

Measurement properties of scales assessing new graduate nurses' clinical competence: A systematic review of psychometric properties

Note: This is the peer-reviewed accepted version of the following article:

Charette, M., McKenna, L. G., Maheu-Cadotte, M.-A., Deschênes, M.-F., Ha, L., & Merisier, S. (2020). Measurement properties of scales assessing new graduate nurses' clinical competence: A systematic review of psychometric properties. *International Journal of Nursing Studies*, 110, 1-13. doi: 10.1016/j.ijnurstu.2020.103734

ABSTRACT

Background: New graduate nurses' (NGNs) competence is a concern for all healthcare organizations. Previous reports show heterogeneous levels of competency among them. As a positive association between competency and quality of care in clinical settings has been suggested, it is essential for researchers and clinicians to select valid, reliable, and responsive scales to assess NGNs' competence. However, a systematic evaluation of the measurement properties of scales measuring NGNs' competence had yet to be published.

Objective: To analyse, evaluate and synthesize the measurement properties of scales used to assess NGNs' clinical competence.

Design: A systematic psychometric review based on the Consensus-based Standards for the selection of health Measurement INstruments (COSMIN) methods.

Data sources: The search strategy included a combination of keywords and thesaurus terms related to new graduate nurses, clinical competence, and competence assessment. Five databases were searched: Embase, CINAHL, MEDLINE, PsycINFO and Web of Science. The search was limited to full-text papers published in English or French, from 2010 to 2019.

Review methods: Two independent reviewers screened eligible papers, extracted data related to validity, reliability, and responsiveness of each scale, and evaluated the quality of their measurement properties as well as risk of bias in their psychometric evaluation. Divergences were solved through discussion.

Results: Ten scales were included: eight original scales, one culturally adapted and one modified. Of these scales, 5 have been used by researchers from 2010 to assess competence of NGNs and 5 were newly developed. Most scales are divided into 6 to 8 subscales and use an adjectival scale with either 4, 5 or 7 points. The content validity study of all scales in this review was deemed to be doubtful or inadequate quality. Reliability was almost exclusively assessed by calculating the

internal consistency with Cronbach's alpha coefficient which gives no information on equivalence or stability of the measure. Responsiveness was never properly assessed in the reviewed studies.

Conclusions: There is little evidence on the measurement properties for each scale regarding their validity, reliability, responsiveness. Therefore, it is impossible to recommend one scale over another.

Prospero registration number: CRD42018109711

TWEETABLE ABSTRACT

Systematic review of scales measuring new nurses' competence: we must do better and conduct more validity/reliability testing of existing scales.

CONTRIBUTION OF THE PAPER

What is already known about this topic?

- Competence is often defined in nursing as a complex and evolutive combination of knowledge, skills, attitudes and values, bound to their context of practice.
- Different scales have been developed to measure nurses' competence, but no systematic review has evaluated and synthesized their measurement properties.

What the paper adds?

- This review provides a summary of evidence of measurement properties of ten scales.
- After their development, few scales were further tested with a different group or in a different context, limiting the evidence of their validity, reliability and responsiveness.

KEY WORDS

clinical competence, competence assessment, new graduate nurse, nursing, psychometric properties, scale, systematic review

1. INTRODUCTION

A positive association between nurses' competence and the improvement of quality of care in clinical settings has been suggested (Church, 2016). The latter can be measured by nursing-sensitive patient outcomes, such as medication administration errors, patient falls, nosocomial infections, patient complaints, pressure ulcers, cardiopulmonary arrests, and death, standardized by patient-days or doses of medication (Twigg et al., 2019). As nursing competence increases, adverse occurrence rates of these outcomes should decrease. To adequately analyse this association, competence assessment scales should be valid (i.e., measure the construct of interest) (i.e., valid), reliable (i.e., be free of measurement error), and responsive (i.e., allow the capture of a valid change that has occurred in a construct of interest).

In 2007, the Quality and Safety Education for Nurses collaborative (QSEN; Cronenwett et al., 2007), proposed that all nurses, from new graduate nurses (NGNs) to senior nurses, should be competent in: 1) patient-centred care; 2) teamwork and collaboration; 3) evidence-based practice; 4) quality improvement; 5) safety; and 6) informatics. These competencies are complex know-how, consistent with Gonczi (1994)'s perspective that competence is a holistic, complex and evolutive combination of knowledge, skills, attitudes and values, and bound to their context of practice. According to this definition, competence should not be viewed only in terms of technical skills and its assessment should reflect that.

Apart from the QSEN framework, other authors have developed competency frameworks specifically to guide nursing undergraduate programs, such as the Competency Outcomes and Performance Assessment (COPA) model (Lenburg, 1999; Lenburg et al., 2011; Lenburg et al., 2009) and the Patient Safety Competency Framework (Levett-Jones et al., 2017), or to guide the professional development of nurses after graduation (Boyer et al., 2016). Most of these frameworks provide either a list of knowledge, skills and attitudes for each competency, or decompose each competency into elements and specific indicators. Even though the competences listed in those frameworks are not exactly the same, they are consistent with the holistic definition previously stated.

Results of previous studies showed that NGNs' competence level varies during the first year of practice, with a significant increase in competence during the first six months that slows down after that point (Lima et al., 2016; Takase, 2013; Takase et al., 2014). While NGNs' competence is mostly assessed as good or adequate, there are still reports of inadequate competence level or lack of preparedness for practice (Hezaveh et al., 2013; Odland et al., 2014). Such heterogeneity in findings could be explained by the numerous scales used to assess NGNs' competence, their measurement properties (validity, reliability, responsiveness) and the complexity of competence assessment in general (Franklin & Melville, 2015). Moreover, some scales are based on a behaviourist conception of competence and were designed to measure performance of nurses in accomplishing specific technical tasks (e.g., venepuncture or endotracheal aspiration). While these scales can be useful in specific situations, they are not coherent with the holistic definition of competence generally adopted in nursing (Garside & Nhemachena, 2013).

To the authors' knowledge, a recent systematic evaluation of the measurement properties of NGNs' competence assessment scales has yet to be published, the last systematic review on this topic having been published almost a decade ago Yanhua and Watson (2011). In an earlier mixed methods review on NGNs' competence assessments (reference omitted), we identified 20 scales used in empirical studies published between 2010-2019 to assess NGNs' competence. This previous review gave a comprehensive overview of the studies' methods and results regarding competence assessment. In this paper, we present the evaluation of the measurement properties of scales to guide researchers' and clinicians' selection process of a scale to evaluate NGN's clinical competence.

2. METHOD

2.1 Aim

The aim of this psychometric review was to analyse, evaluate and synthesize the measurement properties of scales used to assess NGNs' clinical competence in clinical settings.

Two research questions were asked:

1. What scales measuring NGNs' clinical competence have been used or developed since 2010?
2. What are the reported measurement properties of these scales?

2.2 Design

We conducted a two-phase systematic review, inspired by Yu and Kirk (2008, 2009). Phase one was a mixed methods systematic review in which we appraised and synthesized evidence of empirical studies reporting assessment of NGNs' clinical competence in clinical settings. (reference omitted). We report in the present paper phase two of the review, in which we reviewed the measurement properties of the scales identified in phase one as well as new scales developed since 2010. To assess each scale, we used the methodological framework developed by the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) (Prinsen et al., 2018). Even though clinical competence scales are not measuring a patient-related outcome, Prinsen et al. (2018) acknowledge that their methods can be adapted to other types of outcome measures. The protocol was prospectively registered in PROSPERO (CRD42018109711), but minor modifications were made, regarding the partition into phases one and two.

2.3 Search Methods

As part of phase one of this review, we identified empirical studies where researchers assessed NGNs' (less than 24 months since graduation) competence in clinical settings. We adopted the definition of competence as a context-bound combination of knowledge, skills, attitudes, and values, presented earlier. As such, we excluded studies in which researchers focused solely on technical skills or where they used scales focusing on a single competence (e.g., clinical decision-making or leadership). The search strategy was developed with the assistance of a librarian, using keywords and thesaurus terms related to: (1) new graduate nurse, (2) clinical competence and (3) competence assessment. The search was undertaken in CINAHL (EBSCO), EMBASE (OVID), MEDLINE, PsycINFO (APA PSycNET) and Web of Science - SCI and SSCI (ISI – Thomson Scientific). The search was limited to full-text papers available in English or French, published between January 2010 and September 2019. Studies reporting the development of new scales were excluded from phase one but set aside to include in phase two.

