Received: 24 November 2021 | Accepted: 22 March 2022

DOI: 10.1111/2047-3095.12366

INTERNATIONAL JOURNAL OF NURSING KNOWLEDGE
The Official Journal of NANDA International, Inc.

ORIGINAL ARTICLE

Evaluation of the NANDA International, Inc. diagnostic classification in Spain: Development and validation of the EVALUAN-I tool

Claudio Alberto Rodríguez-Suárez PhD Candidate, BSc, MSc, RN [1] |
María Naira Hernández-De Luis BSc, MSc, RN [2] |
María Isabel Mariscal-Crespo PhD, BSc, RN [3] |
Rafaela Camacho-Bejarano PhD, BSc, MSc, RN [3]

[1]RN Maternal and Child Insular University Hospital Complex (Canary Islands Health Service), Associate Professor Faculty of Health Sciences, University of Las Palmas de Gran Canaria, Canary Islands, Las Palmas, Spain

[2]RN Doctoral Primary Care Facility, Primary Healthcare Management, Gran Canaria (Canary Islands Health Service), Canary Islands, Las Palmas, Spain

[3]Full Professor, Faculty of Nursing, University of Huelva, Huelva, Andalusia, Spain

Correspondence
Claudio Alberto Rodríguez-Suárez, RN Maternal and Child Insular University Hospital Complex (Canary Islands Health Service), Associate Professor Faculty of Health Sciences, University of Las Palmas de Gran Canaria, Las Palmas, Canary Islands, Spain.
Email: c.rodriguez@celp.es

Funding information
Fundación Canaria Instituto de Investigación Sanitaria de Canarias (FIISC), Grant/Award Number: ENF-19/33

Abstract

Purpose: To develop and validate a tool to evaluate the NANDA International, Inc. diagnostic classification.

Methods: The tool, EVALUAN-I, was validated in a non-probabilistic sample (N = 460) on Spanish Registered Nurses (September–December 2019) in two phases. First, design and construct the instrument in three steps: (1) literature review to define the construct focusing on the orientation toward nursing concepts and theoretical foundations, the level of scientific evidence, the structural configuration, the applicability, the nurses' clinical reasoning skills, and the attitudes toward nursing diagnosis, (2) substantiation of the questionnaire items and design according to the criteria for a diagnostic classification, (3) expert test to establish the face validity and content validity. The second phase revolved around (4) conducting a pilot test and measuring the temporal stability (test-retest) and Cohen's kappa coefficient; assessing psychometric properties by measuring (5) reliability (internal consistency using Cronbach alpha and interfactor correlation) and (6) construct validity (exploratory and confirmatory factor analysis). The manuscript follows the STROBE checklist. The study was approved by the Research Ethics Committee with registration number 2019-190-1.

Findings: EVALUAN-I displayed moderate test-retest stability, adequate construct validity, and excellent reliability. The confirmatory factor analysis provided evidence about the configuration of EVALUAN-I in relation to nine analytical dimensions: clinical competence, nurses' reasoning skills, attitudes towards nursing diagnosis, discipline's central concepts, classification's contents, physiopathological attributes, level of scientific evidence, diagnostic precision, and conceptual correspondence between terminologies.

Conclusions: EVALUAN-I is a valid and reliable instrument, which can be used to improve the epistemological, normative, and intuitive configuration of NANDA International, Inc. in a structured, systematic manner.

Implications for Nursing Practice: Comprehensive evaluation of NANDA International, Inc. in different clinical settings around the world using a validated instrument, like EVALUAN-I, would allow strengths and weaknesses to be identified and contribute to the classification's development and practical application.

KEYWORDS
nursing diagnosis, nursing methodology research, psychometrics, standardized nursing terminology, surveys and questionnaires, validation study

Resumen

Objetivo: Diseñar y validar un instrumento de medida para evaluar la taxonomía diagnóstica NANDA International, Inc.

Métodos: La herramienta, EVALUAN-I, ha sido validada con una muestra no probabilística (N = 460) de enfermeras españolas (septiembre-diciembre 2019) en dos fases. Primera, diseño y construcción del instrumento en tres etapas: (1) definición del constructo a través de una revisión bibliográfica centrada en la orientación hacia los conceptos y fundamentos teóricos de la enfermería, el nivel de evidencia científica, la configuración estructural, la aplicabilidad, las habilidades de razonamiento clínico de las enfermeras y las actitudes hacia los diagnósticos enfermeros, (2) fundamentación normativa de los criterios que debe reunir una clasificación diagnóstica y (3) prueba de expertos para establecer la validez aparente y validez de contenido. La segunda fase se centra en (4) el desarrollo de un pilotaje y medición de la estabilidad temporal a través del test-retest y cálculo del coeficiente Kappa de Cohen; evaluación de las propiedades psicométricas a través de la (5) fiabilidad (consistencia interna mediante Alfa de Cronbach y correlación interfactorial) y (6) validez de constructo (análisis factorial exploratorio y confirmatorio). El manuscrito se ajusta a los criterios STROBE. El estudio fue aprobado por el Comité de Ética de la Investigación con el número de registro 2019-190-1. [Correction added on 20 April 2022, after first online publication: In point (4) of the 'Methods' section, "Kappa de ohen" has been corrected to "Kappa de Cohen" in this version.]

Resultados: EVALUAN-I mostró moderada estabilidad test-retest, adecuada validez de constructo y excelente fiabilidad. El análisis factorial confirmatorio mostró evidencias acerca de la configuración de EVALUAN-I en 9 dimensiones de análisis: competencia clínica, aptitudes para el razonamiento diagnóstico, actitudes ante el diagnóstico enfermero, conceptos centrales de la disciplina, contenidos de la clasificación, atributos fisiopatológicos, nivel de evidencia científica, precisión diagnóstica y comparativa conceptual entre terminologías.

Conclusiones: EVALUAN-I constituye un instrumento válido y fiable para mejorar la configuración epistemológica, normativa e intuitiva de NANDA International, Inc. de una manera estructurada y sistematizada.

Implicaciones para la práctica clínica: La valoración global de la taxonomía NANDA International, Inc. en los distintos contextos clínicos a nivel internacional a través de una herramienta validada como EVALUAN-I permitiría identificar las fortalezas y debilidades necesarias para contribuir a su desarrollo y aplicación práctica.

INTERNATIONAL JOURNAL OF NURSING KNOWLEDGE
The Official Journal of NANDA International, Inc.

BACKGROUND

Clarifying and refining nursing languages is a central pillar in the development of nursing as a scientific discipline. NANDA International, Inc. (NANDA-I) helps nurses to develop and refine a terminology that precisely reflects their clinical judgment with the aim of improving patient care by encouraging the development, dissemination, and use of a standardized lexicon for clinical decision-making (NANDA International, 2019).

Although NANDA-I is the most widely used classification internationally (Tastan et al., 2014), it has not yet undergone formal evaluation. To improve NANDA-I as a classification system, it is important to carry out a comprehensive analysis, including the following aspects: epistemological basis (orientation toward nursing concepts and theoretical foundations), normative facets (structural configuration and scientific evidence), and intuitive considerations (applicability, nurses' clinical reasoning, and attitudes) (Beckstead, 2009; Juvé i Udina, 2012; Müller-Staub et al., 2007; Von Krogh, 2008).

Epistemological basis

According to Gordon's precepts (1996), coherence must be maintained between the labels and their definitions in a diagnostic classification, and the diagnostic indicators must be refined. This includes the defining characteristics describing the concepts, and adaptation of the related factors in etiological and competency-based terms to allow medical problems to be solved. Mere knowledge of the label and definition of the nursing diagnosis (ND) is insufficient for a clinical judgment to be established; diagnostic indicators (defining characteristics and related factors, or risk factors) must also be identified.

From this perspective, nurses must be able to solve problems identified as ND autonomously. In addition, their ability to describe and categorized the discipline's conceptual phenomena from the perspective of different theoretical currents in nursing must be examined (Von Krogh, 2008).

