NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** 3450

Measure Title: Practice Environment Scale - Nursing Work Index (PES-NWI) (composite and five subscales)

Date of Submission: 7/31/2018

Type of Measure:

☐ Outcome (*including PRO-PM*)	☐ Composite – ***STOP – use composite testing form***
☐ Intermediate Clinical Outcome	☐ Cost/resource
☐ Process *(including Appropriate Use)*	☐ Efficiency
☒ Structure	

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. ***If there is more than one set of data specifications or more than one level of analysis, contact NQF staff*** about how to present all the testing information in one form.
- **For all measures, sections 1, 2a2, 2b1, 2b2, and 2b4 must be completed.**
- **For outcome and resource use measures**, section **2b3** also must be completed.
- If specified for **multiple data sources/sets of specificaitons** (e.g., claims and EHRs), section **2b5** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b1-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 25 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). ***Contact NQF staff if more pages are needed.***
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for version 7.1 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing [10] demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For **instrument-based measures** (including PRO-PMs) **and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b1. Validity testing [11] demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **instrument-based measures (including PRO-PMs) and composite performance measures**, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure; [12]

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). [13]

2b3. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; [14,15] and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** [16] **differences in performance**;

OR

there is evidence of overall less-than-optimal performance.

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. Analyses identify the extent and distribution of **missing data** (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions.

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v.

$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: **(*must be consistent with data sources entered in S.17*)**	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ claims	☐ claims
☐ registry	☐ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☒ other: nurse survey	☒ other: nurse survey

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

Nurse survey data from research projects and the National Database of Nursing Quality Indicators were used to derive and confirm the instrument subscales and composite and to provide ongoing psychometric performance.

1.3. What are the dates of the data used in testing? 1985 to 2018

1.4. What levels of analysis were tested? (*testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: **(*must be consistent with levels entered in item S.20*)**	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☐ other:

1.5. How many and which measured entities were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Most measured entities were acute care hospitals. Some measured entities were home care agencies. Per Warhsawsky & Havens (2010), 37 samples of 4 to 4,783 units over the years 1998 to 2010, and per Swiger, et. al 2017, 46 samples of 2 units to 5322 units and 519 hospitals over the years 2010 to 2016. In addition, per Lake et al. 2018, 212 separate research articles were published through March 2016 that included empirical data on the PES-NWI; some of these articles were included in the two systematic reviews noted previously. From April 2016 through June 2018, 35 separate research articles were published that included empirical data on the PES-NWI from 7 to 489 hospitals. These hospital samples included representative samples of hospitals from multiple U.S. states, including hospitals of all sizes, ownership, and teaching status.

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

There are no patient data. Here we report about data from nurses. Per Warhsawsky & Havens (2010), 37 samples of 31 to 72,889 nurses over the years 1998 to 2010, and per Swiger, et. al 2017, 46 samples of 133 to 33,845 over the years 2010 to 2016. In addition, per Lake et al. 2018, 212 separate research articles were published through March 2016 that included empirical data on the PES-NWI; some of these articles were included in the two systematic reviews noted previously. From April 2016 through June 2018, 37 separate research articles were published that included empirical data on the PES-NWI. In these 35 articles, data from samples of from 87 to 33,000 nurses were reported. The nurse characteristics in many samples resembled nurse characteristics for age, sex, and educational level as described in national nurse surveys.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.
There are no differences for different aspects of testing.

1.8 What were the social risk factors that were available and analyzed? For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.
There is no basis for adjusting for social factors of nurses, such as educational level. There is no contextual reason to think that social factors of nurses would impact their answers or impact being able to compare facilities fairly.

____

2a2. RELIABILITY TESTING
<u>Note</u>*: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)
☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)
☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)
Reliability testing of critical data elements was conducted by computing Cronbach's alpha, which measures internal consistency of the items in a scale. This method was computed and reported in all studies noted above in the Warshawsky & Havens (2010) and Swiger et. al, 2017 papers.

Reliability testing of performance measure score was conducted by assessing inter-rater reliability, which focuses on whether nurses give consistent responses within a hospital or nursing unit, as compared to across hospitals or nursing units in a sample. Performance measure score reliability is assessed using the intraclass correlation (ICC) (1,*k*), which is a function of the number of nurse respondents per hospital and the intraclass correlation coefficient from a one-way analysis of variance of the subscales and composite across hospitals or nursing units. In order to assure reliability, the ICC (1,*k*) should exceed .60 (Glick, 1985).

Glick, W. H. (1985). Conceptualizing and measuring organizational and psychological climate: Pitfalls in multilevel research. *Academy of Management Review*, 10(3), 601-616.

2a2.3. For each level of testing checked above, what were the statistical results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)
Reliability testing of Critical data elements: Cronbach's Alpha Statistics
Of the 46 articles reviewed in Swiger et al (2017) published from 2010 to 2016, 37 reported Cronbach's alphas; coefficients ranged from .71 – .96, with the exception of one .67, and one .53 in a small sample size. These results support the coherence of the different subscales and the composite. Additional internal consistency reliability data are displayed in Table 2b1.3D, from the 35 newest articles. This table is presented at the end of the document due to its length of several pages.

Distribution of Reliability Statistics from a Signal-To-Noise Analysis: Statistics on Organizational Reliability:

Table 2a2.3A.
Analysis for 2018 NQF measure maintenance using 2015 National Database of Nursing Quality Indicators nurse survey data

Measure	ICC (1,k)
Subscale	
Collegial Nurse-Physician Relations	.936
Nursing Foundations for Quality Care	.966
Nurse Manager Ability, Leadership, and Support	.949
Nurse Participation in Hospital Affairs	.973
Staffing and Resource Adequacy	.967
Composite	.966

Note. N = 451 hospitals and from 157,481 to 157,522 staff nurses. ICC (1,k) estimated in one-way ANOVA.

