Instruments evaluating the quality of the clinical learning environment in nursing education: A systematic review of psychometric properties

Instruments evaluating the quality of the clinical learning environment in nursing education: a systematic review of psychometric properties

Irene Mansutti
MNS, RN, Udine University, Italy

Luisa Saiani
MNS, RN, Professor in Nursing Science, Verona University, Italy

Luca Grassetti
PhD, Lecture in Statistics, Department of Economics and Statistics, University of Udine, Italy

Alvisa Palese
MNS, RN, Associate Professor in Nursing Science, Udine University, Italy

Corresponding Author

Alvisa Palese

Associate Professor in Nursing Science

Udine University

Viale Ungheria, 20

33100 Udine (Italy)

0039 (0)432 590926

alvisa.palese@uniud.it

Funding: None

Conflict of interest: None

World count: 6131

What is already known about the topic?

- Clinical learning effectiveness is affected by the environment in which nursing student placement takes place.
- Higher education institutions should systematically evaluate the quality of the clinical learning environments.
- To date, different instruments have been developed to evaluate nursing clinical environments but no systematic review has evaluated their psychometric properties and methodological quality.

What the paper adds?

- Eight instruments evaluating the clinical learning environments as perceived by nursing students have been evaluated for their psychometric properties.
- Not all relevant psychometric properties have been considered in the validation studies and often the methodological approaches used are poor or fair.
- Studies estimating psychometric properties, using increased quality of methodologies in the validation processes, are needed urgently.

Instruments evaluating the quality of the clinical learning environment in nursing education: a systematic review of psychometric properties

ABSTRACT

Background: The clinical learning environment is fundamental to nursing education paths, capable of affecting learning processes and outcomes. Several instruments have been developed in nursing education, aimed at evaluating the quality of the clinical learning environments; however, no systematic review of the psychometric properties and methodological quality of these studies has been performed to date.

Objectives: The aims of the study were: 1) to identify validated instruments evaluating the clinical learning environments in nursing education; 2) to evaluate critically the methodological quality of the psychometric property estimation used; and 3) to compare psychometric properties across the instruments available.

Design: A systematic review of the literature (using the Preferred Reporting Items for Systematic Reviews and Meta-Analysis guidelines) and an evaluation of the methodological quality of psychometric properties (using the COnsensus-based Standards for the selection of health Measurement Instruments guidelines).

Data Sources: The Medline and CINAHL databases were searched. Eligible studies were those that satisfied the following criteria: a) validation studies of instruments evaluating the quality of clinical learning environments; b) in nursing education; c) published in English or Italian; d) before April 2016.

Review methods: The included studies were evaluated for the methodological quality of the psychometric properties measured and then compared in terms of both the psychometric properties and the methodological quality of the processes used.

Results: The search strategy yielded a total of 26 studies and eight clinical learning environment evaluation instruments. A variety of psychometric properties have been estimated for each instrument, with differing qualities in the methodology used. Concept and construct validity were poorly assessed in terms of their significance and rarely judged by the target population (nursing students). Some properties were rarely considered (e.g., reliability, measurement error, criterion validity), whereas others were frequently estimated, but using different coefficients and statistical analyses (e.g., internal consistency, structural validity), thus rendering comparison across instruments difficult. Moreover, the methodological quality adopted in the property assessments was poor or fair in most studies, compromising the goodness of the psychometric values estimated.

Conclusions: Clinical learning placements represent the key strategies in educating the future nursing workforce: instruments evaluating the quality of the settings, as well as their capacity to promote significant learning, are strongly

recommended. Studies estimating psychometric properties, using an increased quality of research methodologies are needed in order to support nursing educators in the process of clinical placements accreditation and quality improvement.

Key words: Clinical learning environment; Learning Environment; Questionnaires; Nursing Student; Nursing Education; Validity and reliability; Systematic Review

1. Background

Becoming a nurse entails a complex educational path promoting several types of learning processes. Nursing students develop theoretical knowledge from lessons and seminars, and it is expected that this theoretical knowledge will be transformed into competences through clinical placement experiences, both at hospital and community levels (Flott and Linden, 2015). During clinical placement students are exposed to real-life situations and called upon to deal with real problems (Benner, 1984). Thus, clinical placements became opportunities to observe clinical nurses, to be exposed to role models, to reflect upon what is seen, heard, sensed or done; to understand personal attitudes and expected professional values, to develop cognitive, psychomotor and communication skills (Chan, 2001), critical thinking and diagnostic reasoning (Papathanasiou et al., 2014), and finally, to become an independent practitioner.

A recent concept analysis has defined the clinical learning environment as any area where nursing students apply theory to practice by conducting actual or simulated patient care to gain the skills, attitudes and decision-making abilities required to become a competent, entry-level nurse. The clinical learning environment includes physical space, psychosocial and interaction factors, the teaching effectiveness of the instructor, student engagement and organisational culture, all of which have an impact on students' capacity to achieve the desired learning outcomes (Flott and Linden, 2015).

Nursing students themselves perceive clinical placement as the most influential context in which they become a nurse (Chan, 2001). Experiencing a positive clinical learning environment increases learning outcomes as well as skill and knowledge acquisition (Flott and Linden, 2015; Henderson et al., 2009). In contrast, experiencing a negative clinical learning environment negatively affects the learning process, satisfaction and self-confidence (Flott and Linden, 2015; Levett-Jones and Lathlean, 2009).

Given its importance, higher educational institutions are recommended to assess clinical learning environments (Flott and Linden, 2015). However, to date only two reviews have been published on the instruments available for evaluating the quality of clinical learning environments. Hooven (2014) conducted an integrative review, analysing the instruments available and identifying the fundamental dimensions used in evaluating the clinical learning environment. Previously, Soemantri and colleagues (2010) performed a systematic literature review, aimed at identifying the tools used to measure the quality of educational environments and understand their practical suitability. Different types of environments were included, e.g. medical schools, college and university classrooms, surgical theatres, and clinical learning environments. Moreover, authors reviewed the available instruments for all health-care students by summarising content validity, criterion validity, construct validity and reliability.

Therefore, no systematic review has been performed to date that specifically focuses on instruments evaluating the quality of nursing clinical learning environments, and no study has assessed and compared the psychometric properties estimated for the instruments available. Thus, the general purpose of this study was to summarise and critically evaluate the instruments that assess the quality of clinical learning environments in nursing education.

2. Aims

In the field of clinical learning environment quality assessment, the aims of the study were: 1) to identify the instruments undergoing validation processes; 2) to evaluate critically the quality of the methods used in ascertaining psychometric properties; and 3) to compare the estimated psychometric properties of the instruments available.

3. Study design and process

A systematic review of the literature was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines (Moher et al., 2009). The included studies were evaluated with respect to their methodological quality using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN, Mokkink et al., 2010), an instrument aimed at evaluating the methodological quality of a validation study by assessing the properties estimated against established standards. Finally, different clinical learning environment instruments were compared, considering both the goodness of the psychometric properties estimated and the quality of the methods used when assessing these properties.

3.1 Search strategy

The search strategy was applied to Medline and CINAHL databases by combining the following MeSH terms: “Clinical Learning Environment” AND “Perception” OR: “Education, Nursing, Baccalaureate”, “Students”, “Students, Nursing”, “Personal Satisfaction”, “Survey and Questionnaires”, “Psychometrics”, “Factor Analysis, Statistical”. For Medline, “Clinical Learning Environment” was replaced with two keywords: “Learning Environment” AND “Educational Environment” in accordance with the MeSH database dictionary definitions.

Eligible studies were those that satisfied the following criteria: a) validation studies of instruments evaluating the quality of clinical learning environments; b) pertaining to nursing education; c) published in English or Italian; d) before April 2016. No limitation for time was introduced while studies were excluded if they a) did not provide instrument data

on validation processes (e.g., investigating students' perceptions), b) involved students enrolled in healthcare programmes other than nursing (e.g., medical students) without differentiating data on nursing students, and/or c) measured different educational settings (e.g., classrooms).

