Published in final edited form as:

J Am Med Dir Assoc. 2013 August ; 14(8): 616–622. doi:10.1016/j.jamda.2013.03.016.

First evidence on the validity and reliability of the Safety Organizing Scale-Nursing home version (SOS-NH)

Dietmar Ausserhofer, PhD, RN,

Institute of Nursing Science, University of Basel, Basel, Switzerland, Bernoulli street 28, 4056 Basel, Switzerland, Tel.: +41 (0)61 269 07 54

Ruth A. Anderson, PhD, RN, FAAN,

Duke University School of Nursing, DUMC 3322, Durham, North Carolina, 27710 USA, Tel: +1 (919) 668-4599;

Cathleen Colón-Emeric, MD, MHSc, and

Durham Veteran Affairs Medical Center, GRECC, 508 Fulton St., Durham, North Carolina, 27710 USA, Tel: +1 (919) 286-0411 x6777

René Schwendimann, PhD, RN

Institute of Nursing Science, University of Basel, Bernoulli street 28, 4056 Basel, Switzerland, Tel.: +41 (0)61 269 07 19

Dietmar Ausserhofer: dietmar.ausserhofer@unibas.ch; Ruth A. Anderson: ruth.anderson@duke.edu; Cathleen Colón-Emeric: Cathleen.Colon-Emeric@va.gov; René Schwendimann: rene.schwendimann@unibas.ch

Abstract

Background—The Safety Organizing Scale is a valid and reliable measure on safety behaviors and practices in hospitals.

Purpose of the study—This study aimed to explore the psychometric properties of the Safety Organizing Scale-Nursing Home version (SOS-NH).

Design and Methods—In a cross-sectional analysis of staff survey data, we examined validity and reliability of the 9-item Safety SOS-NH using American Educational Research Association guidelines.

Subjects and Setting—This sub-study of a larger trial used baseline survey data collected from staff members (n=627) in a variety of work roles in 13 NHs in North Carolina and Virginia, USA.

Results—Psychometric evaluation of the SOS-NH revealed good response patterns with low average of missing values across all items (3.05%). Analyses of the SOS-NH's internal structure (e.g., comparative fit indices = 0.929, standardized root mean square error of approximation = 0.045) and consistency (composite reliability = 0.94) suggested its one-dimensionality. Significant between-facility variability, intraclass correlations, within-group agreement and design effect confirmed appropriateness of the SOS-NH for measurement at the NH level, justifying data

© 2013 American Medical Directors Association. Published by Elsevier Inc. All rights reserved.

Funding/potential competing interests

None of the authors has a potential conflict of interest regarding this sub-study.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

aggregation. The SOS-NH showed discriminate validity from one related concept, communication openness.

Implications—Initial evidence regarding validity and reliability of the SOS-NH supports its utility in measuring safety behaviors and practices among a wide range of NH staff members, including those with low literacy. Further psychometric evaluation should focus on testing concurrent and criterion validity, using resident outcome measures (e.g., patient fall rates).

Keywords

Resident safety climate; safety behaviors; nursing homes; psychometric evaluation

Promoting patient safety is a challenge for healthcare organizations, including nursing homes (NHs).^{1–3} Adverse drug events,⁴⁵ falls,⁶ pressure ulcers,⁷ urinary tract infections,⁸ physical restraint use,⁹ and weight loss,¹⁰ are common safety issues in long term care settings. Older NH residents are especially vulnerable to adverse clinical events due to reduced functional status, multiple co-morbidities, polypharmacy, and cognitive impairment. As in acute-care, “safety culture” in NHs is a potential determinant of patient safety and quality of care.¹¹¹² Principles of a safety culture include (1) shared beliefs that healthcare is a high-risk industry; (2) errors are opportunities to improve the system and prevent harm; and (3) behavioral choices leading to errors are managed in an open, fair and transparent way.⁶ However, a safety culture must be “lived;” all organizational members must engage in safety behaviors¹³. Such behaviors are visible and observable safety culture features that reflect the so-called “safety climate”^{14–16} and can be assessed using questionnaires.¹⁷¹⁸

Despite the number of studies examining safety climate in acute-care, research on safety climate in NHs is scarce.^{1119–23} Instruments have been developed and tested in hospitals,²⁴²⁵ but only the Agency for Healthcare Research and Quality (AHRQ) Hospital Survey on Patient Safety Culture²⁶ has been studied in NHs.¹⁹²⁰ AHRQ’s instrument has 42 items reflecting 12 dimensions (e.g. communication, teamwork, management support).²⁶ Studies in hospitals provided evidence of the internal structure and consistency,²⁷²⁸ but its reliability has not been confirmed in NH settings. A study testing the psychometric properties of the Patient Safety Culture survey in a Swiss NH sample failed to confirm the internal structure of the translated version; different factor structures were found raising questions and concerns about the meaningfulness, validity and reliability of this multidimensional measure in NHs.²⁹