In this review (phase two), we extracted the scales identified in phase one ($n=20$) as well as new scales developed since 2010. We searched the literature in November 2019 to retrieve studies in English or French reporting the measurement properties or psychometric evaluation of these scales, without setting publication date limit. We searched in CINAHL (EBSCO), EMBASE (OVID) PsycINFO (APA PSycNET), and MEDLINE.

The following inclusion criteria were used to identify relevant studies:

- 1) The scale reported being either used to assess clinical competence of NGNs in clinical settings between 2010-2019 OR developed during the same time frame;
- 2) The report of initial measurement properties (validity, reliability or responsiveness) is available in English or French (regardless of the year of publication).

Exclusion criteria:

- 1) Studies in which the scale was only used as an outcome measurement instrument, and where no measurement properties are reported (as recommended by Prinsen et al. (2018)). We extended this recommendation to studies only reporting an internal consistency measure (usually with Cronbach's alpha coefficient), as it provides very limited evidence of the reliability of the scale.

2.4 Search Outcomes

In total, 1566 records were screened in phase one. The subsequent database search yielded 135 new records. A total of 177 full articles were assessed for eligibility and 170 papers were excluded. Additional records ($n = 3$) were identified through reference lists. Finally, 10 studies describing nine scales were included in this review (Figure 1).

2.5 Data Extraction

Data extraction was guided by the evaluation of measurement properties proposed by the COSMIN methods. Data was extracted using an ad hoc Excel form developed and pilot tested by two review authors (initials omitted) using the same two articles. Adjustments were subsequently made. These authors independently extracted data from all scales that met the inclusion criteria and any disagreements were solved through discussion. The following information was extracted: (1) scale name and original authors, (2) the process used to generate items; (3) dimensions of the scale; (4) number of items and type of measurement scale (e.g., Likert, adjectival); (5) the process used to evaluate the validity, reliability, and responsiveness of the scales as well as the psychometric data associated with these measurement properties; and (6) time to complete the scale (as an indicator of the complexity of using the scale).

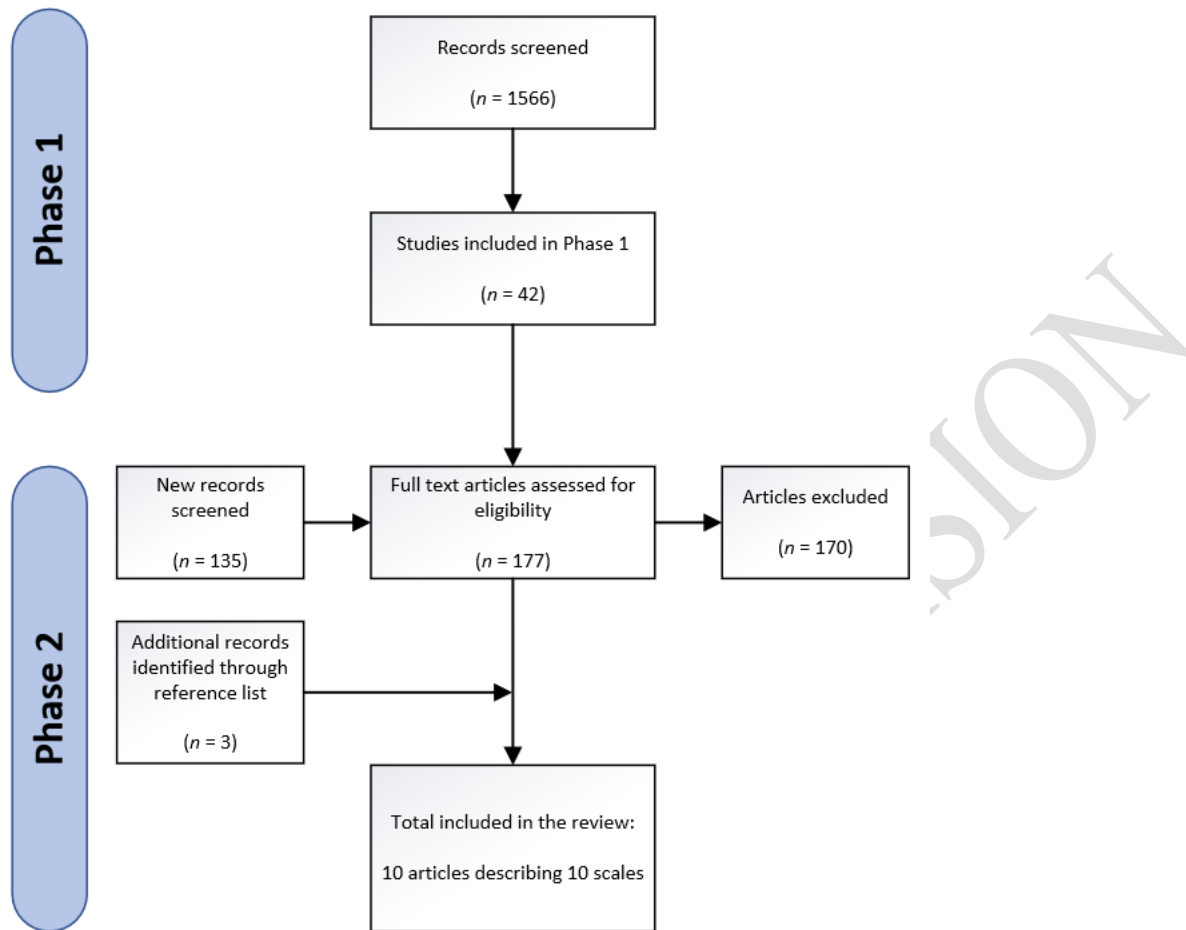


Figure 1. Details of literature search and study selection

2.6 Quality Appraisal

Quality was appraised with the COSMIN checklist. First, included studies were evaluated for risk of bias and then the measurement properties were assessed against a set of criteria developed by Prinsen et al. (2018).

2.6.1 Risk of Bias in Included Studies

Two reviewers (initials omitted) independently used the COSMIN checklist (Prinsen et al., 2018; Terwee et al., 2018) to assess risks of bias of included studies. Risk of bias refers to the influence that methodological elements may have had on the measurement properties of the scales (i.e., were the selected methods adequate regarding the purpose intended and were they correctly used?). The COSMIN checklist is composed of nine elements assessed on a four-point rating scale (very good, adequate, doubtful, inadequate). Each element (or box) represents a measurement property and is comprised of 5-18 items, assessed on a “worst score counts” principle.

2.6.2 *Quality Assessment of Measurement Properties*

Two reviewers (initials omitted) independently used the updated criteria for good measurement properties (Prinsen et al., 2018) to assess the quality of included scales regarding their measurement properties (see Supplementary Table A). Here, quality refers to the actual result of the measurement property (i.e., is the result obtained satisfactory). Each property is rated as either sufficient (+), insufficient (-) or indeterminate (?).

2.7 Synthesis

Synthesis of the evidence of measurement properties for each instrument was undertaken and presented according to their validity, reliability and responsiveness properties (Prinsen et al., 2018).

Validity is defined as the ability of a scale to measure the construct it was designed to measure. There are three major components to validity: face/content validity, construct validity and criterion validity (Polit & Beck, 2017; Streiner et al., 2015). Face/content validity is usually performed with a panel of participants from the target population and/or by a group of experts. It allows researchers to evaluate if the scale appears reasonable to measure the construct of interest and its different dimensions. Construct validity is mostly performed using structural analysis, such as exploratory or confirmatory factor analysis, or by corroborating hypotheses with scores of a similar or correlated construct. It allows researchers to evaluate the internal relationships between the scale items as well their external relationships with other concepts and through hypothesis testing. Criterion validity is assessed by comparing the scale score with scores from a gold standard measurement. It allows researchers to evaluate how well the scale score predicts the score from another measurement.