Normative facets

NANDA-I's development, promotion, adaptation, and international recognition have been made possible by the approval expressed by a range of organizations and institutions (Herdman, 2011) such as the International Organization for Standardization, the Reference Terminology Model for Nursing, the United States National Library of Medicine, the Committee for Nursing Practice Information Infrastructure (Rutherford, 2008), the United States Department of Health and Human Services, the United States Consolidated Health Informatics initiative, the United Kingdom's National Health Service, and Health Level Seven International. Given this broad approval, it is possible that NANDA-I meets most of the criteria applicable to a classification system (Müller-Staub et al., 2007).

In other words, it should feature a hierarchical structure with internal coherence (Beckstead, 2009) deriving from the interwoven nature of its components. This morphology allows phenomena to be grouped, ordered, and codified during the development of first-level theories (Gordon, 1996) with a logical system for distributing the concepts and content that is determined by affinity. An adequately structured taxonomic system should be made up of mutually exclusive, exhaustive, and homogeneous categories (Von Krogh, 2008). In the case of NANDA-I, the hierarchical structure encompasses *domains* where phenomena with common characteristics are grouped in an attempt to classify them clearly, concisely, and consistently (Herdman & Kamitsuru, 2015). The *classes* order the phenomena in a single domain into groups with common attributes and characteristics. Finally, the *concepts* correspond to the individual phenomenon itself, requiring precision, expressivity, and semantic consistency (Rosenbloom, et al., 2008) to minimize any ambiguity (Stallinga et al., 2015).

Overall, NANDA-I is made up of 13 domains organized into 47 classes. The 2018–2020 edition includes 244 labels (Herdman & Kamitsuru, 2019), whereas the 2021–2023 edition contains 267 labels (Herdman et al., 2021) covering three types of diagnoses: *Problem-focused diagnosis*, *Risk diagnosis*, and *Health promotion diagnosis*.

Intuitive considerations

In order for users to apply a taxonomy successfully, it must be internally coherent, intuitive, and reflect its intended universe of situations. Without these characteristics, the taxonomy will be of limited value (Beckstead, 2009). The cognitive suitability of NANDA-I is determined by acceptance from the professionals who use it and its functional utility for clinical practice through training and research.

The diagnostic phase is complex, giving rise to divergences and discrepancies (Frazão et al., 2015). Exploring opinions of these taxonomies is key to improving them. For example, Krenz and Lunney developed an instrument to measure attitudes toward ND. Position on ND (PND) is a questionnaire that uses the semantic differential technique to measure and encourage the expression of attitudes toward the concepts (Romero-Sánchez et al., 2013). It has been applied in a variety of cultural contexts including Italy, USA, Brazil, Japan, Jordan, and Spain (Abed El-Rahman et al., 2017; Collins, 2013; D'Agostino et al., 2016; de Guedes et al., 2013; Hasegawa et al., 2007; Oliva et al., 2005; Romero-Sánchez et al., 2013). Similar attitudes toward its application have been observed in the different contexts: when ND are used more frequently, positive attitudes and diagnostic precision become more prevalent (Collins, 2013).

ND may be viewed as a solution to the urgent need to identify, prioritize, and evaluate medical situations that may be dealt with thoroughly and autonomously by nurses. In this sense, NANDA-I is a world-leading diagnostic classification that requires comprehensive evaluation to identify its strengths and weaknesses. Using validated tools, such as EVALUAN-I (Supporting information File S1), to evaluate NANDA-I could improve its clinical applicability.

Therefore, the following research question has been formulated: Is EVALUAN-I a valid tool to evaluate NANDA-I clinical applicability?

AIM

To develop and validate a measurement tool to evaluate the NANDA-I diagnostic classification.

METHODS

A psychometric validation study of the EVALUAN-I tool was carried out in two phases with a total of six steps. The manuscript follows the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement (Supporting information File S2).

Phase 1. Design and construction of the tool (three steps) .

Step 1. Definition of the construct: A literature review was carried out in PubMed, Web of Science, Scopus, CINAHL, and LILACS using the descriptors MeSH/DeCS: Standardized Nursing Terminology and ND, without time limit and establishing linguistic filter to English, Spanish, and Portuguese. Searches were conducted by the principal investigator and subsequently reviewed by the research team. Studies mentioning NANDA-I in their title or abstract were included for full reading without excluding types of scientific methodology. The selection of articles was performed using the Critical Appraisal Skills Programme tools. Information extracted from the included studies addressed the conceptual correspondence of NANDA-I with the orientation toward nursing concepts and theoretical foundations (Goh et al., 2020; Hine-Sanabria et al., 2018; Mendes et al., 2015; Von Krogh, 2008), the level of scientific evidence and structural configuration (Beckstead, 2009; De Souza Oliveira-Kumakura et al., 2018; Miguel et al., 2016), the applicability of ND and nurses' clinical reasoning skills (Cruz et al., 2009; Mynaříková & Žiaková, 2014; Paans et al., 2012), and attitudes toward ND (Bittencourt & Crossetti, 2013; D'Agostino et al., 2018).

Step 2. Substantiation of the questionnaire items and design: The measurement tool's configuration was based on the criteria established by Müller-Staub et al. (2007) with regard to the standards to be met by a diagnostic classification to: exhaustively describe the knowledge base explaining nursing responsibilities, classify diagnoses using transparent procedures and clearly established criteria, and establish a precise description with valid diagnostic criteria, key defining characteristics, and etiologies allowing diagnosis to be distinguished from one another. The criteria established by Juvé i Udina (2012) for evaluating scientific vocabularies for nursing practice were also taken into consideration including orientation toward nursing concepts, theoretical rationale, scientific evidence, content validity, criterion validity, real use in practical settings, homogeneity and structural coherence, contextualization of knowledge, synonyms, attributes, coding system, and equivalence with other lexicons. Based on the criteria described above for diagnostic classifications, the instrument was developed with 30 items structured in six sections: (1) orientation toward nursing concepts and theoretical foundations, (2) level of scientific evidence, (3) structural configuration, (4) applicability, (5) nurses' clinical reasoning skills, and (6) attitudes toward ND on an adaptation of the abbreviated version of the PND Spanish version scale (Romero-Sánchez et al., 2013). As the reliability increases when the number of items rise from five to seven without any significant impact on the statistical properties or introduction of centrality bias (Matas, 2018), the measuring scale is a likert response graduated into 6 items, 1 (strongly disagree) is the most negative opinion and 6 (strongly agree) is the most positive one; the average value is 3.5 points.

Step 3. Expert test: The face validity and content validity (Carvajal, Centeno, Watson, Martínez, & Sanz Rubiales, 2011) were evaluated by a group of experts on the study topic; for this purpose, five PhD nurses working in academia and research at different Universities in Spain, with knowledge and experience in the design of measurement instruments, were purposively selected. Two of these experts also had professional links and experience in the field of clinical practice. Each expert was sent a draft of the questionnaire and was asked individually whether each item was understandable and relevant, asking them suggestions for improvement. Subsequently, the construct was modified based on the experts' suggestions. After accepting the suggestions for improvement made by the experts, a second round was carried out through a focus group discussion via a joint online meeting to reach a consensus on the implemented modifications. The construct was finally modified and the analytical dimensions were redefined.

Phase 2. Pilot test and evaluation of the tool's psychometric properties (three steps).

Once the final version of the instrument had been compiled, a variety of statistical tests were used to evaluate its psychometric properties (Martín, 2004).

Step 4. Pilot test: A pilot test was carried out with a representative sample of 27 nurses with different sociodemographic profiles, reflecting the characteristics of the study population. They all gave reasoned answers without substantial changes, so they were subsequently included in the population sample. Fifteen days after the first interview, the temporal stability was assessed through test-retest (n = 14) reliability using Cohen's kappa coefficient.

Step 5. Reliability: First, the internal consistency for the instrument as a whole and for each of the constructs was determined using Cronbach alpha. The consistency of internal homogeneity was then established using interfactor Pearson's correlation coefficient.

