

# Implementation Leadership in the Point of Care Nursing Context: A Systematic Review Comparing Two Measurement Tools

SAGE Open Nursing  
Volume 9: 1–18  
© The Author(s) 2023  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/23779608231216161  
journals.sagepub.com/home/son



Sonia Angela Castiglione, RN, MScA, PhD Candidate<sup>1</sup> ,  
Julie Frechette, RN, PhD<sup>1,2</sup> and V. Ramanakumar Agnihotram, PhD<sup>3</sup>

## Abstract

**Introduction:** Implementation leadership (IL) are effective point of care (POC) nursing leadership behaviors that facilitate contexts conducive to the successful implementation of evidence-based practices (EBPs). However, no systematic evaluation of IL tools validated for the nursing context existed.

**Aims:** The purpose of this systematic review was to compare iterations of two IL measurement tools, the Implementation Leadership Scale (ILS) and the iLEAD, for application in a nursing context; and to critically appraise and summarize the methodological quality of studies assessing their psychometric properties.

**Methods:** A comprehensive search was conducted in four databases. Two reviewers independently screened titles and abstracts, reviewed full-text articles, and performed extraction into data tables. Statisticians appraised the quality control aspects. Findings were narratively summarized.

**Results:** A total of 247 records were included, where 10 for the ILS (including different versions) and one for the iLEAD met the inclusion criteria. Three studies evaluated the psychometric properties of the ILS in nursing, and its translations into Chinese and Greek. Content validity was deemed to be doubtful for both tools, but the ILS had adequate rating for comprehensiveness; methodological quality was very good for structural validity, internal consistency, hypothesis testing, and responsiveness where applicable for both scales, with the exception of cross-cultural validity which had ratings of adequate and inadequate for versions of the scales. Several study findings met the criteria for good measurement properties. No studies for either tool formally assessed feasibility.

**Conclusion:** Applying validated and contextually relevant tools to evaluate the capacity of nursing leadership to engage in IL in real-world contexts are needed. The ILS shows promise but requires further validation for contexts with diverse and multiple nursing leaders at the POC. Feasibility needs to be further studied.

## Keywords

implementation leadership, psychometric properties, systematic review, implementation leadership scale, iLEAD, nursing

Received 1 June 2023; Revised 10 October 2023; accepted 5 November 2023

## Introduction

Implementing evidence-based practices (EBPs) is essential for improved patient outcomes and ensuring safe, effective, and patient-centered nursing care (Connor et al., 2023), but has had limited success in real-world settings due to a lack of facilitative contexts (Dryden-Palmer et al., 2020; Wensing & Grol, 2019). Effective leadership is arguably the most critical factor in creating conducive environments for the successful implementation of EBPs (Li et al., 2018). While in scholarship and practice fields, the importance of leadership in the change process has long been accepted,

research to illuminate how leadership influences context and implementation outcomes (e.g., moderators) is emerging (Meza et al., 2021; Nilsen & Bernhardtsson, 2019).

<sup>1</sup>Ingram School of Nursing, McGill University, Montréal, Canada

<sup>2</sup>Professional Development Directorate, Ordre des infirmières et infirmiers du Québec, Montréal, Canada

<sup>3</sup>Research Institute McGill University Health Center, Montréal, Canada

### Corresponding Author:

Sonia Angela Castiglione, Ingram School of Nursing, 680 rue Sherbrooke, Suite #1800, McGill University, Montréal, Quebec, H3A 0G4, Canada.  
Email: sonia.castiglione@mcgill.ca



Diverse leader types characterize nursing leadership at the point of care (POC); nurse managers, advanced practice nurses (clinical nurse specialists and nurse practitioners), nursing professional development educators, and nurse champions have role expectations to improve care, develop nursing practice and oversee local implementation planning, communication, and monitoring (Bianchi et al., 2018; Pol-Castañeda et al., 2020). POC nurse leaders employ enforcing and enabling strategies to influence nursing staff in adopting EBPs, enacting task and relational behaviors depending on the leader's formal authority over staff (Harvey et al., 2020). Several leadership interventions to strengthen nursing leaders' competencies to support the implementation of EBPs at the POC have been developed and tested for feasibility (Gifford et al., 2013, 2019; Kvist et al., 2014; Tistad et al., 2016; Valimaki et al., 2018).

Implementation leadership (IL) is a concept that recently emerged in implementation science literature describing a set of strategic, effective behaviors enacted by POC leaders to promote the successful uptake of EBPs (Castiglione, 2020). These behaviors are distinguished from and build on notions of general leadership theory and propose unique knowledge and skills leaders need to effectively facilitate implementation, including general leadership abilities, identifying EBPs gaps in the setting, setting goals, selection of implementation strategies, and active engagement of staff (Gifford et al., 2017). Recent work by Shuman et al. (2020, 2023) has linked unit nurse managers' IL behaviors to positive implementation climates, a mediator to implementation that reflects the perception that EBPs are rewarded, supported, and expected by adopters in a given setting (Williams et al., 2020).

Conceptually valid and reliable measures of IL are required to predict and determine whether leaders enact IL behaviors and have the requisite competencies. Pragmatic measures of IL may be particularly helpful in planning for implementation, and specifically for determining if and what strategies can be targeted at leaders to strengthen implementation climate (Powell et al., 2021). Based on a recent concept analysis of IL (Castiglione, 2020), the authors are aware of two tools that have already been developed: The Implementation Leadership Scale (ILS) (Aarons et al., 2014) and the iLEAD (Mosson et al., 2018). What is not known is whether these tools are psychometrically sound and feasible for use in nursing practice and professional development with a diversity of POC nursing leaders. To the best of our knowledge, our review is the first to evaluate the methodological quality and adequacy of psychometric properties of IL-specific measures. We will also nuance our synthesis to address the quality and suitability of the two IL tools for use in real-world nursing POC contexts.

## Aim

This review aims to compare, critically appraise, and summarize two self-report IL measures, the ILS and the iLEAD, and their iterations for use in the nursing context.

## Methods

### Design

A systematic review of outcome measures was carried out according to the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN), a rigorous methodological and evaluative framework for identifying and selecting tools for use in research and clinical practice (Prinsen et al., 2018). The Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) guided this report (Page et al., 2021) (Supplemental Table S1).

### Eligibility Criteria

Relevant studies were included if: (1) the study described the development, assessment of validity or feasibility of the original or adapted ILS or iLEAD tools; (2) the study sample consisted of any healthcare leaders and professionals/workers at the POC (in nursing or not); (3) the report was peer-reviewed; (4) written in English or French, and (5) the full text was retrievable. Titles were excluded if the ILS or iLEAD were used as outcome measures or to validate another instrument, as per COSMIN guidelines.