Reliability is the extent to which a measurement is free of measurement error and is reproducible, i.e., is the score obtained by the scale consistent and precise (Polit & Beck, 2017; Streiner et al., 2015). It is also assessed by three components: internal consistency, stability and equivalence. Internal consistency is a measure of the average correlation of items within the scale, mostly reported by Cronbach's alpha. Stability is measured over time and equivalence is between different raters, both reported by reliability coefficients.

The third domain of measurement properties is responsiveness, which is a scale's ability to capture a valid change in a construct that has changed. Unlike validity and reliability, components and language related to responsiveness do not make consensus (Streiner et al., 2015). According to these authors, responsiveness should not be regarded as a third domain, as it relates to validity (i.e., the validity of the score change). Nonetheless, Prinsen et al. (2018) point out that the amount of change should be assessed and compared to a gold-standard measure or by measuring the area under the curve of change.

3. RESULTS

Two research questions guided this systematic review: 1) what scales measuring NGNS' clinical competence have been used or developed since 2010?; 2) what are the reported measurement properties of these scales?

Following the analysis of the studies identified, various observations emerged in this systematic review, namely: the origin of the development of the scale, the theoretical elements leading to the writing of items, the number of items and sub-scales as well as descriptive elements to assess nursing competence.

We also noted in which studies, the scales for evaluating nursing competency were used.

First, we will present the general results of the systematic review regarding scales used to assess clinical competence of NGNs in clinical settings. Next, we will detail the characteristics and report the psychometric qualities of each of the scale.

3.1 Descriptive Results

Ten scales were included in this review (see Table 1): eight original scales (Becker et al., 2018; Kahya & Oral, 2018; Ko & Yu, 2019; Meretoja et al., 2004; Nilsson et al., 2014; Prion et al., 2015; Schwirian, 1978; Takase & Teraoka, 2011), one culturally adapted (Aydin & Hıçdurmaz, 2019) and one modified (Nilsson et al., 2018). Of these scales, eight were developed or adapted in the 2010s decade and the other two scale were developed earlier.

Item generation was based on literature reviews (including concept analysis) for five of the scales. Benner's work on professional expertise development (also known as novice-to-expert model) was used to develop one scale and QSEN competencies were used to develop two scales.

Scales ranged from 19 to 88 items in length (median = 37, interquartile range = 29.5). Seven scales were divided into six to eight subscales; three scales had four subscales. An analysis of the frequency of subscales showed that despite different designations, similar competences were included: patient-centred care ($n = 8$), collaboration/teamwork and communication ($n = 7$), evidence-based practice and nursing process ($n = 7$), professionalism and professional development ($n = 7$), teaching/mentoring ($n = 6$), quality improvement and safety ($n = 6$).

To measure competency levels, eight scales use an adjectival scale with either four, five or seven points, one scale uses a visual analogue scale (VAS) converted from 0 to 100 for analysis and the measurement scale was unknown for the last scale. Only Prion et al. (2015) reported the time to complete their scale (10 minutes).

Table 1: Characteristics of included scales and recent use (2010-2019)

Name of scale	Reference; country	Item generation	Characteristics and subscales (number of items)	Response options	Recent use in nursing
Six-Dimension Scale of Nursing Performance (6-D Scale)	Schwirian (1978); USA	Based on literature review, consultation and collaboration with expert nurses	56 items divided into 6 subscales: 1. Leadership (5) 2. Critical care (7) 3. Teaching and collaboration (11) 4. Planning and evaluation (7) 5. Interpersonal relations and communication (12) 6. Professional development (10)	4-point adjectival scale (1=not very well, 2=satisfactorily, 3=well, 4=very well)	Aggar et al. (2017); Aggar et al. (2018)
Nurse Competence Scale (NCS)	Meretoja et al. (2004); Finland	Based on Benner's work	73 items divided into 7 subscales: 1. Helping role (7) 2. Teaching-coaching (16) 3. Diagnostic functions (7) 4. Managing situations (8) 5. Therapeutic interventions (10) 6. Ensuring quality (6) 7. Work role (19)	VAS scale (0=very low; 100=very high). Items also assessed on the frequency of use on a 4-point Likert scale.	Kuokkanen et al. (2016); Lima et al. (2016); Lima et al. (2014); Numminen et al. (2014); Numminen et al. (2015a, 2015b); Numminen, Leino-Kilpi, et al. (2016); Numminen et al. (2017); Numminen, Ruoppa, et al. (2016); Wangenstein et al. (2012)
Holistic Nursing Competence Scale (HNCS)	Takase and Teraoka (2011); Japan (original version)	Based on a concept analysis of international literature	Part A: 7 items on general aptitude Part B: 29 items divided into 4 subscales: 1. Staff education and management (9) 2. Ethically oriented practice (9) 3. Nursing care in team (7) 4. Professional development (4)	7-point adjectival scale. Part A: 1=not at all, 2=seldom, 3=occasionally, 4=sometimes, 5=frequently, 6=nearly always, and 7=always. Part B: 1=not competent, 4=reasonably competent, 7=extremely competent	Jung et al. (2017); Takase et al. (2014)
	Aydin and Hiçdurmaz (2019)*; Turkey (translation)	Based on the original HNCS			
Nurse Professional Competence (NPC) scale	Original: Nilsson et al. (2014); Sweden	Based on national competence standards	88 items divided into 2 overarching themes and 8 subscales: Theme 1: Patient-related nursing 1. Nursing care (15) 2. Value-based nursing care (8) 3. Medical and technical care (10) 4. Teaching/learning and support (11) 5. Documentation and information technology (4) 6. Legislation in nursing and safety planning (9) Theme 2: Nursing care organisation and development 6. Legislation in nursing and safety planning (9) 7. Leadership in and development of nursing (26) 8. Education and supervision of staff and students (5)	4-point adjectival scale (1=to a very low degree; 2=to a relatively low degree; 3=to a relatively high degree; and 4=to a very high degree) with a 5 th option of "cannot take a standpoint"	Holowaychuk (2018)

Name of scale	Reference; country	Item generation	Characteristics and subscales (number of items)	Response options	Recent use in nursing
	Short form (NPC-SF): Nilsson et al. (2018); Sweden	Based on the NPC	35 items divided into 6 subscales: 1. Nursing care (5) 2. Value-based nursing care (5) 3. Medical and technical care (6) 4. Care pedagogies (5) 5. Documentation and administration of nursing care (8) 6. Development, leadership and organization of nursing care (6)	4-point adjectival scale (1= to a very low degree; 2 = to a relatively low degree; 3 = to a relatively high degree; and 4 = to a very high degree) with a 5 th option of “cannot take a standpoint”	N/A
Appraisal of Nursing Practice (ANP)	Becker et al. (2018); USA	Based on QSEN framework, items generated in collaboration with academic nurses	37 items divided into 7 subscales: 1. Person-centred care 2. Evidence-based practice/quality improvement 3. Teamwork and collaboration 4. Safety 5. Professionalism 6. Informatics 7. Overall satisfaction	Subscales 1-6 scored on a 5-point adjectival scale: 0=no competency, 1=beginning, 2=developing, 3=independent, 4=advanced. Subscale 7 scored on a 6-point Likert scale from strongly disagree to strongly agree	N/A
New Graduate Registered Nurse Transition Program Competency Scale (NGRNTPCT)	Prion et al. (2015); USA	Based on QSEN framework, items generated in collaboration with clinical nurse leaders	35 items divided into 7 subscales: 1. Patient-centred care (9) 2. Safety (5) 3. Evidence-based practice (3) 4. Teamwork and collaboration (8) 5. Professionalism (6) 6. Quality (2) 7. Informatics (2) Item 36: Overall competency level	4-point adjectival scale: 0=not applicable, 1=beginning, 2=developing, 3=accomplished	Oblea et al. (2019)
Clinical Nurse Performance Scale (CNPT)	Kahya and Oral (2018); Turkey	Based on literature review	38 items divided into 8 subscales 1. Contextual (11) 2. Professional skill (4) 3. Clinical skill (6) 4. Interpersonal communication (3) 5. Problem solving (3) 6. Professional ethic (3) 7. Teamwork (4) 8. Leadership (4)	No mention of the measurement scale	N/A
Core Nursing Competency Assessment	Ko and Yu (2019); Korea	Based on a literature review	19 items divided into 4 subscales: 1. Critical thinking and application of the nursing process (5)	4-point scale (unknown measurement)	N/A

Name of scale	Reference; country	Item generation	Characteristics and subscales (number of items)	Response options	Recent use in nursing
Scale (CNCAT)			2. Self-management, coordination, and collaboration (6) 3. Patient care (3) 4. Ensuring quality and communication (5)		

Note: *The Turkish translation of the HNCS has the same number of items and categories as the original version.