Step 6. Construct validity: An exploratory factor analysis (EFA) was carried out with the first half of the sample (n = 230) after testing its feasibility through the adequacy of the sample size with the Kaiser-Meyer Olkin index (KMO) and Bartlett's test of sphericity. The EFA aims to discover an underlying structure that defines common latent dimensions, which explain most of the variance observed in the set of variables, using principal components analysis (PCA) (López-Aguado & Gutiérrez-Provecho, 2019). Second, a confirmatory factor analysis (CFA) was carried out with the total sample (n = 460). The purpose of the CFA is to verify the item structure by means of the factor correlation and the variance explained by each factor, as well as the measurement errors. In this sense, three factor models have been developed: the first one derives from the principal components structure, the second one estimates the measurement error without altering the factor structure, and the third one proposes a modification to the initial structure of the EFA (the variable *10_Contextualises care* was added to Construct 7). To verify the fit of each factorial model, the ratio between the number of degrees of freedom and Chi-square was used (good fit if <4.00). However, given the limitations of Chi-square, the following fit indicators have been developed (Aldás & Uriel, 2017):

- *Root mean square residual* (RMR): measures the probability that the sample and estimated variance and covariance matrices are the same (good fit if <0.08).
- *Root mean square error of approximation* (RMSEA): correction to RMR and Chi-square when introducing the degrees of freedom in the calculation (good fit if <0.08).
- *Goodness of fit index* (GFI): determines the proportion of total variance explained by the model (good fit if >0.90).
- *Adjusted goodness of fit index* (AGFI): introduces a correction factor through the degrees of freedom of the model to the GFI indicator (good fit if >0.90).
- *Non-normed fit index* (NNFI): evaluates the proportion of improvement in fit achieved with the model considered respect to the improvement that could be achieved with the good fit model cited (good fit if >0.90).
- *Normed fit index* (NFI): evaluates the decrease of the Chi-square value in the proposed model respect to a base model (good fit if >0.90).
- *Tucker-Lewis index* (TLI): incorporates a correction factor through the degrees of freedom of the model to the NFI indicator (good fit if >0.90).
- *Comparative fit index* (CFI): stabilizes the correction of the NFI and TLI values (good fit if >0.90).

Samples

The study targeted all nurses (N = 316,094) in Spain in 2019 (Instituto Nacional de Estadística [Spanish National Institute of Statistics], 2019), who were distributed across 17 autonomous communities and 2 autonomous cities. The instrument was applied to a sample of nurses with different professional profiles from 16 of these autonomous communities.

Nurses with official university qualifications recognized in Spain (Diploma of Higher Education and Bachelor's degree in Nursing) who worked in different professional settings and contexts (clinical, management, and academic) were included. Nurses with less than 1 year of experience and professionals who had retired were excluded.

Data collection

Sociodemografic details were also collected including the following variables: sex, age, work experience, level of education, professional setting, professional role, and intensity of use of classification systems. A self-report questionnaire was used to distribute the instrument. This was disseminated as a form accessed via a web link or a QR code via email, mobile applications, and social media during September and December 2019.

Data analysis

Descriptive analysis, mean, and standard deviation were calculated for quantitative variables; frequency and percentage were calculated for qualitative variables. The normality of the data was checked using the Kolmogórov-Smirnov test. A p-value <0.05 was significant. R® version 3.6.3 (Lavaan package) was used for the statistical analysis.

Ethical considerations

The study was approved by the Research Ethics Committee of University Hospital of Gran Canaria Dr. Negrín (Las Palmas, Canary Islands, Spain), registration number 2019-190-1. All participants gave their informed consent. Data were handled with utmost confidentiality and anonymity.

FINDINGS

Pilot test

A pilot test (n = 27) was conducted on 81.5% women with mean age of 45.15 years and professional experience of 21.74 years. A total of 59.2% were Diploma of Higher Education and Bachelor's degree in Nursing, while 29.6% were having Master's degree, and rest 11.1% were PhD. A total of 55.5% of them worked in the clinical setting. Consistency of response was tested by means of test-retest temporal stability tests (n = 14) with Cohen's kappa coefficient. Interobserver agreement showed $k = 0.080$ (95% CI = 0.071–0.089); and the values range from $k = 0.031$ (95% CI = −0.289–0.350) to $k = 0.893$ (95% CI = 0.611–1.174).

TABLE 1 Sociodemographic characteristics of the sample

Sociodemographic characteristics		Mean	SD
Age		43.52	10.09
Work experience		20.19	10.5
Intensity of use of classification systems	NANDA-I	3.58	1.02
	ATIC	1.39	1.02
	ICNP	1.18	0.66
	CCC	1.11	0.48
	OMAHA System	1.08	0.43
		n	%
Sex	Female	345	75
	Male	115	25
Level of education	Diploma of Higher Education	195	42.39
	Bachelor's degree	84	18.26
	Master's degree	127	27.61
	PhD	54	11.74
Professional setting	Hospital departments	139	30.2
	Primary care team	121	26.3
	Management	48	10.4
	Academic	13	2.8
	Nursing home	9	2
	Emergency department	9	2
	Research	8	1.7
	Outpatient emergency department	4	0.9
Professional role	Clinical	227	49.3
	Supervisory	36	7.8
	Teaching	20	4.3
	Management	18	3.9

Abbreviations. ATIC, Arquitectura, Terminología, Interfase, Información, enfermerIa y Conocimiento; ICNP, International Classification for Nursing Practice; CCC, Clinical Care Classification.

A total of N = 460 participants were included in the psychometric validation of the questionnaire; sociodemographic characteristics are shown in Table 1.

Reliability

Reliability studies were focused on the internal consistency analysis. First, the internal consistency of the whole instrument was measured using Cronbach alpha ($\alpha = 0.957$). The internal consistency of each of the factors in the instrument was then determined for a scenario in which an item was removed. If an item was consistent with the rest of the scale, the Cronbach alpha value would be lower than observed once the item was removed; if an item was not consistent with the rest of the scale, the scale's α value would increase once it was removed. In factor 2 ($\alpha = 0.936$) for 13 items (population α value between 0.928 and 0.938; 95% confidence), only the removal of one item (*29_Diagnostic label: Deficient knowledge*) increased the α value ($\alpha = 0.938$). With regard to factor 5 ($\alpha = 0.896$) for four items (population α value between 0.819 and 0.969; 95% confidence), only one item (*19_Each concept is identified with an unambiguous code*) was found whose removal increased the α value ($\alpha = 0.969$). In factor 6 ($\alpha = 0.669$) for three items (population α value between 0.336 and 0.828; 95% confidence), only one item (*18_It is possible to find duplicate concepts or contents*) was found whose removal increased the α value ($\alpha = 0.828$). Finally, in factor 8 ($\alpha = 0.562$) for three items (population α value between 0.844 and 0.057; 95% confidence), only one item (*26_Generally, the degree of abstraction of* ND *is high*) increased the α value after it was removed ($\alpha = 0.844$). Table 2 shows Cronbach alpha and descriptive results of the EVALUAN-I application.

To check the interfactor correlation, the scores of all the items of each factor were summarized and a new variable with the total values of the nine factors was created (Total factor); the Pearson's correlation coefficients were calculated, as shown in Table 3.

INTERNATIONAL JOURNAL OF NURSING KNOWLEDGE
The Official Journal of NANDA International, Inc.