### Information Sources and Search Strategy

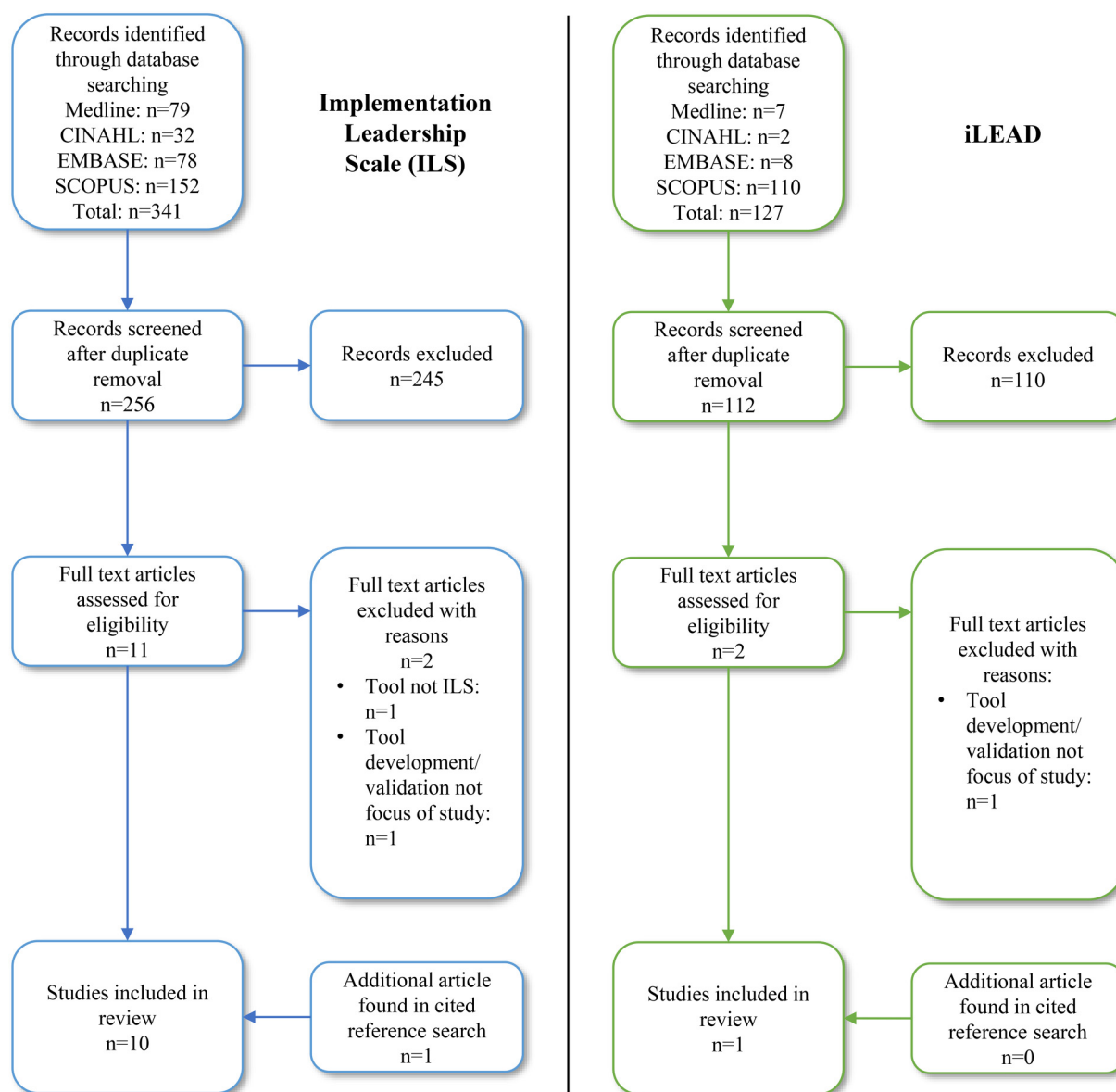
In consultation with a medical librarian, two independent searches (2020, 2021) were performed using keywords and tool names (Supplemental Table S2) in four online health-related bibliographic databases: Medline via OVID, EMBASE via OVID, CINAHL via Ebsco, and Scopus. No additional titles were located through ancestral and dependency searching, and in the COSMIN database for existing systematic reviews.

### Selection Process

Two reviewers (SAC and JF) independently screened the articles and abstracts, considering the inclusion criteria. Classification discrepancies were discussed and resolved among the two reviewers. Titles were classified as included, unsure, or excluded. Titles classified as "included" or "unsure" underwent full-text screening. Full text articles that met the inclusion criteria were included in this review.

### Data Extraction and Items

Data were extracted from each study into five separate tables to describe: (1) study characteristics; (2) psychometric properties; (3) tool characteristics; (4) results of the assessed measurement properties, and (5) the translation and linguistic validation process. Study authors (G. Aarons, A. Richter, J. Hu, P. Sarafis, C. Shuman, J.D. Smith) were contacted for additional information not reported in their study to



**Figure 1.** PRISMA flow diagram.

clarify study methods and included samples. Extraction was conducted independently and then reviewed by SAC and JF to reach a consensus on the accuracy of the data presented in the tables.

### Quality Appraisal and Data Synthesis

Methodological quality was evaluated using COSMIN guidelines to assess content validity, the internal structure (validity and consistency), and the properties related to the quality of the ILS and iLEAD at the full scale and subscale levels (Mokkink et al., 2018). (See Supplemental Table S3 for a description of outcomes.)

The standards for development and content validity assessment were adapted to reflect the construct of interest (e.g., patients replaced by “employees and POC leaders”; professionals replaced by “experts, researchers, and senior leaders”). Measurement properties were evaluated by three independent raters (SAC, JF, and a joint rating between VRA and another statistician, except for content validity, which SAC and JF assessed) using the very good, adequate, doubtful, or inadequate scoring standard. For each rater, an overall risk of bias score for the psychometric property for each study was determined by selecting the lowest quality rating. The overall score for each psychometric property per study was determined by a consensus of at least two scores, which was reached after the

initial rating or following a discussion among the raters. Data on measurement properties were then compared for each study against the criteria rating for good measurement properties stipulated by the COSMIN guidelines (Prinsen et al., 2018).

The accuracy of translation and linguistic validation, defined as the extent to which the construct is equivalent in a different language (WHO, 2020), was also compared to the guidelines by Sousa and Rojjanasrirat (2011). This supplementation was necessary since COSMIN removed guidance on scale translations (Mokkink et al., 2018). Similarly, the ease of administration and resources required to complete the measure in the intended context was considered feasible (Smart, 2006). Though not considered a property of the tool, feasibility significantly affects the implementation of the tool in the nursing practice context and therefore was reported on in this review.

Data were narratively synthesized across tools and findings related to applicability of the tools in the nursing context were discussed. Data were collapsed, combined, and summarized from the extraction and evaluation tables into reporting tables for conciseness and clarity.

## Results

### Study Selection

Search results are presented as a PRISMA diagram (Moher et al., 2009) (Figure 1). Following duplicate removal, 247 records for the ILS and 40 for the iLEAD tools were screened for eligibility. From these, 11 articles met the inclusion criteria and were included in this review. A total of five versions of the ILS tool (described in  $n = 10$  studies) were found. These included the two original versions of the ILS tool: (1) the ILS Leader (leader self-report), (2) the ILS Staff (staff report) (Aarons et al., 2014, 2016; Finn et al., 2016; Gifford et al., 2017; Shuman et al., 2020; Torres et al., 2018), and three culturally adapted ILS tools: (3) ILS Staff-Chinese (Hu et al., 2019, 2021), (4) ILS Staff-Greek (Mandrou et al., 2020), and (5) ILS (Adapted) (Smith et al., 2020). Only one version of the iLEAD tool ( $n = 1$  study) was found (Mosson et al., 2018).

### Study Characteristics

Study characteristics are detailed in Table 1. Included studies were published from 2014 to 2021 and described: Theoretical scale development, adaptation, and content validation of ILS versions and iLEAD ( $n = 4$ ); validation with POC healthcare employees and supervisors in community mental health ( $n = 2$ ), child welfare service providers ( $n = 1$ ), and hospital centers ( $n = 6$ ). Validation research was mainly conducted within the USA, except for four studies in China, Greece, and

Sweden. Three studies included samples of nurse managers and nursing staff.

### Tool Characteristics

The ILS Staff and Leader tools aim to capture the perspective of staff or leaders' self-perceptions of IL. Each consists of 12 items divided into four subscales (3 items per subscale) measuring the extent to which frontline managers are proactive, knowledgeable, supportive, and perseverant in EBP implementation (Table 2). Item generation was based on a literature review of leader behaviors related to performance, organizational climate, and culture change, yet the results of the review were not explicitly detailed or cited in the report (Aarons et al., 2014). Furthermore, seven experts reviewed the items, but findings (e.g., additions, modifications, or deletions) were not reported. The ILS (Adapted) scale retained six items from the original ILS, determined following feedback and pilot testing with stakeholders. The syntax for ILS items within and across scale versions is consistent. An example of an item includes "[Name of supervisor] has developed a plan to facilitate implementation of evidence-based practice." Word changes that maintained the sense of the original word used in the scale were made to mitigate comprehension issues found during the linguistic validation of the Chinese ILS. Similarly, the term "employee" was replaced with "nurse" during the validation of the ILS staff in a nursing context. Responses are measured on an ordinal 5-item scale (response range from 0 to 4), with scores calculated as an average of items per subscale and a total scale score. The ILS Staff and Leader were initially developed to evaluate a specific leadership development intervention and are intended for use in research, practice, and training contexts. The ILS (Adapted) differed in the item scale responses, which ranged from 1 to 5, and only the perseverant leadership subscale score was calculated.

In comparison, the iLEAD aims to describe the type and quality of leadership behaviors for EBP implementation. The authors adapted items and the factor structure from a tool describing transformation and transactional leadership based on active and passive leadership domains of the FRLM (Kelloway et al., 2006). Items were also adapted from another tool describing line managers' attitudes and actions (Randall et al., 2009). To further identify leadership behaviors,  $n = 45$  experts participated in a "co-created program logic" process. The methods were vague, consisting of "group sharing," then an analytic approach to transform behaviors into themes. The authors reported congruence between existing literature and emerging themes, but details were not forthcoming. The resulting iLEAD scale contains two factors measuring active (13 items) and passive (3 items) IL from the perspective of employees of first or second-line healthcare managers. The Active factor is further divided into four subscales: Exemplary behaviors (6), individualized consideration (2),

Table 1. Study Characteristics.