3.2 Methodological quality of included scales

No psychometric data was found regarding the responsiveness of any of the included scales. Therefore, these elements are omitted in our presentation of the results. We present below the results of the methodological quality and measurement ratings for each scale (see Table 2).

3.2.1 *Six-Dimension Scale of Nursing Performance (6-D Scale)*

The 6-D Scale (Schwirian, 1978) was considered the most rigorously tested scale for nursing performance appraisal until the early 2000s, before other scales became available (Meretoja et al., 2004). This scale was specifically developed to be used both as a self-appraisal scale by NGNs and as an evaluation scale by supervisors/employers.

Validity

Methodological quality of the development/content validity study of the 6-D Scale has been deemed doubtful because some information on the method is missing. However, content validity was assessed both by experts from nursing schools and by a pilot test with a sample of graduate nurses, which was rated sufficient. Regarding structural validity, Schwirian (1978) used principal component analysis (PCA) to determine the factor structure of the 6-D Scale, without stating how much variance was explained by each factor. This was judged to be adequate, but the rating is therefore indeterminate. Finally, Schwirian (1978) reported that the 6-D scale was able to differentiate between pre-identified successful and less successful graduates (known-groups validity), without giving sufficient detail on this testing.

Reliability

Internal consistency was rated as sufficient: Cronbach's alphas were calculated for each subscale and varied from 0.901 to 0.978 with graduate nurses self-appraisals and from 0.844 to 0.907 with supervisor appraisals (Schwirian, 1978).

3.2.2 *Nurse Competence Scale (NCS)*

This scale was developed by Meretoja et al. (2004) in Finland, based on Benner's novice-to-expert model.

Validity and reliability

Since a psychometric review of the NCS was published recently (Flinkman et al., 2017), we refer the reader to that publication for a detailed overview of the validity and reliability of the NCS. The authors concluded that the NCS showed good content validity but limited structural validity as many translated versions have not been properly tested. They also concluded that only the internal consistency was used to report on the reliability of the scale.

3.2.3 *Holistic Nursing Competence Scale (HNCS)*

The HNCS was developed by Takase and Teraoka (2011) as they deemed the existing international scales could not be used in the Japanese context. In the development study, the HCNS was used as a self-appraisal scale and results seem to indicate the absence of a ceiling effect. The HCNS was translated into Turkish (Aydin & Hiçdurmaz, 2019) and this translated version kept the same number of items and categories.

Validity

Content validity of the HNCS was established by a group of nursing experts ($n=6$; managers or educators) as well as a pilot test with 23 nurses. While the authors do not report the values of the content validity index (CVI), they mention that the cut-off value was 0.83, which was rated as sufficient. The pilot test was deemed of doubtful methodological quality since it is unclear if participants were asked about comprehensiveness of the scale. Structural validity testing was performed by an exploratory factor analysis (EFA) and a confirmatory factor analysis (CFA). While more than 59% of the variance was explained by the 5-factor model, the methodological quality of the factor analysis was deemed inadequate since the CFA testing the goodness of fit seemed to have been done on the same sample as the EFA on which the factors were first extracted. Finally, further validity testing was done by correlating the HCNS mean score to length of clinical experience, which revealed a weak positive correlation (0.363; insufficient rating). The content validity of the Turkish translation was deemed of doubtful methodological quality because only psychiatric nursing experts were recruited. A CFA confirmed the original structure model of the HNCS (cross-cultural validity).

Reliability

Internal consistency was rated sufficient: Cronbach's alphas were calculated for subscale levels and varied from 0.868 to 0.934 (Takase & Teraoka, 2011) and 0.89 to 0.94 for the Turkish version (Aydin & Hiçdurmaz, 2019). A test-retest with the Turkish version showed a good correlation (0.83) after three weeks, which was rated sufficient and methodologically adequate.

3.2.4 Nurse Professional Competence (NPC)

Nilsson et al. (2014) developed the NPC scale based on Swedish national standards for practice to assess the competence of both nursing students at the point of graduation and practicing nurses. It was developed as a self-appraisal measure. In the development study where participants were nursing students at the point of graduation, Nilsson et al. (2014) reported a mean score range of 2.40 to 3.88 (SD 0.34 to 0.80), with 15% of participants selecting the highest response option for 86% of items; this could indicate a ceiling effect on some of the items.

Validity

Content validity of the HNCS was established by nursing academics ($n=12$) who reviewed the relevance of items, as well as face validity, with 27 practicing nurses who tested the scale and

comprehensibility of items. Methodological quality was deemed inadequate since the authors did not mention if comprehensiveness of the scale was assessed. However, content validity was deemed sufficient since both experts and target population were involved in the process. Nilsson et al. (2014) assessed structural validity of the scale by EFA (which was deemed methodologically adequate), which resulted in an 8-factor structure explaining 48% of the variance. A subsequent EFA revealed two overarching themes that explained 65% of the variance. Finally, the authors report some evidence of known-group validity which was deemed methodologically doubtful, as groups were only compared on two items of the scale.

Reliability

Internal consistency was rated sufficient: Cronbach's alphas were calculated for each subscale and varied from 0.75 to 0.94 and was calculated at 0.97 for the overall scale (Nilsson et al., 2014).

3.2.5 NPC short form (NPC-SF)

As the NPC is lengthy (88 items), Nilsson et al. (2018) developed a shorter version of the scale with the same aim of assessing the competence of both nursing students at the point of graduation and practicing nurses, as a self-report measure. As with the NPC, the NPC-SF seems to exhibit a ceiling effect for some items; Nilsson et al. (2018) reported mean scores of 2.65 to 3.71 (SD 0.50-0.84) with a sample of graduating nursing students ($n=1810$).

Validity

Content validity of the NPC-SF was deemed methodologically inadequate, since only the authors participated in reduction from 88 to 35 items. However, structural validity was deemed of very good quality, as Nilsson et al. (2018) divided their sample in two to perform an EFA which resulted in a 6-factor solution (explaining 53.6% of the variance, which is deemed of sufficient quality) and then a CFA to confirm that structure.

Reliability

Internal consistency was rated as sufficient: Cronbach's alphas were calculated for each subscale and varied from 0.71 to 0.86 (Nilsson et al., 2018).

3.2.6 New Graduate Registered Nurse Transition Program Competency Scale (NGRNTPCT)

Prion et al. (2015) developed the NGRNTPCT, but it should be noted that the scale has no specific name in the development paper; the acronym NGRNTPCT was used by other authors who subsequently used the scale (Oblea et al., 2019). This scale was developed to assess the clinical competence of NGNs undergoing a nurse residency program and was meant to be used by their preceptors, even though the authors acknowledge that it could be used as a self-report scale by NGNs. The authors do not discuss a ceiling effect; however, junior-level nursing students (who

represented third-year students in a four-year program) had a mean score of 2.23/3. Mean score for NGNs was not reported.

Validity

Face validity of the NGRNTPCT was rated as indeterminate because only a small sample of experts ($n=4$) were involved in the validation process. Due to the small sample, no CVI could be calculated. Construct validity was further tested with a sample of junior-level nursing students ($n=94$) and faculty members ($n=17$) and the NGRNTPCT was successful in discriminating between their competence level ($M=2.23$ vs. 2.84).