A new, promising tool is the uni-dimensional Safety Organizing Scale (SOS). While the AHRQ instrument reflects safety attitudes, the SOS measures safety behaviors reflecting “collective mindfulness” in healthcare organizations,¹³ a patient safety climate feature that might facilitate high-quality performance over time.³⁰ Collective mindfulness is based on research in “high-reliability organizations,” such as nuclear power plants, and emphasizes 5 principles that allow frontline employees to stay aware of and immediately address potential adverse events: tracking small failures, resisting over simplification, remaining sensitive to operations, maintaining capabilities for resilience and taking advantage of shifting locations of expertise.³⁰³¹ The items in SOS capture these five principles. On a 7-point Likert scale (“not at all” = 1 to “a very great extent” = 7) staff members assess the extent they and their colleagues engaged in nine safety behaviors and practices (Table 2); it has good psychometric properties in acute-care making it a strong candidate for use in NHs.¹³³²³³ Thus, replicating the SOS’s evidence on validity and reliability in NHs would increase the generalizability of high-reliability organizational theory in health care.³⁴

NH managers and researchers interested in improving resident safety need brief, valid and reliable safety climate measures.²⁰³⁵ The AHRQ questionnaire has not been validated in

NHs and it includes multiple sub-dimensions with a large number of items; its' length might make it impractical for the regular monitoring of patient safety climate in clinical practice, resulting in low response rates and missing data. In the NH setting with a wide range of workers, including employees with low literacy, the SOS might be a short and comprehensible instrument with low survey burden. Also, understanding safety behaviors using the SOS, rather than attitudes using the AHRQ questionnaire, provides a new lever for change because in the end, improving safety is about behavior change. With this study we aimed to explore the psychometric properties of the SOS by testing its reliability and validity in NHs.

Methods

Design

This sub-study used survey data collected at baseline during pilot testing of a randomized trial of a staff education interventions for fall prevention a (Clinical Trials.gov: NCT00636675; NCT00836433) conducted between 2009–2012. Detailed methods of the parent study, which aimed to reduce falls in NHs using a multicomponent approach (e.g. relationship networks for creative problem solving, mentorship to sustain newly acquired behaviors), have been published else where.³⁶

Setting and sample

For this sub-study we used baseline data from the first 13 participating NHs. This sampling frame was chosen because it contained an adequate staff sample size to estimate the psychometric properties with sufficient precision, and also included a broad range of NH types to enhance generalizability. All 69 NH in participating in Medicare and Medicaid within a 100-mile radius of [blinded] were eligible for participation in the larger studies and were contacted in random order to recruit nine NHs. Also, four Veterans Administration Community Living Centers (heretofore referred to as nursing homes) were recruited from North Carolina and Virginia. Given the interdisciplinary focus of the intervention, all staff members in clinical roles (e.g., registered nurses, licensed practice nurses, social workers, etc.) and support staff that had resident contact (e.g., dietary, housekeeping, maintenance) and were English speaking were invited to participate. Staff members were surveyed using the SOS-NH, which was administered as part of a larger battery of questionnaires.

Revisions to the Safety Organizing Scale-Nursing home version (SOS-NH)

Because the SOS was originally developed for use in acute-care settings,³⁶ we adapted the language for the NH setting. For example, we asked staff members about their perceptions of the staff safety behaviors in the NH as whole, rather than at the unit level. This choice is consistent with other studies in NHs that suggest that unit-level boundaries in NHs are not always clear.³⁷ In this minor revision we replaced the word “patient” with “resident.”

Validity and reliability testing

Our general research question was whether the SOS-NH would reveal psychometric properties similar to those of the original version tested in acute-care hospitals.¹³ Our validation strategy was followed the American Educational Research Association guidelines. Accordingly we defined validity as the degree to which evidence supports the theoretical interpretation of test scores, while reliability refers to the consistency of measurement.³⁸ We developed specific hypotheses and research questions (Table 1).

Validity testing—Evidence based on response processes (research question 1 & 2, Table 1) was compiled through assessment of distribution and skewing of data, missing responses

for each item and overall scale, and acceptability (number of respondents omitting no items). The internal consistency of the measurements and the precision of test results were tested (research question 3, Table 1) by calculating Cronbach's Alpha. Because the alpha coefficient might underestimate or even overestimate the population reliability due to violation of the tau-equivalence assumption, we also calculated the composite reliability according to Raykov (2010)^{39, p. 161} (see <http://www.uoguelph.ca/~scolwell/cr.html>). To test whether the SOS-NH reliably reflect a NH level construct – making aggregation of data from all staff in the NH appropriate – we computed five measures. Using F-statistic from a one-way variance analysis we calculated between-group variance. We applied two types of intraclass correlations, to calculate the proportion of variance explainable by NH membership (intraclass correlation 1) and the reliability of NH means (intraclass correlation 2). Both measures describe how strongly responses from staff members in the same NH resemble each other.¹³ We also calculated design effects to account for within-group sample size, which might inflate intraclass correlations.²⁷ Degree to which responses of individuals within a group are interchangeable was calculated with the within-group agreement.⁷

To provide evidence based on internal structure we aimed to confirm the uni-dimensional structure of the original SOS¹³ (Hypothesis 1, Table 1) and conducted confirmatory factor analysis to test how well our model would fit the data. We assessed discriminant validity to provide evidence of construct validity. Both staff members' engagement in safety behaviors and perception of communication patterns are considered visible features of a safety culture.^{13,25,40} and are therefore closely related concepts. Communication patterns served as a comparison concept to explore whether the SOS is discriminant from this feature (Hypothesis 2, Table 1) using confirmatory factor analyses^{13,41} and the method on average variance estimates reported by Fornell and Larcker (1981).⁴² To assess communication patterns we used the "*communication openness scale*".⁴³ Psychometric properties of this scale were reported in previous studies.⁴³ Cronbach's alpha for the communication openness scale was 0.841.

