



Corresponding author



Marcos Venícios
de Oliveira Lopes
E-mail: marcos@ufc.br

From validity to the construction of levels of evidence: a contemporary history of a methodological shift

Da validação à construção de níveis de evidência: história contemporânea de uma inflexão metodológica

De la validación a la construcción de niveles de evidencia: una historia contemporánea de un cambio metodológico

Marcos Venícios de Oliveira LOPES¹
Viviane Martins da SILVA¹

¹ Universidade Federal do Ceará – UFC, Faculdade de Farmácia, Odontologia e Enfermagem – FFOE, Departamento de Enfermagem. Fortaleza, CE, Brasil.

How to cite this article (Vancouver):

Lopes MVO, Silva VM. From validity to the construction of levels of evidence: a contemporary history of a methodological shift. *Hist Enferm Rev Eletr.* 2026;17:e003. <https://doi.org/10.51234/here.2026.v17.532>.

ABSTRACT

Objective: to historically analyze the process of constructing and developing the levels of evidence of nursing diagnoses within the international classification system, with emphasis on the conceptual, methodological, and institutional shifts that integrated validity, expertise, causality, and nursing diagnosis research. **Method:** a qualitative historical-documentary study grounded in contemporary history, intellectual history, the social history of science, and the history of concepts. The *corpus* comprised unpublished narrative sources, scientific publications, diagnostic classification system editions, methodological studies, literature on validity, measurement theory, expertise, and judgment aggregation, as well as available institutional documents. The sources underwent external and internal criticism, were organized into a chronological-thematic matrix, and were triangulated to identify milestones, key actors, academic networks, continuities, discontinuities, and conceptual changes. **Results:** the documentary reconstruction showed that the revision of levels of evidence resulted from a historical and methodological maturation process that integrated critiques of classical validity methods, the reassessment of experts' role, theories of validity and measurement, diagnostic causality, middle-range theories, and the international dissemination of Brazilian research groups. **Final considerations:** levels of evidence constitute scientific, methodological, institutional, and historical artifacts characterized by continuities and discontinuities in nursing diagnosis research.

Descriptors: History of Nursing; Nursing Diagnosis; Nursing Research; Standardized Nursing Terminology; Validation Study.

RESUMO

Objetivo: analisar historicamente o processo de construção e desenvolvimento dos níveis de evidência dos diagnósticos de enfermagem da classificação internacional, com ênfase nas inflexões conceituais, metodológicas e institucionais que articularam a validade, a *expertise*, a causalidade e a pesquisa em diagnósticos de enfermagem. **Método:** estudo histórico-documental qualitativo fundamentado na história contemporânea, na história intelectual, na história social da ciência e na história dos conceitos. O *corpus* reuniu fontes narrativas inéditas, publicações científicas, edições da classificação diagnóstica, estudos metodológicos, literatura sobre validade, teoria da medida, *expertise* e agregação de julgamentos, além de documentos institucionais disponíveis. As fontes foram submetidas à crítica externa e interna, organizadas em matriz cronológico-temática e trianguladas para identificar marcos, atores, redes, continuidades, rupturas e mudanças conceituais. **Resultados:** a reconstrução documental mostrou que a revisão dos níveis de evidência decorreu de um amadurecimento histórico-metodológico, que articulou críticas aos métodos clássicos de validação, à revisão do papel do especialista, às teorias de validade e medida, à causalidade diagnóstica, às teorias de médio alcance e à circulação internacional de grupos de pesquisa brasileiros. **Considerações finais:** os níveis de evidência constituem artefatos científicos, metodológicos, institucionais e históricos, marcados por continuidades e rupturas na pesquisa em diagnósticos de enfermagem.

Descritores: História da Enfermagem; Diagnóstico de Enfermagem; Pesquisa em Enfermagem; Terminologia Padronizada em Enfermagem; Estudo de Validação.

RESUMEN

Objetivo: analizar históricamente el proceso de construcción y desarrollo de los niveles de evidencia de los diagnósticos de enfermería de la clasificación internacional, con énfasis en las inflexiones conceptuales, metodológicas e institucionales que articularon la validez, la experticia, la causalidad y la investigación en diagnósticos de enfermería. **Método:** estudio histórico-documental, cualitativo, fundamentado en la historia contemporánea, la historia intelectual, la historia social de la ciencia y la historia de los conceptos. El *corpus* reunió una fuente narrativa inédita, publicaciones científicas, ediciones de la clasificación diagnóstica, estudios metodológicos, literatura sobre validez, teoría de la medición, experticia y agregación de juicios, además de documentos institucionales disponibles. Las fuentes fueron sometidas a crítica externa e interna, organizadas en una matriz cronológico-temática y trianguladas para identificar hitos, actores, redes, continuidades, rupturas y cambios conceptuales. **Resultados:** la reconstrucción documental mostró que la revisión de los niveles de evidencia fue resultado de un proceso de maduración histórico-metodológica que articuló críticas a los métodos clásicos de validación, la revisión del papel del experto, las teorías de la validez y la medición, la causalidad diagnóstica, las teorías de rango medio y la circulación internacional de grupos brasileños de investigación. **Consideraciones finales:** los niveles de evidencia constituyen artefactos científicos, metodológicos, institucionales e históricos, marcados por continuidades y rupturas en la investigación sobre diagnósticos de enfermería.

Descriptores: Historia de la Enfermería; Diagnóstico de Enfermería; Investigación en Enfermería; Terminología Normalizada de Enfermería; Estudio de Validación.