Table 2a2.3B.
Compilation of entity-level reliability statistics across 14 studies published from 2002 to 2017.

References:

Reference	# organizational units (hospitals or nursing units)	# nurses	ICC (1,k) statistics reported or summarized	Page reference
Lake (2002)	16 magnet hospitals proportionate by regions of the country	1,610	.88 to .97	Pg 183
Lake et al (2006)	156 adult community hospitals in Pennsylvania	10,962	.67 to .82	Pg 4
Clarke (2007)	188 Pennsylvania general acute care hospitals	11,512	.70 to .90	Pg 303
Flynn et al (2010)	63 Medicare and Medicaid certified nursing homes in New Jersey	897	Composite: .68 Subscales range: .55 to .75	Pg 4, 9
Brooks-Carthon et al (2011)	429 hospitals across four states (Florida, Pennsylvania, New Jersey and California)	98,000	Subscales range: .73 to .90	Pg 303
McHugh et al (2012)	396 adult, non-federal acute care hospitals across four states (CA, FL, NJ, PA)	16,241	.61	Pg 3
Kelly et al (2013)	320 hospitals across four states (CA, FL, NJ, PA)	3,217	.69	Pg 484
McHugh et al (2013)	564 Magnet and non-Magnet hospitals across four states (CA, FL, NJ, PA)	100,000	.81	Pg 4
Kelly et al (2014)	303 adult care hospitals across four states (CA, FL, NJ, PA)	55,159	.71	Pg 4
McHugh et al (2014)	534 hospitals across four states (CA, FL, NJ, PA)	26,005	.85	Pg 74
Carthon et al (2015)	419 acute care hospitals across three states (CA, FL, NJ, PA)	20,605	.74 to .91	Pg 257
Ma et al (2015)	373 hospitals from 44 states	33,845	Ranged from .80 to .87	Pg 3
Lake et al (2016)	171 hospitals across four states (CA, FL, NJ, PA)	1,247	4 subscales >.60; 5th = .58	Pg 3
Swiger et al (2018)	45 acute care units in 10 Army hospitals	180	ICC (1,*k*) reported as satisfactory	Pg 134, 136

Lake, E. T. (2002). Development of the practice environment scale of the Nursing Work Index. Research in nursing & health, 25(3), 176-188.

Lake, E. T., & Friese, C. R. (2006). Variations in nursing practice environments: relation to staffing and hospital characteristics. Nursing research, 55(1), 1-9.

Clarke, S. P. (2007). Hospital work environments, nurse characteristics, and sharps injuries. American Journal of Infection Control, 35(5), 302-309.

Flynn, L., Liang, Y., Dickson, G. L., & Aiken, L. H. (2010). Effects of nursing practice environments on quality outcomes in nursing homes. Journal of the American Geriatrics Society, 58(12), 2401-2406.

Brooks-Carthon, J. M., Kutney-Lee, A., Sloane, D. M., Cimiotti, J. P., & Aiken, L. H. (2011). Quality of care and patient satisfaction in hospitals with high concentrations of black patients. Journal of Nursing Scholarship, 43(3), 301-310.

McHugh, M. D., & Stimpfel, A. W. (2012). Nurse reported quality of care: a measure of hospital quality. Research in nursing & health, 35(6), 566-575.

Kelly, D., Kutney-Lee, A., Lake, E. T., & Aiken, L. H. (2013). The critical care work environment and nurse-reported health care–associated infections. American Journal of Critical Care, 22(6), 482-488.

McHugh, M. D., Kelly, L. A., Smith, H. L., Wu, E. S., Vanak, J. M., & Aiken, L. H. (2013). Lower mortality in magnet hospitals. Medical care, 51(5), 382.

Kelly, D. M., Kutney-Lee, A., McHugh, M. D., Sloane, D. M., & Aiken, L. H. (2014). Impact of critical care nursing on 30-day mortality of mechanically ventilated older adults. Critical care medicine, 42(5), 1089.

McHugh, M. D., & Ma, C. (2014). Wage, work environment, and staffing: effects on nurse outcomes. Policy, Politics, & Nursing Practice, 15(3-4), 72-80.

Carthon, J. M. B., Lasater, K. B., Sloane, D. M., & Kutney-Lee, A. (2015). The quality of hospital work environments and missed nursing care is linked to heart failure readmissions: a cross-sectional study of US hospitals. BMJ Qual Saf, 24(4), 255-263.

Ma, C., & Park, S. H. (2015). Hospital magnet status, unit work environment, and pressure ulcers. Journal of Nursing Scholarship, 47(6), 565-573.

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i*.e., what do the results mean and what are the norms for the test conducted?*)
The measure exhibits high internal consistency reliability as well as high performance score reliability, exemplified through satisfactory ICC(1,*k*) values in 14 samples over 16 years, plus recent 2015 national data from 157,500 nurses in 451 hospitals analyzed for NQF measure maintenance.

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☐ **Systematic assessment of face validity of <u>performance measure score</u> as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

The method of validity testing was by statistical association between the measure and hypothesized related constructs, to demonstrate construct, concurrent, and predictive validity.

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

The PES-NWI was developed in 2002 to measure nursing practice environments through factor analysis of 1986 survey data from staff nurses in 16 original magnet hospitals, and confirmed in 1999 data from 11,636 nurses throughout Pennsylvania (Lake, 2002). The five PES-NWI subscales can be combined into a composite measure of the practice environment, as either a continuous variable or a three-category variable indicating favorable, mixed, or unfavorable practice environments (Lake & Friese, 2006).

Lake, E. T. (2002). Development of the practice environment scale of the Nursing Work Index. Research in nursing & health, 25(3), 176-188.

Lake, E. T., & Friese, C. R. (2006). Variations in nursing practice environments: relation to staffing and hospital characteristics. Nursing research, 55(1), 1-9.