Data collection and data management

The Institutional Review Boards of [blinded] and four VA Medical Centers approved all data collection and management procedures. Research assistants introduced the study in staff meetings or one-on-one and answered questions after which participants provided written informed consent. NH staff members on all shifts were invited to complete the surveys, whether or not they also participated in the intervention that was part of the larger study. Data collection occurred over a 4 week period during which participants returned surveys to a locked drop box in the NH. Research assistants approached non-respondents weekly, reminding them if they had not yet returned the survey. To avoid harassing people who did not want to participate after initially agreeing, we included a statement in the instructions that they could return the survey without completing it and we would not contact them again. Surveys were written at an eight-grade reading level. People separately completed demographic forms, which enabled us to compare respondents to non-respondents that did not complete the survey.

Trained study personnel entered data into an Access database and double entered a random 20% for quality control. Missing data were coded separately. When respondents circled more than one contiguous number on a single item, we averaged and rounded toward the mid-point of the scale. We coded data as missing if respondents circled more than 1 non-contiguous number.

Statistical analysis

Analyses for questions and hypotheses are in Table 1. Descriptive, correlation, reliability and ANOVA statistics were performed using IBM SPSS Statistics (version 21.0.0; SPSS Inc., Chicago, IL). ICCs, design effects and within-group agreement were calculated with Microsoft Office Excel 2011[®] using data from the SPSS outputs. To evaluate the SOS-NH's internal structure we performed confirmatory factor analysis with MPlus (version 7.0) using maximum likelihood with robust standard errors and chi-square setting level of significance at $P < 0.05$.

Results

Eighty-eight percent of surveys were returned ($N=644$). Table 2 displays characteristics of participant and non-participants. Staff members less likely to complete the survey were those aged 21 and 25 years, 36 and 45 years, of Black or African American race, RNs, social workers, housekeeping or laundry personal, part-time workers and staff in the probationary period.

Of returned questionnaires, frequencies of missing values were relatively low, ranging from 2.2% (item 8) to 3.4% (item 5; average = 3.05%). Of respondents, 91.5% ($n=589$) filled out the SOS-NH with no missing values. Because characteristics of respondents with no missing items were equally distributed throughout the sample, for the remaining analyses, we used data from completed questionnaires.

Mean item values on the 7-point scale ranged from 3.89 ($SD = 1.87$; item 3) to 4.97 ($SD = 1.56$, item 5 and $SD = 1.64$, item 9; see Table 3). The mean for the entire scale (SOS score) was 4.46 ($SD = 1.41$) and the median score was 4.56 (25th–75th percentile = 3.44, 4.56, 5.56, see Table 3). We did not observe a ceiling or floor effect in any of the nine items or the SOS-NH scale score and q-q plots demonstrated relatively normal distributed data.

Both the alpha and the composite reliability coefficient for this one-dimensional construct were 0.94, indicating scale reliability. The reliability of the SOS as an aggregate NH measure was shown by the significant between-NH variance [$F(12,567) = 3.60$, $p < 0.001$], within group agreement ($r_{wg(j)} = 0.76$), intraclass correlation 1 ($ICC1 = 0.05$) and 2 ($ICC2 = 0.72$), and design effect (= 2.25).

As Table 4 (column 1) shows, our CFA measurement model demonstrated acceptable construct validity. Although all items load highly significantly ($P < 0.001$) on the intended factor, we failed to demonstrate acceptable fit across all fit-indices. Comparative fit index ($CFI = 0.913$) and the Standardized Root Mean Square Residual ($SRMR = 0.045$) confirmed the one-dimensional structure of the SOS-NH, but Root Mean Square Error of Approximation was insufficient ($RMSEA = 0.121$; 90% CI: 0.107–0.135).

We evaluated discriminant validity by comparing the fit of the 2-factor model (Table 3, column 3; SOS-NH/communication openness item load on two factors) with the 1-factor solution (column 2; all SOS-NH and communication openness items load on one factor). Results on the two measurement models, including chi-square difference ($\chi^2 = 592.811$, $df = 1$, $P < 0.001$) and fit-indices suggested that communication openness is distinct from the SOS-NH. Discriminant validity was also supported by Average Variance Extracted (AVE), as both the AVE for the SOS-NH (0.63) and communication openness (0.52) were higher than the shared variance (i.e., square of the correlation), which was 0.35.

Discussion

This study provides first evidence on the validity and reliability of the SOS-NH, which we tested in 13 NHs using international questionnaire validation standards.³⁸ Overall, our results indicate good psychometric properties of the SOS-NH, similar to those for the original SOS tested in acute-care hospitals, supporting its' use in the NH setting.