One researcher (IM) conducted the literature search and two researchers (IM, AP) worked independently to evaluate study eligibility on the basis of the title and contents of each abstract retrieved. Any difference was discussed with a third researcher (LS). Then, the full text of those studies eligible were retrieved. Two researchers (IM, AP) independently evaluated the eligibility of each study by reading the full text carefully; decisions on article inclusion were based upon joint agreement. The reference list of the studies included were also evaluated aiming at retrieving new studies. In addition, studies included were also matched with those referenced in the available reviews (Hooven 2014; Soemantri et al., 2010). Figure 1 shows the process of study inclusion.

3.2 Data extraction

Data extraction was performed by two researchers (IM, AP) considering: author; year of publication; country where the study was performed; year of data collection; study design; sample characteristics; setting (e.g., hospital); instrument validated, any tool modification when a re-validation process of an original version of an instrument was performed; number and conceptual definitions of factors emerged; number of items included in the tool and metrics used. The estimated values of the psychometric properties were then extracted.

Researchers worked independently and then compared the extracted data. In cases where an included study had been conducted by one of the researchers, in order to avoid bias an independent researcher (LG) was involved in evaluating the studies. Any differences were discussed and agreement among researchers was achieved.

3.3 Methodological quality evaluation

The quality of an instrument is based on its estimated psychometric properties and on how these properties have been investigated. In order to be valid, studies evaluating instrument measurement properties should be grounded in high standards of methodological quality (Mokkink et al., 2010).

The COSMIN tool enables the evaluation of both the psychometric properties and the research methods used through different dimensions categorised into boxes named in accordance with the property under evaluation: internal consistency, reliability (including test-retest reliability, inter-rater reliability and intra-rater reliability), measurement error, content validity (including face validity), structural validity, hypotheses testing (including convergent validity), criterion validity,

and cross-cultural validity. In addition, COSMIN procedures also require the evaluation of responsiveness, interpretability and generalisability of the findings. Each box includes a pool of items (from five to 18 based on which property is considered) scored on a 4-point scale (1 poor, 2 fair, 3 good, 4 excellent). The overall score of a given estimated psychometric property is obtained by taking the lowest score indicated by the items in the box (Mokkink et al., 2010). Therefore, a final score based on a 4-point scale is given for each psychometric property, ranging from "poor" to "excellent".

Two authors (IM, AP) underwent training on the COSMIN tool and then critically applied the procedures to the included studies. They worked independently and then compared their evaluations. In the case of studies conducted by the same researchers (AP, LS) an independent researcher was involved (LG) with the aim of avoiding any bias.

Author(s) of the included studies were also involved in the process of evaluation on the basis of the following considerations: a) in several nursing journals the space allowed for reporting the findings emerged from validation studies is typically limited, thus some data may be missed; b) in some cases, relevant data may have been published at the country level and in a different language and this may have preceded publication at an international level; c) thus, researchers may have avoided self-plagiarism by not publishing at international levels data already published in national circuits; and d) the COSMIN tool has been published recently (Mokkink et al., 2010) and authors may have performed analyses that are not analytically reported in their publications. Therefore, the authors were contacted and asked to cooperate with the researchers. They were then sent the COSMIN guidelines and the grid with the evaluations emerged from their studies. They were required to consider the evaluations obtained by researchers and to express their agreement. In the case of disagreement, they were asked to send data/evidence for their disagreement, and researchers debated with authors via email until an agreement was reached. A total of 14 authors responded to the email request; the remaining six were contacted three times leaving at least two weeks from one attempt to the next.

Finally, researchers compared the available instruments evaluating the quality of the clinical learning environment, their psychometric properties and their research methodological quality.

4. Results

A total of 27 articles were included, reporting the validation processes and findings of eight clinical learning environment instruments (Table 1). A total of 26 articles were considered, given that one author [Chan] published two articles (2001, 2003) reporting equal data regarding participants, methods and values of psychometric properties; therefore, we considered this to be one study.

4.1 Clinical learning environment instruments

The Clinical Learning Environment scale (CLE scale) by Dunn and Burnett (1995) was the first instrument developed on the basis of Bloom's (1964) and Orton's (1981) theories. Considering the ward learning climate survey performed by Orton (1983), authors modified the 124 items in accordance with the cultural and professional changes occurred in health settings since the 1980s. Experts were involved in item evaluation while through factor analysis, authors obtained an instrument composed of 23 items categorised into five factors as reported in Table 1.

The Clinical Learning Environment Inventory (CLEI) was developed by Chan (2001) based upon Knowles's (1990) and Moos's (1974) theories. The CLEI was developed through an in-depth literature review on classroom and other educational learning environments, as well as on the basis of the College and University Classroom Environment Inventory (Fraser et al., 1986). Semi-structured interviews with 21 randomly selected 2nd-year students were also performed, obtaining qualitative data on perceptions of hospital learning environments. The concept of clinical learning environment was then discussed with experts in nursing education. The final instrument was developed into two formats: the first contained 35 items divided into five factors (personalisation, student involvement, task orientation, innovation, and individualisation); the second was an integral form containing 42 items with one additional factor: satisfaction, as reported in Table 1. After about 10 years, Newton et al. (2010) confirmed the structure by re-validating the tool with a more consistent sample size. A short version of the instrument composed of 19 items on a 5-point Likert scale (CLEI-19), assessing only two domains (satisfaction and personalisation), was also validated (Salamonson et al., 2011) as reported in Table 1.

The Clinical Learning Environment and Supervision scale (CLES) was then developed by Saarikoski et al. (2002a) taking the theories of Quinn (1995), Wilson-Barnett et al. (1995), and Moss and Rowles (1997) into account. From a literature review focused on clinical learning environments and the supervisory relationship (Saarikoski et al., 2002a, 2002b), authors categorised and summarised those items capable of reflecting the construct which was then tested in a pilot study. Subsequently, the number and scope of items were changed and reviewed by a panel of expert clinical teachers (Saarikoski et al., 2005). The final version of the CLES scale consists of 27 items and five factors, as reported in Table 1. The CLES instrument was translated and validated in several countries: Belgium (De Witte et al., 2011), Cyprus (Papastavrou et al., 2010), and Italy (Burrai et al., 2012; Tomietto et al., 2009) and also through an international comparative validation study (Finland and the UK; Saarikoski et al., 2002b).

Over the following years, the Clinical Learning Environment, Supervision and Nurse Teacher scale (CLES+T) was developed on the basis of the revised version of the above-mentioned CLES, including an additional sub-dimension aimed at evaluating the quality of the nurse-teachers' cooperation with clinical practice (Saarikoski et al., 2008). The CLES+T

scale consists of 34 items and five factors (Table 1) and is currently the most translated and validated instrument across countries, specifically in Cyprus (Papastavrou et al., 2015), Germany (Bergjan et al., 2012), Italy (Tomietto et al., 2012), New Zealand (Watson et al., 2014), Norway (Henriksen et al., 2012), Spain (Vizcaya-Moreno et al., 2015) and Sweden (Gustafsson et al., 2015; Johansson et al., 2010). The instrument has been validated also in primary healthcare settings (Bos et al., 2012) and was used in the explorative comparative validation study involving Belgium, Cyprus, Finland, Italy, Spain, Sweden, The Netherlands, and the UK (Warne et al., 2010).

The Clinical Learning Environment Diagnostic Inventory (CLEDI) was developed by Hosoda (2006), on the basis of the following theories: Dewey (1933), Kolb and Fry (1975), Schön (1983), Oliver and Endersby (1994) and Dunn and Burnett (1995). Hosoda (2006) designed a hypothetical clinical learning environment model capable of linking five different environmental components: affective, perceptual, symbolic, behavioural and reflective. On the basis of these components, through semi-structured interviews involving students and preceptors, the author initially identified 96 items which were subsequently reduced to 35 through pilot-test and factor analysis, as reported in Table 1.