Ref	Tool	Sample size (n)	Discipline	Role	Participant selection	Setting	Country (region)	Language	Response rate	Study design
<b>Aarons et al. (2014)</b>	ILS (Staff)	n = 459	Community mental health service providers	Marriage and family therapists (47%) Social workers (26%) Psychologists (16%) Child development specialists (3%) Human relations specialists (2%) Nurses (1%) Other (e.g., drug/alcohol counselors, counselors, probation officers, psychiatrists) (4.8%)	Frontline clinician team members selected with agency administrators	93 different outpatient mental health programs	USA (Southern California)	English	80.1%	Cross-sectional survey
<b>Aarons et al. (2016)</b>	ILS (Staff)	n = 323	Community mental health service providers	Drug/Alcohol counselors (69.7%) Marriage and family therapists (11.1%) Psychologists (8%) Social workers (8%) Other (not specified) (3.2%)	Eligibility included prior experience with one or more EBPs and had been working with their supervisor for 3 months or longer	Alcohol and other drug (AOD) use treatment agencies (community) (n = not known)	USA (California and New York)	English	90.1%	Cross-sectional survey
<b>Finn et al. (2016)</b>	ILS (Staff)	n = 214	Child welfare service providers	Roles not specified	Providers (grouped into teams of ~5) who reported to the same supervisor	12 community-based organizations (CBOs) that provide child welfare services	USA (California, Illinois, Oklahoma, Washington)	English	93.0%	Cross-sectional survey
<b>Gifford et al. (2017)</b>	ILS (Staff)	n = 5	Academia (Research)	Model and scale developers from: School of Nursing (n = 1) School of Epidemiology, Public Health and Preventive Medicine (n = 1) Department of Psychology (n = 1), Department of Psychiatry (n = 1) Child and Adolescent Services	N/A	University of Ottawa (n = 2) San Diego State University (n = 1) University of California (n = 2)	Canada and USA (n/s)	English	100%	Modified template analysis to conceptually map items of ILS (staff) onto concepts of the O-MLE.
<b>Hu et al. (2019)</b>	ILS (staff) (leader)	Sample 1 & 2 (Translation) n = 5 Sample 3 & 4 (Linguistic validation) n = 10	Sample 1 & 2: Academia (research) Sample 3 & 4: Nursing	Sample 1: Nursing scholars Sample 2: School of Nursing (n = 3) Department of Psychology (n = 1) Child and Adolescent Services Research Centre (n = 1) Sample 3: Senior nurse leader (n = 1)	Sample 3 & 4: Nursing staff and leaders having more than 3 years working experience in their current positions in China	A university-affiliated hospital	China and Canada (n/s)	English to Chinese	Sample 1 & 2: 100% Sample 3: 71% Sample 4: 83%	Translation, linguistic validation with cognitive interviews

(continued)

Table 1. Continued.

Ref	Tool	Sample size (n)	Discipline	Role	Participant selection	Setting	Country (region)	Language	Response rate	Study design
<b>Hu et al. (2021)</b>	ILS (Leader) for content validation (Staff) for content validation and other psychometric properties *Content validation was evaluated for both ILS leader and staff in sample 1.	Sample 1 (Content Validation): n = 10 Sample 2 (Content validation and psychometric properties): n = 234	Sample 1: Nursing Medicine (leaders), Academia (research & education) Sample 2: Nursing	Clinical manager (n = 1) Head nurse (n = 2) Nurse educator (n = 1) Care facilitator (n = 1) Bedside nurse (n = 4) Sample 4: Senior nurse leader (n = 1) Clinical manager (n = 1) Head nurse (n = 3) Nurse educator (n = 1) Care facilitator (n = 1) Bedside nurse (n = 2) Sample 1: Hospital administrative leaders (n = 2) Physician senior leaders (n = 3) Nurse senior leaders (n = 2) Professors (n = 3) Sample 2: Nursing staff (n = 234)	Sample 1: Experts knowledgeable about healthcare leadership and EBP Sample 2: Nursing staff who worked for more than 1 year in their current positions	Sample 1: A university-affiliated hospital in China Sample 2: Inpatient units of an affiliated university teaching hospital (Surgical, Emergency, Internal medicine, Pediatrics, Gynecology & obstetrics units)	China (n/s)	Chinese	Sample 1: 71.4% Sample 2: 85.3%	Cross-sectional survey design
<b>Mandrou et al. (2020)</b>	ILS (Staff)	Sample 1 (Translation): n = 6 Sample 2 (Pilot study): n = 9 Sample 3 (Face validity): n = 3 Sample 4 (Content validity): n = 7 Sample 5 (psychometric evaluation): n = 143 Total n = 143	Sample 1: Academia Sample 2: Nursing and midwifery Sample 3: Academia Sample 4: Academia Sample 5: Nursing & midwifery	Sample 1: Translators (n = 2) Study authors (n = 3) Independent researcher (n = 1) Sample 2: Frontline nurses (n = 6) Midwives (n = 3) Sample 3: Study authors Sample 4: Study authors (n = 3) Other researchers with relevant experience (n = 4) Sample 5: Nurses (n = 58) Midwives (n = 83) Did not respond (n = 2) Registered nurses providing direct care	No description provided	Samples 1–4: Unclear Sample 5: A private, general obstetrics and gynecology hospital	Greece (n/s)	English to Greek	89.4%	Translation, pilot study and cross-sectional survey
<b>Shuman et al. (2020)</b>	ILS (Staff)	Sample 1: n = 200 Sample 2: n = 284	Nursing	Registered nurses providing direct care	Sample 1: Nurses providing direct care Sample 2:	Sample 1: "Units" from a large community	USA (Sample 1: California;	English	Sample 1: unknown Sample 2: 53.3%	Cross-sectional survey

(continued)

Table 1. Continued.

Ref	Tool	Sample size (n)	Discipline	Role	Participant selection	Setting	Country (region)	Language	Response rate	Study design
<b>Smith et al. (2020)</b>	ILS (Adapted)	n = 785	Nursing Medicine Allied health Research Leadership Support staff	Clinicians n = 544 Researcher n = 100 Support staff n = 52 Leaders n = 79	Nurses from units. Eligibility included nurses working a minimum of 0.40 FTE, designated as staff, and provided direct patient care. For units with less than 30 eligible nurses, all nurses were invited.	hospital system (Units not described) Sample 2: 24 medical-surgical units across 7 acute care community hospitals	Sample 2: Iowa, Minnesota, New Hampshire & Vermont.)	English	Total: 65% For each group: Clinicians = 63% Researcher = 58% Leaders = 92% Support staff = 64%	Cross-sectional survey
<b>Torres et al. (2018)</b>	ILS (Leader)	n = 119	Mental health leadership	Clinic supervisors	All current employees of the organization determined by active email addresses in the human resources system.	Transition of a major urban academic hospital, the Rehabilitation Institute of Chicago (RIC) to a new building: The Shirley Ryan AbilityLab (SRAIab) 31 mental health organizations	USA (Chicago & Pennsylvania)	English	87.5%	Cross-sectional survey
<b>Mosson et al., (2018)</b>	iLEAD	Time 1: n = 336 Time 2: n = 443	Healthcare professionals	Not specified	Eligible employees whose managers were taking part in the iLead intervention and who were not on leave of absence, parental leave, had quit their job, (etc) were invited to participate in the study	Regional Healthcare Organization (primary, rehabilitation, acute care, psychiatric care)	Sweden (Stockholm)	Swedish	Time 1: 41% Time 2: 90%	Cross-sectional survey at two time points (different measured administered at each time.)

Abbreviations: ILS = Implementation Leadership Scale; n/s = not specified.

Table 2. Tool Characteristics.