Validity testing since the measure was developed entails evaluating hypothesized relationships by computing correlation coefficients, ANOVAs, t-tests and estimating regression coefficients. Here we describe these associations as summarized from two systematic reviews.

Warshawsky & Havens (2011) report that the majority of the 37 studies associated the PES-NWI with organization (n = 16 studies), nurse outcomes (n = 23 studies), or patient outcomes (n = 16 studies). Studies reported nurse outcomes including, job satisfaction, intent to leave, burnout, and work engagement. Articles reported patient related outcomes including, patient satisfaction, and medication errors. Moreover, studies investigated organizational outcomes such as safety climate and morale. The results of these analyses are displayed in Warshawsky & Havens Table 4 on article pages 10 & 11. Swiger et al. (2017) report that the majority of the 46 studies they reviewed associated the PES-NWI with organization (n = 8 studies), nurse outcomes (n = 24 studies), or patient outcomes (n = 14 studies).

<u>Scores in Magnet and Non-Magnet Hospitals Demonstrating Discriminant Validity</u>

We hypothesize that work environments in Magnet hospitals, recognized for achieving excellent nursing standards, will have higher scores than work environments in non-Magnet hospitals. In this table, present the score ranges by Magnet status. In Table 2b1.3A we show studies where data were collected from nurses working in Magnet hospitals and non-Magnet hospitals. We show that scores were significantly higher in the Magnet facilities, demonstrating the continued discriminant ability of the instrument.

Table 2b1.3A.
Replication of Swiger et al., 2017 (Table 1):
Reported Score Ranges (n = 3 articles)

PES-NWI Measure	**Reported Mean Score Range (SD)**	
Subscale	Non-magnet scores	Magnet hospital scores
Nurse participation in hospital affairs	2.34-2.87	2.76-3.01
Nursing foundations for quality of care	2.82-3.11	3.09-3.20
Nurse manager ability, leadership, & support of nurses	2.41-3.00	2.72-3.07
Staffing and resource adequacy	2.07-2.62	2.65-2.88
Collegial nurse-physician relations	2.78-2.99	2.99-3.07
Composite	2.51-2.92	2.92-3.00

Additionally, of the 13 publications that reported PES-NWI composite scores studied by Warshawsky and Havens, the lowest score reported (2.48) was by acute care nurses working in non-Magnet hospitals in Pennsylvania (Lake, 2002). Furthermore, three studies reported positive correlations between PES-NWI scores and Magnet hospital recognition (Friese et al., 2005; Lake, 2002; Lake & Friese, 2006).

Studies noted above:

Kelly, L. A., McHugh, M. D., & Aiken, L. H. (2011). Nurse outcomes in Magnet® and non-Magnet hospitals. The Journal of nursing administration, 41(10), 428.

Kutney-Lee, A., Stimpfel, A.W., Sloane, D.M., Cimiotti, J.P., Quinn, L.W., Aiken, L.H., 2015. Changes in patient and nurse outcomes associated with magnet hospital recognition. Med. Care 53 (6), 550–557.

Ma, C., Park, S.H., 2015. Hospital magnet status, unit work environment, and pressure ulcers. J. Nurs. Scholarsh. 47 (6), 565–573

McHugh, M. D., Kelly, L. A., Smith, H. L., Wu, E. S., Vanak, J. M., & Aiken, L. H. (2013). Lower mortality in magnet hospitals. Medical care, 51(5), 382.

Walker, K., Middleton, S., Rolley, J., Duff, J., 2010. Nurses report a healthy culture: results of the Practice Environment Scale (Australia) in an Australian hospital seeking Magnet recognition. Int. J. Nurs. Pract. 16 (6), 616–623.

In Table 2b1.3B we note the hypothesized relationship with the various outcomes and report the studies linking the PES-NWI to those outcomes from the two systematic reviews. The last column shows the direction of the association (- or +) and the value of the coefficients. Evidence from the 35 studies published since the later systematic review is presented in Table 2b1.3D at the end of the document for ease of viewing.

Table 2b1.3B
Statistical evidence of associations between the PES-NWI and related constructs

Outcomes	Hypothesized relationship with PES-NWI	Research study	Statistical test value
Patient Record Outcomes			

30 day inpatient mortality	negative		
		Aiken et al (2008)	(-, OR = 0.91)
		Aiken et al (2011) b	(-, OR = 0.93)
		Cho et al (2014)	(-, OR = 0.52)
		Friese et al (2008)	- Mortality
		Nicely et al (2013)	(-, OR = 0.89)
		Kelly (2014)	(-, OR = 0.97)
30 day hospital readmission	negative		
		Gardner et al (2007)	— Hospitalizations
		Ma & Park (2015)	(-, OR = 0.97)
		McHugh et al (2016)	(-, OR = 0.84)
Complications	negative		
		Friese et al (2008)	-
Failure to rescue	negative		
		Aiken et al (2008)	(-, OR = 0.91)
		Aiken et al (2011) b	(-, OR = 0.93)
		Friese et al (2008)	- Failure to rescue
		Nicely et al (2013)	(-, OR = 0.90)
Discharged without breastmilk	negative		
		Lake (2016)	(-, OR= 0.92)
Percent of infants on unit discharged on breastmilk	positive		
		Hallowell et al (2016)	(+, β = 0.04); Adjusted R2 = 0.37
Nurse-reported (NR) Adverse Outcomes			
NR nosocomial infection	negative		
		Kutney-Lee et al (2009)	-
		Lake et al (2015)	(-, OR= 0.85)
		Spence Laschinger and Leiter (2006)	-
NR patient falls	negative		
		Cho et al (2016)	Falls with injury (-, OR = 0.68)
		Kutney-Lee, Lake, et al (2009)	- Falls with injury
		Prezerakos et al (2015)	All falls (-, OR= .02)
		Spence Laschinger & Leiter (2006)	- All falls
NR medication errors	negative		
		Cho et al (2016)	(-, OR=0.55)
		Manojlovich & DeCicco (2007)	-