The SOS measures healthcare professionals' engagement in nine safety behaviors reflecting on "collective mindfulness", which is considered an important safety climate feature in health care.¹³³⁰ To date, a valid, reliable, and brief measure reflecting and measuring safety climate in NHs has been lacking. The AHRQ's NH Survey on Patient Safety Culture has been used in this setting,¹⁹²⁰ but has undergone only preliminary psychometric testing. This frequently used instrument includes multiple sub-dimensions and a large number of items, characteristics which impeded confirmation of their internal structure after cross-cultural translation in acute-care⁴⁴ and NH setting.²⁹ This might also impair cross-national comparison and benchmarking. The SOS-NH's content of crucial patient safety behaviors that might directly influence resident outcomes makes it a valuable tool for monitoring the safety climate of NHs. The low proportions of missing values indicate it's practicability and acceptability for a wide range of staff members with direct resident contact. However, we found that staff member' professional characteristics influenced whether or not they completed the survey. It is not clear whether staff members' willingness to complete the questionnaire, including the SOS-NH was related to the SOS or other surveys in the questionnaire. Nevertheless, the SOS-NH score demonstrated variability between and within NHs, which is an important requirement in research for computing inferential analyses, e.g. analyses of variance or regression analyses.

The uni-dimensional structure of the SOS-NH was supported by highly significant item loadings on one factor and by two out of four fit indices (comparative fit index and standardized root mean square residual). The 2-statistics and root mean square error of approximation resulting from our confirmatory factor analyses falsified our hypotheses **H1**. It was not surprising that 2 was statistically significant (indicating a poor fit), as it is very sensitive to sample size.⁴⁵ A large sample is necessary to increase precision of parameter estimation, 2 results are almost always significant even with only modest sample sizes.⁴⁶ We were surprised, however, that in contrast to the SOS studies in the acute care setting the root mean square error of approximation suggested poor model fit for the SOS-NH.¹³³²³³ The root mean square error of approximation - defined as "misfit per degree of freedom" ^{39, p. 71} - is currently the most popular measure of model fit and is reported in nearly all papers that use confirmatory factor analyses.⁴⁶ However, this index is positively biased, i.e. it tends to be too large, because it depends on sample size and degrees of freedom.⁴⁵ Simpler structural models, such as the SOS-NH have small degrees of freedom (i.e. resulting from low numbers of items/variables) in combination with low to modest sample size might be disadvantaged to result in acceptable values of the root mean square error of approximation. For this reason, some authors do not calculate the root mean square error of approximation for low degrees of freedom models,⁴⁷ and suggest reporting the standardized root mean square residual only. This index is relatively less sensitive to such issues⁴⁶ and showed excellent model fit for the SOS-NH in our study.

Ideally, different fit indices should lead to the same conclusion. If not, the most conservative choice is to reject the hypothesized model.⁴⁵ Fit indices are a useful guide, but should not drive the research process, as it moves away from the original theory-testing purpose of confirmatory factor analyses. As there is an ongoing debate about which fit indices should be reported⁴⁶ and which pre-defined cut-off values are appropriate,⁴⁸ strictly adhering to recommended cutoff values might lead to Type I errors, i.e. the incorrect rejection of an

acceptable model. To avoid such error for our model, we undertook a post-hoc exploration on the factor structure of the SOS-NH and computed principal axis factoring using IBM SPSS Statistics. Our analyses revealed one strong factor accounting for 63.7% of the variance in the SOS-NH items and communalities ranging from 0.506 (item 1) to (0.741 item 2). This result indicates that more than 50% of the variance in the single items is explained by the single factor.

In summary, we are confident that these results provide sufficient evidence of the one-dimensional structure of the SOS-NH version and justify the aggregation of the nine individual items to a single SOS-NH score. However, for further evaluation on the internal structure of the SOS-NH, item response theory might be approach to evaluate the contribution of each of the nine items if there is a latent trait behind the responses of staff members and thus if there are fewer number of items accounting for the variance of this construct³⁹.

Limitations

This article provides first evidence on the validity and reliability of the SOS-NH. According to international guidelines, researchers should evaluate psychometric properties as much as possible “to develop a scientifically sound validity argument to support the intended interpretation of test scores and their relevance to the proposed use.”³⁸ Although only minor adaptations of the items (patients to residents) were made and face validity was established, there was no quantitative content validity testing on the SOS-NH (e.g. content validity index⁴⁹), showing the relevance of its items for the NH setting. According to studies in NHs the unit of analysis was set at the NH level, because boundaries at the department- or unit-level, such as in hospitals, are not clear in NHs.³⁷ This limits the descriptions/explorations of staffs’ engagement in safety behaviors to this organizational level and might be associated with a loss of information for the unit or team level within NHs. Because we did not test criterion validity, i.e. higher safety climate levels are associated with improved resident outcomes, we cannot state whether variability in NHs’ safety climate scores is related resident safety outcomes. We pooled data from both community and VA NHs. VA NHs may differ from community NHs in organizational factors such as staffing ratios, administrative structure, and resident case-mix. However, VA centers have to deal with the same safety and quality issues as community NHs. From a theoretical perspective the SOS’ safety behaviors are important and applicable in VA centers. Furthermore, we did not observe differences or patterns for the responses on the SOS-NH items from VA centers (e.g. ceiling or floor effects, higher number of missing values), therefore we considered it appropriate to pool the data to increase statistical power and generalizability of findings.