INTRODUCTION

Nursing diagnoses occupy a central place in the contemporary history of nursing because they embody two complementary aspirations: to make nursing-specific clinical judgments visible and to establish a shared language for education, research, documentation, and clinical practice. Throughout the second half of the twentieth century, the consolidation of the nursing process, the development of nursing theories, and the computerization of clinical records increased the need for standardized nursing language systems capable of representing human responses, health problems, risks, interventions, and outcomes across diverse clinical and cultural contexts⁽¹⁻³⁾.

The International Nursing Knowledge Association (INKA), which originated from the North American tradition of nursing diagnosis classification and progressively expanded internationally, has become one of the principal institutional settings for this endeavor. Its classification system organizes diagnostic labels, definitions, defining characteristics, related factors, risk factors, at-risk populations, and associated conditions. At the same time, it functions as an instrument of epistemological authority by establishing criteria for the recognition, revision, and use of nursing diagnoses across different care settings^(1,2).

The history of levels of evidence should be understood within this broader context. These levels are not merely a technical mechanism; they reflect debates concerning validity, methodological quality, diagnostic inference, causality, expertise, and the cumulative production of knowledge through a disciplinary terminology. The transition from a classification system grounded primarily in consensus, established practice, and isolated validity studies to a framework more responsive to conceptual, clinical, theoretical, and methodological evidence may be understood as a historical turning point in nursing diagnosis research⁽⁴⁻⁸⁾.

In Brazil, this process acquired its own distinctive characteristics. From the earliest adoption of the nursing process and nursing diagnoses in centers located in São Paulo, Paraíba, and Ceará to the establishment of research networks within federal and state universities, Brazilian researchers have actively contributed to debates on validity, clinical reasoning, concept analysis, middle-range theories, diagnostic accuracy, and advanced quantitative methods⁽⁹⁻¹²⁾. The narrative source that motivated the present study documents, within its historical context, the emergence of a Brazilian research agenda that evolved from the application of classical validity methods toward an epistemological critique of criteria for validity and evidence⁽¹³⁾.

Although a substantial body of methodological literature addresses the validity of nursing diagnoses, historiographical studies devoted to reconstructing the development of levels of evidence remain scarce. Viewing them solely as technical criteria obscures their nature as scientific, institutional, methodological, and historical artifacts. This gap justifies examining the continuities, discontinuities, key actors, academic networks, and conceptual changes associated with their development, without assuming a linear trajectory or attributing the process to a single individual.

A historical understanding of levels of evidence also has potential implications for the teaching of nursing diagnoses, the education of clinical reasoning, the planning of validity studies, the assessment of methodological quality of evidence, clinical practice, and nursing documentation. It may also foster a more critical perspective on the processes of submitting, reviewing, and updating nursing diagnoses within international classification systems. These contributions are presented here as analytical possibilities derived from historical inquiry rather than as empirically tested outcomes of this study.

The historical problem guiding this article may be formulated as follows: how did the intellectual, methodological, and institutional trajectory that led to the revision of levels of evidence for nursing diagnoses within the international classification system develop? Addressing this question requires integrating intellectual biographies, academic networks, scientific events, methodological publications, conceptual transformations, and institutional changes while avoiding the reduction of this process to individual trajectories or to a teleological narrative of progress.

OBJECTIVE

To historically analyze the process of constructing and developing the levels of evidence of nursing diagnoses within the international classification system, with emphasis on the conceptual, methodological, and institutional shifts that integrated validity, expertise, causality, and nursing diagnosis research.

METHODS

Study design

This is a qualitative historical-documentary study informed by approaches from contemporary history, intellectual history, the social history of science, and the history of concepts. Contemporary history provided the framework for examining a recent process whose key actors, institutions, and impacts remain partially accessible. Intellectual history guided the contextual interpretation of arguments and conceptual uses. The social history of science directed attention to networks, research communities, the circulation of knowledge,

and the institutional conditions under which knowledge was produced. Finally, the history of concepts supported the analysis of semantic changes in terms such as validity, evidence, causality, and expertise⁽¹⁴⁻¹⁷⁾. Accordingly, levels of evidence were examined as situated methodological artifacts shaped by research practices, institutional decision-making, and the transnational circulation of knowledge.

Temporal and spatial scope

The study considered developments during the 1960s and 1970s to contextualize the expansion of the nursing process, nursing theories, and the early debates on nursing diagnoses. The primary analytical focus extended from the 1990s through the publication of revised criteria in 2024 and their subsequent developments. Earlier periods were included as essential historical antecedents for understanding the study object. Spatially, the analysis followed the circulation of knowledge between INKA and Brazilian academic networks without assuming that this axis encompasses the entire international history of the process.

Documentary corpus constitution

The *corpus* consisted of an initial narrative source (a doctoral dissertation) produced in 2024 on the content validity of nursing diagnoses, together with the scientific publications cited therein, editions of INKA classification, particularly those published in 2018–2020 and 2021–2023 and, when consistent with the confirmed temporal scope, those published in 2024–2026, methodological studies on diagnostic validity, literature addressing validity, content validity, measurement theory, psychometrics, expertise, consensus, and judgment aggregation, studies on causality and middle-range theories, available institutional documents concerning the submission and revision of nursing diagnoses, and historiographical references used to support interpretation^(1-13,18-38). A total of 34 sources were identified and analyzed.