		Spence Laschinger & Leiter (2006)	-
NR catheter-associated sepsis	negative		
		Manojlovich & DeCicco (2007)	-
NR pressure ulcer	negative		
		Cho et al (2016)	(-, OR = 0.61)
		Choi and Staggs (2014)	Unit acquired pressure ulcers SRA (-, OR = 0.78)
		Flynn et al (2010)	(-, β=0.37)
		Ma and Park (2015)	(-, OR= 0.73)
NR urinary tract infection	negative		
		Kelly (2013)	(-, OR= 0.80)
NR bloodstream infection	negative		
		Kelly (2013)	(-, OR=0.77)
NR pneumonia	negative		
		Kelly (2013)	(-, OR= 0.80)
NR central line infection	negative		
		Lake (2016)	(-, OR= 0.89)
Patient Satisfaction			
Patient safety climate	positive		
		Armstrong and Laschinger (2006)	+
		Armstrong et al (2009)	+
Perceived quality of care	positive		
		Gardner et al (2009)	+
Nurses communicated well	positive		
		Aiken et al (2012)	(+, OR=1.11)
		You et al (2013)	(+, OR= 1.30)
Patient rates hospital highly	positive		
		Aiken et al (2012)	(+, OR= 1.16)
		Kutney-Lee et al (2015)	(+, OR= 1.17)
		You et al (2013)	(+, OR= 1.29)
Patient satisfaction	positive		
		Boev (2012)	Patient satisfaction and Nurse Manager Ability and Support of Nurses (+, β= 0.424)
		Kutney-Lee, McHugh, et al (2009)	+ (HCAHPS)

	Tei-Tominaga and Sato (2016)	and NPR (+, OR= 0.144)

Table 2b1.3C below reports mean and range for percentage of patients who reported on the variables indicated and regression coefficient from a linear regression of the HCAHPS variable on the PES-NWI composite score.

Table 2b1.3C
Analysis for 2018 NQF measure maintenance Linking 2015 hospital-level data from the Hospital Consumer Assessment of Health Providers and Systems HCAHPS to the PES-NWI from the National Database of Nursing Quality Indicators (n = 390).

HCAPHS Measure	Measure definition	M	Range	β coefficient and 95% CI
Composite Measures				
Communication with nurses	Patients who reported that their nurses "Always" communicated well	79	63 – 93	9.75*** (7.65-11.86)
Responsiveness of hospital staff	Patients who reported that they "Always" received help as soon as they wanted	65	44 – 86	14.30*** (10.76-17.83)
Pain management	Patients who reported that their pain was "Always" well controlled	70	56 – 84	11.21*** (9.05-13.37)
Communication about medicines	Patients who reported that staff "Always" explained about medicines before giving it to them	64	53 – 81	11.27*** (8.88-13.67)
Discharge information	Patients who reported that YES, they were given information about what to do during their recovery at home	87	79 – 97	3.87*** (2.11-5.63)
Care Transition	Patients who "Strongly Agree" they understood their care when they left the hospital	52	33 – 69	14.60*** (11.74-17.45)
Global measures				
Overall rating of hospital	Patients who gave their hospital a rating of 9 or 10 on a scale from 0 (lowest) to 10 (highest)	70	50 - 95	18.98*** (15.06–22.90)
Willingness to recommend the hospital	Patients who reported YES, they would definitely recommend the hospital	73	44 – 98	20.73*** (16.20-25.27)

Note. N = 390 hospitals except for overall rating of hospital (n = 377 hospitals); ***p < .001

2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The results demonstrate that the measure exhibits satisfactory validity across a wide range of related constructs in many international samples across 16 years as well as in national 2015 data analyzed for measure re-endorsement.

2b2. EXCLUSIONS ANALYSIS
NA ☒ no exclusions — *skip to section 2b3*

2b2.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

____

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

Risk adjustment is not applicable

2b3.1. What method of controlling for differences in case mix is used?

☒ **No risk adjustment or stratification**

☐ **Statistical risk model with risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

Not applicable

2b3.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

The conceptual rationale for not controlling for differences in nurse characteristics is that nurse capacity to assess aspects of the work environment does not depend on nurse age, sex, or educational level. All nurses in direct clinical care positions are ideally positioned to make these assessments.

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

☐ **Published literature**

☐ **Internal data analysis**

☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b3.9

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

2b3.9. Results of Risk Stratification Analysis:

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

2b3.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

____

2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE
2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*
The method was to provide descriptive statistics at the level of the measured entities (hospitals or nursing units) showing mean, standard deviation, and range.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly* *different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Warshawsky & Havens (2011) reported PES-NWI scores on a 4-point Likert Scale across 22 studies. The theoretical range is from 1.00 to 4.00. The composite score range was reported as 2.48 to 3.17. The subscale score ranges are demonstrated in Table 3 of Warshawsky & Havens, replicated here:

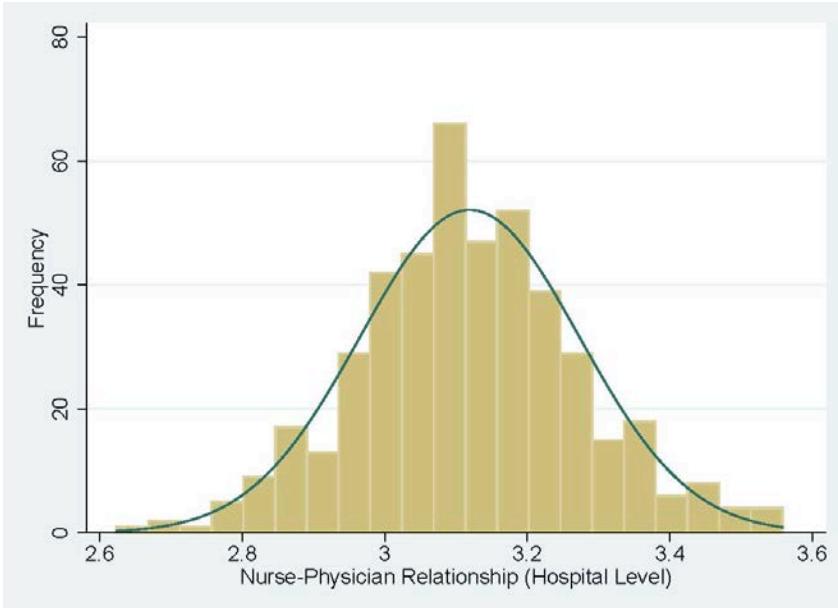
Measure	Score Range
Subscale	
Collegial Nurse-Physician Relations	2.32-3.26
Nursing Foundations for Quality Care	2.20-3.35
Nurse Manager Ability, Leadership, and Support	2.08-3.42
Nurse Participation in Hospital Affairs	1.98-2.90
Staffing and Resource Adequacy	1.87-2.90
Composite	2.48-3.17

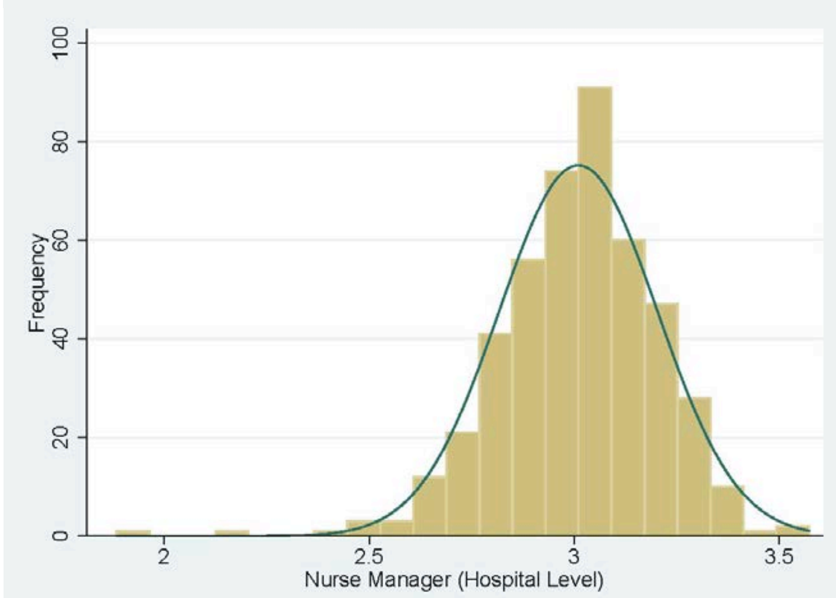
Warshawsky, N. E., & Havens, D. S. (2011). Global use of the practice environment scale of the nursing work index. Nursing research, 60(1), 17.

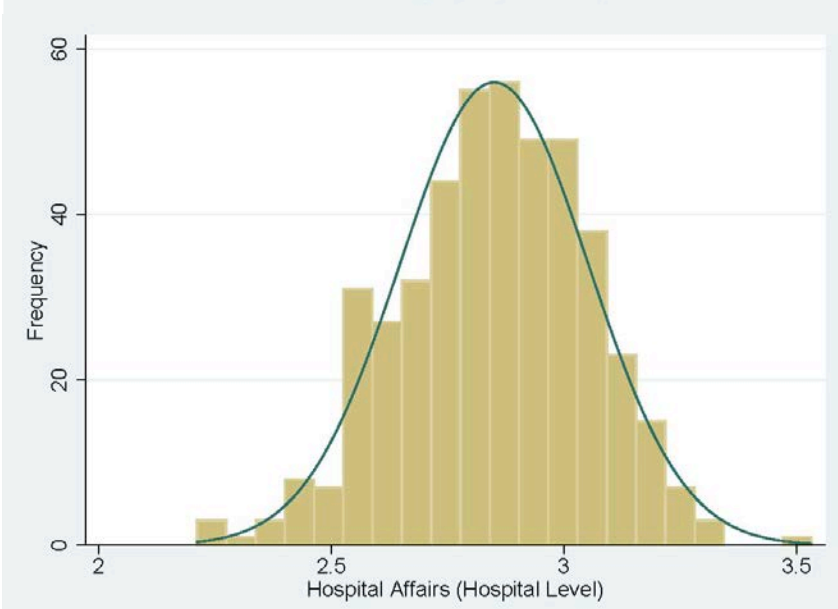
In a 2017 review of the PES-NWI measure (Swiger), sixteen articles reported composite scores ranging from 2.30 to 3.07 based on the 4-point Likert scale. Composite scores showed meaningful variation. Like in Warshawsky & Havens (2011), the Staffing and Resource Adequacy subscale remains the lowest range for hospitals.

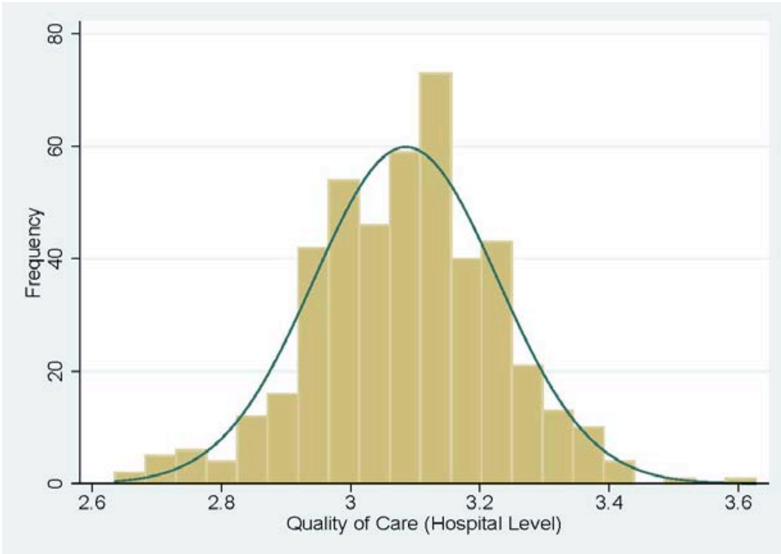
Swiger, P. A., Patrician, P. A., Miltner, R. S. S., Raju, D., Breckenridge-Sproat, S., & Loan, L. A. (2017). The Practice Environment Scale of the Nursing Work Index: an updated review and recommendations for use. International journal of nursing studies, 74, 76-84.