Conclusions

The SOS-NH is a valuable tool to measure staff members “collective mindfulness” about resident safety issues in NHs. Initial evidence on validity and reliability support its use in NHs. In clinical practice the SOS-NH might be used to monitor staff members’ engagement in safety behaviors. “Collective mindfulness” might be an important factor (1) for determining resident safety and quality sustainable over time, and (2) for the successful implementation of quality improvement activities, such as a patient fall prevention program. As the individual items entail information on changeable behaviors and care processes, the SOS-NH allows NH leaders to plan, implement and evaluate activities to improve resident safety climate. Further studies on the SOS-NH are needed to evaluate criterion validity. It will be crucial to explore what antecedents, such as leadership, staffing levels, explain variability of staff members’ engagement in safety behaviors and practices. Criterion validity of the SOS-NH with resident safety outcomes, such as falls, should be established to verify this relevance of this concept for safety and quality in NHs.

Acknowledgments

We wish to acknowledge the support of our funders, National Institutes of Health, National Institute of Nursing Research (R56NR003178 and R01NR003178) Anderson and Colón-Emeric, PIs and VA HSR&D, EDU 08-417, Colón-Emeric, PI

References

1. Quigley PA, Campbell RR, Bulat T, Olney RL, Buerhaus P, Needleman J. Incidence and cost of serious fall-related injuries in nursing homes. *Clin Nurs Res*. 2012; 21(1):10–23. [PubMed: 21788448]
2. Gross L, Barbieri A, Carnevale L. Prevalence of infections in nursing homes in the Vercelli area (Piemonte, Italy). *Ig Sanita Pubbl*. 2012; 68(1):29–48. [PubMed: 22507991]
3. Jylha V, Saranto K, Bates DW. Preventable adverse drug events and their causes and contributing factors: the analysis of register data. *Int J Qual Health Care*. 2011; 23(2):187–97. [PubMed: 21242162]
4. Handler SM, Wright RM, Ruby CM, Hanlon JT. Epidemiology of medication-related adverse events in nursing homes. *The American journal of geriatric pharmacotherapy*. 2006; 4(3):264–72. [PubMed: 17062328]
5. Gurwitz JH, Field TS, Judge J, Rochon P, Harrold LR, Cadoret C, et al. The incidence of adverse drug events in two large academic long-term care facilities. *Am J Med*. 2005; 118(3):251–8. [PubMed: 15745723]
6. Gruneir A, Mor V. Nursing home safety: current issues and barriers to improvement. *Annu Rev Public Health*. 2008; 29:369–82. [PubMed: 18173385]
7. Horn SD, Bender SA, Bergstrom N, Cook AS, Ferguson ML, Rimmasch HL, et al. Description of the National Pressure Ulcer Long-Term Care Study. *J Am Geriatr Soc*. 2002; 50(11):1816–25. [PubMed: 12410900]
8. Nicolle LE. Urinary tract infection in geriatric and institutionalized patients. *Current Opinion in Urology*. 2002; 12(1):51–5. [PubMed: 11753134]
9. Hamers JPH, Huizing AR. Why do we use physical restraints in the elderly? *Z Gerontol Geriatr*. 2005; 38(1):19–25. [PubMed: 15756483]
10. Pauly L, Stehle P, Volkert D. Nutritional situation of elderly nursing home residents. *Z Gerontol Geriatr*. 2007; 40(1):3–12. [PubMed: 17318726]
11. Bonner AF, Castle NG, Men A, Handler SM. Certified nursing assistants' perceptions of nursing home patient safety culture: is there a relationship to clinical outcomes? *J Am Med Dir Assoc*. 2009; 10(1):11–20. [PubMed: 19111848]
12. Institute of Medicine. Achieving a new standard for care. Washington D.C: The National Academies Press; 2004. Patient safety.
13. Vogus TJ, Sutcliffe KM. The Safety Organizing Scale: development and validation of a behavioral measure of safety culture in hospital nursing units. *Med Care*. 2007; 45(1):46–54. [PubMed: 17279020]
14. Guldenmund FW. The nature of safety culture: a review of theory and research. *Safety Science*. 2000; 34(1–3):215–57.
15. Guldenmund, FW. Understanding and Exploring Safety Culture. Technische Universiteit Delft; 2010.
16. Denison DR. What is the Difference between Organizational Culture and Organizational Climate? A Native's Point of View on a Decade of Paradigm Wars. *The Academy of Management Review*. 1996; 21(3):619–54.
17. Zohar D. Thirty years of safety climate research: reflections and future directions. *Accid Anal Prev*. 2010; 42(5):1517–22. [PubMed: 20538108]
18. Guldenmund FW. The use of questionnaires in safety culture research - an evaluation. *Safety Science*. 2007; 45(6):723–43.
19. Castle NG, Sonon KE. A culture of patient safety in nursing homes. *Qual Saf Health Care*. 2006; 15(6):405–8. [PubMed: 17142587]