Inclusion and exclusion criteria

Sources were included if they were directly relevant to the history of levels of evidence, the validity of nursing diagnoses, the participation of actors, research groups, or institutions identified in the narrative, the understanding of conceptual or methodological changes, or the comparison of narrative accounts with scientific publications and institutional documentation. Duplicate records, documents unrelated to the study object, abbreviated versions when complete versions were available, sources lacking identifiable authorship, date, or provenance, texts that merely mentioned levels of evidence without contributing to their historical reconstruction, and materials that were not fully accessible were excluded.

Source identification and selection

The search strategy began with the references and historical milestones identified in the initial narrative source, followed by backward citation tracking of the selected publications, consultation of available editions of the classification system, and review of accessible institutional documents. Documents were retrieved between March and August 2025. Source selection was progressively refined according to each document's relevance to the chronological and conceptual reconstruction of the study object.

Data extraction procedures

The sources were organized into a chronological-thematic matrix. The matrix included, among other fields, source identification and type, authorship, year, production context, institution, event or historical milestone, central concept, actors and academic networks, evidence of continuity and discontinuity, institutional changes, contributions to the development of levels of evidence, disagreements or ambiguities, and interpretative notes. The sources were independently examined by the authors, and interpretative disagreements were resolved through discussion.

Documentary criticism, analysis, and category development

External criticism examined authorship, date, purpose, format, provenance, and conditions of production. Internal criticism assessed narrative coherence, silences, emphases, contradictions, conceptual vocabulary, and the relationship between retrospective memory and documentary evidence. In practice, narrative statements concerning the chronology of the institutional revision were compared with successive editions

of the classification system, scientific publications, and the chronology of the events documented therein. When documentary convergence was insufficient, interpretations were expressed cautiously. The findings were grouped according to historical milestones, concepts, actors, events, continuities, and discontinuities, resulting in thematic cores addressing the antecedents of standardized nursing language, the development of the Brazilian research agenda, critiques of classical validity methods, validity and measurement theory, causality and middle-range theories, expertise and predictive diversity, and the institutional revision of levels of evidence. These groupings were not designated as a formal thematic analysis or content analysis method.

Source triangulation

Source triangulation integrated the primary narrative source, scientific publications, successive editions of the classification system, and available institutional documents (Chart 1)⁽¹⁻³⁸⁾. Its purpose was to verify chronological and conceptual convergence, contextualize retrospective accounts, and avoid unwarranted causal inferences. Convergence across different types of sources strengthened the interpretative basis of the findings. Divergences or the absence of supporting documentation were treated as limitations of the historical reconstruction and justified the use of cautious expressions such as “the sources suggest”, “the documentation indicates”, and “it may be interpreted”. When ambiguities arose, the original sources were consulted and, whenever necessary, confirmation was sought from a third independent source.

Chart 1 - Documentary *corpus* composition and analytical purpose. Fortaleza, Ceará, Brazil, 2026

Source type	Number	Period	Source or place of retrieval	Selection criteria	Analytical purpose
NANDA International Editions	3	2018–2026	Published editions of the classification ⁽¹⁻³⁾	Record of criteria, terminology, and editorial changes	Compare the institutional evolution of levels of evidence.
Publications on diagnostic validity	10	1979–2013	References identified in the narrative source and backward citation searching ^(4-6, 9-12, 36-38)	Direct methodological contribution	Reconstruct the continuities and critiques of classical methods.
Studies on causality and middle-range theories	4	2009–2017	Scientific publications and reference works ^(7, 8, 34, 35)	Contribution to etiological relationships and diagnostic structure	Examine the causal and theoretical inflection.
Primary narrative source	1	2024	Thesis for full professor ⁽¹³⁾	Direct relationship with the object and the record of milestones, actors, and networks	Identify the temporalities, methodological problems, and clues for triangulation.
Historiographical references	4	1979–2012	Published academic works ⁽¹⁴⁻¹⁷⁾	Explicit analytical function in design and interpretation	Support the history of concepts, the intellectual history, and the social history of science.
Publications on validity and measurement theory	4	1955–2014	Methodological literature cited in the manuscript ⁽¹⁸⁻²¹⁾	Relevance to the concepts of validity, content, and inference	Contextualize conceptual shifts.
Publications on expertise and aggregation of judgments	12	1975–2008	Literature cited in the manuscript ⁽²²⁻³³⁾	Relationship with selection, performance, and aggregation of evaluators	Analyze the revision of experts' role.

Source: prepared by the authors (2026).

Ethical aspects

As this was a documentary study, approval by a Research Ethics Committee was not required. Nevertheless, the authors complied with the principles established in Resolution 510 of April 7, 2016. Because the study drew upon personal recollections, proper names, and individual trajectories, biographical details that were not essential to the interpretation were minimized, and the authors were not identified within the body of the manuscript. The primary narrative source was treated as a situated, retrospective, and self-referential document rather than as a transparent account of historical events. The prominence of the Brazilian experience, particularly that of the state of Ceará, reflects the *corpus* composition and should be understood as an analytical delimitation rather than as an exhaustive representation of the international process.

RESULTS

Antecedents: standardized nursing language, the nursing process, and nursing diagnoses

The emergence of nursing diagnoses should be understood within the broader movement toward the formalization of professional knowledge. The nursing process, nursing theories, and classification systems shared the common goal of transforming dispersed practices into knowledge that could be communicated, taught, and investigated. In Brazil, the work of Wanda de Aguiar Horta played an important role in disseminating the nursing process and in recognizing that nursing care could be organized into interconnected stages of data collection, diagnosis, planning, intervention, and assessment⁽⁹⁾.