TABLE 2 Descriptive analysis of the sample and Cronbach's alpha, if remove item

Items	Mean	SD	α Factor	α if remove item
F1. Clinical competence			0.957	
22 NANDA-I is useful	4.12	1.56		0.952
27 Concepts facilitate organization of specific, autonomous knowledge	3.97	1.49		0.953
7 NANDA-I is essential for nursing science to progress academically	4.09	1.68		0.953
17 The structure of NANDA-I makes it easier to learn	3.66	1.51		0.954
28 The content in NANDA-I helps nurses to decide what care to deliver	4.05	1.48		0.954
4 Using NANDA-I is compulsory to ensure that nurses are able to deliver care reflecting exclusive roles and responsibilities	3.76	1.81		0.955
24 NANDA-I is applied in clinical practice	3.53	1.61		0.955
23 NANDA-I is intuitive	3.42	1.49		0.954
25 Nursing professionals accept NANDA-I as part of their clinical practice	3.09	1.49		0.956
6 NANDA-I encourages development of nursing discipline exclusive concepts	4.30	1.43		0.954
5 NANDA-I represents and identifies theoretical currents in nursing	3.65	1.44		0.955
15 There is internal coherence between the concepts	4.23	1.11		0.954
13 The diagnostic labels are clear and descriptive	3.88	1.41		0.954
11 The defining characteristics are relevant to diagnostic judgments	4.56	1.25		0.954
16 The configuration of the hierarchical and taxonomic structure is precise	4.17	1.16		0.955
12 Related factors or risk factors are relevant to diagnostic judgments	4.48	1.25		0.955
10 Using NANDA-I contextualises the care to be delivered to individuals	4.09	1.39		0.955
F2. Nurses' reasoning skills			0.936	
29 Diagnostic label: Risk for infection	4.77	1.58		0.931
29 Diagnostic label: Chronic pain	4.31	1.67		0.929
29 Diagnostic label: Deficient knowledge	5.18	1.33		0.938
29 Diagnostic label: Risk for perioperative positioning injury	4.73	1.49		0.936
29 Diagnostic label: Nausea	3.78	1.66		0.929
29 Diagnostic label: Impaired gas exchange	3.67	1.65		0.931
29 Diagnostic label: Activity intolerance	4.70	1.43		0.932
29 Diagnostic label: Insomnia	4.23	1.55		0.929
29 Diagnostic label: Acute confusion	3.87	1.64		0.928
29 Diagnostic label: Risk for bleeding	4.24	1.67		0.929
29 Diagnostic label: Anxiety	4.43	1.54		0.928
29 Diagnostic label: Diarrhea	3.88	1.68		0.929
29 Diagnostic label: Decreased cardiac output	3.24	1.70		0.931
F3. Attitudes towards nursing diagnosis			0.956	
30 Attitude: Significance	4.05	1.30		0.949
30 Attitude: Realism	3.81	1.31		0.948

(Continues)

TABLE 2 (Continued)

Items	Mean	SD	α Factor	α if remove item
30 Attitude: Obstruction	3.75	1.33		0.947
30 Attitude: Validity	3.97	1.29		0.950
30 Attitude: Relevance	3.93	1.36		0.949
30 Attitude: Gratification	3.75	1.32		0.948
30 Attitude: Creativity	3.63	1.39		0.954
F4. Discipline's central concepts			0.935	
3 NANDA-I identifies the central concept of person	4.62	1.31		0.910
3 NANDA-I identifies the central concept of health	4.55	1.29		0.903
3 NANDA-I identifies the central concept of environment	4.29	1.32		0.930
3 NANDA-I identifies the central concept of nursing	4.69	1.28		0.919
F5. Classification's contents			0.896	
14 The content is clear and descriptive: defining characteristics	4.14	1.32		0.829
14 The content is clear and descriptive: related factors	4.18	1.30		0.819
14 The content is clear and descriptive: risk factors	4.22	1.31		0.823
19 Each concept is identified with an unambiguous code	4.48	1.43		0.969
F6. Physiopathological attributes			0.669	
2 Some related/risk factors may correspond to medical diagnoses or diseases	4.26	1.51		0.376
1 Some nursing diagnoses may correspond to medical diagnoses or diseases	4.07	1.65		0.336
18 It is possible to find duplicate concepts or content in NANDA-I	4.13	1.25		0.828
F7. Scientific evidence			0.888	
8 NANDA-I diagnoses are based on scientific evidence	4.57	1.25		
9 The designs of studies offer high levels of scientific evidence	4.33	1.28		
F8. Diagnostic precision			0.562	
26 Generally, the degree of abstraction of NANDA-I diagnoses is high	4.22	1.32		0.844
29 Diagnostic label: Reflex urinary incontinence	3.68	1.75		0.057
29 Diagnostic label: Sexual dysfunction	3.71	1.75		0.088
F9. Conceptual correspondence between terminologies			0.639	
21 Each concept has been mapped in other classifications	3.82	1.28		
20 The concepts may be linked to synonyms in other classifications	3.98	1.23		

Exploratory factor analysis

The EFA was viable (KMO = 0.931; χ^2 = 16,786.24; 1485 degrees of freedom; $p < 0.001$). Using PCA, the eigenvalue saturation curve stabilized from Factor 9 onwards with 70.862% explanatory power of the variance. The Varimax orthogonal rotation displayed saturations that facilitated interpretation of the theoretical construct, as shown in Table 4. This EFA was considered an approximation to the CFA and was intended to reveal evidence of the existence of the following nine factors:

F1. Evaluation of clinical competence
F2. Evaluation of nurses' reasoning skills
F3. Evaluation of attitudes towards nursing diagnosis
F4. Evaluation of the discipline's central concepts
F5. Evaluation of the classification's contents

INTERNATIONAL JOURNAL OF NURSING KNOWLEDGE
The Official Journal of NANDA International, Inc.

TABLE 3 Interfactor correlations

	Pearson	*p*-value	95% CI
Factor 1 vs. Factor 2	0.31	<0.001	0.22–0.39
Factor 1 vs. Factor 3	0.52	<0.001	0.45–0.59
Factor 1 vs. Factor 4	0.63	<0.001	0.57–0.68
Factor 1 vs. Factor 5	0.73	<0.001	0.69–0.77
Factor 1 vs. Factor 6	0.07	0.124	−0.02–0.16
Factor 1 vs. Factor 7	0.63	<0.001	0.57–0.68
Factor 1 vs. Factor 8	0.24	<0.001	0.15–0.32
Factor 1 vs. Factor 9	0.32	<0.001	0.24–0.40
Factor 1 vs. Total_factor	0.86	<0.001	0.84–0.89
Factor 2 vs. Factor 3	0.32	<0.001	0.24–0.40
Factor 2 vs. Factor 4	0.23	<0.001	0.14–0.31
Factor 2 vs. Factor 5	0.31	<0.001	0.22–0.39
Factor 2 vs. Factor 6	0.01	0.91	−0.09- 0.10
Factor 2 vs. Factor 7	0.28	<0.001	0.19–0.36
Factor 2 vs. Factor 8	0.76	<0.001	0.72–0.80
Factor 2 vs. Factor 9	0.05	0.256	−0.04–0.14
Factor 2 vs. Total_factor	0.69	<0.001	0.64–0.74
Factor 3 vs. F actor4	0.30	<0.001	0.22–0.38
Factor 3 vs. Factor 5	0.39	<0.001	0.31–0.46
Factor 3 vs. Factor 6	0.03	0.527	−0.06–0.12
Factor 3 vs. Factor 7	0.28	<0.001	0.19–0.36
Factor 3 vs. Factor 8	0.28	<0.001	0.19–0.36
Factor 3 vs. Factor 9	0.12	0.011	0.03–0.21
Factor 3 vs. Total_factor	0.65	<0.001	0.60–0.70
Factor 4 vs. Factor 5	0.56	<0.001	0.49–0.62
Factor 4 vs. Factor 6	0.13	0.005	0.04–0.22
Factor 4 vs. Factor 7	0.49	<0.001	0.41–0.55
Factor 4 vs. Factor 8	0.16	<0.001	0.07–0.25
Factor 4 vs. Factor 9	0.24	<0.001	0.15–0.32
Factor 4 vs. Total_factor	0.64	<0.001	0.58–0.69
Factor 5 vs. Factor 6	0.00	0.938	−0.10–0.09
Factor 5 vs. Factor 7	0.49	<0.001	0.42–0.56
Factor 5 vs. Factor 8	0.20	<0.001	0.11–0.28
Factor 5 vs. Factor 9	0.26	<0.001	0.18–0.35
Factor 5 vs. Total_factor	0.72	<0.001	0.68–0.77
Factor 6 vs. Factor 7	0.03	0.454	−0.06–0.13
Factor 6 vs. Factor 8	0.04	0.357	−0.05–0.13
Factor 6 vs. Factor 9	0.29	<0.001	0.20–0.37
Factor 6 vs. Total_factor	0.15	0.001	0.06–0.24
Factor 7 vs. Factor 8	0.21	<0.001	0.13–0.30
Factor 7 vs. Factor 9	0.24	<0.001	0.16–0.33
Factor 7 vs. Total_factor	0.62	<0.001	0.56–0.67
Factor 8 vs. Factor 9	0.14	0.002	0.05–0.23
Factor 8 vs. Total_factor	0.57	<0.001	0.51–0.63
Factor 9 vs. Total_factor	0.33	<0.001	0.25–0.41