Over the last 10 years, Sand-Jecklin (2009) has adopted the Cognitive Apprenticeship theory (Brown et al., 1989) as a conceptual framework for developing the Student Evaluation of Clinical Education Environment tool (SECEE). A literature review and the data emerged from faculty and student focus groups, were used and factors impacting on student learning in the clinical environments were identified. The SECEE final version is composed of 32 items and three factors, as reported in Table 1.

More recently, the Clinical Learning Environment instrument (CLE) was developed and validated by Chuan and Barnett (2012). Considering the clinical learning environment as an interactive network of forces influencing student learning outcomes, the authors performed a literature review identifying six characteristics of the clinical learning environment considered important in the Malaysian context: ward atmosphere, supervision by staff nurses, the clinical teachers, student satisfaction, the theory-practice gap and peer support. These were identified as the factors of the instrument, which is composed of 44 items, including those available in previous instruments (e.g., Chan, 2002; Dunn and Burnett, 1995; Hosoda, 2006; Saarikoski et al., 2002a; Sand-Jecklin, 2000).

The most recent instrument validated is the Modified Clinical Learning Environment, Supervision and Nurse-Teacher scale (Modified CLES+T) based on Saarikoski et al. (2008) and the CLE scale (Dunn and Burnett, 1995). The modified CLES+T contains a total of 57 items and is composed of 11 factors as reported in Table 1. The instrument was pre-tested in a pilot-study and validated using Structural Equation Modeling (D'Souza et al., 2015).

4.2 Population and settings involved in the validation processes

All studies involved nursing students, the majority of whom were female (from 61.3% in Papastavrou et al., 2015, to 99% in Bos et al., 2012). The average age of students was between 20.4 (Papastavrou et al., 2010) and 30.3 years (Salamonson et al., 2011). The research involved students who were in their 2nd-year of study (Burrai et al., 2012; Chan 2001, 2003), their 3rd-year (Chuan and Barnett, 2012; Watson et al., 2014), or were students from both years of study (Newton et al., 2010; Saarikoski et al., 2002a, 2002b, 2005; Tomietto et al. 2009). As reported in Table 1, ten studies involved students from all years of nursing programmes and no authors involved only 1st-year students. Response rate ranged from 41% (Watson et al., 2014) to 100% when students were volunteers (Chuan and Barnett, 2012; D'Souza et al., 2015), but 14 studies did not report data regarding response rates.

In most cases the validation was performed at the hospital level and only one study validated the instrument in a primary healthcare setting (Bos et al., 2012). Four studies considered hospitals and long-term care settings (e.g., nursing homes or private institutions for aged care: Henriksen et al., 2012; Hosoda, 2006; Saarikoski et al., 2002a; Watson et al., 2014), whereas a further four studies did not specify the setting (De Witte et al., 2011; Dunn and Burnett, 1995; Salamonson et al., 2011; Sand-Jecklin, 2009). In addition to students, three studies involved staff nurses and preceptors or university tutors (Chuan and Barnett, 2012; Dunn and Burnett, 1995; Hosoda, 2006). In Hosoda's study (2006), students and preceptors were involved in order to detect hypothesised differences between scores, whereas Chuan and Barnett (2012) and Dunn and Burnett (1995) failed to report the rationale for this sampling decision.

4.3 Methodological quality evaluation and comparison of the psychometric properties

As reported in Table 2, not all of the studies have estimated all of the psychometric properties prescribed by the COSMIN guidelines: internal consistency and structural validity were mostly evaluated while measurement error, convergent and criterion validity was evaluated in only a few studies. Moreover, in the majority of studies the quality of the methodologies used in evaluating the psychometric properties ranged from poor to fair.

4.3.1 Content Validity

As reported in Table 2, the content validity (including face validity) that is considered to be the first step in validating instruments — given that it identifies the degree to which the tool measures the construct that it is intended to measure (Mokkink et al., 2010) — was estimated by 10 studies using a poor quality methodological approach. Students were not involved in this process, with the exception of Watson et al. (2014) who involved some nursing students, clinical supervisors, managers and nurse teachers.

4.3.2 Internal consistency

Internal consistency, the psychometric property evaluating the interrelatedness among items, was estimated in 22 studies; however, as reported in Table 2, in 16 of these a poor or fair methodological quality emerged, mostly because missing items were not specified in their handling while there were insufficient sample sizes. In the remaining studies the methodological quality ranged from good (Saarikoski et al., 2002a, 2005; Salamonson et al., 2011; Tomietto et al., 2012) to excellent (Newton et al., 2010; Watson et al., 2014). Specifically, CLES+T reported the highest Cronbach's α of 0.82-0.93 with excellent quality in Watson et al. (2014) and 0.95 (0.80-0.96) with good quality in Tomietto et al. (2012). CLES and SECEE have demonstrated similar internal consistency coefficients but with poor methodological quality (De Witte et al., 2011; Saarikoski et al., 2002a).

4.3.3 Reliability

Taking reliability as the proportion of the total variance in the data that is due to true differences among learning environments as well as the extent to which scores are the same for repeated measurements (Mokkink et al., 2010), only three studies (Hosoda, 2006; Gustafsson et al., 2015; Tomietto et al., 2009) performed a test-retest evaluation as reported in Table 2, with poor or fair methodological quality.

4.3.4 Measurement error

Measurement error, which is the systematic and random error of a respondent score not attributed to true changes in the construct under measurement (Mokkink et al., 2010), was reported only by Gustafsson et al. (2015) in validating the CLES+T, as reported in Table 2.

4.3.5 Structural validity

The majority of studies (21 out of 26) assessed the structural validity as required by the COSMIN procedures. More precisely, the structural validity is the degree to which the scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured (Mokkink et al., 2010). The structural validity findings, when reported, were concordant with the construct (dimensions) of the instrument, but the methodological quality that emerged was poor or fair in 14 of 21 studies due to insufficient sample size, no explanation with regard to the treatment of the missing items and lack of precision in reporting the performed analysis.

The highest explained variance estimated by studies included in this systematic review was 76.9% (Burrai et al., 2012) and 71.2% (De Witte et al., 2011) for the CLES tool, and 72.8% (Bergjan et al., 2013) for the CLES+T tool. Moreover, when considering only the original versions of tools, the majority reported good explained variance (from 50% to 60%)

but the methodological quality was poor or fair; in fact, only six studies estimated this psychometric property by adopting good methodology: Newton et al. (2010) reported an explained variance of 51% in validating the CLEI; Salamonson et al. (2011) reported an explained variance of 63.3% in the CLEI-19; Saarikoski et al. (2002a, 2005) reported 64% for the CLES instrument, while Tomietto et al. (2012) and Papastavrou et al. (2015) reported 67.2% and 67.4% respectively, in validating the CLES+T. Finally, only one study estimated the structural validity using an excellent quality methodological approach, achieving 58.2% of variance in validating the CLES+T (Watson et al., 2014).

As reported in Table 2, some authors (e.g., Hosoda, 2006; Papastavrou et al., 2010; 2015; Saarikoski et al., 2002a) used exploratory factor analysis (EFA); others (e.g., Newton et al., 2010; Salamonson et al., 2011) used Principal Component Analysis (PCA) or both (e.g., Newton et al., 2010; Salamonson et al., 2011), whilst still others (e.g., Bos et al., 2012; Tomietto et al., 2012; Vizcaya-Moreno et al., 2015) used Confirmatory Factor Analysis in addition to EFA and PCA. More recently, D'Souza et al. (2015) used Structural Equation Modeling.

Moreover, Hosoda (2006) considered only students' questionnaires when performing the EFA, whereas students and staff nurses' data were considered together by Dunn and Burnett (1995). Differently, Chuan and Barnett (2012) did not specify whether they had considered the data collected from students and from educators differently. Thus, findings regarding the structural validity estimations are not comparable given the differences in methodological evaluation analyses performed and in the quality of the methods used.