20. Handler SM, Castle NG, Studenski SA, Perera S, Fridsma DB, Nace DA, et al. Patient safety culture assessment in the nursing home. *Qual Saf Health Care*. 2006; 15(6):400–4. [PubMed: 17142586]
21. Hughes CM, Lapane KL. Nurses' and nursing assistants' perceptions of patient safety culture in nursing homes. *Int J Qual Health Care*. 2006; 18(4):281–6. [PubMed: 16855298]
22. Wagner LM, Capezuti E, Rice JC. Nurses' perceptions of safety culture in long-term care settings. *J Nurs Scholarsh*. 2009; 41(2):184–92. [PubMed: 19538703]
23. Wisniewski AM, Erdley WS, Singh R, Servoss TJ, Naughton BJ, Singh G. Assessment of safety attitudes in a skilled nursing facility. *Geriatric nursing (New York, NY)*. 2007; 28(2):126–36.
24. Halligan M, Zecevic A. Safety culture in healthcare: a review of concepts, dimensions, measures and progress. *BMJ Qual Saf*. 2011; 20(4):338–43.
25. Flin R, Burns C, Mearns K, Yule S, Robertson EM. Measuring safety climate in health care. *Qual Saf Health Care*. 2006; 15(2):109–15. [PubMed: 16585110]
26. Sorra JS, Nieva VF. Final Report to Agency for Health Care Research and Quality. Washington: Rockville, MD: AHRQ; 2004. Psychometric Analysis of the Hospital Survey on Patient Safety. Prepared by Westat uCN--, editor
27. Sorra JS, Dyer N. Multilevel psychometric properties of the AHRQ hospital survey on patient safety culture. *BMC Health Serv Res*. 2010; 10:199. [PubMed: 20615247]
28. Blegen MA, Gearhart S, O'Brien R, Sehgal NL, Alldredge BK. AHRQ's hospital survey on patient safety culture: psychometric analyses. *J Patient Saf*. 2009; 5(3):139–44. [PubMed: 19920453]
29. [blinded]
30. Weick, KE.; Sutcliffe, KM. *Managing the Unexpected: Resilient Performance in an Age of Uncertainty*. 2. Jon Wiley & Sons, Inc; 2009.
31. Chassin MR, Loeb JM. The ongoing quality improvement journey: next stop, high reliability. *Health Aff (Millwood)*. 2011; 30(4):559–68. [PubMed: 21471473]
32. [blinded]
33. [blinded]
34. McKinley W. Organizational Theory Development: Displacement of Ends? *Organization Studies*. 2010; 31(1):47–68.
35. Nieva VF, Sorra J. Safety culture assessment: a tool for improving patient safety in healthcare organizations. *Qual Saf Health Care*. 2003; 12(Suppl 2):ii17–23. [PubMed: 14645891]
36. [blinded]
37. Estabrooks CA, Morgan DG, Squires JE, Bostrom AM, Slaughter SE, Cummings GG, et al. The care unit in nursing home research: evidence in support of a definition. *BMC medical research methodology*. 2011; 11:46. [PubMed: 21492456]
38. American Educational Research Association. *Standards for Educational and Psychological Testing*. 1999.
39. Raykov, T.; Marcoulides, GA. *Introduction to Psychometric Theory*. Taylor & Francis; 2010.
40. Colla JB, Bracken AC, Kinney LM, Weeks WB. Measuring patient safety climate: a review of surveys. *Qual Saf Health Care*. 2005; 14(5):364–6. [PubMed: 16195571]
41. DeVellis, RF. *Scale Development: Theory and Applications*. 2. SAGE Publications; 2011.
42. Fornell C, Larcker DF. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*. 1981; 18(1):39–50.
43. O'Reilly CA, Roberts KH. Task group structure, communication, and effectiveness in three organizations. *J Appl Psychol*. 1977; 62(6):674–81.
44. Pfeiffer Y, Manser T. Development of the German version of the Hospital Survey on Patient Safety Culture: Dimensionality and psychometric properties. *Safety Science*. 2010; 48(10):1452–62.
45. Kline, RB. *Principles and Practice of Structural Equation Modeling*. 3. New York: The Guilford Press; 2011.
46. Iacobucci D. Structural equations modeling: Fit Indices, sample size, and advanced topics. *Journal of Consumer Psychology*. 2010; 20(1):90–98.

47. Kenny DA, McCoach DB. Effect of the Number of Variables on Measures of Fit in Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*. 2003; 10(3):333–51.
48. Hopper D, Coughlan J, Mullen MR. Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods*. 2008; 6(1):53–60.
49. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Res Nurs Health*. 2007; 30(4):459–67. [PubMed: 17654487]
50. Hu, Lt; Bentler, PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*. 1999; 6(1):1–55.