Tool	# Studies and ref	Conceptual model used	Target population	Year of development	Total number of items, (sub) scale(s), and number of items per subscale	Response options	Range of scores/scoring	Original language	Available translations	Feasibility
<b>ILS (Leader)</b>	3 Hu et al. (2019) Hu et al. (2021) Torres et al. (2018)	Implementation theories (n/s) Leadership (FRLM) Organizational theories (n/s)	First level leader (manager/supervisor) linked with frontline clinicians/employees	2014	12 items 4 subscales: 1. Proactive leadership (3) 2. Knowledgeable leadership (3) 3. Supportive leadership (3) 4. Perseverant leadership (3)	5-point scale: 0 (not at all) to 4 (very great extent)	Total ILS score ranges from 0 to 48, while score for each sub-scale ranges from 0 to 12 Scoring scales: Mean score for each set of items that load onto the subscale is computed. Scoring total scale: Mean of the scale scores are computed to yield the mean score.	English	Chinese	Not formally evaluated. Relatively short (< 20 items). 5 min to complete. English version freely available on the internet with scoring instructions. Interpretability of scores not established.
<b>ILS (Staff)</b>	8 Aarons et al. (2014) Aarons et al. (2016) Finn et al. (2016) Gifford et al. (2017) Hu et al. (2019) Hu et al. (2021) Mandrou et al. (2020) Shuman et al. (2020)	Implementation theories (n/s) Leadership (FRLM) (O-MILe) Organizational theories (n/s)	Frontline clinicians/staff perspective on direct manager/supervisor (e.g., frontline manager)	2014	12 items 4 subscales: 1. Proactive leadership (3) 2. Knowledgeable leadership (3) 3. Supportive leadership (3) 4. Perseverant leadership (3) Wording adjustment made in Shuman et al., 2020 (i.e., "employee replaced with "nurse")	5-point scale: 0 (not at all) to 4 (very great extent)	Total ILS score ranges from 0 to 48, while score for each sub-scale ranges from 0 to 12 Scoring scales: Mean score for each set of items that load onto the subscale is computed. Scoring total scale: Mean of the scale scores are computed to yield the mean score.	English	Greek Chinese	Not formally evaluated. Relatively short (< 20 items). 5 min to complete. English version freely available on the internet with scoring instructions. Interpretability of scores not established.

(continued)



Table 2. Continued.

Tool	# Studies and ref	Conceptual model used	Target population	Year of development	Total number of items, (sub) scale(s), and number of items per subscale	Response options	Range of scores/scoring	Original language	Available translations	Feasibility
<b>ILS (Adapted)</b>	1 Smith et al. (2020)	Same as ILS No other conceptual model used.	Clinicians (e.g., physicians, nurses, allied health therapists) Researchers Leaders Support staff	2020	6 items 4 subscales: 1. Proactive leadership (1), 2. Knowledgeable subscale, 3. Supportive leadership (3) (As per original ILS scale) 4. Perseverant leadership (1)	5-point scale: 1 (not at all) to 5 (to a very great extent)	They are not clearly described. Subscale scores were only computed when at least three items from the original subscale were used (i.e., Supportive leadership subscale only)	English	None	Not formally evaluated. Relatively short (< 20 items). Time to complete not established. Not freely available. Interpretability of scores not established.
<b>iLEAD</b>	1 Mosson et al. (2018)	FRLM	Employees	2018	16 item scale Two major factors with 5 subscales in total: 1. Active IL - Exemplary behaviors (6) - Individualized consideration (2) - Intellectual stimulation (3) - Contingent reward (2) 2. Passive IL (3)	Likert scale: 1 (strongly disagree) to 5 (strongly agree)	The mean score for each overarching factor (active or passive) or each subscale within the active factor. Scores range from 1–5.	Swedish	Swedish with English phrasing used for purposes of the publication.	Not formally evaluated. Relatively short (<20 items). Original version not freely available. Interpretability of scores not established.

Abbreviations: FRLM = Full Range Leadership Model; ILS = Implementation Leadership Scale; n/s = not specified; O-MILe = Ottawa Model of Implementation Leadership.

**Table 3.** Methodological Quality and Measurement Property Ratings.

Tool	Reference Country Language	Content Validity		Structural Validity		Internal Consistency		Cross-cultural Validity		Hypothesis Testing		Responsiveness	
		n	M	Result	n	M	Result	n	M	Result	n	M	Result
<b>ILS Leader</b>	Torres et al., 2018 USA English	119	V	CFA	119	V	Cronbach's alpha: Subscales = 0.92–0.96 Total scale = 0.95	119	V	Convergent (ICS): Total score correlation r = 0.72	119	V	Convergent (ICS): Total score correlation r = 0.72
<b>ILS Staff</b>	Aarons et al., 2014 USA English	230	V	CFA	229	V	Cronbach's alpha: Subscales = 0.95–0.96 Total scale = 0.98	459	V	Convergent (MLQ): Subscale correlations range r = 0.62–0.72 Total score correlation range r = 0.715–0.753	459	V	Convergent (MLQ): Subscale correlations range r = 0.62–0.72 Total score correlation range r = 0.715–0.753
	Aarons et al., 2016 USA English	323	V	CFA	323	V	Cronbach's alpha: Subscales = 0.93–0.97 Total scale = 0.97	323	V	Convergent (MLQ): Subscale correlation range r = 0.57–0.77 Total score correlation range r = 0.74–0.77	323	V	Convergent (MLQ): Subscale correlation range r = 0.57–0.77 Total score correlation range r = 0.74–0.77
	Finn et al., 2016 USA English	214	V	CFA	214	V	Cronbach's alpha: Subscales = 0.93–0.97 Total scale: 0.97	SI = 214 S2 = 230	V	First order model compared to configurable model: ΔCFI = -0.007; ΔRMSEA = -0.03			
	Gifford et al., 2017 Canada English	5	D	Tool and model developers evaluated the relevance of items to the O-MILE									

(continued)

Table 3. Continued.

Tool	Reference		Content Validity		Structural Validity		Internal Consistency		Cross-cultural Validity		Hypothesis Testing		Responsiveness												
	Country	Language	n	M Result	n	M Result	n	M Result	n	M Result	n	M Result	n	M Result											
<b>ILS Staff-Greek</b> Shuman et al., 2020 USA English	30	+	A Tool and model developers evaluated the comprehensiveness of the items to the O-MILE D NMs (n = 4 + unknown), educators (n = 7) and staff nurses (n = 26) reviewed relevance	S1 = V CFA 200 S2 = Second-order model fit indices: CFI = 0.978 SRMR = 0.030 + S2: Second-order model fit indices: CFI = 0.986 RMSEA = 0.051 SRMR = 0.022 +	S1 = V Cronbach's alpha: 0.93–0.97 S2 = Subcales = Total scale = 0.98 + S2: Cronbach's alpha: Subcales = 0.89–0.91 Total scale = 0.97 +	S1 = A $\Delta X^2$ Model 0–1 comparison = 200 S2 = 2.75 (reported not statistically significant) $\Delta X^2$ Model 1–2 = 1.87 (reported not statistically significant) $\Delta CFI = 0.000$ was found across the models. +	I43 V Convergent (Quality of leadership dimensions of COPSOQ II): subscale score correlation $r = 0.59$ – 0.73 Total score correlation $r = 0.79$ +	Discriminant (OCM): Subscale score correlation OCM subscales range $r = 0.05$ – $0.33$ OCM subscale score to total ILS-Staff Greek score range $r = 0.09$ – 0.36 +	n M Result	n M Result	n M Result	n M Result	n M Result	n M Result											
															Mandrou et al., 2020 Greece Greek	7	D Study authors (n = 3) and independent researchers with “relevant research experience” (n = 4) evaluated relevance. Authors reported that I-CVI scores ranged from 0.80–1.00 for each of the 12 items. +	I43 V CFA High order model fit indices: CFI = 0.930 RMSEA = 0.060 +	I43 V Cronbach's alpha: Subcales = 0.85–0.91 Total scale = 0.94 +	I43 V Convergent (Quality of leadership dimensions of COPSOQ II): subscale score correlation $r = 0.59$ – 0.73 Total score correlation $r = 0.79$ +	Discriminant (OCM): Subscale score correlation OCM subscales range $r = 0.05$ – $0.33$ OCM subscale score to total ILS-Staff Greek score range $r = 0.09$ – 0.36 +	n M Result	n M Result	n M Result	n M Result
Hu et al., 2021 China Chinese	S1 = 10	D Experts (n = 10) evaluated relevance (for 10 of 12 items, I-CVI = 1.00, $k^* \leq 0.90$ ); for 2 of 12 items (I-CVI $\leq 0.90$ , $k^* \leq 0.90$ ) +	V CFA First and second-order fit indices: CFI = 0.98 TLI = 0.97 +	V Cronbach's alpha: Subcales = 0.86–0.95 Total scale = 0.93 +	V Convergent (MLQ): Subscale correlation range $r = 0.40$ – $0.63$ Total scale $r = 0.50$ – $0.63$ +	V Convergent (MLQ): Subscale correlation range $r = 0.40$ – $0.63$ Total scale $r = 0.50$ – $0.63$ +	n M Result	n M Result	n M Result	n M Result	n M Result														

(continued)

Table 3. Continued.