F6. Evaluation of physiopathological attributes
F7. Evaluation of level of scientific evidence
F8. Evaluation of diagnostic precision
F9. Evaluation of conceptual correspondence between terminologies

Confirmatory factor analysis

To confirm the unidimensional structure of EVALUAN-I, three factor models were developed to test the goodness of fit indicators, as shown in Table 5.

Model 1 was based on the structure proposed in the PCA (χ^2 = 4248.069; 1394 degrees of freedom; $p < 0.001$). Its optimal goodness of fit indicators were RMR = 0.115; RMSEA = 0.067; GFI = 0.736; AGFI = 0.708; NNFI = 0.853; NFI = 0.809; CFI = 0.862; TLI = 0.853.

Given the unsatisfactory indicators found in model 1, a second model was developed by estimating the correlation error measure. Following these modifications, model 2 (χ^2 = 3257.12; 1386 degrees of freedom; $p < 0.001$) showed improvements in several indicators compared to the previous model (RMSEA = 0.054; NNFI = 0.903; CFI = 0.910; TLI = 0.903).

To improve the CFA, a further change was made to create a third model: the variable *10_Using the classification contextualises the care to be delivered to individuals* to construct 7. In model 3 (χ^2 = 3270.19; 1386 degrees of freedom; $p < 0.001$), the goodness of fit indicators behaved similarly to the previous model (RMSEA = 0.054; NNFI = 0.903; CFI = 0.909; TLI = 0.903).

In the three proposed models, all variables were statistically significant for factors 1, 2, 3, 4, 5, 6, 7, and 9 ($p < 0.001$); however, factor 8 was not significant in any of the models: model 1 ($p = 0.180$), model 2 ($p = 0.200$), and model 3 ($p = 0.199$).

With regard to the variance between the variables in the instrument and between the latent variables, factor 8 was the only factor that was not significant in any of the three models: model 1 ($p = 0.503$), model 2 ($p = 0.521$), and model 3 ($p = 0.521$). Moreover, no statistical significance was observed with regard to the covariances in factors 6 and 8 in relation to the other latent variables in the three models proposed.

Once the factor model with the best fit had been determined, the adequacy of convergence of the altered variables in the alternative models was checked using Pearson's correlation coefficient; all convergences were significant ($p < 0.001$) with values ranging from $r = 0.73$ and $r = 0.86$ (CI = 0.68–0.89).

The variations in the coefficients of determination were also calculated for the two alternative models. The maximum difference was found in the variable *P10_Using the classification contextualises the care to be delivered to individuals*, in model 2 ($R^2 = 0.507$) and model 3 (R^2 = 0.572).

DISCUSSION

Using ND classification systems improves health records, benefiting patients and professionals in the correct clinical decision-making,

TABLE 4 Saturations of the matrix of rotated components

	Components								
	1	2	3	4	5	6	7	8	9
22_Is useful	0.819								
27_Knowledge organization	0.795								
7_Academic performance	0.769								
17_Facilitates learning	0.765								
28_Facilitates decisions on care	0.747								
4_Exclusive responsibilities	0.741								
24_Is applied in clinical practice	0.739								
23_Is intuitive	0.727								
25_Professionals accept it	0.716								
6_Exclusive concepts	0.685								
5_Nursing theories	0.685								
15_Internal coherence	0.671								
13_Clear, descriptive labels	0.663								
11_Relevant defining characteristics	0.642								
16_Hierarchical structure	0.638								
12_Relevant related factors and risk factors	0.620								
10_Contextualises care	0.614								
29_Risk for infection		0.840							
29_Chronic pain		0.834							
29_Deficient knowledge		0.823							
29_Risk for perioperative positioning injury		0.815							
29_Nausea		0.814							
29_Impaired gas exchange		0.796							
29_Activity intolerance		0.792							
29_Insomnia		0.774							
29_Acute confusion		0.758							
29_Risk for bleeding		0.748							
29_Anxiety		0.731							
29_Diarrhoea		0.652							
29_Decreased cardiac output		0.643							
30_Significance			0.851						
30_Realism			0.846						
30_Obstruction			0.845						
30_Validity			0.841						
30_Relevace			0.837						
30_Gratification			0.833						
30_Creativity			0.785						
3_Concept of person				0.830					
3_Concept of health				0.822					
3_Concept of environment				0.812					
3_Concept of nursing				0.760					
14_Clear, descriptive defining characteristics					0.694				
14_Clear, descriptive related factors					0.681				
14_Clear, descriptive risk factors					0.680				

(Continues)

INTERNATIONAL JOURNAL OF NURSING KNOWLEDGE
The Official Journal of NANDA International, Inc.

TABLE 4 (Continued)

	Components								
	1	2	3	4	5	6	7	8	9
19_Unequivocal codes					0.414				
2_Related factors corresponding diseases						0.851			
1_Nursing diagnoses corresponding diseases						0.846			
18_Contains duplicate concepts						0.439			
8_Evidence-based nursing diagnoses							0.685		
9_Research designs providing evidence							0.639		
26_Abstraction								0.649	
29_Reflex urinary incontinence								0.555	
29_Sexual dysfunction								0.491	
21_Mapped concepts									0.781
20_Synonyms									0.657

Extraction method: principal components analysis.
Rotation method: Varimax with Kaiser normalization.
Rotation converged in 10 iterations.

TABLE 5 Fit indicators of CFA models

	X^2	*p*-value	RMR	RMSEA	GFI	AGFI	NNFI	NFI	CFI	TLI
Model 1	4248.07	<0.001	0.115	0.067	0.736	0.708	0.853	0.809	0.862	0.853
Model 2	3257.12	<0.001	0.109	0.054	0.787	0.764	0.903	0.853	0.910	0.903
Model 3	3270.19	<0.001	0.114	0.054	0.790	0.766	0.903	0.853	0.909	0.903
Good fit if			<0.08	<0.08	>0.90	>0.90	>0.90	>0.90	>0.90	>0.90

Abbreviations. AGFI, Adjusted goodness of fit index; CFI, Comparative fit index; GFI, Goodness of fit index; NFI, Normed fit index; NNFI, Non-normed fit index; RMR, Root mean square residual; RMSEA, Root mean square error of approximation; TLI, Tucker-Lewis index.

leading to greater quality, consistency, and safety in the provision of health care. In this sense, EVALUAN-I reflects the need for comprehensive evaluation of NANDA-I as the most widely used classification system in the world (Tastan et al., 2014) to improve its usefulness and intuitiveness. The aim of this new tool is to improve upon several aspects of NANDA-I, which have been described above. During the process of designing EVALUAN-I, the different dimensions were organized and structured to enable systematic analysis, creating a configuration that addresses theoretical aspects from an empirical perspective (Marrs & Lowry, 2006; Weaver & Olson, 2006), and promotes the logical, scientific development of the nursing discipline.

The initial structure proposed in the theoretical framework for evaluating a diagnostic system, such as NANDA-I, has been reorganized after statistical validation into nine new dimensions of analysis.