According to the trajectory reconstructed from the primary narrative source, the intellectual development of the researcher who produced the document was shaped by this transitional context. The first contact with the nursing process, the study of Horta's work, the introduction to Roy's Adaptation Model, and the encounter with diagnostic lists then associated with the North American Nursing Diagnosis Association (NANDA, the institutional designation at that time) are described as milestones that situated the adoption of nursing diagnoses in the state of Ceará during the mid-1990s⁽¹³⁾. This information is historically relevant not because of its autobiographical value, but because it illustrates the uneven chronology of taxonomy dissemination. While discussions in North America had already accumulated congresses and scientific publications throughout the 1970s and 1980s, the incorporation of nursing diagnoses into some Brazilian contexts was still relatively recent. Even among Brazilian states, interest in the subject developed unevenly, with the earliest discussions emerging in Paraíba and São Paulo before expanding to other states.

This dissemination did not occur through simple importation. The taxonomies were reinterpreted by graduate programs, research groups, nursing assessment courses, clinical experiences, and mentoring networks. As INKA expanded its presence beyond North America, it encountered in Brazil a rapidly developing research environment characterized by the consolidation of graduate education, the growth of scientific output, and the search for methodological approaches capable of supporting nursing diagnoses across different populations and healthcare settings.

Development of a Brazilian research agenda on nursing diagnoses

The primary narrative source describes the development of a Brazilian research agenda on nursing diagnoses through geographically situated academic networks. Institutions such as *Universidade de São Paulo*, *Universidade Federal da Paraíba*, *Universidade Federal de São Paulo*, *Universidade Federal Fluminense*, *Universidade Federal do Ceará*, *Universidade Federal do Rio Grande do Norte*, *Universidade Federal do Rio Grande do Sul*, and other centers are identified in the document as hubs for the circulation of theories, methods, classification systems, and research practices⁽¹³⁾. This enumeration reflects the perspective of the narrative source and should not, by itself, be interpreted as an exhaustive mapping of national scientific production.

In Ceará, the incorporation of nursing diagnoses was associated with graduate nursing education, the work of faculty members who had trained in centers already engaged in this field, and the establishment of research groups. According to the narrative, by the late 1990s nursing diagnoses were still a novel topic in local education, clinical practice, and research, making the experience in Ceará particularly sensitive to

a methodological question: how can reliable diagnostic inferences be produced in contexts where nursing assessment, clinical reasoning, and diagnostic language themselves were still being consolidated?⁽¹³⁾

In São Paulo and Paraíba, the contributions of researchers engaged in nursing care systematization, the nursing process, and diagnostic validity promoted the national dissemination of the work of Marjory Gordon, Patricia Risner, Richard Fehring, Lois Hoskins, Daniel Pesut, Lorraine Walker, and Kay Avant. The historical relevance of these individuals should be understood as part of scientific translation networks rather than as a sequence of individual or institutional tributes. The sources indicate the development of an epistemic community dedicated to stabilizing research objects, methods, and evaluative criteria surrounding nursing diagnoses^(4,5,10,11,37,38).

This Brazilian research agenda also expanded through scientific meetings, thesis and dissertation examination committees, educational courses, interinstitutional collaborations, and publications in national and international journals. The participation of Brazilian researchers in INKA activities and discussions on diagnostic revision strengthened the Latin American presence within a classification system that aspired to international scope. This aspect is historically significant because it shifts the narrative from one of passive reception to one of methodological co-production.

From the use of classical methods to methodological critique

Classical validity methods, particularly those associated with Richard Fehring, exerted a decisive influence on nursing diagnosis research. The diagnostic content validity model, the clinical validity model, and the emphasis placed on expert judges provided a relatively straightforward operational framework capable of guiding studies in a field seeking methodological standardization^(4,5).

The same simplicity that facilitated the dissemination of these methods also exposed their limitations. The critical literature began to question the reliance on expert judges, the weakness of formal expertise criteria, the difficulty of addressing disagreement among evaluators, the possibility of leniency bias, the tautological nature of some diagnostic inference processes, and the risk of incorporation bias when the same elements used to infer a diagnosis were subsequently assessed as evidence supporting its validity^(6,7,10,11).

The documentary narrative indicates that, between 2000 and 2010, Brazilian research groups explored alternative approaches based on evaluator training, performance monitoring, measures of diagnostic accuracy, sensitivity, specificity, and the analysis of clinical indicators. Publications from this period document part of this methodological shift^(7,10). Chronological comparison suggests a transition from simple consensus toward greater concern with inference, error, reproducibility, and the discriminatory power of clinical indicators.

Thus, the methodological crisis was not merely technical. It also affected the epistemological foundations of nursing diagnoses. If a nursing diagnosis is understood as an inference regarding a human response, its validity requires distinguishing manifestations, causes, associated conditions, context, clinical judgment, and practical consequences. Failure to make these distinctions may produce diagnoses that are formally accepted but conceptually unstable or of limited clinical usefulness.

Validity, measurement theory, and conceptual reconfiguration

The need to define validity led nursing diagnosis research to engage with psychometrics, measurement theory, and test validity theory. Beginning with the work of Lee Cronbach and Paul Meehl, validity came to be understood not merely as a property of an instrument but as the inferences supported by evidence⁽¹⁸⁾. Samuel Messick expanded this perspective by integrating content validity, internal structure, relationships with other variables, response processes, and the consequences of test use^(19,20). Contemporary educational and psychological testing standards further consolidated the view that validity refers to the degree to which evidence and theory support the interpretation of test scores for their intended uses⁽²¹⁾.