Tool	Reference Country Language	Content Validity		Structural Validity		Internal Consistency		Cross-cultural Validity		Hypothesis Testing		Responsiveness	
		n	M Result	n	M Result	n	M Result	n	M Result	n	M Result	n	M Result
		S2 = 234	D Staff nurses (n = 234) rated relevance where >88.9% found items were relevant	336	V CFA SRMR = 0.034	324	V Cronbach alpha by factor and subscales for active leadership: Active leadership = 0.95	336	V Convergent (Global transformational leadership scale): Active leadership subscale score correlation range r = 0.70–0.78	336	V Hypothesis met for active leadership subscale (ICS as gold standard)	336	V Hypothesis met for active leadership subscale (ICS as gold standard)
iLEAD	Mosson et al., 2018 Sweden Swedish	II	D Employees and senior organizational leaders (exact n of each unknown) evaluated relevance and comprehensibility of the items in the scale.	336	V CFA Fit model indices (for model 1): RMSEA = 0.059	336	V Subscale = 0.80–0.92	336	V Discriminant: (Global transformational leadership scale): Passive leadership subscale score r = -0.22	336	V Discriminant: (Transactional subscale MLQ): Passive leadership subscale score r = -0.18	336	V Hypothesis met for active leadership subscale (with ICS)

Abbreviations: + = Criteria met; - = Criteria not met; A = Adequate; CFA = Confirmatory Factor Analysis; CFI = Comparative Fit Index; COPSOQ II = Copenhagen Psychosocial Questionnaire version 2; EBPAAS-15 = Evidence-Based Practice Attitudes Scale; I = Inadequate; I-CVI = Item-Content Validity Index; ICS = Implementation Climate Scale; k\* = Modified Cohen's Coefficient Kappa; M = Methodological quality rating; MLQ = Multifactorial Leadership Questionnaire; n = Sample Size; OCM = Organizational Climate Measure; O-MILe = Ottawa Model of Implementation Leadership; r = Pearson Correlation Coefficient; RMSEA = Root Mean Square Error of Approximation; S1 = Sample 1; S2 = Sample 2; SRMR = Standardized Root Mean Square Residual; TLI = Tucker-Lewis Index; V = Very good; X<sup>2</sup> = Chi-Squared.

intellectual stimulation (3), and contingent reward (2). Items describe the quality of leader behaviors in terms of positive and negative behaviors for IL (e.g., “My closest manager has behaved in a way that explicitly displays a commitment to working according to the new working method”). Responses are measured on a Likert-type scale (1-strongly disagree to 5-strongly agree), where a mean score is calculated for each overarching factor (“active” or “passive”) or for the “active” subscale. The iLEAD was designed to assess a generic leadership training intervention, validated in Swedish and translated into English for publication.

### **Methodological Quality and Measurement Property Ratings**

Methodological quality and measurement property ratings for content validity and psychometric properties are presented in Table 3 and summarized below. Translation and linguistic validation for the ILS is also evaluated.

**Content Validity. ILS.** The methodological quality rating for determining item relevance and comprehensiveness was doubtful as there were sparse process details reported and missing information concerning the number, type, and background of the stakeholders involved. Hu et al. (2021) and Mandrou et al. (2020) calculated I-CVIs ranging from 0.80–1.00, demonstrating high agreement on the relevance of the items. Exceptionally, an adequate rating was found evaluating the comprehensiveness of the ILS Staff from the perspective of experts on the topic, as reported in the study by Gifford et al. (2017). A committee mapped 100% of the items onto one or more components of the Ottawa Model of Implementation Leadership (O-MILe), a model developed independently of the ILS describing effective leadership behaviors for implementation (Gifford et al., 2017). Additionally, 77% of the O-MILe concepts corresponded to at least one ILS item (directly or indirectly), indicating a strong level of congruency. Comprehensibility was not assessed for any version of the ILS tool.

**iLEAD.** As with the ILS, there were sparse details of the process reported for content validation, and the resultant findings from the content validation process were lacking. Content validation was rated as doubtful for relevance and comprehensibility. The iLEAD items were not assessed for comprehensiveness.

**Internal Structure. ILS.** The methodological quality of structural validity assessments was rated as very good for all versions of the ILS tool. Based on standards described by LeBreton and Senter (2008), a lack of agreement was found using the intraclass correlation statistic (ICC(1)) ranging from 0.19 to 0.29 (total ILS score) and 0.08 to 0.30 (subscale and item scores) across the studies. However, the authors reported moderate to strong agreement

using the  $A_{wg}$  statistic (range of 0.70–0.80 for total ILS score; and 0.67–0.80 for subscale and items scores), suggesting that team/unit level data could be aggregated to measure the higher-level construct (Aarons et al., 2014, 2016; Finn et al., 2016; Hu et al., 2021; Shuman et al., 2020).

To determine the items’ latent structure and underlying relationships, the ILS Staff was subjected to exploratory factor analysis in the initial ILS study (Aarons et al., 2014). This resulted in a 4-factor, 12-item scale with an equal number of items per factor. The four subscales were also significantly correlated (range = 0.73–0.80), which supported a higher-order factor for IL (Aarons et al., 2014). This is an appropriate test for newly developed tools and the unavailability of an underlying model at the time of its development.

Subsequently, a series of confirmatory factor analyses (CFA) were performed based on the initial factor structure to determine a higher-order factor model where each subscale was indicative of an overall IL score. An exception was the CFA conducted using the ILS (Adapted), which included only a single-factor structure. Sample sizes met the minimum number ( $n = 119–785$ ), and an appropriate selection of goodness of fit indicators was used in each of these studies. In each study, when completed, at least one indicator met criteria suggesting support for the four-factor IL model (CFI = 0.930–1.000; RMSEA = 0.030–0.060; TLI = 0.964–0.970; SRMR = 0.022–0.050). For the ILS (Adapted), CFI and RMSEA fit indices met the single-factor IL model criteria.

For internal consistency, methodological quality was rated as very good for all versions of the ILS, except for the ILS (Adapted), which was rated as adequate. Cronbach’s alpha for each of the four subscales for the ILS versions was greater than 0.85 (range 0.85–0.97), which is acceptable. The results for the total ILS score were high, ranging from 0.94 to 0.98, and acceptable for the ILS (Adapted) with a score of 0.89.

**iLEAD.** The quality of the structural validation process was rated as very good. Several goodness of fit indicators were assessed, with the RMSEA indicator just meeting the criteria for good measurement property (0.059) for a 16-item model proposed for four first-order factors (with the active leadership subscale) and two second-order factors (active and passive leadership). The full range leadership model structure for the iLEAD was confirmed with factor analysis and evaluated in an adequate sample size (20 items,  $n = 336$ ) of healthcare employee respondents though their professional discipline was not reported.

As with the ILS, the methodological quality of internal consistency evaluations of the iLEAD tool was deemed very good, as the subscale and factors within the active leadership subscale Cronbach’s alpha were calculated. Acceptable results were found using Cronbach’s alpha per factor within the active leadership subscale (0.80–0.92), the total active leadership subscale (0.95), and the passive leadership subscale (0.91).

**Other Properties Related to Quality. ILS.** The methodological quality ratings varied for cross-cultural validity as studies provided limited details of how sample groups were similar and only described the variable distinguishing each group (Terwee et al., 2018). Results of cross-cultural validity was variable, with minor changes in model fit indices (meeting weak invariance by study authors) found between mental health and child welfare samples but not among nursing samples in evaluating the ILS Staff. The cross-cultural validity of ILS (Adapted) was acceptable.

The methodological quality for convergent and discriminant validation assessments was rated very good across the applicable studies. Convergent validity was most frequently determined using the transformational and transactional subscales of the Multifactor Leadership Questionnaire (MLQ), a well-recognized, validated, and reliable tool to assess organizational leadership (Bass & Avolio, 1995). Transformational and transactional leadership subscales are expected to correlate positively with IL as they are theoretically similar and form the basis of IL (Castiglione, 2020). Pearson product correlations supported this hypothesis between subscale ILS and MLQ subscale scores ( $r=0.40-0.77$ ) and between ILS total and MLQ subscales ( $0.50-0.77$ ).