4.3.6 Hypotheses testing

Hypotheses testing assessment as expected mean differences between groups or as expected correlations between instrument scores and other variables, such as the scores of other instruments (Mokkink et al., 2010) were estimated in eight studies out of 26 with poor or fair methodological quality (Table 2). According to the findings, three levels of variables were considered in the hypotheses testing to date:

- Individual variables (CLEI-19): as differences between worker and non-worker students (Salamonson et al., 2011);
- Educational variables (CLEDI; CLEI-19; CLE instrument; CLES; CLES+T; SECEE): as differences with regard to the academic year attended, placement duration, types of shifts, types of supervisory relationships, number of briefing and debriefing meetings with the nurse teacher, and differences between students' and clinical tutors' scores (Chuan and Barnett, 2012; Hosoda, 2006; Papastavrou et al., 2010; Saarikoski et al., 2002b; Salamonson et al., 2011; Sand-Jecklin, 2009; Warne et al., 2010);

- Macro-variables (CLES; CLES+T): as differences in the perceptions of students in different European countries, types of higher educational institutions (university colleges vs. polytechnics), or in higher educational institutions established for more or less than 20 years — thus with a different experience in nursing education (Saarikoski et al., 2002b; Warne et al., 2010).

4.3.7 Convergent validity

Convergent validity, defined by the COSMIN tool as hypotheses testing measured with regard to the expected relations with other instruments (Mokkink et al., 2010), was estimated only by Chan (2001, 2003) who did not specify which comparative instruments were considered and therefore had poor methodology quality (Table 2). Moreover, the correlation values that emerged were poor, from 0.39-0.45 (Chan 2001, 2003).

4.3.8 Criterion validity

Criterion validity, a comparison of the tool under validation with an acknowledged gold standard instrument (Mokkink et al., 2010) was estimated in only two studies reporting good correlations, 0.93 between CLES and the CLE scale (Saarikoski et al., 2005) and 0.76 between CLEDI and CLES (Hosoda, 2006), both applying fair methodological quality (Table 2).

4.3.9 Cross-cultural validity

Although 14 translated instruments were used, only seven (Bergjan et al., 2013; De Witte et al., 2011; Henriksen et al., 2012; Johansson et al., 2010; Tomietto et al., 2009, 2012; Vizcaya-Moreno et al., 2015) assessed cross-cultural validity by adopting methodological quality from poor to fair (Table 2). In all studies, tools were forward- backward- translated only once, and only De Witte et al. (2011), Henriksen et al. (2012) and Vizcaya-Moreno et al. (2015) performed a pre-test with the translated instrument.

5. Discussion

5.1 Clinical learning environment instruments

To the best of our knowledge this is the first psychometric systematic review of instruments evaluating clinical learning environment quality in nursing education. In our systematic review, a total of 26 studies emerged that estimated the reliability and validity of eight instruments in 16 different countries, mainly across Europe.

The first instrument underwent the validation process with data collected in 1993 (Dunn and Burnett, 1995) whereas the latest was based on data collected from 2011 to 2012 (Vizcaya-Moreno et al., 2015), indicating that this research field spans over 20 years, a period during which there has been a tremendous amount of change in nursing programmes, hospital environments and student profiles (Anderson, 2010).

Two different strategies of tool development, thus a first and second generation instrument, can be identified. The first were conceptually-based (CLE, CLEDI, CLES, CLES+T, SECEE) and developed from prominent learning theories mainly established in the 1980s and 1990s. The second-generation instruments were developed from previously well-established instruments in clinical environments (e.g., modified CLES+T based on CLE and CLES+T, D'Souza et al., 2015) or in other learning environments (e.g., CLEI based on the University Classroom Environment Inventory). In addition, assessing the validity and reliability of well-established instruments in different countries, as occurred for the CLES+T scale which was validated in >10 countries (Bos et al., 2012; Warne et al., 2010) has emerged as a trend in recent years, thus developing an international framework capable of accumulating evidence on instrument validity and of comparing data.

The instruments emerged are composed of two (Salamonson et al., 2011) to eleven factors (D'Souza et al., 2015) and from 19 (Salamonson et al., 2011) to 57 items (D'Souza et al., 2015). Some factors are similar across instruments, such as 'Supervisory relationship' and 'Ward atmosphere', whereas the 'Hierarchy/ritual' factor has appeared only in the recently modified CLES+T, thus reflecting cultural commonalities and differences in healthcare settings that may affect the perceptions of students (D'Souza et al., 2015).

The shortest instrument emerged is the CLEI-19 (Salamonson et al., 2011) composed of two factors ('Satisfaction' and 'Personalisation') including 19 items, whilst the modified CLES+T is the most complex, composed of 57 items and 11 factors (D'Souza et al., 2015). In general, instruments have increased the number of factors and items over the years, possibly due to the increased complexity of the clinical learning environments (Palese et al., 2016).

Furthermore, homogeneity has emerged in the metrics: the majority have used a 5-point Likert scale to express the evaluation from strongly disagree to strongly agree (e.g., Hosoda, 2006; D'Souza et al., 2015; Dunn and Burnett, 1995) and from totally disagree to totally agree (De Witte et al., 2011). However, Likert scales with a mid-point may introduce a central tendency bias in that participants may avoid extreme response categories. Only Chan (2001, 2003) and Newton et al. (2010) used a 4-point Likert scale, whereas Burrai and colleagues (2012) used a 6-point Likert scale. In addition,

agree/disagree Likert scales may introduce an acquiescence bias (participants may agree with statements as presented), social desirability bias, and lack of reproducibility (Jamieson et al., 2004; Nadler et al., 2015).

5.2 Population and settings

The studies involved from 42 (Gustafsson et al., 2015) to 1,903 (Warne et al., 2010) students; participants were recruited from a single nursing programme (e.g., Hosoda, 2006) or different programmes located in different countries (eight in Warne et al., 2010). The largest study involved 2,768 participants (Sand-Jecklin et al., 2009) but the amount of the sample composed exclusively of nursing students was not declared.

Participants were mainly female and this may have introduced a gender bias that should be addressed in the future as recent changes documented in several countries show an increased proportion of males among nursing students (Loughrey et al., 2008). The majority of students were in the 2nd or 3rd-year of their programme and no studies involved 4th-year students who have an intense experience in clinical practice, as occurs in Spain (Zabalegui and Cabrera, 2009). There is a need for future research to include entire cohorts of students, who may have different expectations and perceptions, and also nursing programmes based on 4 years of education.

The sampling method used in the studies was not always reported and the response rate was varied, from 41.6% to 100% when students were volunteers. Although greater accuracy in the sampling methods are suggested, the low response rates may reflect dissatisfaction among students and the lack of desire to participate, due to fear of the consequences (e.g., impact on the clinical competences evaluation). Given that this may affect the perceptions, future studies should specify also when students completed the instrument, before or after their clinical competence evaluation.

In addition, current instruments have mostly been subjected to validation processes in public hospitals, in specific wards, such as medical units and surgery. With the transition in the focus of nursing education from hospitals to communities and primary health care settings, more emphasis should be given to validating instruments that are capable of measuring clinical learning environments across different settings with different missions (private/public, academic or not) and different patient profiles (e.g., Accident and Emergency department vs. nursing homes).

5.3 Methodological quality evaluation and comparison of the psychometric properties

A varying number of psychometric properties have been estimated in the included studies, from one to six. Furthermore, the methodological quality of these estimations was heterogeneous, with the majority from poor to fair.

Therefore, limited comparison is possible across the estimated properties of the available instruments, threatening the identification of the most reliable and valid tool in evaluating clinical learning environments.