Applied to nursing diagnoses, this perspective raises the following question: what evidence supports the interpretation that a particular set of defining characteristics, related factors, and associated conditions represents a human response relevant to nursing practice? In this context, validity cannot be reduced to the opinion of expert judges, the frequency of clinical indicators, or the existence of clinical studies. Rather, it depends on the coherence among the concept, definition, indicators, causal structure, methods of observation, and consequences for intervention and outcomes.

Content validity, frequently employed in nursing diagnosis research, has also undergone reinterpretation. The work of Charles Lawshe, Mary Lynn, Stephen Haynes, and Denise Polit contributed to clarifying that content validity encompasses the domain relevance, representativeness, clarity, and comprehensiveness, as well as systematic judgment processes^(22–25). The central question therefore shifted from simply asking “how many experts agree?” to considering “who is making the judgment?”, “how was their performance established?”, “which criteria were applied?”, “which rating scale was used?”, “what sources of bias may be present?”, and “how should the resulting indices be interpreted?”.

This conceptual shift helps explain the revision of INKA levels of evidence as the result of a broader conceptual transformation. The revised levels of evidence came to require more clearly differentiated forms of evidence, including concept analysis, operational definitions, content validity studies, clinical studies, assessment of etiological factors, the use of appropriate methodological approaches, and coherence between the diagnostic structure and the accumulated body of evidence^(1,2).

Causality, middle-range theories, and the structure of nursing diagnoses

The discussion of causality represented another major conceptual shift. Nursing diagnoses are not merely labels, but they organize the relationships among human responses, defining characteristics, related factors, risk factors, at-risk populations, and associated conditions. Distinguishing among these elements has direct implications for diagnosis, intervention, and outcomes. If a causal factor is treated as a manifestation, or a manifestation is treated as a cause, both clinical inference and therapeutic decision-making become weakened.

The documentary narrative reports that, beginning in 2013, discussions regarding related factors and the scarcity of etiological studies contributed to the development of a middle-range theory on causality and the validity of nursing diagnoses. The publication of this proposal sought to integrate diagnosis, intervention, and outcomes, emphasizing that causal models can guide both diagnostic validity and knowledge translation into clinical practice⁽⁸⁾.

The terminological reorganization observed across successive editions of INKA classification, with greater emphasis on related factors, risk factors, at-risk populations, and associated conditions, may be interpreted as part of an institutional movement toward greater causal clarity. The available sources do not permit attributing this development directly to a single research group or publication. Rather, triangulation of scientific publications, successive editions of the classification, and the documentary narrative suggests convergence between Brazilian concerns regarding causality and broader changes within the classification system^(1,2,8,13).

Middle-range theories played an important historical role in this process because they made it possible to connect abstract concepts, diagnostic components, and empirically testable research questions. Unlike purely taxonomic approaches, they encouraged inquiry into mechanisms, antecedents, consequences, and the clinical application of nursing diagnoses.

Expertise, the wisdom of crowds, and predictive diversity

The revision of the concept of expertise was fundamental to rethinking content validity. In classical models, expertise was generally defined according to formal criteria such as academic qualifications, professional experience, scientific publications, or clinical practice. Although these criteria are useful, they do not necessarily guarantee superior performance in judging diagnostic components. The literature on expertise indicates that performance, calibration, consistency, and discriminative ability are more robust indicators than formal authority alone^(26–29).

This critique is particularly relevant when judges are asked to assess the relevance and representativeness of diagnostic components. Individual authority may amplify bias, produce leniency, and obscure disagreement among evaluators. The Dunning–Kruger effect, described as the discrepancy between actual competence and self-perceived performance, reinforces the need for caution when accepting self-reported expertise as sufficient evidence of competence⁽³⁰⁾.

Within this context, the wisdom of crowds and the diversity prediction theorem provided an alternative perspective. The central argument is not that any crowd necessarily produces better judgments, but rather that diverse, independent, and appropriately aggregated groups may outperform individual judges, particularly when individual errors offset one another and meaningful cognitive diversity is present^(31–33). Scott Page

demonstrated that predictive diversity can play a mathematically significant role in the quality of collective judgments, provided that participants possess a minimum level of competence and that appropriate aggregation methods are used^(32,33).

Applied to the content validity of nursing diagnoses, this perspective shifts the question from “who is the expert?” to “how should expertise, diversity, training, performance, and independence of judgment be combined?” weighting according to the level of expertise, the use of simulation methods, comparison of estimators, and assessment of precision became part of a methodological agenda aimed at improving inference regarding the relevance and representativeness of diagnostic components.

Revision of levels of evidence of the international classification as a historical product

The primary narrative source identifies participation in an INKA committee beginning in 2014 as an institutional milestone in the revision of levels of evidence. According to this document, the work was conducted through collaboration among Brazilian and international researchers and continued until the publication of revised levels of evidence in the 2021–2023 edition of the classification^(2,13). This chronology is consistent with the maturation of discussions on causality, validity methods, diagnostic accuracy, content validity, and validity theory. However, temporal consistency alone does not establish a direct causal relationship.

The 2018–2020 edition had already documented concerns regarding diagnoses for which no evidence level had been assigned or which required substantial updating, in addition to introducing new requirements for the submission of nursing diagnoses⁽²⁾. The 2021–2023 edition consolidated the revised evidence level criteria and included Brazilian editorial participation, reflecting the internationalization of taxonomic governance⁽²⁾. Chart 2 summarizes the major chronological milestones in the historical and methodological development of INKA nursing diagnosis levels of evidence.