Reliability

Internal consistency was rated as indeterminate, as Cronbach's alpha was only calculated for the total scale and not any subscale. Alphas varied from 0.82 when self-reported by students and faculty members to 0.92 when preceptors reported NGNs' competence levels (Prion et al., 2015).

3.2.7 Appraisal of Nursing Practice (ANP)

Becker et al. (2018) developed the ANP to assess the clinical competence of NGNs undergoing a nurse residency program and was meant to be used by their preceptors. Individual items of the scale were not available in the paper. Despite training raters and providing a clear definition of each score of the 5-point scale, Becker et al. (2018) noted a tendency for some raters to score NGNs highly, which could indicate a ceiling effect.

Validity

The ANP underwent several steps during the validation process but was rated as indeterminate for two reasons: the individual items are not freely accessible in the development paper and no CVI or ratio is provided. Despite this, the overall methodological quality of the process was judged adequate, since Becker et al. (2018) tested the content through cognitive interviews (with managers and educators) and undertook follow-up interviews (with preceptors) to later refine the scale.

Reliability

Internal consistency was rated as sufficient: Cronbach's alphas were calculated for each subscale and varied from 0.86 to 0.97 (Becker et al., 2018). Equivalence was tested in the same study by interrater agreement and also rated as sufficient, with a mean interrater agreement of 0.80 (ranging from 0.44 to 0.93).

3.2.8 Clinical Nurse Performance Scale (CNPT)

Kahya and Oral (2018) developed the CNPT because they felt that no existing scale appropriately integrated contextual items while measuring nurse performance. They report that the CNPT is designed to be an assessment scale completed by a manager/supervisor while acknowledging that

self-report has a certain value in understanding the competence of nurses. In the development study, Kahya and Oral (2018) do not report the actual performance of the scale, but rather a large-scale content validation, with participants reporting appropriateness of each item.

Validity

Content validity of the CNPT was assessed by experts who reviewed the comprehensiveness and comprehensibility of the items and by a mixed sample who reviewed the importance and appropriateness of items. This was deemed methodologically inadequate and of insufficient quality. It is worth noting that the structure of the CNPT was not obtained by a structural analysis.

Reliability

Since it appears Kahya and Oral (2018) did not collect data by testing the actual scale but rather testing the appropriateness of items, the internal consistency coefficient they reported it was not deemed an accepted measurement property.

3.2.9 Core Nursing Competency Assessment Scale (CNCAT)

Ko and Yu (2019) developed the CNCAT, but it should be noted that this abbreviation is not mentioned by the authors. This scale was developed to assess the clinical competence of NGNs who completed a specific Korean outcome-based education program (called KABONE). The authors do not mention if the target population was NGNs or their preceptors/supervisors, even though they recruited NGNs to evaluate the construct validity and reliability of the scale, but not the face or content validity.

Validity

The CNCAT underwent several steps during the validation process. Methodological quality of the content validity was deemed inadequate and rated as insufficient, since this step was performed with two expert groups (managers and professors). Structural validity of the scale was tested by a factor analysis, resulting in four factors explaining 54.1% of the variance; the methodological quality was judged to be adequate and the result is sufficient. Finally, Ko and Yu (2019) assessed the convergent validity by looking at the association between the CNCAT and two other measures (one other performance measure and a leadership measure). The quality of this study was deemed inadequate since the authors did not provide the measurement properties of both comparators. The correlation between the CNCAT and the performance measure was 0.78 (rated sufficient) and only 0.47 with the leadership measure (rated insufficient).

Reliability

Internal consistency was calculated for the overall scale with Cronbach's alpha calculated at 0.87 (Ko & Yu, 2019), which is deemed of inadequate methodological quality and the result is indeterminable, since alpha should be calculated for each subscale.

Table 2: Methodological quality and ratings of measurement properties

Instrument	Reference Country	Content validity			Structural validity			Hypotheses testing			Internal consistency			Reliability		
		<i>n</i>	<i>M</i>	<i>Results (Q)</i>	<i>n</i>	<i>M</i>	<i>Results (Q)</i>	<i>n</i>	<i>M</i>	<i>Results (Q)</i>	<i>n</i>	<i>M</i>	<i>Results (Q)</i>	<i>n</i>	<i>M</i>	<i>Results (Q)</i>
6-D Scale	Schwirian (1978); USA	160	D	Experts (<i>n</i> =151) reviewed representativeness. Nurses (<i>n</i> =9) tested the questionnaire (?)	1501	A	No mention of the variance (?)	587	I	Differentiation between successful and less successful graduates (+)	1501	V	Cronbach's alpha (subscales) 0.844-0.978 (+)			
HNCS	Takase and Teraoka (2011); Japan	29	D	Expert (<i>n</i> =6) reviewed relevance (CVI not reported, but cut-off value of 0.83 reported). Nurses (<i>n</i> =23) reviewed comprehensibility (+)	331	I	59,739% of the variance was explained (+)	331	V	Correlation 0.363 with length of clinical experience (+)	331	V	Cronbach's alpha (subscales) 0.868-0.934 and (overall tool) 0.967 (+)			
HNCS - Turkish version	Aydin and Hiçdurmaz (2019); Turkey	11	D	Experts (<i>n</i> =11) reviewed relevance, comprehensibility and comprehensiveness (CVI ≥ 0.80 for all items). No target population involved (-)	288	V	Cross-cultural validity: Original structure confirmed by CFA (?)				288	V	Cronbach's alpha (subscales) 0.89-0.94 and (overall tool) 0.97 (+)	72	A	Equivalence: Spearman correlation 0.83 after 3 weeks (+)
NPC	Nilsson et al. (2014); Sweden	39	D	Academic (<i>n</i> =12) reviewed relevance of items. Nursing students (<i>n</i> =27) reviewed comprehensibility (+)	1086	A	48% of the variance explained by the subscales, but 65% explained by 2 overarching themes (+)	1086	D	Differentiation between disaster-prepared and unprepared students on 2 items (+)	1086	V	Cronbach's alpha (subscales) 0.75-0.94 and (overall tool) 0.97 (+)			
NPC-SF	Nilsson et al. (2018); Sweden	4	I	Authors reviewed relevance of highly correlated items. No target population involved (-)	1810	V	53,6% of the variance was explained. CFA showed good fit of the original factor structure (+)				1810	V	Cronbach's alpha (subscales) 0.71-0.86 (+)			

Instrument	Reference Country	Content validity			Structural validity			Hypotheses testing			Internal consistency			Reliability		
		<i>n</i>	<i>M</i>	<i>Results (Q)</i>	<i>n</i>	<i>M</i>	<i>Results (Q)</i>	<i>n</i>	<i>M</i>	<i>Results (Q)</i>	<i>n</i>	<i>M</i>	<i>Results (Q)</i>	<i>n</i>	<i>M</i>	<i>Results (Q)</i>
NGRNTPT	Prion et al. (2015); USA	4	I	Experts (<i>n</i> =4) reviewed relevance and comprehensibility. No target population involved (-)				111	A	Students' mean score (2.23, SD 0.29) lower than faculty's (2.86, SD 0.27) (+)	304	I	Cronbach's alpha (overall tool) 0.82-0.92 (?)			
ANP	Becker et al. (2018); USA	11+	D	Experts (<i>n</i> =unknown) reviewed relevance, then cognitive interview with experts (<i>n</i> =4) to review comprehensibility and follow-up interviews with nurses (<i>n</i> =7) were done to review relevance and comprehensibility (+)							100	V	Cronbach's alpha (subscales) 0.86-0.97 (+)	23	I	Stability: interrater agreement 0.44-0.93 (M=0.8) (+)
CNPT	Kahya and Oral (2018);	197	I	Experts (<i>n</i> =34) reviewed comprehensiveness and comprehensibility; Mixed sample (<i>n</i> =163) reviewed the importance of items (?)												
CNCAT	Ko and Yu (2019); Korea	20	I	Nurse managers (<i>n</i> =10) and professors (<i>n</i> =10) reviewed items relevance, comprehensibility and comprehensiveness. CVI respectively 0.89 and 0.98. No target population involved (-)	141	A	54.1% of the variance was explained (+)	141	I	Correlation 0.78 with a performance comparator (+); Correlation 0.47 with a leadership scale (-)	141	I	Cronbach's alpha (overall tool) 0.87 (?)			