With regard to content validity, concepts and constructs were rarely assessed for their significance (e.g., only two studies calculated the Content Validity Index: De Witte et al., 2011; D'Souza et al., 2015) and were rarely judged for their relevance and comprehensiveness of the target population (Mokkink et al., 2010). Nursing students were not involved in the majority of the studies, thus resulting in a fundamental flaw. In the process of tool development all authors took account of expert opinions (e.g., nurse educators), thus failing to consider that the learning clinical experience is subjective and that it is important to elicit elements that influence the quality of the experience as perceived by students. This gap should be addressed in future studies.

A few studies estimated reliability, although test-retest procedures may be easier with nursing students given their availability. However, the duration of the clinical rotations, as well as their frequency, may have threatened the potential for undertaking a second evaluation for the same unit after one or two weeks when students have already moved on to their next learning experience. Furthermore, measurement error was estimated only in one study (Gustafsson et al., 2015): as a consequence, comparisons of reliability and measurement error across different instruments are limited.

Internal consistency and structural validity have been estimated for the majority of the tools, but with different quality of methodological approaches, compromising also in this case comparisons across instruments. Structural validity was evaluated using different statistical analyses. Specifically, Bergjan et al. (2013), Burrai et al. (2012) and De Witte et al. (2011) obtained the highest proportion of explained variance when validating the CLES and the CLES+T. Nevertheless, they all used modified instruments, changing, removing or adding some items and using a different Likert scale, thus threatening the ability to make comparisons with the original tools. Moreover, their structural validity values were also affected by the poor methodological quality adopted.

Finally, convergent and criterion validity have rarely been assessed. Whereas in the case of the first generation of instruments (e.g., CLE scale: Dunn and Burnett, 1995) the lack of available knowledge in the field possibly threatened comparison with gold standards, since valid and reliable tools have been documented, an increased tendency to evaluate convergent and criterion validity is expected. Without criterion validity evaluation, it is not certain that instruments evaluate the same constructs used by other tools (McDowell, 2006).

5.4 Limitations

Several limitations affect this systematic review. Aiming to develop a focused search strategy, only two databases were searched (MedLine, CINAHL) in accordance with their relevance to nursing literature; only those MeSH terms

accepted in the database dictionary of the above-mentioned databases were considered, thus relevant text words such as "scale", "tool", "measurement" were not considered; in addition, only studies published in English or Italian were included. Moreover, the Boolean operators OR/AND were not used within each element of the Population/Intervention and Outcome elements. Thus, other instruments may have been developed and circulated as grey literature, as well as in different languages, therefore introducing a potential publication bias.

Second, the assessment of the studies was based on COSMIN guidelines (2010) that were developed for health status measures and not specifically for nursing education instruments. In addition, the guidelines have only recently been established — when the majority of the tools were validated — thus, in reporting their findings, authors may not have been supported by the methodological quality recommendations included in these guidelines. However, multiple contacts with authors aiming to collect unpublished data with regard to some properties, have been performed.

Moreover, the COSMIN guidelines apply the "worst score counts" method (Mokkink et al., 2010), thus, instead of an average evaluation of the trends, it emphasises problems in the measurement of psychometric properties. In addition, responsiveness, or the ability of an instrument to detect change in the measured construct over time (as required by the COSMIN procedure), was not evaluated in this review due to the absence of longitudinal studies among those included.

6. Conclusions

Eight instruments evaluating the quality of clinical learning environments in nursing education have been exposed to a validation process to date. First-generation instruments have been developed from different learning theories, whereas second-generation instruments have been developed from the first-generation, mixing, revising, and integrating different instruments already validated. In the studies included in this review, not all relevant psychometric properties have been estimated and often the methodological approaches used are poor or fair. In addition, a lack of homogeneity in reporting participants and setting data, with a large amount of missing data within the studies, has emerged thus threatening the external validity of the instruments.

There is a need to address future research in the field by completing the processes of validation undertaken to date for the available instruments; by using higher quality of methods. New instruments developed should also estimate all psychometric properties with increasing quality of the methodologies. A minimum data set regarding students (e.g., duration of the clinical placement, tutorial models — for example, one-to-one or peer education with other students), their status (supernumerary or not, paid or not, alone or with other students), and settings (private, public, hospital, community units), is also strongly recommended in future studies, aiming to increase the external validity of the findings.

References

Anderson B. A perspective on changing dynamics in nursing over the past 20 years. Br J Nurs. 2010 Oct 14-27;19(18):1190-1.

Benner P. (2003). From novice to expert: Excellence and power in clinical nursing practice. McGraw-Hill: Milan.

Bergjan M, Hertel F. Evaluating students' perception of their clinical placements - testing the clinical learning environment and supervision and nurse teacher scale (CLES+T scale) in Germany. Nurse Educ Today. 2013 Nov;33(11):1393-8.

Bloom BS, 1964. Stability and Change m Human Characteristics. John Wiley & Sons, New York.

Bos E, Alinaghizadeh H, Saarikoski M, Kaila P. Validating the 'clinical learning environment, supervision and nurse teacher' CLES+T instrument in primary healthcare settings using confirmatory factor analysis. J Clin Nurs. 2012 Jun;21(11-12):1785-8.

Brown JS, Collins A, Duguid P. Situated cognition and the culture of learning. Educational Researcher. 1989;18:32-42.

Burrai F, Cenerelli D, Sebastiani S, Arcoleo F. Reliability analysis and structure factorial exploration of Clinical Learning Environment of Supervision (CLES). SCENARIO. Dec 2012; 29(4):41-47.

Chan D. Combining qualitative and quantitative methods in assessing hospital learning environments. Int J Nurs Stud. 2001 Aug;38(4):447-59.

Chan D. Validation of the Clinical Learning Environment Inventory. West J Nurs Res. 2003 Aug;25(5):519-32.

Chuan OL, Barnett T. Student, tutor and staff nurse perceptions of the clinical learning environment. Nurse Educ Pract. 2012 Jul;12(4):192-7.

Dewey J, 1933. How We Think. D.C. Heath and Company, Lexington, MA.

De Witte N, Labeau S, De Keyzer W. The clinical learning environment and supervision instrument (CLES): validity and reliability of the Dutch version (CLES+NL). Int J Nurs Stud. 2011 May;48(5):568-72.

D'Souza MS, Karkada SN, Parahoo K, Venkatesaperumal R. Perception of and satisfaction with the clinical learning environment among nursing students. Nurse Educ Today. 2015 Jun;35(6):833-40.

Dunn SV, Burnett P. The development of a clinical learning environment scale. J Adv Nurs. 1995 Dec;22(6):1166-73.

Flott EA, Linden L. The clinical learning environment in nursing education: a concept analysis. J Adv Nurs. 2016 Mar;72(3):501-13.

Fraser BJ, Treagust DF, Dennis NC. Development of an instrument for assessing classroom psychosocial environment at universities and college. Studies in Higher Education. 1986;11(1):43-54.

Gustafsson M, Blomberg K, Holmefur M. Test-retest reliability of the Clinical Learning Environment, Supervision and Nurse Teacher (CLES + T) scale. Nurse Educ Pract. 2015;15(4):253-7.

Henderson A, Briggs J, Schoonbeek S, Paterson K. A framework to develop a clinical learning culture in health facilities: ideas from the literature. In Nurs Rev. 2011;58:196-202.

Henriksen N, Normann HK, Skaalvik MW. Development and testing of the Norwegian version of the Clinical Learning Environment, Supervision and Nurse Teacher (CLES+T) evaluation scale. Int J Nurs Educ Scholarsh. 2012 Sep;18;9.

Hooven K. Evaluation of instruments developed to measure the clinical learning environment: an integrative review. Nurse Educ. 2014 Nov-Dec;39(6):316-20.

Hosoda Y. Development and testing of a Clinical Learning Environment Diagnostic Inventory for baccalaureate nursing students. J Adv Nurs. 2006 Dec;56(5):480-90.

Jamieson S. Likert Scales: how to (Ab)Use them. Med Ed. 2004;38:1217-1218.