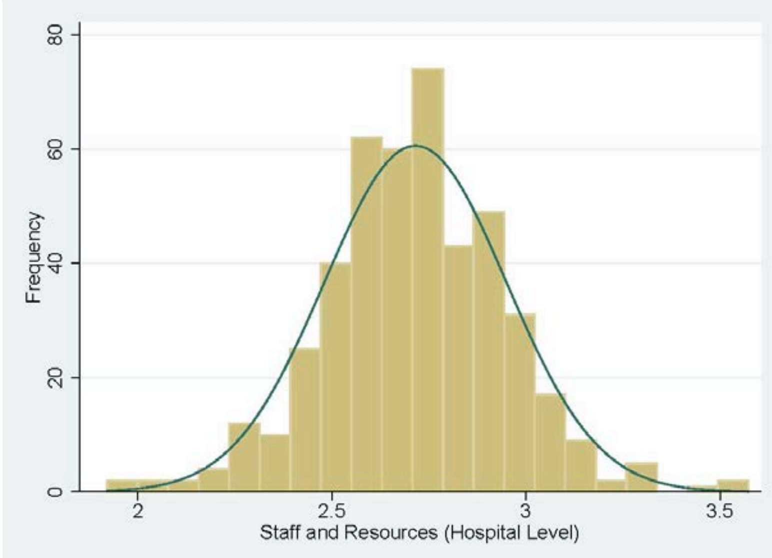
Analysis for 2018 NQF measure maintenance: Density plots displayed below for each subscale and composite measure of the PES-NWI provided from 2015 NDNQI data of 452 hospitals provide further insight to the meaningful differences in measure scores. The differences in the distributions across subscales show that they provide meaningful measures for comparison across hospitals of constructs that may be targets for institutional improvements.

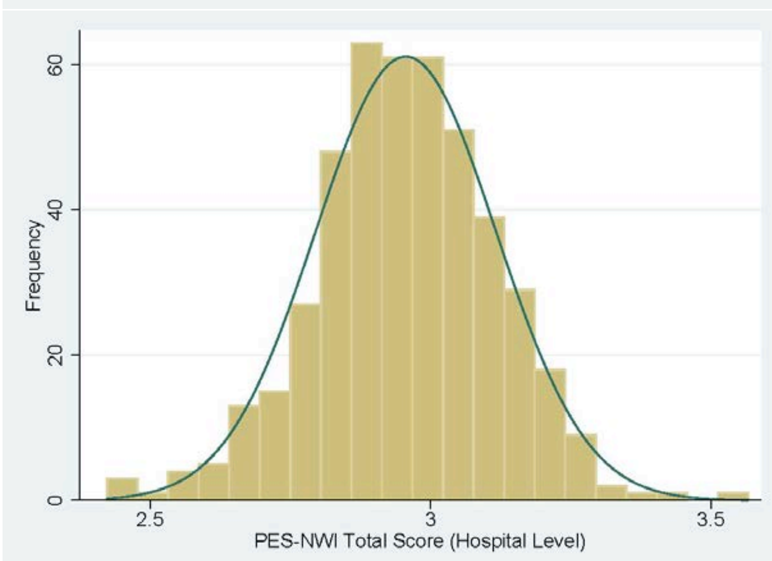

Frequency
Nurse-Physician Relationship (Hospital Level)


Frequency
Nurse Manager (Hospital Level)


Frequency
Hospital Affairs (Hospital Level)


Frequency
0
20
40
60
80
2.6
2.8
3
3.2
3.4
3.6
Quality of Care (Hospital Level)


Frequency
0
20
40
60
80
2
2.5
3
3.5
Staff and Resources (Hospital Level)


Frequency
0
20
40
60
2.5
3
3.5
PES-NWI Total Score (Hospital Level)

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i*.e., what do the results mean in terms of statistical and meaningful differences?*)

There are consistent statistically significant and clinically meaningful differences in performance across measured entities.

A unique study measured changes in the PES-NWI composite score in a panel of hospitals from 1999 to 2006 (Kutney-Lee et al. 2013). This study demonstrates that work environments can change over time, which provides the basis for improving work environments in order to enhance quality of care and patient outcomes. The study also demonstrated that improvements in work environments had a strong negative association with changes in rates of job dissatisfaction, nurse burnout, and intention to leave the job. These are the relationships that have been observed in cross-sectional studies. The finding in a longitudinal design enhances the causal basis for this structural element to influence care quality and nurse and patient outcomes.

Kutney-Lee, A., Wu, E. S., Sloane, D. M., & Aiken, L. H. (2013). Changes in hospital nurse work environments and nurse job outcomes: an analysis of panel data. *International journal of nursing studies*, *50*(2), 195-201.

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS
If only one set of specifications, this section can be skipped.

Note*: This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***
Not applicable

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i*.e., what do the results mean and what are the norms for the test conducted*)

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Analysis for 2018 NQF measure maintenance: Missing data were calculated for the 31 items that comprise the measure.

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Statistics from 2015 NDNQI nurse survey data: For each of the 31 items: At the respondent level: less than 1% of respondents have missing data. At the hospital level, about 90% of hospitals have less than 4% of their respondents with missing data.

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Our interpretation is that missing data is minimal and appears to be at random. Therefore, performance results would be non-biased.