Note: Empty cell means no reported testing. *n* = Sample size; *M* = Methodological quality: V=Very Good, A=Adequate, D=Doubtful, I=Inadequate; *Q* = Quality of results: + = sufficient; - = insufficient; ? = indeterminate; 6-D Scale: Six-Dimension Scale of Nurse Performance; HNCS: Holistic Nursing Competence Scale; NPC: Nurse Professional Competence; NPC-SF: Nurse Professional Competence – Short form; NGRNTPT: New Graduate Registered Nurse Transition Program Competency Tool; ANP: Appraisal of Nursing Practice; CNPT: Clinical Nurse Performance Scale; CNCAT: Core Nursing Competency Assessment Scale

4. DISCUSSION

This systematic review identified ten scales used by researchers or developed from 2010 to assess the competence of NGNs in clinical settings. We summarized the validity and reliability testing performed for each scale. We found that no authors reported any form of responsiveness testing. Moreover, our systematic review shows that there is little evidence on the measurement properties for each scale and therefore, it is impossible to recommend one scale over another.

Streiner et al. (2015) mentioned that researchers tend to quickly dismiss existing scales to develop a new measure without properly justifying the underlying reason; this systematic review confirms this perspective. While a scale might have sufficient initial content and structural validity, it is in the interest of the nursing community to further explore the validity of existing scales in other populations and/or contexts (Streiner et al., 2015). Such steps of testing and revising scales are necessary to enable deeper understanding of the concept being measured and appropriate measurement of it. Scales measuring dynamic constructs are meant to be dynamic as well, which means they should be revised and evolve over time. Moreover, contrary to beliefs, scales do not have intrinsic measurement properties; rather, their validity and reliability are bound to the context and population in which they were tested (Streiner et al., 2015). For this reason, after a scale is developed, it should undergo further testing. Other than reporting the internal consistency coefficient (Cronbach's alpha), very few studies in nursing are undertaken to bring further evidence regarding measurement properties.

The content validity study of all scales in this review was deemed to be of doubtful or inadequate quality. This does not mean the scales were not rigorously developed, rather than the standards defined by the COSMIN initiative might not apply to every scale being developed. Content validity is usually assessed by questioning two groups of participants: one from the target population and the second from experts (or professionals) (Boateng et al., 2018; Polit & Beck, 2017; Terwee et al., 2018). Boateng et al. (2018) point out that the use of a sample from a target population is ideal but not essential. When a self-appraisal scale is developed, the target population is easily identifiable. However, this distinction is not as limpid when the scale is to be used by a proxy assessor, such as a preceptor, supervisor or manager, which is often the case in nursing competence assessment. For this review, we considered nurses whose competence was being assessed as the target population for all scales. However, following COSMIN standards, the absence of the target population in the development of a scale is rated as inadequate (Terwee et al., 2018), which contradicts Boateng et al. (2018).

In a previous review (reference omitted), we identified that the NCS was the most used scale since 2010, but even if it has many translated versions and some evidence of cross-cultural validity, it still lacks strong evidence of structural validity (Flinkman et al., 2017), as many studies were not able to confirm the 7-factor structure. Other than the NCS, two other scales in this review had their factor structure confirmed, namely the NPC (CFA done on the NPC-SF) and HNCS (CFA

performed on the Turkish version). Other studies would be necessary to confirm the factor structure of other scales included in this review.

Reliability was almost exclusively assessed by calculating the internal consistency with Cronbach's alpha coefficient. As recommended by Streiner et al. (2015), reliability should at least be assessed in terms of stability (e.g., test-retest) which was only done by Aydin and Hiçdurmaz (2019) when they developed the Turkish version of the HNCS. Becker et al. (2018) assessed the equivalence of the ANP by comparing two independent raters assessing the same nurse. Flinkman et al. (2017), in their review of the NCS, did not find studies that reported stability or equivalence testing. Further studies should be done to investigate the reliability of the scales presented in this systematic review.

One of the reasons most researchers rely widely on Cronbach's alpha coefficient is its accessibility: it is the only reliability coefficient that can be obtained with a single administration of the scale (Streiner et al., 2015). However, it is also the least adequate coefficient, as it gives no information on the equivalence or stability of the measure (Streiner et al., 2015). Deng and Chan (2017) also argue that "alpha is well known *but poorly understood* by many applied researchers" (emphasis added). The findings from this review support that statement. Even though competence assessment scales included in this review were multidimensional scales, some authors did not report the coefficient for each unidimensional subscale, rather only for the overall scale, which is inadequate (Prinsen et al., 2018; Streiner et al., 2015). This leads to another major problem with interpretation of items increases (or if a dimension is added), alpha will also increase even if dimensions are modestly correlated (Streiner et al., 2015). Therefore, multidimensional scales will almost always have a high alpha. Other measures have been advocated, such as McDonald's omega (Deng & Chan, 2017), but are scarcely used in nursing. While it was not our objective to recommend one particular reliability measure more than another, we believe the rigor of nursing education research could be improved by deeper understanding of the limits of these measures.

The fact that responsiveness was never properly assessed in the reviewed studies is worrisome as competence assessment scales are often used longitudinally (Aggar et al., 2018; Jung et al., 2017; Lima et al., 2016; Oblea et al., 2019). This reflects much of the research undertaken to date in nurse education. Assessment of responsiveness would provide further information regarding strength and sensitivity of the scales in measuring the components they set out to, and hence, effective evaluation of competence.

Surprisingly, authors of two new scales (Ko & Yu, 2019; Prion et al., 2015) included in this review did not use a specific designation, title or acronym for their new scale. While this might seem trivial, it can have an impact on the indexation of future studies using those scales, creating confusion around their use. We recommend that any researchers developing a scale or working on a new version of a scale should include a title and acronym for the scale in their report.

Finally, while scales in this review were included if the development study was reported in English or French, many authors do not specifically report the language of the original scale and if (or how) the English version presented in the paper was translated. For example, the HNCS was developed in Japan (Takase & Teraoka, 2011) by Japanese authors and involved the recruitment of Japanese nurses. Takase and Teraoka (2011) do not mention in which language was the HNCS developed, but they present an English version in their paper. However, when presenting the Turkish version of the HNCS, Aydin and Hiçdurmaz (2019) explicitly mention that the original version is in Japanese language and that the English version was translated by one of the original authors alone. This is of importance, since flaws in the translation process can have an impact on the validity and reliability of a scale (Squires et al., 2020). The choice of words in a scale is linked to the cultural context in which it was developed. Many guidelines exist for the translation and cultural adaptation of scales (McKenna & Doward, 2005; Sousa & Rojjanasrirat, 2011; Wild et al., 2005), which not only include the translation, but most importantly, the validation of translated scales. The COSMIN guideline (Prinsen et al., 2018; Terwee et al., 2018) explicitly states that the language in which an instrument is developed or used should be clearly stated, but this was not the case for most scales in this review.

4.1 Limitations

The major limitation of this systematic review is the restrictive inclusion criteria of the scales: they must have been used or developed since 2010, to assess NGNs' competence. Furthermore, while effort was made to locate subsequent validation studies for the included scales, screening by title and abstract only permitted adding three references. Other relevant scales might exist but were not included in this review. The fact that the NCS was the subject of a systematic review in 2017 is also a limitation in our review, since we did not extract measurement data of the NCS but rather relied on Flinkman et al. (2017)'s analysis. Another limitation is that we made the conscious decision of not contacting authors for missing information. While a scale might have gone through rigorous testing, the rigor and precision of how the development study is reported are crucial. The validity of the results of this review is therefore limited by the scale development methods and how they were reported. Furthermore, the assessment of the overall quality of the evidence using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach could not be conducted as it requires more than one psychometric evaluation of a scale to be available in order to pool study results and express level of confidence in them. Despite these limitations, this review provides insightful findings on the measurement properties of scales to assess NGNs' competence in clinical settings.