Johansson UB, Kaila P, Ahlner-Elmqvist M, Leksell J, Isoaho H, Saarikoski M. Clinical learning environment, supervision and nurse teacher evaluation scale: psychometric evaluation of the Swedish version. J Adv Nurs. 2010 Sep;66(9):2085-93.

Knowles M, 1990. The Adult Learner: A Neglected Species. 4th Edition. Gulf Publishing, Houston.

Kolb DA, Fry R, 1975. Towards an applied theory of experiential learning. In Theories of Group Processes (Cooper C.L., ed.), John Wiley & Sons, London, pp. 33–57.

Levett-Jones T, Lathlean J. The ascent to competence conceptual framework: an outcome of a study of belongingness. J Clin Nurs. 2009 Oct;18(20):2870-9.

Loughrey M. Just how male are male nurses..? J Clin Nurs. 2008;17:1327-1334.

McDowell I, 2006. Measuring health. A Guide to Rating Scales and Questionnaires. 3rd ed. Oxford University Press, New York.

Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009 Jul 21;339:b2535.

Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, de Vet HC. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010 May;19(4):539-49.

Moos RH, Trickett EJ, 1974. Classroom Environment Scale Manual, 1st Edition. Consulting Psychologists Press, Palo Alto, CA.

Moss R, Rowles CJ. Staff nurse job satisfaction and management style. Nurs Manage. 1997 Jan;28(1):32-4.

Nadler JT, Weston R, Voyles EC. Stuck in the middle: the use and interpretation of mid-points in items on questionnaires. J Gen Psychol. 2015;142(2):71-89.

Newton JM, Jolly BC, Ockerby CM, Cross WM. Clinical learning environment inventory: factor analysis. J Adv Nurs. 2010 Jun;66(6):1371-81.

Oliver R, Endersby C, 1994. Teaching and Assessing Nurses: A Handbook for Preceptors. Bailliere Tindall, London.

Orton HD, 1981. Ward Learning Climate. Royal College of Nursing, London.

Orton HD, 1983. Ward learning climate and student nurse response. In: Davis, B.D. (Ed.), Research into Nurse Education. Croom Helm, London.

Palese A, Destrebecq A, Terzoni S, Grassetti L, Altini P, Bevilacqua A, Brugnolli A, Benaglio C, Dalponte A, De Biasio L, Dimonte V, Gambacorti B, Fasci A, Grosso S, Mansutti I, Mantovan F, Marognolli O, Montalti S, Nicotera R, Perli S, Randon G, Stampfl B, Tollini M, Canzan F, Zannini L, Saiani L. [Validation of the Italian Clinical Learning Environment Instrument (SVIAT). Assist Inferm Ric. 2016 Jan-Mar;35(1):29-35.

Papathanasiou IV, Tsaras K, Sarafis P. Views and perceptions of nursing students on their clinical learning environment: teaching and learning. Nurse Educ Today. 2014 Jan;34(1):57-60.

Papastavrou E, Dimitriadou M, Tsangari H. Psychometric Testing of the Greek Version of the Clinical Learning Environment-Teacher (CLES+T). Glob J Health Sci. 2015 Sep 1;8(5):49573.

Papastavrou E, Lambrinou E, Tsangari H, Saarikoski M, Leino-Kilpi H. Student nurses experience of learning in the clinical environment. Nurse Educ Pract. 2010 May;10(3):176-82.

Quinn FM, 1995. The Principles and Practice of Nurse Education, 3rd Edition. Chapman and Hall, London.

Saarikoski M, Isoaho H, Leino-Kilpi H, Warne T. Validation of the clinical learning environment and supervision scale. Int J Nurs Educ Scholarsh. 2005;2:Article 9.

Saarikoski M, Isoaho H, Warne T, Leino-Kilpi H. The nurse teacher in clinical practice: developing the new sub-dimension to the Clinical Learning Environment and Supervision (CLES) Scale. Int J Nurs Stud. 2008 Aug;45(8):1233-7.

Saarikoski M, Leino-Kilpi H, Warne T. Clinical learning environment and supervision: testing a research instrument in an international comparative study. Nurse Educ Today. 2002 May;22(4):340-9. (b)

Saarikoski M, Leino-Kilpi H. The clinical learning environment and supervision by staff nurses: developing the instrument. Int J Nurs Stud. 2002 Mar;39(3):259-67. (a)

Salamonson Y, Bourgeois S, Everett B, Weaver R, Peters K, Jackson D. Psychometric testing of the abbreviated Clinical Learning Environment Inventory (CLEI-19). J Adv Nurs. 2011 Dec;67(12):2668-76.

Sand-Jecklin K. Assessing nursing student perceptions of the clinical learning environment: refinement and testing of the SECEE inventory. J Nurs Meas. 2009;17(3):232-46.

Sand-Jecklin, K. Evaluating the student clinical learning environment: development and validation of the SECEE inventory. Southern Online Journal of Nursing Research. 2000;1(4).

Schön DA, 1983. The Reflective Practitioner: How Professionals Think in Action. Basic Books, New York.

Soemantri D, Herrera C, Riquelme A. Measuring the educational environment in health professions studies: a systematic review. Med Teach. 2010;32(12):947-52.

Tomietto M, Saiani L, Palese A, Cunico L, Cicolini G, Watson P, Saarikoski M. Clinical learning environment and supervision plus nurse teacher (CLES+T) scale: testing the psychometric characteristics of the Italian version. G Ital Med Lav Ergon. 2012 Apr-Jun;34(2 Suppl B):B72-80.

Tomietto M, Saiani L, Saarikoski M, Fabris S, Cunico L, Campagna V, Palese A. Assessing quality in clinical educational setting: Italian validation of the clinical learning environment and supervision (CLES) scale. G Ital Med Lav Ergon. 2009 Jul-Sep;31(3 Suppl B):B49-55.

Vizcaya-Moreno MF, Pérez-Cañaveras RM, De Juan J, Saarikoski M. Development and psychometric testing of the Clinical Learning Environment, Supervision and Nurse Teacher evaluation scale (CLES+T): the Spanish version. Int J Nurs Stud. 2015 Jan;52(1):361-7.

Warne T, Johansson UB, Papastavrou E, Tichelaar E, Tomietto M, Van den Bossche K, Moreno MF, Saarikoski M. An exploration of the clinical learning experience of nursing students in nine European countries. Nurse Educ Today. 2010 Nov;30(8):809-15.

Watson PB, Seaton P, Sims D, Jamieson I, Mountier J, Whittle R, Saarikoski M. Exploratory factor analysis of the Clinical Learning Environment, Supervision and Nurse Teacher Scale (CLES+T). J Nurs Meas. 2014;22(1):164-80.

Wilson-Barnett J, Butterworth T, White E, Twinn S, Davies S, Riley L. Clinical support and the project 2000 nursing student: factors influencing this process. J Adv Nurs. 1995;21:1152-1158.

Zabalegui A, Cabrera E. New nursing education structure in Spain. Nurse Educ Today. 2009 Jul;29(5):500-504.