Discriminant validation was conducted for the ILS-Staff and the ILS-Staff Greek scales for using the Organizational Climate Measure (OCM; Patterson et al., 2011). Based on the competing values framework, the OCM items reflect both positive and negative perceptions of organizational climate, which is distinct from measuring leadership behaviors and distantly related to the climate favoring the implementation of EBPs (Aarons et al., 2016). Though criteria for low and moderate correlation were not reported in the studies, the resulting correlations supported the hypothesis for the ILS-Staff and the ILS-Staff Greek subscale to subscale correlations ( $r=0.05-0.33$ ) and total ILS score to OCM subscales ( $0.09-0.36$ ). Finally, a low to moderate correlation hypothesis was confirmed ( $r=0.24$ ) for the ILS-Leader with the Evidence-Based Practice Attitude Scale (EBPAS; Aarons, 2004), which measures the attitudes of healthcare providers in adopting EBPs in mental health clinic supervisors.

Translation and validation of the ILS-Leader and Staff versions into Chinese (Hu et al., 2019) and Greek (Mandrou et al., 2020) both closely followed the process for translation and linguistic validation outlined by Sousa & Rojjanasrirat (2011). This included forward and back translations by independent translators, cognitive interviews with stakeholders, and integration of feedback to the final translated tool prior to further psychometric testing.

**iLEAD.** Methodological quality for convergent and discriminant validity hypothesis testing was deemed very good, as with the ILS versions. Two scales were used for both convergent and discriminant validation: The Global Transformational Leadership Scale (Carless et al., 2000),

which is a brief measure of transformational leadership behaviors, and the Transactional subscale of the MLQ (reported above). Both tools were hypothesized to be highly correlated with the active leadership subscale and negatively correlated with the passive leadership subscale of the iLEAD. Both hypotheses for convergent validity ( $r=0.61-0.78$ ) and discriminant validity ( $r=0.18$ ,  $r=0.22$ ) were supported.

Methodological quality for tool responsiveness was rated as very good. Multilevel regressions supported the authors' hypothesis that active IL significantly predicted implementation climate ( $B=0.4$ ,  $SE=0.08$ ) when age, gender and educational background were controlled, whereas passive IL did not, although this result was not significant ( $B=-0.07$ ,  $SE=0.05$ ).

## Discussion

In this systematic review of outcome measures, 11 studies that evaluated six versions of two IL tools, the ILS and iLEAD were summarized, critically appraised and their psychometric properties were compared. Overall, the studies poorly reported item development and content validation according to the COSMIN guidelines, which can affect the internal validity of the tools despite adequate and consistent performance in further validation studies. Neither measure was psychometrically evaluated for all eight properties stipulated in COSMIN and feasibility, thus, they require additional testing. However, both tools were brief and could be feasible in practice. Additionally, the ILS tool is freely available and has been translated and linguistically validated in Greek and Chinese, increasing accessibility to diverse populations. Of both tools, the ILS is currently the most promising for use in nursing since versions of the tool were validated in hospital-based nursing contexts. ILS items were mapped against concepts of the O-MILE, which was partly informed through empirical studies in nursing. Additionally, the ILS staff was subjected to structural, cross-cultural, and construct validation within acute care nursing samples of nurse managers and nursing staff. Finally, the ILS was translated and linguistically validated in Chinese and Greek hospital-based nursing samples.

From this assessment, both tools lacked a solid theoretical base from which to develop or adapt items. The iLEAD was established on a general leadership model (FRLM) rather than IL, considered domain-specific leadership. A recent concept analysis found that strong available leadership (such as transformation and transactional leadership) is antecedents of IL, but IL remains a distinct concept (Castiglione, 2020). The ILS was similarly void of a theoretical foundation upon initial development, but its validity was strengthened when mapped with the O-MILE. However, while all scale items mapped onto the theoretical concepts, several concepts were not represented by a corresponding item. Thus, Gifford et al. (2017) contend that greater specificity is needed to capture IL behaviors in the ILS tool.

Remarkably, there was a general lack of reporting concerning item development procedures. Best practice emphasizes strong deductive or inductive approaches to item generation (Boateng et al., 2018). However, neither tool presented formal qualitative or quantitative studies. Though the iLEAD authors sought input from apparent experts, the process was inadequately described and consequently deemed methodologically doubtful. Similarly, content validation lacked important detail to judge the rigor of methods in both tools. Considered the most critical phase in tool development (Almanasreh et al., 2019), this validity refers to how relevance and representative the items reflect concept of interest for content experts (Boateng et al., 2018). Considerations of who the experts are and what constitutes their evaluation are therefore important to report. Background of the experts or a formal evaluation process (quantitative or qualitative) to determine relevance, comprehensiveness or comprehensibility was not described for either tool. Content validity is often underreported compared to reliability and other validation processes (Connell et al., 2018) but is critical for complex constructs such as leadership. These concerns over development and content validity place some doubt in the remaining psychometric properties evaluated despite the strong performance of both tools. Future studies should re-visit whether items for these IL tools are relevant and representative from the perspective of appropriate and bonafide experts, especially with nursing leaders at the POC. Additionally, future research on these tools can include formal quantitative content validation methods to compare across tools.

The findings of this review contribute to several implications when considering IL tools for use within a nursing context. In a recent systematic review of leadership scales by Carlson et al. (2021), the authors endorse both the ILS and iLEAD for use within nursing despite gaps in validation. As stated above, the authors of the present paper agree that the ILS is adequate for nursing. However, the ILS would be strengthened with further validation within a context more reflective of the nature of nursing leadership at the POC in hospital settings. The ILS was assessed with hospital based POC nurse managers in espousing behaviors for IL. However, the traditional hierarchical conception of leaders as managers is incompatible with nursing (Kok et al., 2023). Nursing leadership has evolved into various formalized leadership roles, especially at the unit level. In addition to the nurse manager, the assistant nurse manager, nurse educator, and clinical nurse specialists have unique and sometimes overlapping responsibilities to support and ensure the strong performance of nursing staff (Mohr & Coke, 2018). This notion holds in EBP implementation.

Though nurse managers have direct control over managing resources, each leader influences EBP implementation (Birken et al., 2018; Kitson et al. 2021). Similarly, change champions are often nursing staff who take on informal leadership roles to influence EBP change (Ferren et al., 2022). Moreover, nursing leadership in acute care units is typically

organized as “nursing leadership teams” (Kalisch & Lee, 2012). This implies a shared leadership approach, where leadership results from group members’ interaction (Pearce & Conger, 2003). Several nursing studies have alluded to the importance of collective nursing leadership to achieve successful EBP implementation (Fleischer et al., 2016; Gifford et al., 2019). The current tools do not consider this relational aspect of IL in nursing. Thus, future studies are needed to validate IL with the POC nursing leaders other than managers and as a collective action of diverse and multiple leaders.

### Limitations

The ILS and iLEAD tools were selected based on an unstructured scan of healthcare literature and the first author’s (SAC) familiarity with the subject. An expanded search to include organizational and management databases may have generated other tools. Some challenges were encountered applying the COSMIN risk of bias tool, specifically the oftentimes vague and subjective nature of some of the evaluation criteria. Engaging in regular team discussions and documenting interpretations of each standard so that it could be equally applied across studies were used to compensate for this limitation. Finally, COSMIN instructs reviewers to include studies that mainly aim to assess psychometric properties, which may limit our findings by excluding studies that evaluated psychometric properties as a secondary aim. Finally, this systematic review on outcome measures was not prospectively registered, potentially introducing bias in the conduct of the review. However, the authors of this paper promoted transparency in reporting the review by adhering to a standard reporting checklist for systematic review.

### Implications for Practice

The findings of this review inform the application of validated and contextually relevant tools to evaluate the capacity of nursing leadership to engage in IL in a real-world context. In this regard, the ILS and iLEAD are promising tools to assess IL in nursing but require further validation for contexts with diverse and multiple leaders at the POC. When selecting measurement tools for use in practice, the feasibility and practicality of their use needs to be considered.

### Conclusions

The developing concept of IL holds promise for understanding the context surrounding the effectiveness of EBP implementation processes. Since nurses constitute the largest percentage of healthcare workers globally (WHO, 2022), assessing and strengthening IL in nursing may substantially impact the quality of overall care delivery. This review is the first to examine the strength of methodological quality and psychometric properties of two IL tools for use in a nursing

context. The results of this review provide a current understanding of tools to gauge IL in POC nurse leaders, to elucidate the robustness and appropriateness of these tools for nurse leaders at the frontlines, and to inform future validation studies for use in determining the effectiveness of nursing leadership development interventions.