Appendix

Covariance Matrix SOS-NH items (n=589)

Items of the SOS-NH	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36	Q37
Q29	2.429								
Q30	1.734	2.766							
Q31	1.636	2.094	2.767						
Q32	1.660	1.749	1.951	3.037					
Q33	1.529	1.789	2.165	2.447	3.319				
Q34	1.619	1.798	2.181	2.189	2.442	3.366			
Q35	1.422	1.568	1.847	1.784	1.948	2.507	3.075		
Q36	1.373	1.549	1.735	1.925	2.129	2.327	2.247	2.958	
Q37	1.429	1.571	1.646	1.611	1.747	1.866	1.930	2.015	2.685

Table 1

Research questions and hypotheses of this validation study

Evidence on validity and reliability	Research questions (R) and hypothesis (H)	Developed research questions and hypothesis	Statistical analyses	Desirable results for decision making
Evidence based on response processes	R1	How many missing values appear in the SOS-NH?	Descriptive statistics (frequencies, medians, interquartile ranges, means, standard deviations, variances, graphs, and cross-tabulations)	<ul style="list-style-type: none"> Floor effects (>50% indicating 1 (not at all) or 2 (to a very limited extent))
	R2	Are there distribution abnormalities in the different items of the SOS-NH?		<ul style="list-style-type: none"> Ceiling effects (>50% indicating 6 (to a great extent) and 7 (to a very great extent)).
	R3 (Reliability)	Is the SOS-NH internally consistent and does it reflect on a nursing home-level construct?	<ul style="list-style-type: none"> Calculations of Intraclass correlations, within-group agreement and design effects based on results of one-way ANOVA with the Safety Organizing score as the dependent variable and hospital units as independent variable Calculation of Cronbach's alpha. 	<ul style="list-style-type: none"> Significant between-group variance using F-statistic ($P < 0.05$) Intraclass correlation 1 between 0.05 and 0.30 and Intraclass correlation 2 above 0.70⁷. Design effects should be • 2 to demonstrate the nesting of the data²⁰. Within-group agreement values should be 0.70 or greater⁷. Cronbach's alpha • 0.70 indicate factor consistency⁴⁶
Evidence based on internal structure	H1 (Construct validity)	The SOS-NH has an unidimensional structure	<ul style="list-style-type: none"> Calculation of composite reliability*. Confirmatory Factor Analysis 	<ul style="list-style-type: none"> Composite reliability • 0.70³⁹ Non-significant chi-square values Comparative fit index close to 0.95⁵⁰ Upper confidence interval of the root mean square error of approximation < 1.00 Standardized Root Mean Square Residual < 0.08⁴⁰
Evidence based on relationship with other variables	H2 (Discriminant validity)	The items from a related concept ("communication openness") are significantly distinct from the SOS-NH.	<ul style="list-style-type: none"> Confirmatory Factor Analysis Calculation of average variance extracted# and correlation analysis 	<ul style="list-style-type: none"> The two-factor measurement model has significantly better fit (2-difference and fit-indices as the one-factor model⁷. Extracted variance for SOS-NH and for communication openness both need to be larger than the shared variance (i.e., square of the correlation) between SOS-NH and communication openness⁴²

Legend:

* Composite reliability = (sum of standardized loading)²/[(sum of standardized loading)² + sum of indicator measurement error]

Average Variance Extracted = (sum of squared standardized loading/sum of squared standardized loading + sum of indicator measurement error)

Table 2

Characteristics of the participants

Staff members	Total sample (N=728)	Responders (n=644)	Non-responders (n=84)
Male - %	13.1.	12.7	16.0
Age (in years) - %			
Less than 21	0.7	0.6	1.2
21–25	6.4	6.1	8.6
26–35	18.2	18.7	14.2
36–45	22.2	21.9	24.7
46–55	31.5	31.5	32.1
56 and over	20.9	21.2	18.5
Ethnicity - %			
Black of African American	45.9	47.0	52.6
White	46.2	48.6	42.3
Hispanic or Latino	1.2	1.3	1.3
Asian	1.9	2.1	1.3
Others (e.g. American Indian, Multi-Racial)	4.8	1.0	2.5
Job title - %			
Certified Nursing Assistant	31.0	30.8	32.9
Licensed Practical Nurse	13.7	13.7	13.9
Registered Nurse	9.9	9.5	12.7
Resident Assessment Nurse	2.8	3.0	1.3
Nurse Manager/Supervisor	3.0	2.9	3.8
Director /Asst Director of Nursing	1.5	1.4	2.4
Dietary Staff	3.4	3.5	2.5
Physical or Occupational Therapist	6.5	7.0	2.5
Activities Coordinator	2.7	2.7	2.5
Social Worker	2.8	2.5	5.1
Medical staff	0.8	0.8	1.3
Nursing Home Administrator	1.1	1.1	0.0
Housekeeping or Laundry	4.1	3.8	6.3
Activities coordinator	2.7	2.7	2.5
Others (e.g. Hall Attendants, Maintenance)	14.0	14.6	10.3
Employment status - %			
Permanent, fulltime	84.0	84.4	81.3
Permanent, part-time	6.1	5.9	7.5
Contract staff	2.3	2.2	2.5
Work on “as-needed” basis, PRN	1.4	1.6	0.0
Probationary period	5.9	5.6	8.8
Other	0.3	0.3	0.0
Professional experience in this facility in years (including all positions) – Median (IQR)	3 (7)	3 (7)	3 (5)

Table 3

Mean, 95% confidence interval (CI), standard deviation (SD), median and interquartile range (IQR), Minimum and Maximum for the single nine items and the SOS-NH score (n=589)