Table 1
Characteristics of the studies included

Instrument validated	Author, publication year	Country, data collection year	Study design	Sample characteristics[a]	Settings	Instrument characteristics
Clinical Learning Environment scale (CLE scale)	Dunn and Burnett, 1995	Australia, 1993	Validation study	381 students; female 87%; mean age 22.4; range 17-52 course year 1st, 2nd, 3rd	Not specified	23 items, five factors: staff-student relationships, nurse manager commitment, patient relationship, interpersonal relationship, student satisfaction 5-point Likert scale (1 strongly disagree, 5 strongly agree)
Clinical Learning Environment Inventory (CLEI)	Chan, 2001 Chan, 2003	Australia, 1998	Validation study	108 students; response rate 67.5% course year 2nd	Government and private Hospitals	1st version: 35 items, five factors: personalisation, student involvement, task orientation, innovation, individualisation 2nd version: 42 item and one factor more: satisfaction 4-point Likert scale (1 strongly disagree, 4 strongly agree)
	Newton et al., 2010	Australia, 2006-2008	Validation study	513 students course year 60.6% 2nd, 39.4% 3rd	Public and private Hospitals Wards: ICU, Accident and Emergency department, Medicine/surgery	
	Salamonson et al., 2011	Australia, 2009	Validation study	231 volunteer; female 87%; mean age 30.3; SD 10.4 course year 1st, 2nd, 3rd	Not specified	Abbreviated CLEI-19 items, two factors: satisfaction, personalisation 5-point Likert scale (1 strongly disagree, 5 strongly agree)
Clinical Learning Environment and Supervision scale (CLES)	Saarikoski et al., 2002a	Finland, 2000	Validation study	416 students; response rate 81% female 91%; mean age 23; range 17-52 course year 2nd, 3rd	33% in University Hospitals; 31% in regional Hospitals; 30% in local health care centres Ward: 8 units (6% in social sector unit)	27 items, five factors: ward atmosphere, leadership style of the ward manager, premises of nursing care on the ward, premises of learning on the ward, supervisory relationship 5-point Likert scale (1 fully disagree, 5 fully agree)
	Saarikoski et al., 2002b	Finland, UK, 2000	Comparative validation study	416 Finnish[b]; 142 UK students female 89%; mean age 24.6 course year 2nd, 3rd	Hospitals	
	Saarikoski et al., 2005	Finland, 2000	Validation study	416 Finnish students[b] course year 2nd, 3rd	Hospitals	
	Tomietto et al., 2009	Italy, not specified	Validation study	117 students; response rate 60.3% female 79.5%; mean age 23.8; SD 4.3; range 20-42; course year 2nd, 3rd	Hospitals	
	Papastavrou et al., 2010	Cyprus, not specified	Exploratory study	559 students mean age 20.4; SD 2.7 course year 40% 1st, 35% 2nd, 25% 3rd	Hospitals Wards: 24% medical; 16% orthopaedic; 13.8% surgery; 56% other	
	De Witte et al., 2011	Belgium, 2007-2008	Validation study	768 students; female 74% course year 35.5% 1st, 31.3% 2nd, 33.2% 3rd	Hospitals, 31 health institutions Wards: 190 different units	Items adjusted (n=32) for the specific country
	Burrai et al., 2012	Italy, not specified	Validation study	59 students; female 73% mean age 22; SD 1.5; range 20-25 course year 2nd	Hospital	6-point Likert scale (1 fully disagree, 6 fully agree)

Instrument validated	Author, publication year	Country, data collection year	Study design	Sample characteristics[a]	Settings	Instrument characteristics
Clinical Learning Environment, Supervision and Nurse Teacher scale (CLES+T)	Saarikoski et al., 2008	Finland, not specified	Validation study	549 students	Hospital	34 items, five factors: ward atmosphere and premises of learning on the ward, role of nurse teacher, leadership style of the ward manager, premises of nursing on the ward 5-point Likert scale (1 fully disagree, 5 fully agree)
	Johansson et al., 2010	Sweden, 2008-2009	Validation study	324 students; female 91% mean age 28.6; range 19-50 course year 9% 1st, 44% 2nd, 47% 3rd	Hospitals (85% University Hospitals) Wards: 36% medical; 31% surgery/orthopaedic; 18% psychiatry; 8% elderly care; 3% gynaecology; 3% others; 1% paediatrics	
	Warne et al., 2010	BE, CY, ES, FI, IT, NL, SE, UK, UK, 2007-2008	Exploratory comparative validation study	1,903 students; female 89% mean age 24.6	Hospitals (57% university colleges, 43% polytechnics)	
	Bos et al., 2012	Sweden, 2008-2010	Validation study	356 students, female 99% mean age 28; range 19-54	Primary Healthcare	Items revised for the specific setting
	Henriksen et al., 2012	Norway, 2009	Validation study	407 students; response rate 41.6% mean age 27.4; SD 7.9 course year 1st, 2nd, 3rd	Hospitals, Nursing Homes Wards: Accident and Emergency department, psychiatry	See Saarikoski et al., 2008
	Tomietto et al., 2012	Italy, not stated	Validation study	855 students; response rate 97% female 74.8% mean age 24; SD 5.7; range 19-54 course year 43.6% 1st, 42.6% 2nd, 13.8% 3rd	Hospitals Wards: 49.5% medical; 34.7% surgery; 10.6% unspecified; 3.1% ICU; 1.8% maternal/paediatrics	See Saarikoski et al., 2008
	Bergjan et al., 2013	German, 2011	Validation study	167 students; response rate 74% age 70% 19-22 years course year 35% 1st, 31% 2nd, 34% 3rd	Hospital Wards: 34% medical; 21% surgery; 20% paediatrics; 11% psychiatry; 7% neurology; 4% acute day; 3% gynaecology	Revised in the "Role of Nurse Teacher" factor
	Watson et al., 2014	New Zealand, 2008-2009	Validation study	416 students; response rate 41% course year 3rd	Governmental and private Hospitals Wards: aged care, mental health, medicine, surgery, Accident and Emergency department, paediatrics	See Saarikoski et al., 2008
	Gustafsson et al., 2015	Sweden, 2011	Validation study	42 students; female 98% mean age 23; range 20-47	Hospitals	
	Papastavrou et al., 2015	Cyprus, Not stated	Validation study	463 students; response rate 70.3% female 61.3% mean age 21.0; SD 2.2; range 18-34	68% public and 32% private universities	
	Vizcaya-Moreno et al., 2015	Spain, 2011-2012	Validation study	370 students; response rate 89.6% female 82.7% mean age 22.3; range 20-43 course year 3rd	Hospitals (56.4% University Hospitals) Wards: 26.6% Accident and Emergency department; 22.8% ICU; 13.9% oncology; 10.1% psychiatry; 9.2% medical; 8.9% surgery; 6.3% haemodialysis; 2.2% others	

Instrument validated	Author, publication year	Country, data collection year	Study design	Sample characteristics[a]	Settings	Instrument characteristics
Clinical Learning Environment and Diagnostic Inventory (CLEDI)	Hosoda, 2006	Japan, 2004	Validation study	312 students; response rate 79.6% female 94.2% age 95.5% 20-24 years	23.1% University Hospitals; 68.3% Hospitals; 5.8% Healthcare facilities for elderlies; 2.2% Mental Hospital; 0.6% other Hospitals Wards: 26.9% surgery; 20.2% medical; 13.5% obstetric; 11.9% paediatrics; 9.9% medical-surgical; 4.8% psychiatry; 12.8% others	35 items, five factors: affective CLE, perceptual CLE, symbolic CLE, behavioural CLE, reflective CLE 5-point Likert scale (1 strongly disagree, 5 strongly agree)
Student Evaluation of Clinical Education Environment (SECEE)	Sand-Jecklin, 2009	USA, 2001-2005	Validation study	2,768 inventories nursing sophomore, junior and baccalaureate students	Not specified	32 items, three factors: instructor facilitation, preceptor facilitation, learning opportunities 5-point Likert scale (1 strongly disagree, 5 strongly agree)
Clinical Learning Environment instrument (CLE instrument)	Chuan and Barnett, 2012	Malaysia, Not specified	Validation study	142 volunteers; response rate 74.7% course year 3rd	Private Hospital Wards: 3 medical, 3 surgery	44 items, five factors: ward atmosphere, supervision by staff nurses and the clinical teacher, student satisfaction, the theory-practice gap, peer support 4-point Likert scale (1 strongly disagree, 4 strongly agree)
Modified Clinical Learning Environment, Supervision and Nurse Teacher scale (modified CLES+T)	D'Souza et al., 2015	Oman, 2011	Cross-sectional validation study	310 students; response rate 100% female 74% age 15% ≤20, 74% 21-25, 12% ≥26 years	Hospitals Wards: ICU, Accident and Emergency department, medical, surgery, maternity and paediatric	57 items, 11 factors: hierarchy/ritual, patient relationships, clinical nurse commitment, staff-student relationships, student satisfaction, ward atmosphere, premises of learning on the ward, supervisory relationship, leadership style of the ward manager, premises of nursing care on the ward, role of the clinical teacher 5-point Likert scale (1 strongly disagree, 5 strongly agree)