Chart 2 - Chronological milestones of the historical-methodological process. Fortaleza, CE, Brazil, 2026

Period	Milestone	Historical interpretation
1960s–1970s	Expansion of the nursing process and nursing theories; emergence of debates regarding nursing diagnoses.	Establishment of the disciplinary foundation for a standardized diagnostic language.
1979	Publication and dissemination of Wanda de Aguiar Horta's work on the nursing process.	Strengthening of the nursing process in Brazil and opening of debates on diagnosis.
1980s–1990s	Dissemination of classic validity methods, particularly Fehring's models.	Initial standardization of content studies and clinical validity.
1994–1998	Initial exposure in Ceará to diagnostic lists, the nursing process, Roy, and NANDA International; dissertation on diagnostic profiles.	Local adoption of a language in international circulation.
2000–2010	Growing criticism of classic methods; incorporation of statistics, accuracy measures, and rater training.	Shift in focus from simple consensus to inference, reliability, and performance.
2009–2013	Brazilian publications on the accuracy of indicators and methodological alternatives for diagnostic validity.	Expansion of the national and international methodological agenda.
2013–2017	Discussions and publication of mid-range theory on causality and validity.	Clarification of the relationship between causal factors, indicators, intervention, and outcomes.
2014–2020	Participation in a NANDA International committee and review of levels of evidence.	Institutional articulation of validity, causality, methods, and submission criteria.
2021–2023	Publication of the classification edition with revised criteria for level of evidence.	Editorial consolidation of a methodological shift in classification.
2018–2024	Development of a thesis on content validity, expertise, wisdom of crowds, and predictive diversity.	Theoretical-methodological elaboration of the diagnostic validity agenda.

Source: prepared by the authors (2026).

Historically, the revision of levels of evidence may be interpreted as an attempt to align the classification system with more rigorous methodological standards. It sought to differentiate types of evidence, emphasize theoretical foundations, distinguish diagnostic components, encourage clinical studies, and reinforce the need for coherence among concepts, methodology, and application. This did not represent a complete break with the previous tradition, as consensus and peer review remained relevant. Rather, it represented a reconfiguration of consensus through more explicit evidence criteria.

Source triangulation suggests that scientific meetings, collaboration among Brazilian research groups, visits by international researchers, methodological publications, and participation in committees formed an ecosystem that supported the development of revised levels of evidence. This interpretation is historiographical and is supported by convergent documentary evidence rather than by a fully established causal chain. The levels of evidence emerged as an institutional response to shared challenges, including insufficient validity, causal inconsistencies, weaknesses in expertise criteria, and the need to support nursing diagnoses with methodologically defensible evidence. The concepts involved in this reconfiguration are summarized in Chart 3.

Chart 3 - Analytical concepts and contribution to levels of evidence. Fortaleza, CE, Brazil, 2026

Concept	Conceptual synthesis	Historical contribution to levels of evidence
Validity	The degree to which evidence and theory support inferences and uses of an interpretation.	Reorients levels of evidence, going beyond the formal acceptance of diagnostic labels.
Content validity	Assessment of relevance, representativeness, clarity, and comprehensiveness of elements within a domain.	Refines judgments regarding defining characteristics, factors, and conditions.
Latent variable	A construct not directly observed but inferred from observable indicators.	Supports the understanding of diagnosis as an inference rather than raw data.
Causality	Relationships among antecedents, conditions, mechanisms, and manifestations.	Distinguishes between related factors, risk factors, populations at risk, associated conditions, and defining characteristics.
Expertise	Situated competence, dependent on performance, experience, calibration, and reliability.	Redefines the selection and weighting of raters in content and consensus studies.
Wisdom of the crowd	Aggregation of independent and diverse judgments to produce more robust estimates.	Offers an alternative to individual authority and restricted expert panels.
Predictive diversity theorem	Formulation stating that cognitive diversity can improve collective performance under specific conditions.	Bases the weighting and simulation of estimators on content validity.
Middle-range theory	Intermediate theoretical framework connecting abstract concepts to testable propositions.	Enables the alignment of diagnosis, intervention, outcome, and empirical validity.

Source: prepared by the authors (2026).

DISCUSSION

Historical interpretation of the findings

The history of levels of evidence demonstrates that professional classification systems are historical constructions. They incorporate values, methods, documentation technologies, expectations of scientific rigor, and forms of authority. From the perspective of the history of concepts, changes in the meanings of validity, evidence, causality, and expertise are not merely terminological; they reshape the problems that can be formulated and the criteria by which a diagnostic statement becomes acceptable⁽¹⁴⁾. Likewise, the

contextualist approach of intellectual history precludes separating methodological arguments from the historical circumstances in which they were formulated and applied⁽¹⁵⁾.

The social history of science makes it possible to understand levels of evidence as the product of research collectives, academic networks, the circulation of knowledge, and institutions, rather than as the outcome of an individual trajectory. The notion of a thought collective helps explain how styles of reasoning become stabilized and transformed through interactions among researchers, editors, and users of the classification system⁽¹⁶⁾. Complementarily, the social history of knowledge draws attention to the processes through which knowledge is produced, validated, and disseminated, making certain criteria more visible and authoritative than others⁽¹⁷⁾.

Continuities, tensions, and counterpoints

The revision did not eliminate expert assessment. Despite criticisms of simple consensus, expert judgment remained essential for defining domains, interpreting evidence, and revising diagnostic components. The change consisted less in abandoning experts than in making explicit the criteria for their selection, performance, diversity, independence, and aggregation. This continuity limits interpretations that portray the revision as a complete rupture.