### Acknowledgements

The authors wish to acknowledge the support of Francesca Frati, Liaison Librarian at McGill University, for reviewing the search strategy, Alexander Torres-Vasquez, Statistician with the Research Institute of the McGill University Health Centre, for support with evaluation, and Dr. Celine Gelin, Professor at the Ingram School of Nursing, McGill University, for consultation regarding the interpretation of the results.

### Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Fonds de recherche du Québec – Santé (FRQ-S), the Quebec Network on Nursing Intervention Research (RRISIQ), the Newton Foundation, the Canadian Nurses Foundation, The McGill Nursing Collaborative and McGill University.

### ORCID iD

Sonia Angela Castiglione  <https://orcid.org/0000-0001-5487-9119>

### Supplemental Material

Supplemental material for this article is available online.

### References

- Aarons, G. A. (2004). Mental health provider attitudes toward adoption of evidence-based practice: The Evidence-Based Practice Attitude Scale (EBPAS). *Mental Health Services Research, 6*(2), 61–74. <https://doi.org/10.1023/b:mhsr.0000024351.12294.65>
- Aarons, G. A., Ehrhart, M. G., & Faraone, L. R. (2014). The Implementation Leadership Scale (ILS): Development of a brief measure of unit level implementation leadership. *Implementation Science, 9*(1), 45. <https://doi.org/10.1186/1748-5908-9-45>
- Aarons, G. A., Ehrhart, M. G., Torres, E. M., Finn, N. K., & Roesch, S. C. (2016). Validation of the Implementation Leadership Scale (ILS) in substance use disorder treatment organizations. *Journal of Substance Abuse Treatment, 68*, 31–35. <https://doi.org/10.1016/j.jsat.2016.05.004>
- Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy, 15*(2), 214–221. <https://doi.org/10.1016/j.sapharm.2018.03.066>
- Bass, B. M., & Avolio, B. J. (1995). Multifactor Leadership Questionnaire Psychological Assessment. *APA PsycTests*. <https://doi.org/10.1037/t03624-000>
- Bianchi, M., Bagnasco, A., Bressan, V., Barisone, M., Timmins, F., Rossi, S., Pellegrini, R., Aleo, G., & Sasso, L. (2018). A review of the role of nurse leadership in promoting and sustaining evidence-based practice. *Journal of Nursing Management, 26*(8), 918–932. <https://doi.org/10.1111/jonm.12638>
- Birken, S., Clary, A., Tabriz, A. A., Turner, K., Meza, R., Zizzi, A., Larson, M., Walker, J., & Charns, M. (2018). Middle managers' role in implementing evidence-based practices in healthcare: A systematic review. *Implementation Science, 13*(1), 149. <https://doi.org/10.1186/s13012-018-0843-5>
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*, 149. <https://doi.org/10.3389/fpubh.2018.00149>
- Carless, S. A., Wearing, A. J., & Mann, L. A. (2000). Short measure of transformational leadership. *Journal of Business and Psychology, 14*, 389–405. <https://doi.org/10.1023/A:1022991115523>
- Carlson, M. A., Morris, S., Day, F., Dadich, A., Ryan, A., Fradgley, E. A., & Paul, C. (2021). Psychometric properties of leadership scales for health professionals: A systematic review. *Implementation Science, 16*(1), 85. <https://doi.org/10.1186/s13012-021-01141-z>
- Castiglione, S. A. (2020). Implementation leadership: A concept analysis. *Journal of Nursing Management, 28*(1), 94–101. <https://doi.org/10.1111/jonm.12899>
- Connell, J., Carlton, J., Grundy, A., Taylor Buck, E., Keetharuth, A. D., Ricketts, T., Barkham, M., Robotham, D., Rose, D., & Brazier, J. (2018). The importance of content and face validity in instrument development: Lessons learnt from service users when developing the Recovering Quality of Life measure (ReQoL). *Quality of Life Research, 27*(7), 1893–1902. <https://doi.org/10.1007/s11136-018-1847-y>
- Connor, L., Dean, J., McNett, M., Tydings, D. M., ShROUT, A., Gorsuch, P. F., Hole, A., Moore, L., Brown, R., Melnyk, B. M., & Gallagher-Ford, L. (2023). Evidence-based practice improves patient outcomes and healthcare system return on investment: Findings from a scoping review. *Worldviews on Evidence Based Nursing, 20*(1), 6–15. <https://doi.org/10.1111/wvn.12621>
- Dryden-Palmer, K. D., Parshuram, C. S., & Berta, W. B. (2020). Context, complexity and process in the implementation of evidence-based innovation: A realist informed review. *BMC Health Services Research, 20*(1), 81. <https://doi.org/10.1186/s12913-020-4935-y>
- Ferren, M., Von Ah, D., & Menachemi, N. (2022). EBP champion responsibilities and sustainability: A scoping review. *Nursing Management, 53*(8), 22–33. <https://doi.org/10.1097/01.NUMA.0000853152.64293.46>
- Finn, N. K., Torres, E. M., Ehrhart, M. G., Roesch, S. C., & Aarons, G. A. (2016). Cross-validation of the Implementation Leadership Scale (ILS) in child welfare service organizations. *Child Maltreatment, 21*(3), 250–255. <https://doi.org/10.1177/1077559516638768>
- Fleischer, A. R., Semenic, S. E., Ritchie, J. A., Richer, M. C., & Denis, J. L. (2016). A unit-level perspective on the long-term sustainability of a nursing best practice guidelines program: An embedded multiple case study. *International Journal of Nursing Studies, 53*, 204–218. <https://doi.org/10.1016/j.ijnurstu.2015.09.004>
- Gifford, W., Graham, I. D., Ehrhart, M. G., Davies, B. L., & Aarons, G. A. (2017). Ottawa Model of Implementation Leadership and Implementation Leadership Scale: Mapping concepts for