5. CONCLUSION

This psychometric review provides a rigorous synthesis of scales designed to measure NGNs' competence in clinical settings. It can help researchers and educators in selecting the best scale

available for their needs, based on critical appraisal using the rigorous COSMIN methodology to evaluate measurement properties of the scales.

There is no perfect scale, especially since competence assessment is a complex, dynamic concept. Every scale evaluated in this review was composed of different subscales representing competencies. Therefore, selection of the most appropriate scale depends on the context and purpose of the assessment. We agree with Franklin and Melville (2015) that future studies looking into NGNs' competence should be multifaceted, not only in terms of assessors, but also in terms of methods. A valid and reliable scale is not only a useful tool but becomes essential to ensure the validity of the findings of such studies.

With this review, we wanted to initiate a broader discussion in nursing on the rigor of the development and use of scales. Further research should focus on generating more evidence regarding the measurement properties of scales with different populations or contexts. Researchers should focus on refining existing scales by cross-culturally validating them and bringing new evidence to the validity and reliability of existing scales. Translation and adaptation of existing scales can also be an option, even if the process can be quite lengthy and strenuous.

Authors contribution

All authors have approved the final version of this article and meet both of the following criteria, based on the recommendations by the ICMJE (<http://www.icmje.org/recommendations>):

- substantial contributions to conception and design, acquisition of data or analysis and interpretation of data;
- drafting the article or revising it critically for important intellectual content.

Conflict of Interest statement

None

Funding statement

In-kind support was provided by La Trobe University, in the form of librarian support.

6. References

- Aggar, C., Bloomfield, J., Thomas, T. H., & Gordon, C. J. (2017). Australia's first transition to professional practice in primary care program for graduate registered nurses: a pilot study. *BMC Nursing*, 16, 1-11. <https://doi.org/10.1186/s12912-017-0207-5>
- Aggar, C., Gordon, C. J., Thomas, T. H. T., Wadsworth, L., & Bloomfield, J. (2018). Evaluation of a community transition to professional practice program for graduate registered nurses

- in Australia. *Nurse Education in Practice*, 32, 101-107. <https://doi.org/10.1016/j.nepr.2018.03.005>
- Aydin, A., & Hiçdurmaz, D. (2019). Holistic nursing competence scale: Turkish translation and psychometric testing. *International Nursing Review*, 66(3), 425-433. <https://doi.org/10.1111/inr.12514>
- Becker, H., Meraviglia, M., Seo, J. E., Ndlovu, C., Kimmel, L., & Rowin, T. (2018). The Appraisal of Nursing Practice: Instrument Development and Initial Testing. *Journal of Nursing Administration*, 48(1), 44-49. <https://doi.org/10.1097/NNA.0000000000000569>
- 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, 1-18. <https://doi.org/10.3389/fpubh.2018.00149>
- Boyer, L., Pepin, J., Dubois, S., Descôteaux, R., Robinette, L., & Déry, J. (2016). *Référentiel de compétences infirmières en milieu de soins hospitaliers québécois de l'enfant à l'adulte*. CHU Sainte-Justine.
- Church, C. D. (2016). Defining Competence in Nursing and Its Relevance to Quality Care. *Journal for Nurses in Professional Development*, 32(5), E9-E14. <https://doi.org/10.1097/nnd.0000000000000289>
- Cronenwett, L., Sherwood, G., Barnsteiner, J., Disch, J., Johnson, J., Mitchell, P., Sullivan, D. T., & Warren, J. (2007). Quality and safety education for nurses. *Nursing Outlook*, 55(3), 122-131. <https://doi.org/10.1016/j.outlook.2007.02.006>
- Deng, L., & Chan, W. (2017). Testing the Difference Between Reliability Coefficients Alpha and Omega. *Educational and Psychological Measurement*, 77(2), 185-203. <https://doi.org/10.1177/0013164416658325>
- Flinkman, M., Leino-Kilpi, H., Numminen, O., Jeon, Y., Kuokkanen, L., & Meretoja, R. (2017). Nurse Competence Scale: a systematic and psychometric review. *Journal of Advanced Nursing*, 73(5), 1035-1050. <https://doi.org/10.1111/jan.13183>
- Franklin, N., & Melville, P. (2015). Competency assessment tools: An exploration of the pedagogical issues facing competency assessment for nurses in the clinical environment. *Collegian*, 22, 25-31. <https://doi.org/10.1016/j.colegn.2013.10.005>

- Garside, J. R., & Nhemachena, J. Z. Z. (2013). A concept analysis of competence and its transition in nursing. *Nurse Education Today*, 33, 541-545. <https://doi.org/10.1016/j.nedt.2011.12.007>
- Gonczi, A. (1994). Competency Based Assessment in the Professions in Australia. *Assessment in Education: Principles, Policy & Practice*, 1(1), 27-44. <https://doi.org/10.1080/0969594940010103>
- Hezaveh, M. S., Rafii, F., & Seyedfatemi, N. (2013). Novice nurses' experiences of unpreparedness at the beginning of the work. *Global journal of health science*, 6(1), 215-222. <https://doi.org/10.5539/gjhs.v6n1p215>
- Holowaychuk, A. S. (2018). *The relationship between the number of hours taken in a senior practicum course in prelicensure nursing education and the score of preception of professional competence by the newly graduated nurse*; [Liberty University]. <https://search.proquest.com/docview/2101891360/2686D7C2E45A4D91PQ/1?accountid=12001>
- Jung, D., Lee, S. H., Kang, S. J., & Kim, J.-H. (2017). Development and evaluation of a clinical simulation for new graduate nurses: A multi-site pilot study. *Nurse Education Today*, 49, 84-89. <https://doi.org/10.1016/j.nedt.2016.11.010>
- Kahya, E., & Oral, N. (2018). Measurement of clinical nurse performance: Developing a tool including contextual items. *Journal of Nursing Education and Practice*, 8(6), 112-123. <https://doi.org/10.5430/jnep.v8n6p112>
- Ko, Y., & Yu, S. (2019). Core nursing competency assessment tool for graduates of outcome-based nursing education in South Korea: A validation study. *Japan Journal of Nursing Science*, 16(2), 155-171. <https://doi.org/10.1111/jjns.12223>
- Kuokkanen, L., Leino-Kilpi, H., Numminen, O., Isoaho, H., Flinkman, M., & Meretoja, R. (2016). Newly graduated nurses' empowerment regarding professional competence and other work-related factors. *BMC Nursing*, 15, 1-8. <https://doi.org/10.1186/s12912-016-0143-9>
- Lenburg, C. B. (1999). The Framework, Concepts and Methods of the Competency Outcomes and Performance Assessment (COPA) Model. *Online Journal of Issues in Nursing*, 4(2), Repéré à <http://www.nursingworld.org/MainMenuCategories/ANAMarketplace/ANAPeriodicals/OJIN/TableofContents/Volume41999/No41992Sep41999/COPAModel.html>.