The internal consistency of EVALUAN-I was $\alpha = 0.957$, indicating an excellent correlation between its variables and demonstrating the reliability of the instrument. Values exceeding 0.900 may be indicative of redundancy in some of the items (Halberstadt, Schmitz, & Sammel, 2012), so these may be reduced in future as long as the main aim of holistically evaluating NANDA-I is maintained. This possible redundancy can be seen in the stratified internal consistency of each factor in a scenario where one item is removed (Oviedo & Campo-Arias, 2005), revealing weaknesses in some of the variables and significantly increasing the scale's internal consistency. This structural weakness may be attributed to the conceptual complexity of these variables in a largely clinical population with a lower level of education. The complexity of these items may require extensive education and knowledge of the study topic, as well as expertise, experience, and training in the use of diagnostic taxonomies, even among supposedly homogeneous populations. It would be appropriate to consider eliminating or modifying these items in future research.

With regard to interfactor correlation, the weighting of each of the factors in the total score for the instrument was as expected. Although the correlation was generally acceptable for most of the scales, the values were lower in factor 6, indicating weaknesses in internal consistency in relation to this latent variable.

The decision to maintain nine factors in the EFA was based on the contribution made by each of their items to the overall scale (Lloret-Segura et al., 2014). Factors with fewer than three items played an essential part in an exhaustive evaluation of NANDA-I due to the high levels of saturation and the importance of these factors identified during the construct definition process in evaluating all attributes of NANDA-I.

The combination of Cronbach's alpha and EFA is entirely inadequate to guarantee the validity and reliability of a questionnaire relating to health (Batista-Foguet et al., 2004), and CFA is an adequate alternative.

An ideal design for a measurement tool consists not of a single battery of items relating to the overall construct of interest but of optimized subsets of specific items for each dimension. Although the proposed model based on the PCA appeared appropriate, its goodness of fit indicators were not entirely satisfactory.

By developing the two alternative factor models, a larger number of these indicators could be optimized. If a model displays good fit via the combined RMSEA and CFI indicators, it is very unlikely that it is inadequate; this is a good rule of thumb in the search for the model best suited to the data. However, the results obtained show that model 3 does not display any substantial improvements to the goodness of fit indicators compared to model 2. These differences may be resolved by using the Occam's razor principle whereby, when faced with two alternative explanations of a single reality under equal conditions, the simplest explanation is always preferable (Aldás & Uriel, 2017). The two models are also equal in terms of their number of degrees of freedom. However, the second model keeps the proposed factor structure unchanged by modifying only the error correlations, while the third model alters the initial structure proposed for item 10, adding it to factor 7. In light of these results, the goodness of fit indicators showed that model 2 had the best statistical behavior.

The coefficient of determination measures the proportion of information that each model is able to explain, indicating which may be the most adequate. The closer the value is to 1, the better the fit, so values exceeding 0.250 are considered adequate (Aldás & Uriel, 2017). The coefficients of determination of the models studied maintained similar results or underwent minimal changes following the modifications made to the explanatory variables.

The results of the CFA indicate the line of work to be followed, as these indicators do not clearly improve in relation to the structure initiated with the EFA. This supports the hypothesis that the weakness of the construct lies in the aforementioned variables, as do the strategies for improving it.

One of the strengths of the study is that evaluation of NANDA-I encourages progress toward a refinement of its diagnostic indicators, which form the basis for interventions aimed at addressing medical problems (Herdman & Kamitsuru, 2015) by linking ND to the Nursing Interventions Classification and the Nursing Outcomes Classification, allowing it to be applied universally (Smith & Craft-Rosenberg, 2010). This leads to improvements in the classification's descriptive capacity, eliminates imprecision (Carrington, 2008), and promotes evidence-based practice (Powers, 2002) in accordance with professionals' cognitive skills in relation to the nursing discipline (Cho & Park, 2006).

Descriptive findings showed that the best-recognized diagnostic label was *Deficient knowledge*, which is a situation closely related to nurses' own roles such as Health Education. In contrast, the application of EVALUAN-I showed that the worst rated items were the acceptability of NANDA-I for clinical practice. In this sense, using EVALUAN-I to evaluate NANDA-I has repercussions on the clinical practice of nurses through improvements by optimizing the precision and specificity of nurses' clinical judgment, favoring the visibility, potential and scientific autonomy of these professionals.

The main limitations of the study include the use of a nonprobabilistic method to recruit the sample which was determined by the study objective, to validate EVALUAN-I in an initial exploration of the tool's psychometric properties in the Spanish context with a population of nursing professionals determined by the selection criteria. In addition, the wording of some of the items was excessively complex, which may have led to memorial disparity or lack of comprehension of the variables among some interviewees. In future applications of EVALUAN-I, these factorial weaknesses will need to be addressed to improve the construct of the instrument, the validity, and reliability of the results.

Different populations with more homogeneous sociodemographic characteristics may be examined in future studies, or population stratification could be carried out in randomized heterogeneous groups including samples with sufficient cases.

Internal improvements to a classification system, like NANDA-I, based on an analysis of the dimensions set out above could improve the quality of clinical records. However, diagnostic precision among professionals must also be improved (Müller-Staub et al., 2006). NANDA-I is currently accepted for use in clinical practice, despite displaying normative, structural, and teleological shortcomings hindering adequate diagnostic judgment (Zanotti & Chiffi, 2015), which is associated with a disciplinary knowledge that contributes to correctly describing the population's health status and is appropriately perceived by nurses in their practice (Juvé i Udina, 2016).

The results of the evaluation of NANDA-I may vary by professional profile, pointing to the need to categorized different profiles into population clusters to establish appropriate strategies for each group (D'Agostino et al., 2018). Equally, different sociocultural contexts can determine the clinical application of ND (Lai et al., 2013), affecting the identification of strengths and weaknesses of NANDA-I in different countries.

Given the underdeveloped nature of certain concepts and labels in the categories identified in NANDA-I (Gordon, 1996), the University of Iowa College of Nursing and the Diagnosis Development Committee, in collaboration with NANDA-I, aim to research and improve the length, scope, and clinical utility of the classification (Craft-Rosenberg & Jirathummakoon, 2003; Herdman, Gordon, & Craft-Rosenberg, 2000). Many ND have yet to be refined, but most studies draw on expert-based methods to validate the content of these diagnoses; it is important that further clinical validation studies are conducted to provide higher levels of evidence (De Souza Oliveira-Kumakura et al., 2018).

Improvements in a diagnostic classification, such as NANDA-I, have repercussions in optimizing the diagnostic accuracy and specificity of nurses' clinical judgment, strengthening their visibility and scientific autonomy.

CONCLUSIONS

The EVALUAN-I tool is a valid, reliable instrument, which can be used to improve the epistemological, normative, and intuitive configuration

of NANDA-I in a structured, systematic manner, although some aspects require on-going improvement to refine their structure and statistical behavior in future applications.

Comprehensive evaluation of the NANDA-I taxonomy in different clinical contexts around the world using a validated tool, such as EVALUAN-I, would allow strengths and weaknesses to be identified and contribute to the classification's development and practical application.

ACKNOWLEDGMENTS

The authors would like to thank Dr José-Manuel Romero-Sánchez for authorizing the rights to use the Position on Nursing Diagnosis-Spanish Version (PND-7-SV) scale. The authors would like to thank Drs. Alejandro García-Aragón and Eleanor Staniforth for translating the manuscript. This research was partially funded by the Fundación Canaria Instituto de Investigación Sanitaria de Canarias (Canary Islands Foundation for Health Research, FIISC), under file number ENF-19/33.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Claudio Alberto Rodríguez-Suárez PhD Candidate, BSc, MSc, RN https://orcid.org/0000-0001-6226-7374

María Naira Hernández-De Luis BSc, MSc, RN https://orcid.org/0000-0001-9720-3661

María Isabel Mariscal-Crespo PhD, BSc, RN https://orcid.org/0000-0003-0736-8695

Rafaela Camacho-Bejarano PhD, BSc, MSc, RN https://orcid.org/0000-0003-0716-5038

REFERENCES

Abed El-Rahman, M., Al Kalaldeh, M. T., & Malak, M. Z. (2017). Perceptions and attitudes toward NANDA-I nursing diagnoses: a cross-sectional study of jordanian nursing students. *International Journal of Nursing Knowledge*, *28*(1), 13–18. https://doi.org/10.1111/2047-3095.12100

Aldás, J., & Uriel, E. (2017). Análisis multivariante aplicado con R *(2ª)*. Paraninfo.