Table 2b1.3D

Reference	# hospitals	# nurses	Outcome measure	Reliability (Cronbach's alpha)
Yan P, Yang Y, Zhang L, et al. Correlation analysis between work-related musculoskeletal disorders and the nursing practice environment, quality of life, and social support in the nursing professionals. Medicine. 2018;97(9):e0026.	12 hospitals	2170 nurses	Work related musculoskeletal disorders	0.91 for composite 0.67-0.79 for the subscale Retest reliability was 0.84 Content validity was 0.94
Wu Y, Zheng J, Liu K, et al. The associations of occupational hazards and injuries with work environments and overtime for nurses in China. Res Nurs Health. 2018.	111 medical/surgical units in 23 hospitals	1517 nurses	Occupational hazards and injuries	0.96 for composite 0.79-0.93 for the subscales
Wan Q, Zhou W, Li Z, Shang S, Yu F. Work engagement and its predictors in registered	10-15 units in 3	1065 registe	Work engagement	0.89 for composite

nurses: A cross-sectional design. Nurs Health Sci. 2018.	specialized hospitals	red nurses		0.60-0.75 for the subscales
Swiger PA, Loan LA, Raju D, Breckenridge-Sproat ST, Miltner RS, Patrician PA. Relationships between Army nursing practice environments and patient outcomes. Res Nurs Health. 2018;41(2):131-144	45 units in 10 Army hospitals	1,710 of all nurse types	Patient outcomes (falls, medication errors, etc.)	0.94-0.95 for the composite 0.79-0.91 for the subscale
Smith JG, Morin KH, Lake ET. Association of the nurse work environment with nurse incivility in hospitals. J Nurs Manag. 2018;26(2):219-226.	5 acute care hospitals	233 staff nurses	Work incivility	0.94 for the composite 0.83-0.86 for the subscales
Newhouse R, Byon HD, Storkman Wolf E, Johantgen M. Multisite Studies Demonstrate Positive Relationship Between Practice Environments and Smoking Cessation Counseling Evidence-Based Practices. Worldviews Evid Based Nurs. 2018;15(3):217-224.	45 hospitals	844 registe red nurses	Nurse smoking cessation counseling practices	no
Nelson-Brantley HV, Park SH, Bergquist-Beringer S. Characteristics of the Nursing Practice Environment Associated With Lower Unit-Level RN Turnover. The Journal of nursing administration. 2018;48(1):31-37.	1002 adult care units in 162 NDNQI hospitals	Does not report	RN turnover	0.82 for the composite α≥ 0.80 for the subscales, with the exception of the interprofessional relations subscale (α= 0.71)
Moreno-Casbas MT, Alonso-Poncelas E, Gomez-Garcia T, Martinez-Madrid MJ, Escobar-Aguilar G. Perception of the quality of care, work environment and sleep characteristics of nurses working in the National Health System. Enferm Clin. 2018.	7 hospitals	635 registe red nurses	Measure relationship between ward and work shift with nurses' perception their work environmen t, and sleep quality	no
Hiler CA, Hickman RL, Jr., Reimer AP, Wilson K. Predictors of Moral Distress in a US Sample of Critical Care Nurses. American journal of critical care : an official publication, American Association of Critical-Care Nurses. 2018;27(1):59-66.	Not reported	328 critical care nurses	Moral distress	0.71-0.84 for the composite α≥ 0.70 for all subscales

Gea-Caballero V, Castro-Sanchez E, Juarez-Vela R, Diaz-Herrera MA, de Miguel-Montoya I, Martinez-Riera JR. Essential elements of professional nursing environments in Primary Care and their influence on the quality of care. Enferm Clin. 2018;28(1):27-35.	Not reported	144 nurses	Evaluates the characteristics of nursing environments in primary care settings	No
Cho H, Han K. Associations Among Nursing Work Environment and Health-Promoting Behaviors of Nurses and Nursing Performance Quality: A Multilevel Modeling Approach. Journal of nursing scholarship : an official publication of Sigma Theta Tau International Honor Society of Nursing. 2018.	57 units in 5 hospitals	432 nurses	Health promoting behaviors of hospital nurses	0.72-0.81 for the subscales
Al-Maaitah R, AbuAlRub RF, Al Blooshi S. Practice environment as perceived by nurses in acute care hospitals in Sharjah and North Emirates. Nursing forum. 2018;53(2):213-222.	10 hospitals	450 nurses	Nurses' perceptions of their practice environment	0.90 for the composite
Akter N, Akkadechanunt T, Chontawan R, Klunklin A. Factors predicting quality of work life among nurses in tertiary-level hospitals, Bangladesh. Int Nurs Rev. 2018;65(2):182-189	6 tertiary-level hospital	288 registered nurses	Level of quality of work life	0.90 for the composite
Zhang L, Wang A, Xie X, et al. Workplace violence against nurses: A cross-sectional study. Int J Nurs Stud. 2017;72:8-14	28 hospitals	3835 clinical nurses	Workplace violence	0.921 for the composite
Swiger PA, Raju D, Breckenridge-Sproat S, Patrician PA. Adaptation of the Practice Environment Scale for military nurses: a psychometric analysis. J Adv Nurs. 2017;73(9):2219-2236	42 US military treatment facilities	2608 nurses	Psychometric analysis	0.96 for the composite 0.81-0.90 for the subscales
Swiger PA, Patrician PA, Miltner RSS, Raju D, Breckenridge-Sproat S, Loan LA. The Practice Environment Scale of the Nursing Work Index: An updated review and recommendations for use. Int J Nurs Stud. 2017;74:76-84			46 articles published were reviewed in study	
Numminen O, Leino-Kilpi H, Isoaho H, Meretoja R. Development of Nurses' Professional Competence Early in Their Career: A Longitudinal Study. Journal of	Not reported	318 nurses	Examine competence development in nurses	0.77 to 0.86 for subscales (reports Lake, 2002)