Legend: SD, Standard Deviation; ICU, Intensive Care units

[a] Only data described in the studies was reported in this table; in addition, only data regarding nursing students have been reported

[b] Same sample as the previous study (Saarikoski et al., 2002a)

Table 2

Instruments evaluating the clinical learning environment quality: psychometric properties and their methodological quality of evaluation

Instrument	Authors, year	Internal Consistency	Reliability	Measurement Error	Content Validity	Structural Validity	Hypotheses Testing	Convergent Validity	Criterion Validity	Cross-cultural Validity
		α Cronbach	ICC	SEM, SDC	Yes	Variance explained %, methods[a]	Yes	*r* Pearson	Tool, *r* Pearson *p* value	Yes
CLE scale	Dunn and Burnett, 1995	0.63 ++			yes +	34.6, EFA, CFA ++				
CLEI	Chan, 2001, 2003	0.73-0.84 +						0.39-0.45 +		
	Newton et al., 2010	0.50-0.88 ++++				51, PCA +++				
	Salamonson et al., 2011	0.93 +++				63.37, PCA +++	yes +			
CLES	Saarikoski et al., 2002a	0.73-0.94 +++			yes +	64, EFA +++	yes +			
	Saarikoski et al., 2002b	0.74-0.95 +					yes ++			
	Saarikoski et al., 2005	0.86 0.73-0.95 +++				64, PCA +++			CLE scale, 0.93 <0.001 ++	
	Tomietto et al., 2009	0.96 +	0.89 +		yes +					Forward-backward translation +
	Papastavrou et al., 2010					67, EFA ++	yes +			
	De Witte et al., 2011	0.970 ++			yes +	71.28, PCA ++				Forward-backward translation +
	Burrai et al., 2012	0.957 +				76.9, PCA +				
CLES+T	Saarikoski et al., 2008	0.77-0.96 ++				67/62-64, EFA, PCA ++				
	Johansson et al., 2010	0.95 ++			yes +	60.2, EFA ++				Forward-backward translation +
	Warne et al., 2010	0.83-0.96 +					yes +			
	Bos et al., 2012					[b]EFA, CFA +				
	Henriksen et al., 2012					64, PCA ++				Forward-backward translation ++

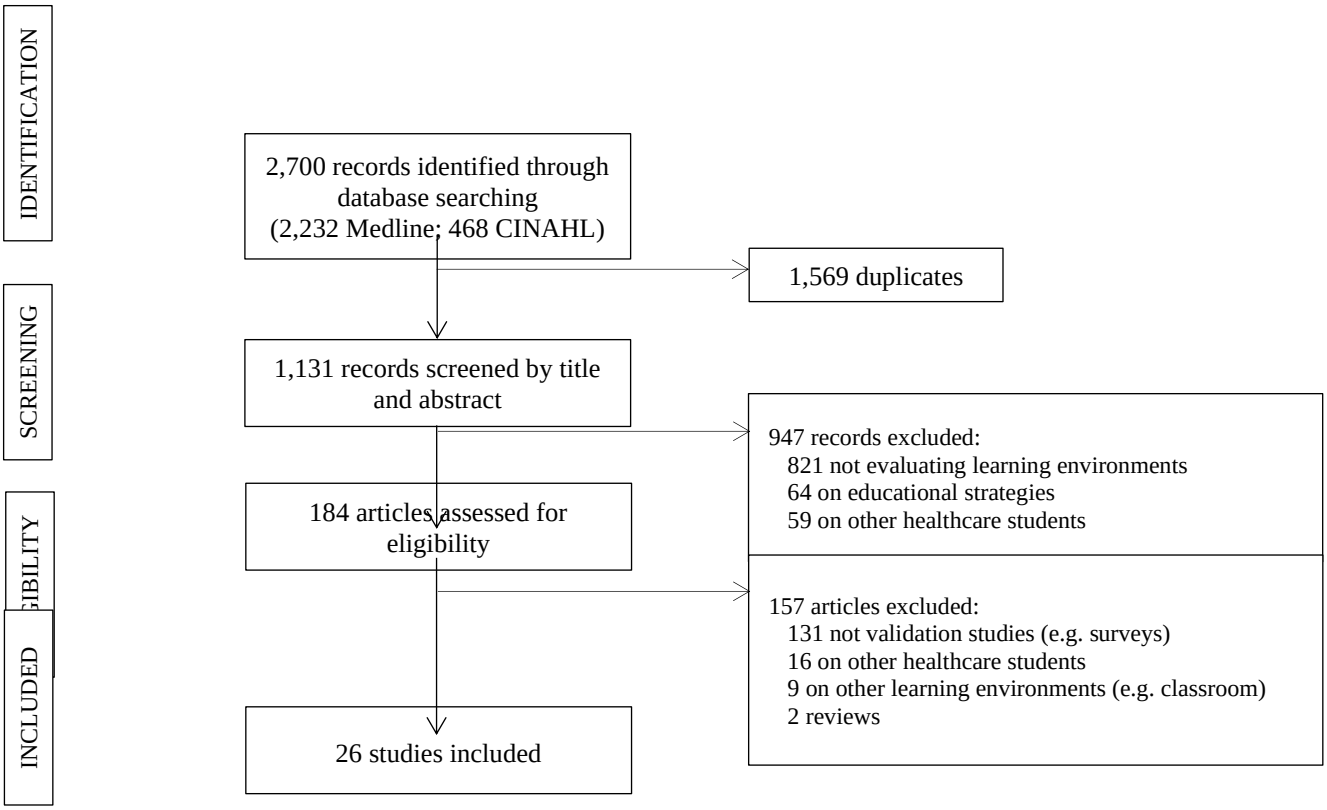
	Tomietto et al., 2012	0.95 0.80-0.96 +++				67.27, EFA, CFA +++			Forward-backward translation +
	Bergjan et al., 2013	0.82-0.96 +				72.82, EFA, PCA +			Forward-backward translation +
	Watson et al., 2014	0.82-0.93 ++++			Yes[c] +	58.28, EFA ++++			
	Gustafsson et al., 2015		0.70-0.96 ++	0.2-0.42 SEM 0.56-1.18 SDC ++					
	Papastavrou et al., 2015	0.95 0.81-0.96 ++			yes +	67.405, EFA +++			
	Vizcaya-Moreno et al., 2015	0.95 0.80-0.97 ++				66.4, EFA, CFA ++			Forward-backward translation ++
CLEDI	Hosoda, 2006	0.84 ++	0.76 ++		yes +	52.45, EFA ++	yes +	CLES, 0.76 <0.01 ++	
SECEE	Sand-Jecklin, 2009	0.94 0.82-0.94 ++				59, EFA, CFA ++	yes ++		
CLE instrument	Chuan and Barnett, 2012	0.867 0.658-0.875 +			yes +	54, PCA +	yes ++		
Modified CLES+T	D'Souza et al., 2015	0.84 +			Yes +	SEqM ++			

Legend. + poor; ++ fair; +++ good; ++++ excellent; CFA Confirmatory Factor Analysis; EFA Explorative Factor Analysis; ICC Intraclass Correlation Coefficient; PCA Principal Component Analysis; SDC Smallest Detectable Change; SEM, Standard Error of Measurement; SEqM, Structural Equation Model

[a] When CFA was used, the data has not been reported here, in the interest of summarization; however, the data is available in the included studies or from the authors of this review

[b] data not reported in the study

[c] in the content validity evaluation, nursing students were also involved



Legend: CINAHL, The Cumulative Index to Nursing and Allied Health Literature

Fig. 1. Study selection and inclusion process: flowchart