Batista-Foguet, J. M., Coenders, G., & Alonso, J. (2004). Análisis factorial confirmatorio. Su utilidad en la validación de cuestionarios relacionados con la salud. *Medicina Clínica*, *122*, Supl, 21–27. Retrieved from https://www.researchgate.net/profile/Joan_manuel_Batista-foguet/publication/277156595_04_SUPLEMENTO_2_21-27/links/55641b4b08ae6f4dcc98c10d.pdf

Beckstead, J. W. (2009). Taxonomies of nursing diagnoses: a psychologist's view. *International Journal of Nursing Studies*, *46*(3), 295–301. https://doi.org/10.1016/j.ijnurstu.2008.08.010

Bittencourt, G. K. G. D., & Crossetti, M. d. G. O. (2013). Critical thinking skills in the nursing diagnosis process. *Revista da Escola de Enfermagem da USP*, *47*(2), 341–347. https://doi.org/10.1590/S0080-62342013000200010

Carrington, J. M. (2008). The effectiveness of electronic health record with standardized nursing languages for communicating patient status related to a clinical event (*Doctoral Dissertation*). Arizona, USA: University of Arizona. Retrieved from https://repository.arizona.edu/handle/10150/195397

Carvajal, A., Centeno, C., Watson, R., Martínez, M., & Sanz Rubiales, Á. (2011). ¿Cómo validar un instrumento de medida de la salud? *Anales del Sistema Sanitario de Navarra*, *34*(1), 63–72. Retrieved from https://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1137-66272011000100007?

Cho, I., & Park, H. (2006). Evaluation of the Expressiveness of an ICNP-based nursing data dictionary in a computerized nursing record system. *Journal of the American Medical Informatics Association*, *13*(4), 456–464. https://doi.org/10.1197/jamia.M1982

Collins, A. (2013). Effect of continuing nursing education on nurses' attitude toward and accuracy of nursing diagnosis. *International Journal of Nursing Knowledge*, *24*(3), 122–128. https://doi.org/10.1111/j.2047-3095.2013.01237.x

Craft-Rosenberg, M., & Jirathummakoon, S. (2003). Nursing diagnosis extension and classification: ongoing phase. *International Journal of Nursing Terminologies and Classifications*, *14*(4), 2. https://doi.org/10.1111/j.1744-618X.2003.001_3.x

Cruz, D. M., Pimenta, C. M., & Lunney, M. (2009). Improving critical thinking and clinical reasoning with a continuing education course. *Journal of Continuing Education in Nursing*, *40*(3), 121–127. https://doi.org/10.3928/00220124-20090301-05

D'Agostino, F., Pancani, L., Romero-Sánchez, J. M., Lumillo-Gutièrrez, I., Paloma-Castro, O., Vellone, E., & Alvaro, R. (2018). Nurses' beliefs about nursing diagnosis: a study with cluster analysis. *Journal of Advanced Nursing*, *74*(6), 1359–1370. https://doi.org/10.1111/jan.13545

D'Agostino, F., Vellone, E., Cerro, E., Di Sarra, L., Juárez-Vela, R., Ghezzi, V., … Alvaro, R. (2016). Psychometric evaluation of the Positions on Nursing Diagnosis scale. *Applied Nursing Research*, *29*, e1–e6. https://doi.org/10.1016/j.apnr.2015.03.012

de Guedes, E. S., de Sousa, R. M. C., Turrini, R. N. T., Baltar, V. T., & de Cruz, D. A. L. M. d. (2013). Adaptation and validation of the instrument positions on the nursing process. *Revista Latino-Americana de Enfermagem*, *21*(1), 404–411. https://doi.org/10.1590/S0104-11692013000100016

De Souza Oliveira-Kumakura, A. R., Caldeira, S., Prado Simão, T., Camargo-Figuera, F. A., de Almeida Lopes Monteiro da Cruz, D., & Campos de Carvalho, E. (2018). The contribution of the Rasch Model to the clinical validation of nursing diagnoses: integrative literature review. *International Journal of Nursing Knowledge*, *29*(2), 89–96. https://doi.org/10.1111/2047-3095.12162

Frazão, C. M. F. Q., De Sá, J. D., De Paiva, M., das, G. M. N., De Carvalho Lira, A. L. B., De Oliveira Lopes, M. V., & Enders, B. C. (2015). Association between nursing diagnoses and socioeconomic/clinical characteristics of patients on hemodialysis. *International Journal of Nursing Knowledge*, *26*(3), 135–140. https://doi.org/10.1111/2047-3095.12051

Goh, S., Leong Tang, M., & Ahmad, A. (2020). Competence-based frameworks in nursing–a concept analysis. *Journal of Applied Learning & Teaching*, *3*(1), https://doi.org/10.37074/jalt.2020.3.1.3

Gordon, M. (1996). Diagnóstico enfermero: Proceso y aplicación *(3ª)*. Mosby/Doyma.

Halberstadt, S. M., Schmitz, K. H., & Sammel, M. D. (2012). A joint latent variable model approach to item reduction and validation. *Biostatistics*, *13*(1), 48–60. https://doi.org/10.1093/biostatistics/kxr018

Hasegawa, T., Ogasawara, C., & Katz, E. C. (2007). Measuring diagnostic competency and the analysis of factors influencing competency using written case studies. *International Journal of Nursing Terminologies and Classifications*, *18*(3), 93–102. https://doi.org/10.1111/j.1744-618X.2007.00057.x

Herdman, T. H. (2011). *Nursing diagnoses 2012–2014: definitions and classification*. Wiley-Blackwell.

Herdman, T. H., & Kamitsuru, S. (2015). *NANDA International. Diagnósticos Enfermeros: Definiciones y Clasificación. 2015–2017*. Elsevier.

Herdman, T. H., Gordon, M., & Craft-Rosenberg, M. (2000). The evolving role of DRC: impact on collaboration with NDEC and other research groups. *International Journal of Nursing Terminologies and Classifications*, *11*(4), 176–178. https://doi.org/10.1111/j.1744-618X.2000.tb00424.x

Herdman, T. H., & Kamitsuru, S. (2019). Diagnósticos enfermeros. Definiciones y clasificación 2018–2020 *(11ª)*. Elsevier.

Herdman, T. H., Kamitsuru, S., & Takáo Lopes, C. (2021). NANDA International Nursing Diagnoses. Definitions and Classification, 2021–2023 *(12ª)*. Thieme.

Hine-Sanabria, A., Mendoza-Monsalve, A., Rojas, L. Z., & Roa-Díaz, Z. M. (2018). Conocimiento e investigación sobre la autonomía profesional en enfermería. Revisión narrativa. *Enfermería Universitaria*, *15*(2), 212–221. https://doi.org/10.22201/eneo.23958421e.2018.2.65177

Instituto Nacional de Estadística [Spanish National Institute of Statistics]. (2019). *Estadística de profesionales sanitarios colegiados. Últimos datos*. Retrieved December 11, 2020. https://www.ine.es/dyngs/INEbase/operacion.htm?c=Estadistica_C&cid=1254736176781&menu=resultados&idp=1254735573175?

Juvé i Udina, M. E. (2012). Evaluación de la validez de una terminología enfermera de interfase (Doctoral Dissertation). Barcelona, España: Universitat de Barcelona. Retrieved from http://www.tdx.cat/handle/10803/108998

Juvé i Udina, M. E. (2016). *La Terminología ATIC: Eje diagnóstico*. Naaxpot-Slu.