- developing and evaluating theory-based leadership interventions. *Journal of Healthcare Leadership*, 9, 15–23. <https://doi.org/10.2147/jhl.S125558>
- Gifford, W., Lewis, K. B., Eldh, A. C., Fiset, V., Abdul-Fatah, T., Aberg, A. C., Thavorn, K., Graham, I. D., & Wallin, L. (2019). Feasibility and usefulness of a leadership intervention to implement evidence-based falls prevention practices in residential care in Canada. *Pilot and Feasibility Studies*, 5(1), 103. <https://doi.org/10.1186/s40814-019-0485-7>
- Gifford, W. A., Davies, B. L., Graham, I. D., Tourangeau, A., Woodend, A. K., & Lefebvre, N. (2013). Developing leadership capacity for guideline use: A pilot cluster randomized control trial. *Worldviews on Evidence Based Nursing*, 10(1), 51–65. <https://doi.org/10.1111/j.1741-6787.2012.00254.x>
- Harvey, G., Kelly, J., Kitson, A., Thornton, K., & Owen, V. (2020). Leadership for evidence-based practice—Enforcing or enabling implementation? *Collegian Journal of the Royal College of Nursing Australia*, 27(1), 57–62. <https://doi.org/10.1016/j.colegn.2019.04.004>
- Hu, J., Gifford, W., Ruan, H., Harrison, D., Li, Q., Ehrhart, M. G., & Aarons, G. A. (2019). Translation and linguistic validation of the implementation leadership scale in Chinese nursing context. *Journal of Nursing Management*, 27(5), 1030–1038. <https://doi.org/10.1111/jonm.12768>
- Hu, J., Gifford, W., Ruan, H., Harrison, D., Li, Q., Ehrhart, M. G., Harrison, M. A., Barrowman, N., & Aarons, G. A. (2021). Validating the Implementation Leadership Scale in Chinese nursing context: A cross-sectional study. *Nursing Open*, 8(6), 3420–3429. <https://doi.org/10.1002/nop2.888>
- Kalisch, B. J., & Lee, K. H. (2012). Congruence of perceptions among nursing leaders and staff regarding missed nursing care and teamwork. *Journal of Nursing Administration*, 42(10), 473–477. <https://doi.org/10.1097/NNA.0b013e31826a1fa4>
- Kelloway, E. K., Mullen, J., & Francis, L. (2006). Divergent effects of transformational and passive leadership on employee safety. *Journal of Occupational Health Psychology*, 11(1), 76–86. <https://doi.org/10.1037/1076-8998.11.1.76>
- Kitson, A. L., Harvey, G., Gifford, W., Hunter, S. C., Kelly, J., Cummings, G. G., Ehrenberg, A., Kislov, R., Pettersson, L., Wallin, L., & Wilson, P. (2021). How nursing leaders promote evidence-based practice implementation at point-of-care: A four-country exploratory study. *Journal of Advanced Nursing*, 77(5), 2447–2457. <https://doi.org/10.1111/jan.14773>
- Kok, E., Weggelaar, A. M., Reede, C., Schoonhoven, L., & Lalleman, P. (2023). Beyond transformational leadership in nursing: A qualitative study on rebel nurse leadership-as-practice. *Nursing Inquiry*, 30(2), e12525. <https://doi.org/10.1111/nin.12525>
- Kvist, T., Tähkä, K., Ruotsalainen, M., & Tervo-Heikkinen, T. (2014). The impact of nursing leadership training on evidence-based leadership and practice. *Journal of Nursing & Care*, 03, <https://doi.org/10.4172/2167-1168.1000181>
- LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, 11(4), 815–852. <https://doi.org/10.1177/1094428106296642>
- Li, S.-A., Jeffs, L., Barwick, M., & Stevens, B. (2018). Organizational contextual features that influence the implementation of evidence-based practices across healthcare settings: A systematic integrative review. *Systematic Reviews*, 7(1), 72. <https://doi.org/10.1186/s13643-018-0734-5>
- Mandrou, E., Tsounis, A., & Sarafis, P. (2020). Validity and reliability of the Greek version of Implementation Leadership Scale (ILS). *BMC Psychology*, 8(1), 49. <https://doi.org/10.1186/s40359-020-00413-5>
- Meza, R. D., Triplett, N. S., Woodard, G. S., Martin, P., Khairuzzaman, A. N., Jamora, G., & Dorsey, S. (2021). The relationship between first-level leadership and inner-context and implementation outcomes in behavioral health: A scoping review. *Implementation Science*, 16(69). <https://doi.org/10.1186/s13012-021-01104-4>
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- Mohr, L. D., & Coke, L. A. (2018). Distinguishing the clinical nurse specialist from other graduate nursing roles. *Clinical Nurse Specialist*, 32(3), 139–151. <https://doi.org/10.1097/NUR.0000000000000373>
- Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, 27(5), 1171–1179. <https://doi.org/10.1007/s11136-017-1765-4>
- Mosson, R., von Thiele Schwarz, U., Hasson, H., Lundmark, R., & Richter, A. (2018). How do iLead? Validation of a scale measuring active and passive implementation leadership in Swedish healthcare. *BMJ Open*, 8(6), e021992. <https://doi.org/10.1136/bmjopen-2018-021992>
- Nilsen, P., & Bernhardsson, S. (2019). Context matters in implementation science: A scoping review of determinant frameworks that describe contextual determinants for implementation outcomes. *BMC Health Services Research*, 19(1), 189. <https://doi.org/10.1186/s12913-019-4015-3>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & McGuinness, L. A., ... D. Moher (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Patterson, M., West, M., Shackleton, V., Dawson, J. F., Lawthom, R., Maitlis, S., Robinson, D., & Wallace, A. (2011). Development and validation of an organizational climate measure. *Journal of Organizational Behavior*, 26, <https://doi.org/10.1037/t62894-000>
- Pearce, C. L., & Conger, J. A. (Eds). (2003). *Shared leadership: Reframing the hows and whys of leadership*. Sage. <https://doi.org/10.4135/9781452229539>
- Pol-Castañeda, S., Rodríguez-Calero, M. Á., Zaforteza-Lallemand, C., Villafila-Gomila, C. J., Blanco-Mavillard, I., Ferrer-Cruz, F., & De Pedro-Gómez, J. (2020). Moving evidence into practice by advanced practice nurses in hospitalization wards. Protocol for a multicentre quasi-experimental study in three hospitals in Spain. *International Journal of Environmental Research and Public Health*, 17(10), 3473. <https://doi.org/10.3390/ijerph17103473>
- Powell, B. J., Mettert, K. D., Dorsey, C. N., Weiner, B. J., Stanick, C. F., Lengnick-Hall, R., Ehrhart, M. G., Aarons, G. A., Barwick, M. A., Damschroder, L. J., & Lewis, C. C. (2021). Measures of organizational culture, organizational climate, and implementation climate in

- behavioral health: A systematic review. *Implementation Research and Practice*, 2, <https://doi.org/10.1177/26334895211018862>
- Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*, 27(5), 1147–1157. <https://doi.org/10.1007/s11136-018-1798-3>
- Randall, R., Nielsen, K., & Tvedt, S. D. (2009). The development of five scales to measure employees' appraisals of organizational-level stress management interventions. *Work & Stress*, 23(1), 1–23. <https://doi.org/10.1080/02678370902815277>
- Shuman, C. J., Ehrhart, M. G., Torres, E. M., Veliz, P., Kath, L. M., VanAntwerp, K., Banaszak-Holl, J., Titler, M. G., & Aarons, G. A. (2020). EBP implementation leadership of frontline nurse managers: Validation of the implementation leadership scale in acute care. *Worldviews on Evidence-Based Nursing*, 17(1), 82–91. <https://doi.org/10.1111/wvn.12402>
- Shuman, C. J., Ehrhart, M. G., Veliz, P. T., & Titler, M. G. (2023). Perceptual differences in nursing implementation leadership and climate: A cross-sectional study. *Implementation Science Communications*, 4(1), 9. <https://doi.org/10.1186/s43058-023-00392-9>
- Smart, A. (2006). A multi-dimensional model of clinical utility. *International Journal for Quality in Health Care*, 18(5), 377–382. <https://doi.org/10.1093/intqhc/mzl034>
- Smith, J. D., Rafferty, M. R., Heinemann, A. W., Meachum, M. K., Villamar, J., Lieber, R. L., & Hendricks Brown, C. (2020). Pragmatic adaptation of implementation research measures for a novel context and multiple professional roles: A factor analysis study. *BMC Health Services Research*, 20, 257. <https://doi.org/10.1186/s12913-020-05118-4>
- Sousa, V. D., & Rojjanasrirat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: A clear and user-friendly guideline. *Journal of Evaluation in Clinical Practice*, 17(2), 268–274. <https://doi.org/10.1111/j.1365-2753.2010.01434.x>
- Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C. W., Mokkink, L. B., & Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: A Delphi study. *Quality of Life Research*, 27(5), 1159–1170. <https://doi.org/10.1007/s11136-018-1829-0>
- Tistad, M., Palmcrantz, S., Wallin, L., Ehrenberg, A., Olsson, C. B., Tomson, G., Holmqvist, L. W., Gifford, W., & Eldh, A. C. (2016). Developing leadership in managers to facilitate the implementation of national guideline recommendations: A process evaluation of feasibility and usefulness. *International Journal of Health Policy and Management*, 5(8), 477–486. <https://doi.org/10.15171/ijhpm.2016.35>
- Torres, E. M., Ehrhart, M. G., Beidas, R. S., Farahnak, L. R., Finn, N. K., & Aarons, G. A. (2018). Validation of the implementation leadership scale (ILS) with supervisors' self-ratings. *Community Mental Health Journal*, 54(1), 49–53. <https://doi.org/10.1007/s10597-017-0114-y>
- Valimaki, T., Partanen, P., & Haggman-Laitila, A. (2018). An integrative review of interventions for enhancing leadership in the implementation of evidence-based nursing. *Worldviews on Evidence-Based Nursing*, 15(6), 424–431. <https://doi.org/10.1111/wvn.12331>
- Wensing, M., & Grol, R. (2019). Knowledge translation in health: How implementation science could contribute more. *BMC Medicine*, 17(1), 88. <https://doi.org/10.1186/s12916-019-1322-9>
- WHO. (2020). *Process of translation and adaptation of instruments*. [https://www.who.int/substance\\_abuse/research\\_tools/translation/en/](https://www.who.int/substance_abuse/research_tools/translation/en/)
- WHO. (2022). *Nursing and midwifery. Key Facts*. <https://www.who.int/news-room/fact-sheets/detail/nursing-and-midwifery>
- Williams, N. J., Wolk, C. B., Becker-Haimes, E. M., & Beidas, R. S. (2020). Testing a theory of strategic implementation leadership, implementation climate, and clinicians' use of evidence-based practice: A 5-year panel analysis. *Implementation Science*, 15(1), 10. <https://doi.org/10.1186/s13012-020-0970-7>