Items of the SOS-NH	Mean (95% CI) $\pm$ SD	Median (IQR)	Min.-Max.
(1) We have a good “map” of each other’s talents and skills	4.20 (4.06 to 4.34) $\pm$ 1.74	4 (3)	1–7
(2) We talk about mistakes and ways to learn from them	4.23 (4.08 to 4.38) $\pm$ 1.83	4 (3)	1–7
(3) We discuss our unique skills with each other so we know who on the unit has relevant specialized skills and knowledge	3.89 (3.74 to 4.04) $\pm$ 1.82	4 (3)	1–7
(4) We discuss alternatives as to how to go about our normal work activities	4.32 (4.18 to 4.45) $\pm$ 1.67	4 (3)	1–7
(5) When giving report to an oncoming nurse, we usually discuss what to look out for	4.97 (4.85 to 5.10) $\pm$ 1.56	5 (2)	1–7
(6) When attempting to resolve a problem, we take advantage of the unique skills of our colleagues	4.39 (4.25 to 4.53) $\pm$ 1.72	4 (3)	1–7
(7) We spend time identifying activities we do not want to go wrong	4.56 (4.43 to 4.69) $\pm$ 1.67	5 (3)	1–7
(8) When errors happen, we discuss how we could have prevented them	4.58 (4.44 to 4.73) $\pm$ 1.76	5 (3)	1–7
(9) When a resident crisis occurs, we rapidly pool our collective expertise to attempt to resolve it	4.97 (4.84 to 5.10) $\pm$ 1.64	5 (2)	1–7
SOS-NH score	4.46 (4.34 to 4.57) $\pm$ 1.40	4.5 (2)	1–7

Table 4
Standardized factor loadings, standard errors and fit-indices of the confirmatory factor analysis for the SOS-NH

Construct and Items	SOS-NH (n=589)		2-Factor Model SOS-NH/Communication openness (n=589)		1-Factor Model SOS-NH and Communication openness (n=589)	
	Factor loadings	Standard Error	Factor loadings	Standard Error	Factor loadings	Standard Error
SOS-NH						
1. We have a good "map" of each other's talents and skills (q32)	0.824	0.028	0.822	0.021	0.812	0.021
2. We talk about mistakes and ways to learn from them (q34)	0.843	0.017	0.845	0.017	0.844	0.017
3. We discuss our unique skills with each other so we know who on the unit has relevant specialized skills and knowledge (q33)	0.843	0.017	0.844	0.017	0.836	0.017
4. We discuss alternatives as to how to go about our normal work activities (q31)	0.852	0.016	0.845	0.016	0.827	0.017
5. When giving report to an oncoming nurse (or CNA), we usually discuss what to look out for (q29)	0.716	0.028	0.712	0.027	0.705	0.026
6. When attempting to resolve a problem, we take advantage of the unique skills of our colleagues (q36)	0.785	0.024	0.793	0.023	0.802	0.021
7. We spend time identifying activities we do not want to go wrong (q30)	0.767	0.026	0.762	0.026	0.748	0.026
8. When errors happen, we discuss how we could have prevented them (q35)	0.761	0.023	0.767	0.022	0.771	0.022
9. When a resident crisis occurs, we rapidly pool our collective expertise to attempt to resolve it (q37)	0.728	0.025	0.733	0.025	0.740	0.024
Communication openness						
1. It is easy for me to talk openly with all workers in this nursing home.	-	-	0.701	0.028	0.426	0.040
2. Communication in this nursing home is very open.	-	-	0.765	0.027	0.551	0.035
3. I find it enjoyable to talk with other workers in this nursing home.	-	-	0.698	0.027	0.418	0.039
4. When people talk to each other in this nursing home, there is a good deal of understanding.	-	-	0.773	0.030	0.535	0.036
5. It is easy to ask advice from any worker in this nursing home.	-	-	0.673	0.029	0.511	0.036
2 baseline model (df)	2501.853 (36)		3935.176 (91)		3935.176 (91)	
2 default model	239.997 (25)		364.055 (74)		808.687 (75)	
• 2 (from 2 factor)	-		-		592.811 (1)	

Construct and Items	SOS-NH (n=589)		2-Factor Model SOS-NH/Communication openness (n=589)		1-Factor Model SOS-NH and Communication openness (n=589)	
	Factor loadings	Standard Error	Factor loadings	Standard Error	Factor loadings	Standard Error
CFI	0.913		0.925		0.809	
RMSEA (90% CI)	0.121 (0.107–0.135)		0.082 (0.074–0.091)		0.130 (0.122–0.138)	
SRMR	0.045		0.046		0.094	

- In all models, the error terms of Q2 and Q8 and Q6 and Q9 of the SOS are correlated.
- All factor loadings and 2-statistics significant at $P<0.001$;
- 2 difference test used the Satorra-Bentler Scaled (mean-adjusted) Chi-Square when using MLR in Mplus⁴⁷.
- SOS-NH indicates Safety Organizing Scale-Nursing Home version, n = the number of staff members respondents included in the analysis, 2 = chi-square, df = degree of freedom, CFI = comparative fit index, RMSEA = root mean square error of approximation, SRMR = chi standardized root mean square residual;