continuing education in nursing. 2017;48(1):29-39				
Nantsupawat A, Kunaviktikul W, Nantsupawat R, Wichaikhum OA, Thienthong H, Poghosyan L. Effects of nurse work environment on job dissatisfaction, burnout, intention to leave. Int Nurs Rev. 2017;64(1):91-98	43 inpatient units in 5 university hospitals	1351 nurses	Association between work environment and nurse reported job dissatisfaction, burnout and intention to leave	0.85-0.91 for subscales (reports Nantsupawt et al, 2011)
Liu J, Zhou H, Yang X. Evaluation and Improvement of the Nurse Satisfactory Status in a Tertiary Hospital using the Professional Practice Environment Scale. Medical science monitor : international medical journal of experimental and clinical research. 2017;23:874-880	Not reported	1050 nurses	Associated factors influencing satisfaction	No
Hussein R, Everett B, Ramjan LM, Hu W, Salamonson Y. New graduate nurses' experiences in a clinical specialty: a follow up study of newcomer perceptions of transitional support. BMC Nurs. 2017;16:42	1 teaching hospital	87 new graduate nurses	Examine change in graduate nurses' perception	0.91 for the composite
Hallowell SG, Rogowski JA, Lake ET. How Nurse Work Environments Relate to the Presence of Parents in Neonatal Intensive Care. Advances in neonatal care : official journal of the National Association of Neonatal Nurses. 2017	104 US NICUs	6060 registered nurses	Infants whose parents were present during the NICU shift	No
Gasparino RC, Guirardello EB. Validation of the Practice Environment Scale to the Brazilian culture. J Nurs Manag. 2017;25(5):375-383	Not reported	209 nurses	Psychometric analysis of Brazilian version	0.86 for the composite 0.76-0.87 for the subscales
Elmi S, Hassankhani H, Abdollahzadeh F, Jafar Abadi MA, Scott J, Nahamin M. Validity and Reliability of the Persian Practice Environment Scale of Nursing Work Index. Iranian journal of nursing and midwifery research. 2017;22(2):106-111	Not reported	350 nurses	Psychometric analysis of Persian version	0.935 for the composite 0.70-0.92 for the subscales
Casalicchio G, Lesaffre E, Kuchenhoff H, Bruyneel L. Nonlinear Analysis to Detect if Excellent Nursing Work Environments Have Highest Well-Being. Journal of nursing	2184 nursing units in 489 hospitals	33731 registered nurses	Burnout	No

scholarship : an official publication of Sigma Theta Tau International Honor Society of Nursing. 2017;49(5):537-547				
Bruyneel L, Li B, Squires A, et al. Bayesian Multilevel MIMIC Modeling for Studying Measurement Invariance in Cross-group Comparisons. Med Care. 2017;55(4):e25-e35	87 nursing units in a single institution	87 nurse managers	Comparing and evaluating measurement invariance	No
Al-Hamdan Z, Manojlovich M, Tanima B. Jordanian Nursing Work Environments, Intent to Stay, and Job Satisfaction. Journal of nursing scholarship : an official publication of Sigma Theta Tau International Honor Society of Nursing. 2017;49(1):103-110.	Not reported	582 registered nurses	Intent to stay and job satisfaction	0.92 for the composite
Yokoyama M, Suzuki M, Takai Y, Igarashi A, Noguchi-Watanabe M, Yamamoto-Mitani N. Workplace bullying among nurses and their related factors in Japan: a cross-sectional survey. J Clin Nurs. 2016;25(17-18):2478-2488	Not reported	825 nurses	Workplace bullying	0.75-0.84 for the subscsales
Schwendimann R, Dhaini S, Ausserhofer D, Engberg S, Zuniga F. Factors associated with high job satisfaction among care workers in Swiss nursing homes - a cross sectional survey study. BMC Nurs. 2016;15:37	162 nursing homes	4,145 care workers	Job satisfaction	0.74-0.89 for subscales
Roche MA, Duffield C, Friedman S, Twigg D, Dimitrelis S, Rowbotham S. Changes to nurses' practice environment over time. J Nurs Manag. 2016;24(5):666-675	6 acute care hospitals	1605 nurses	To examine changes in the practice environment	0.82 for the composite 0.70-0.85 for the subscales
Hussein R, Everett B, Hu W, et al. Predictors of new graduate nurses' satisfaction with their transitional support programme. J Nurs Manag. 2016;24(3):319-326	Not reported	109 new graduate nurses	Satisfaction with transitional support program	0.91 for the composite
Gomez-Garcia T, Ruzafa-Martinez M, Fuentelsaz-Gallego C, et al. Nurses' sleep quality, work environment and quality of care in the Spanish National Health System: observational study among different shifts. BMJ Open. 2016;6(8):e012073	7 hospitals	635 registered nurses	Nurses sleep quality and quality of care	No

Duffield C, Roche M, Twigg D, Williams A, Clarke S. A protocol to assess the impact of adding nursing support workers to ward staffing. J Adv Nurs. 2016;72(9):2218-2225	20 pairs of matched wards	No	Protocol to asses the impact of adding nurse support workers	No
Brzyski P, Kozka M, Squires A, Brzostek T. How Factor Analysis Results May Change Due to Country Context. Journal of nursing scholarship : an official publication of Sigma Theta Tau International Honor Society of Nursing. 2016;48(6):598-607	30 hospitals	2605 registered nurses	PES-NWI changes in the country context	0.72-0.89 for the subscales
Brooks-Carthon, J. M., Lasater, K. B., Rearden, J., Holland, S., & Sloane, D. M. (2016). Unmet nursing care linked to rehospitalizations among older Black AMI patients: A cross-sectional study of US hospitals. Medical care, 54(5), 457.	253 acute care hospitals	14879 registered nurses	Variable all-cause readmissions	No