Another tension concerns the relationship between international standardization and cultural diversity. Although common criteria facilitate comparability and editorial governance, they may also privilege problems, methods, and terminology developed in centers with greater research capacity and international visibility. The participation of researchers from Brazil and Latin America broadens the diversity of the process but does not eliminate inequalities related to language, funding, editorial access, and scientific recognition. The sources examined do not allow these asymmetries to be mapped comprehensively; however, they recommend caution regarding universalizing narratives.

More complex evidence criteria may also produce ambivalent effects. Although they strengthen the submission and review process, they increase the demands placed on study design, sample size, methodological competence, and available resources. Recently proposed, culturally situated, or empirically underdeveloped nursing diagnoses may encounter greater difficulty progressing, even when they are clinically relevant. Moreover, there is a risk that levels of evidence may become a rigid hierarchy or a bureaucratic mechanism detached from the substantive quality of research questions and the appropriateness of methodological approaches.

The coexistence of methodological innovation and institutional continuity suggests that changes in classification systems occur through negotiation, gradual incorporation, and resistance. The available documentation does not allow the identification of specific opponents or the reconstruction of all editorial debates. The absence or inaccessibility of meeting minutes, correspondence, and preliminary versions prevents attributing direct causality between the work of a particular research group and institutional decisions. Therefore, the chronological proximity among publications, committee participation, and editorial changes should be interpreted as plausible convergence rather than as conclusive evidence of influence.

STUDY LIMITATIONS

In its initial phase, the analysis was based on an autobiographical, retrospective, and self-referential narrative source. Although this approach is useful for identifying temporalities, key actors, and methodological issues, it may also be influenced by selective memory and retrospective knowledge. The unequal availability of documentary sources may result in documentary survival bias and in asymmetry between published sources and institutional records. The centrality of the Brazilian experience, particularly that of Ceará, broadens the understanding of an important research network but does not represent the entire international circulation of nursing diagnoses.

The absence of complete meeting minutes, internal correspondence, and preliminary editorial versions made it difficult to reconstruct collective decision-making processes and prevented the attribution of direct causality between publications, individual participation, and editorial changes. It was also not possible to eliminate entirely the risk of interpreting past events according to current methodological standards. These limitations were addressed through documentary criticism, source triangulation, and the use of cautious interpretative formulations.

CONTRIBUTIONS TO THE HISTORY OF NURSING, HEALTH, AND EDUCATION

Historicizing levels of evidence may promote a more critical understanding of validity methods and help educators and students recognize that classification systems are historically produced, negotiated, and revised. For the education of clinical reasoning, this perspective makes the distinctions among manifestations, causes, associated conditions, risks, and consequences more explicit, preventing taxonomy from being viewed as a static list.

In research, this historical reconstruction may guide the selection of study designs that are appropriate for specific diagnostic components, strengthen the methodological assessment of evidence, and increase the transparency of submission and review processes. In clinical practice and nursing documentation, it may encourage a more critical use of nursing diagnoses and reinforce the need for continuous updating of their components. These are presented as potential implications of historical knowledge rather than as outcomes directly assessed in the present study.

CONCLUSIONS

This study historically analyzed the construction and development of levels of evidence of nursing diagnoses within the NANDA International classification. The historical reconstruction identified continuities reflected in the persistence of expert judgment, peer review, and discipline-specific diagnostic language, as well as discontinuities related to the expansion of validity requirements, criticism of simple consensus, and the demand for coherence among concepts, causal structure, methodology, and empirical evidence.

Levels of evidence should be understood simultaneously as scientific, methodological, institutional, and historical constructions. Their development integrated validity theory, content validity, diagnostic inference, causality, expertise, judgment aggregation, and middle-range theories. Brazilian and Latin American participation became part of this process through academic networks, scientific publications, scholarly events, and institutional engagement, although the available documentation does not allow these changes to be attributed to a single researcher or research group.

Future research should expand access to institutional archives, meeting minutes, correspondence, and editorial drafts; conduct interviews with different participants; compare national trajectories; and map networks of scientific circulation and translation. Such investigations may clarify tensions, silences, and decision-making mechanisms that remain insufficiently documented, thereby deepening the historical understanding of standardized nursing languages.

REFERENCES

1. Herdman TH, Kamitsuru S, Lopes CT, editores. NANDA International nursing diagnoses: definitions and classification 2018-2020. 11th ed. New York: Thieme; 2018.
2. Herdman TH, Kamitsuru S, Lopes CT, editores. NANDA International nursing diagnoses: definitions and classification 2021-2023. 12th ed. New York: Thieme; 2021.
3. Herdman TH, Kamitsuru S, Lopes CT, editores. NANDA International nursing diagnoses: definitions and classification 2024-2026. 13th ed. New York: Thieme; 2024.
4. Fehring RJ. Methods to validate nursing diagnoses. *Heart Lung*. 1987 [cited 2025 June 28];16(6):625-9. Available from: https://epublications.marquette.edu/nursing_fac/27/
5. Fehring RJ. Validation: validating diagnostic labels. In: Hurley ME, editor. *Classification of nursing diagnoses: proceedings of the sixth conference of North American Nursing Diagnosis Association*. St. Louis: Mosby; 1986. p. 183-90.
6. Lopes MVO, Silva VM, Araujo TL. Validação de diagnósticos de enfermagem: desafios e alternativas. *Rev Bras Enferm*. 2013;66(5):649-55. <https://doi.org/10.1590/S0034-71672013000500002>.
7. Lopes MVO, Silva VM, Araujo TL. Methods for establishing the accuracy of clinical indicators in predicting nursing diagnoses. *Int J Nurs Knowl*. 2012;23(3):134-9. <https://doi.org/10.1111/j.2047-3095.2012.01213.x>.
8. Lopes MVO, Silva VM, Herdman TH. Causation and validation of nursing diagnoses: a middle range theory. *Int J Nurs Knowl*. 2017;28(1):53-9. <https://doi.org/10.1111/2047-3095.12104>.