Lai, W.-S., Chao, C.-S. C., Yang, W.-P., Liu, H.-C., & Chen, C.-H. (2013). "Does one size fit all?" Exploring the cultural applicability of NANDA nursing diagnoses to chinese nursing practice. *Journal of Transcultural Nursing*, *24*(1), 43–50. https://doi.org/10.1177/1043659612462403

Lloret-Segura, S., Ferreres-Traver, A., Hernández-Baeza, A., & Tomás-Marco, I. (2014). El análisis factorial exploratorio de los ítems: Una guía práctica, revisada y actualizada. *Anales de Psicología*, *30*(3), 1151–1169. https://doi.org/10.6018/analesps.30.3.199361

López-Aguado, M., & Gutiérrez-Provecho, L. (2019). Cómo realizar e interpretar un análisis factorial exploratorio utilizando SPSS. *Revista d Innovació i Recerca en Educació*, *12*(2), 1–14. https://doi.org/10.1344/reire2019.12.227057

Marrs, J.-A., & Lowry, L. W. (2006). Nursing theory and practice: connecting the dots. *Nursing Science Quarterly*, *19*(1), 44–50. https://doi.org/10.1177/0894318405283547

Martín, M. (2004). Diseño y validación de cuestionarios. *Matronas Profesión*, *5*(17), Retrieved from http://ebevidencia.com/wp-content/uploads/2014/07/validacion_cuestionarios.pdf

Matas, A. (2018). Diseño del formato de escalas tipo Likert: Un estado de la cuestión. *Revista Electrónica de Investigación Educativa*, *20*(1), 38. https://doi.org/10.24320/redie.2018.20.1.1347

Mendes, M. A., da Cruz, D. A., & Angelo, M. (2015). Clinical role of the nurse: concept analysis. *Journal of Clinical Nursing*, *24*(3-4), 318–331. https://doi.org/10.1111/jocn.12545

Miguel, S., Caldeira, S., & Vieira, M. (2016). The adequacy of the Q methodology for clinical validation of nursing diagnoses related to subjective foci. *International Journal of Nursing Knowledge*, https://doi.org/10.1111/2047-3095.12163

Müller-Staub, M., Lavin, M. A., Needham, I., & van Achterberg, T. (2006). Nursing diagnoses, interventions and outcomes ? Application and impact on nursing practice: systematic review. *Journal of Advanced Nursing*, *56*(5), 514–531. https://doi.org/10.1111/j.1365-2648.2006.04012.x

Müller-Staub, M., Lavin, M. A., Needham, I., & van Achterberg, T. (2007). Meeting the criteria of a nursing diagnosis classification: evaluation of ICNP®, ICF, NANDA and ZEFP. *International Journal of Nursing Studies*, *44*(5), 702–713. https://doi.org/10.1016/j.ijnurstu.2006.02.001

Mynaříková, E., & Žiaková, K. (2014). The use of nursing diagnoses in clinical practice. *Central European Journal of Nursing and Midwifery*, *5*(3), 117–126. https://doi.org/10.15452/CEJNM.2014.05.0006

NANDA International. (2019). *NANDA International. About our name*. Retrieved 1 April, 2019. http://nanda.host4kb.com/article/AA-00219/23/English-/About-NANDA-International/About-NANDA-International.html

Oliva, A. P. V., de Lopes, D. A., Volpato, M. P., & Hayashi, A. A. M. (2005). Atitudes de alunos e enfermeiros frente ao diagnóstico de enfermagem. *Acta Paulista de Enfermagem*, *18*(4), 361–367. https://doi.org/10.1590/S0103-21002005000400004

Oviedo, H. C., & Campo-Arias, A. (2005). Aproximación al uso del coeficiente alfa de Cronbach. *Revista Colombiana de Psiquiatría*, *34*(4), 572–580. Retrieved from https://www.redalyc.org/articulo.oa?id=80634409%250AC%25C3%25B3mo

Paans, W., Sermeus, W., Nieweg, R. M., Krijnen, W. P., & van der Schans, C. P. (2012). Do knowledge, knowledge sources and reasoning skills affect the accuracy of nursing diagnoses? A randomised study. *BMC Nursing*, *11*(1), 11. https://doi.org/10.1186/1472-6955-11-11

Powers, P. (2002). A discourse analysis of nursing diagnosis. *Qualitative Health Research*, *12*(7), 945–965. https://doi.org/10.1177/104973202129120377

Romero-Sánchez, J. M., Paloma-Castro, O., Paramio-Cuevas, J. C., Pastor-Montero, S. M., O'Ferrall-González, C., Gabaldón-Bravo, E. M., … Martínez-Sabater, A. (2013). Desarrollo y evaluación psicométrica de una forma abreviada de la escala de posicionamiento ante el diagnóstico enfermero. *Revista da Escola de Enfermagem da USP*, *47*(3), 591–599. https://doi.org/10.1590/S0080-62342013000030 0010

Romero-Sánchez, J. M., Paramio-Cuevas, J. C., Paloma-Castro, O., Pastor-Montero, S. M., O'Ferrall-González, C., Gabaldón-Bravo, E. M., … Frandsen, A. J. (2013). The Spanish version of the position on nursing diagnosis scale: cross-cultural adaptation and psychometric assessment. *Journal of Advanced Nursing*, *69*(12), 2759–2771. https://doi.org/10.1111/jan.12128

Rosenbloom, S. T., Miller, R. A., Johnson, K. B., Elkin, P. L., & Brown, S. H. (2008). A model for evaluating interface terminologies. *Journal of the American Medical Informatics Association*, *15*(1), 65–76. https://doi.org/10.1197/jamia.M2506

Rutherford, M. A. (2008). Standarized nursing language: what does it mean for nursing practice? *OJIN: The Online Journal of Issues in Nursing*, *13*(1), https://doi.org/10.3912/OJIN.Vol13No01PPT05

Smith, K. J., & Craft-Rosenberg, M. (2010). Using NANDA, NIC, and NOC in an undergraduate nursing practicum. *Nurse educator*, *35*(4), 162–166. https://doi.org/10.1097/NNE.0b013e3181e33953

Stallinga, H. A., ten Napel, H., Jansen, G. J., Geertzen, J. H., de Vries Robbé, P. F., & Roodbol, P. F. (2015). Does language ambiguity in clinical practice justify the introduction of standard terminology? An integrative review. *Journal of Clinical Nursing*, *24*(3-4), 344–352. https://doi.org/10.1111/jocn.12624

Tastan, S., Linch, G. C. F., Keenan, G. M., Stifter, J., McKinney, D., Fahey, L., … Wilkie, D. J. (2014). Evidence for the existing American Nurses Association-recognized standardized nursing terminologies: a systematic review. *International Journal of Nursing Studies*, *51*(8), 1160–1170. https://doi.org/10.1016/j.ijnurstu.2013.12.004

Von Krogh, G. (2008). An examination of the NANDA international taxonomy for domain completeness, ontological homogeneity, and clinical functionality. *International Journal of Nursing Terminologies and Classifications*, *19*(2), 65–75. https://doi.org/10.1111/j.1744-618X.2008.00083.x

Weaver, K., & Olson, J. K. (2006). Understanding paradigms used for nursing research. *Journal of Advanced Nursing*, *53*(4), 459–469. https://doi.org/10.1111/j.1365-2648.2006.03740.x

Zanotti, R., & Chiffi, D. (2015). Diagnostic frameworks and nursing diagnoses: a normative stance. *Nursing Philosophy*, *16*(1), 64–73. https://doi.org/10.1111/nup.12074

How to cite this article: Rodríguez-Suárez, C. A., Hernández-De Luis, M. N., Mariscal-Crespo, M. I., & Camacho-Bejarano, R. (2023). Evaluation of the NANDA International, Inc. diagnostic classification in Spain: Development and validation of the EVALUAN-I tool. *International Journal of Nursing Knowledge*, *34*, 21–34. https://doi.org/10.1111/2047-3095.12366