus M, *et al*. Evaluation of psychometric properties of the German Hospital survey on patient safety culture and its potential for cross-cultural comparisons: a cross-sectional study. *BMJ Open* 2017;7:e018366.
7. Reis CT, Paiva SG, Sousa P. The patient safety culture: a systematic review by characteristics of hospital survey on patient safety culture dimensions. *Int J Qual Health Care* 2018;30:660–77.
8. 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: Agency for Healthcare Research and Quality, 2004.
9. Hammer A, Manser T. The Use of the Hospital Survey on Patient Safety Culture in Europe. In: *"Patient Safety Culture: Theory, methods and application" Waterson, ed*. Achgate Publishing, 2014.
10. Government of Georgia. *ORDINANCE No 724, On Approval of the 2014-2020 State Concept of Healthcare System of Georgia for "Universal Health Care and Quality Control for the Protection of Patients' Rights"*, 2014.
11. The healthcare and social issues Committee of the Parliament of Georgia. *Vision for developing the healthcare system in Georgia by 2030*. Tbilisi, Georgia, 2017.
12. Wirtz M. On the problem of missing data: how to identify and reduce the impact of missing data on findings of data analysis]. *Rehabilitation* 2004;43:109–15.
13. Hair JF, Black WC, Babin BJ, *et al*. *Multivariate data analysis*. Upper Saddle River, N.J: Pearson Education, 2014.
14. Boussat B, François O, Viotti J, *et al*. Managing missing data in the hospital survey on patient safety culture. *J Patient Saf* 2019:1.
15. Field A. *Discovering statistics using SPSS: (and sex and drugs and rock'n'roll)*. 3rd edn. Los Angeles: SAGE, 2012.
16. Giai J, Boussat B, Occelli P, *et al*. Hospital survey on patient safety culture (HSOPS): variability of scoring strategies. *Int J Qual Health Care* 2017;29:685–92.

17. Terwee CB, Bot SDM, de Boer MR, *et al*. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
18. Weijters B, Baumgartner H, Schillewaert N. Reversed item bias: an integrative model. *Psychol Methods* 2013;18:320–34.
19. Waterson P, Griffiths P, Stride C, *et al*. Psychometric properties of the hospital survey on patient safety culture: findings from the UK. *BMJ Qual Saf* 2010;19:e2.
20. 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.
21. Tereanu C, Smith SA, Ghelase MS, *et al*. Psychometric properties of the Romanian version of the hospital survey on patient safety culture (HSOPS). *Maedica* 2018;13:34–43.
22. Sarac C, Flin R, Mearns K, *et al*. Hospital survey on patient safety culture: psychometric analysis on a Scottish sample. *BMJ Qual Saf* 2011;20:842–8.
23. Robida A. Hospital survey on patient safety culture in Slovenia: a psychometric evaluation. *International Journal for Quality in Health Care* 2013;25:469–75.
24. 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.
25. 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.
26. 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.
27. Moghri J, Akbari Sari A, Yousefi M, *et al*. Is scores derived from the most internationally applied patient safety culture assessment tool correct? *Iran J Public Health* 2013;42:1058–66.
28. Blegen MA, Gearhart S, O'Brien R, *et al*. AHRQ's Hospital survey on patient safety culture: psychometric analyses. *J Patient Saf* 2009;5:139–44.
29. Famolaro T, Yount N, Hare R, *et al*. Hospital survey on patient safety culture: 2018 user database report. *Part I. Rockville, MD* 2018.
30. Freeth D, Sandall J, Allan T, *et al*. A methodological study to compare survey-based and observation-based evaluations of organisational and safety cultures and then compare both approaches with markers of the quality of care. *Health Technol Assess* 2012;16:1–184.