9. Horta WA. Processo de enfermagem. São Paulo: EPU; 1979.
10. Gordon M. Nursing diagnosis: process and application. 3rd ed. St. Louis: Mosby; 1994.
11. Walker LO, Avant KC. Strategies for theory construction in nursing. 5th ed. Boston: Pearson; 2011.
12. Fawcett J. Contemporary nursing knowledge: analysis and evaluation of nursing models and theories. 2nd ed. Philadelphia: F.A. Davis; 2005.
13. Lopes MVO. Medidas de agregação para estudos sobre validade de conteúdo de diagnósticos de enfermagem [tese]. Fortaleza: Universidade Federal do Ceará, Faculdade de Farmácia, Odontologia e Enfermagem; 2024. 563 p.
14. Koselleck R. Futures past: on the semantics of historical time. New York: Columbia University Press; 2004.
15. Skinner Q. Visions of politics. Vol. 1, regarding method. Cambridge: Cambridge University Press; 2002.
16. Fleck L. Genesis and development of a scientific fact. Chicago: University of Chicago Press; 1979.
17. Burke P. A social history of knowledge II: from the Encyclopaedia to Wikipedia. Cambridge: Polity Press; 2012.
18. Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull. 1955;52(4):281-302. <https://doi.org/10.1037/h0040957>.
19. Messick S. Validity. In: Linn RL, editor. Educational measurement. 3rd ed. New York: Macmillan; 1989. p. 13-103.
20. Messick S. Validity of psychological assessment: validation of inferences from persons responses and performances as scientific inquiry into score meaning. Am Psychol. 1995;50(9):741-9. <https://doi.org/10.1037/0003-066X.50.9.741>.
21. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for educational and psychological testing. Washington, DC: American Educational Research Association; 2014.
22. Lawshe CH. A quantitative approach to content validity. Pers Psychol. 1975;28(4):563-75. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>.
23. Lynn MR. Determination and quantification of content validity. Nurs Res. 1986;35(6):382-5. <https://doi.org/10.1097/00006199-198611000-00017>.
24. Haynes SN, Richard DCS, Kubany ES. Content validity in psychological assessment: a functional approach to concepts and methods. Psychol Assess. 1995;7(3):238-47. <https://doi.org/10.1037/1040-3590.7.3.238>.
25. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity?: Appraisal and recommendations. Res Nurs Health. 2007;30(4):459-67. <https://doi.org/10.1002/nur.20199>.
26. Ericsson KA, Krampe RT, Tesch-Romer C. The role of deliberate practice in the acquisition of expert performance. Psychol Rev. 1993;100(3):363-406. <https://doi.org/10.1037/0033-295X.100.3.363>.
27. Shanteau J. Competence in experts: the role of task characteristics. Organ Behav Hum Decis Process. 1992;53(2):252-66. [https://doi.org/10.1016/0749-5978\(92\)90064-E](https://doi.org/10.1016/0749-5978(92)90064-E).
28. Cooke RM. Experts in uncertainty: opinion and subjective probability in science. New York: Oxford University Press; 1991.
29. Cooke RM, Goossens LLHJ. TU Delft expert judgment data base. Reliab Eng Syst Saf. 2008;93(5):657-74. <https://doi.org/10.1016/j.ress.2007.03.005>.
30. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing ones own incompetence lead to inflated self-assessments. J Pers Soc Psychol. 1999;77(6):1121-34. <https://doi.org/10.1037/0022-3514.77.6.1121>.
31. Surowiecki J. The wisdom of crowds. New York: Doubleday; 2004.
32. Hong L, Page SE. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. Proc Natl Acad Sci U S A. 2004;101(46):16385-9. <https://doi.org/10.1073/pnas.0403723101>.
33. Page SE. The difference: how the power of diversity creates better groups, firms, schools, and societies. Princeton: Princeton University Press; 2007.

34. Pearl J. Causality: models, reasoning, and inference. 2nd ed. Cambridge: Cambridge University Press; 2009.
35. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
36. Carpenito-Moyet LJ. Nursing diagnosis: application to clinical practice. 13th ed. Philadelphia: Lippincott Williams & Wilkins; 2010.
37. Gordon M. Manual of nursing diagnosis. 12th ed. Sudbury: Jones & Bartlett Learning; 2010.
38. Risner PB. Diagnosis: analysis and synthesis of data. In: Christensen PJ, Kenney JW, editors. Nursing process: application of conceptual models. 4th ed. St Louis: Mosby; 1995. p. 132-57.

Submission: 2026 May 29

Reformulation: 2026 July 23

Approval: 2026 July 27

Editor-in-Chief: Deybson Borba de Almeida

Associate editor: Maria Itayra Padilha

***Ad hoc* reviewers:**

Shirley da Rocha Afonso

Maria Itayra Padilha

AUTHORS' CONTRIBUTION

Conception of the study: MVOL, VMS

Data collection: MVOL, VMS

Data analysis: MVOL, VMS

Manuscript writing: MVOL, VMS

Critical revision for important intellectual content: MVOL, VMS