[www.nursingworld.org/MainMenuCategories/ANAMarketplace/ANAPeriodicals/OJIN/T
ableofContents/Volume41999/No2Sep1999/COPAModel.aspx](http://www.nursingworld.org/MainMenuCategories/ANAMarketplace/ANAPeriodicals/OJIN/TableofContents/Volume41999/No2Sep1999/COPAModel.aspx)

- Lenburg, C. B., Abdur-Rahman, V. Z., Spencer, T. S., Boyer, S. A., & Klein, C. J. (2011). Implementing the COPA Model in Nursing Education and Practice Settings: Promoting Competence, Quality Care, and Patient Safety. *Nursing Education Perspectives*, 32(5), 290-296. <https://doi.org/10.5480/1536-5026-32.5.290>
- Lenburg, C. B., Klein, C., Abdur-Rahman, V., Spencer, T., & Boyer, S. (2009). The COPA Model: a comprehensive framework designed to promote quality care and competence for patient safety. *Nursing Education Perspectives*, 30(5), 312-317.
- Levett-Jones, T., Dwyer, T., Reid-Searl, K., Heaton, L., Flenady, T., Applegarth, J., Guinea, S., & Andersen, P. (2017). *Patient Safety Competency Framework (PSCF) for Nursing Students*. <https://patientsafetyfornursingstudents.org/>
- Lima, S., Newall, F., Jordan, H. L., Hamilton, B., & Kinney, S. (2016). Development of competence in the first year of graduate nursing practice: a longitudinal study. *Journal of Advanced Nursing*, 72(4), 878-888. <https://doi.org/10.1111/jan.12874>
- Lima, S., Newall, F., Kinney, S., Jordan, H. L., & Hamilton, B. (2014). How competent are they? Graduate nurses self-assessment of competence at the start of their careers. *Collegian*, 21, 353-358. <https://doi.org/10.1016/j.colegn.2013.09.001>
- McKenna, S. P., & Doward, L. C. (2005). The translation and cultural adaptation of patient-reported outcome measures. *Value in Health*, 8(2), 89-91. <https://doi.org/10.1111/j.1524-4733.2005.08203.x>
- Meretoja, R., Isoaho, H., & Leino-Kilpi, H. (2004). Nurse Competence Scale: development and psychometric testing. *Journal of Advanced Nursing*, 47(2), 124-133. <https://doi.org/10.1111/j.1365-2648.2004.03071.x>
- Nilsson, J., Engström, M., Florin, J., Gardulf, A., & Carlsson, M. (2018). A short version of the nurse professional competence scale for measuring nurses' self-reported competence. *Nurse Education Today*, 71, 233-239. <https://doi.org/10.1016/j.nedt.2018.09.028>
- Nilsson, J., Johansson, E., Egmar, A. C., Florin, J., Leksell, J., Lepp, M., Lindholm, C., Nordstrom, G., Theander, K., Wilde-Larsson, B., Carlsson, M., & Gardulf, A. (2014). Development and validation of a new tool measuring nurses self-reported professional competence--the

- nurse professional competence (NPC) Scale. *Nurse Education Today*, 34, 574-580. <https://doi.org/10.1016/j.nedt.2013.07.016>
- Numminen, O., Laine, T., Isoaho, H., Hupli, M., Leino-Kilpi, H., & Meretoja, R. (2014). Do educational outcomes correspond with the requirements of nursing practice: educators' and managers' assessments of novice nurses' professional competence. *Scandinavian Journal of Caring Sciences*, 28(4), 812-821. <https://doi.org/10.1111/scs.12115>
- Numminen, O., Leino-Kilpi, H., Isoaho, H., & Meretoja, R. (2015a). Ethical climate and nurse competence – newly graduated nurses' perceptions. *Nursing Ethics*, 22(8), 845-859. <https://doi.org/10.1177/0969733014557137>
- Numminen, O., Leino-Kilpi, H., Isoaho, H., & Meretoja, R. (2015b). Newly Graduated Nurses' Competence and Individual and Organizational Factors: A Multivariate Analysis. *Journal of Nursing Scholarship*, 47(5), 446-457. <https://doi.org/10.1111/jnu.12153>
- Numminen, O., Leino-Kilpi, H., Isoaho, H., & Meretoja, R. (2016). Newly graduated nurses' occupational commitment and its associations with professional competence and work-related factors. *Journal of Clinical Nursing*, 25(1-2), 117-126. <https://doi.org/10.1111/jocn.13005>
- Numminen, O., Leino-Kilpi, H., Isoaho, H., & Meretoja, R. (2017). Development of Nurses' Professional Competence Early in Their Career: A Longitudinal Study. *Journal of Continuing Education in Nursing*, 48(1), 29-39. <https://doi.org/10.3928/00220124-20170110-08>
- Numminen, O., Ruoppa, E., Leino-Kilpi, H., Isoaho, H., Hupli, M., & Meretoja, R. (2016). Practice environment and its association with professional competence and work-related factors: perception of newly graduated nurses. *Journal of Nursing Management*, 24(1), E1-E11. <https://doi.org/10.1111/jonm.12280>
- Oblea, P. N., Berry-Caban, C. S., Dumayas, J. Y., Adams, A. R., & Beltran, T. A. (2019). Evaluation of Clinical Nurse Transition Program at US Army Hospitals. *Military Medicine*, 08(0/0), 1-8. <https://doi.org/10.1093/milmed/usz108>
- Odland, L.-H., Sneltvedt, T., & Sörlie, V. (2014). Responsible but unprepared: Experiences of newly educated nurses in hospital care. *Nurse Education in Practice*, 14(5), 538-543. <https://doi.org/10.1016/j.nepr.2014.05.005>

- Polit, D. F., & Beck, C. T. (2017). *Nursing research : generating and assessing evidence for nursing practice* (10th ed.). Wolters Kluwer Health.
- 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 [journal article]. *Quality of Life Research*, 27(5), 1147-1157. <https://doi.org/10.1007/s11136-018-1798-3>
- Prion, S., Berman, A., Karshmer, J., Van, P., Wallace, J., & West, N. (2015). Preceptor and Self-Evaluation Competencies Among New RN Graduates. *Journal of Continuing Education in Nursing*, 46(7), 303-310. <https://doi.org/10.3928/00220124-20150619-01>; 20150619-15
- Schwirian, P. M. (1978). Evaluating the performance of nurses: a multidimensional approach. *Nursing Research*, 27(6), 347-351.
- 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>
- Squires, A., Sadarangani, T., & Jones, S. (2020). Strategies for overcoming language barriers in research. *Journal of Advanced Nursing*, 76(2), 706-714. <https://doi.org/10.1111/jan.14007>
- Streiner, D. L., Norman, G. R., & Cairney, H. (2015). *Health measurement scales: A practical guide to their development and use* (5 ed.). Oxford University Press.
- Takase, M. (2013). The relationship between the levels of nurses' competence and the length of their clinical experience: a tentative model for nursing competence development. *Journal of Clinical Nursing*, 22, 1400-1410. <https://doi.org/10.1111/j.1365-2702.2012.04239.x>
- Takase, M., Nakayoshi, Y., Yamamoto, M., Teraoka, S., & Imai, T. (2014). Competence development as perceived by degree and non-degree graduates in Japan: A longitudinal study. *Nurse Education Today*, 34, 451-456. <https://doi.org/10.1016/j.nedt.2013.04.017>
- Takase, M., & Teraoka, S. (2011). Development of the Holistic Nursing Competence Scale. *Nursing & Health Sciences*, 13, 396-403. <https://doi.org/10.1111/j.1442-2018.2011.00631.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. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study [journal article]. *Quality of Life Research*, 27(5), 1159-1170. <https://doi.org/10.1007/s11136-018-1829-0>
- Twigg, D. E., Kutzer, Y., Jacob, E., & Seaman, K. (2019). A quantitative systematic review of the association between nurse skill mix and nursing-sensitive patient outcomes in the acute care setting. *Journal of Advanced Nursing*, 75, 3404-3423. <https://doi.org/10.1111/jan.14194>
- Wangenstein, S., Johansson, I. S., Björkström, M. E., & Nordström, G. (2012). Newly Graduated Nurses' Perception of Competence and Possible Predictors: A Cross-sectional Survey. *Journal of Professional Nursing*, 28(3), 170-181. <https://doi.org/10.1016/j.profnurs.2011.11.014>
- Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., & Erikson, P. (2005). Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value in Health*, 8(2), 94-104. <https://doi.org/10.1111/j.1524-4733.2005.04054.x>
- Yanhua, C., & Watson, R. (2011). A review of clinical competence assessment in nursing. *Nurse Education Today*, 31, 832-836. <https://doi.org/10.1016/j.nedt.2011.05.003>
- Yu, J., & Kirk, M. (2008). Measurement of empathy in nursing research: systematic review. *Journal of Advanced Nursing*, 64(5), 440-454. <https://doi.org/10.1111/j.1365-2648.2008.04831.x>
- Yu, J., & Kirk, M. (2009). Evaluation of empathy measurement tools in nursing: systematic review. *Journal of Advanced Nursing*, 65(9), 1790-1806. <https://doi.org/10.1111/j.1365-2648.2009.05071.x>