

# Accepted Manuscript

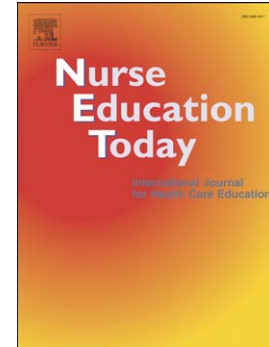
A review of mentorship measurement tools

Yanhua Chen, Roger Watson, Andrea Hilton

PII: S0260-6917(16)00047-2  
DOI: doi: [10.1016/j.nedt.2016.01.020](https://doi.org/10.1016/j.nedt.2016.01.020)  
Reference: YNEDT 3187

To appear in: *Nurse Education Today*

Accepted date: 26 January 2016



Please cite this article as: Chen, Yanhua, Watson, Roger, Hilton, Andrea, A review of mentorship measurement tools, *Nurse Education Today* (2016), doi: [10.1016/j.nedt.2016.01.020](https://doi.org/10.1016/j.nedt.2016.01.020)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**A review of mentorship measurement tools**

Yanhua CHEN, RN, PhD Candidate, the Infectious Disease Department of the First Affiliated Hospital of Lu Zhou Medical College, China

Email: chen\_yanhua25@163.com

Roger WATSON PhD RN, Professor of Nursing, The University of Hull, UK

Email: r.watson@hull.ac.uk

Andrea HILTON, PhD, Senior Lecturer, The University of Hull, UK

Email: a.hilton@hull.ac.uk

Word count :3784

**Objectives**

To review mentorship measurement tools in various fields to inform nursing educators on selection, application and developing of mentoring instruments.

**Design**

A literature review informed by PRISMA 2009 guidelines.

**Data sources**

Six databases: CINHAL, Medline, PsycINFO, Academic Search Premier, ERIC, Business premier resource

**Review methods**

Search terms and strategies used: mentor\* N3 (behav\* or skill? or role? or activit? or function\* or relation\*) and (scale or tool or instrument or questionnaire or inventory). The time limiter was set from January 1985 to June 2015. Extracted data was content of instruments, samples, psychometrics, theoretical framework and utility. An integrative review method was used.

**Results**

Twenty-eight papers linked to 22 scales were located, seven from business and industry, 11 from education, three from health science and one focused on research mentoring.

Mentorship measurement was pioneered by business with a universally accepted theoretical framework, i.e. career function and psychosocial function, and the trend of scale development is developing: from focusing on the positive side of mentorship shifting to negative mentoring experiences and challenges. Nursing educators mainly used instruments from business to assess mentorship among nursing teachers. In education and nursing, measurement has taken to a more specialized focus: researchers in different contexts have

developed scales to measure different specific aspects of mentorship. Most tools show psychometric evidence of content homogeneity and construct validity, but lack more comprehensive and advanced tests.

### **Conclusion**

Mentorship is widely used and conceptualized differently in different fields and is less mature in nursing than in business. Measurement of mentorship is heading to a more specialized and comprehensive process. Business and education provided measurement tools to nursing educators to assess mentorship among staff, but a robust instrument to measure nursing students' mentorship is needed.

### **Key words**

Nursing education, mentor, behaviour, measurement tool, theoretical framework, psychometrics

## **1. Introduction**

### **1.1 Mentorship**

Mentorship flourished after the work of Levenson et al (1978) in business and organization. It has been used as a strategy to nurture new leaders, new staff, to raise morale and reduce turn-over rate. It has also been applied in social science, mainly to youth development, and the most famous organization is Big Brother and Big Sister to help problematic children to get proper social skills and academic achievements (Ferro et al, 2013). Furthermore mentorship is extensively employed in higher education to reduce drop-out rate; in doctoral student education to enhance research productivity; and to nurture new teaching staff and leaders. It has also been applied in varying areas, such as nursing.

### **1.2 Mentorship in nursing education**

Mentorship has been adopted in many nursing fields for more than 30 years (Berk et al. 2005). It is generally accepted that mentoring has advantages for mentees (Andrews and Wallis, 1999) and mentors (Dibert and Goldenberg, 1995) in nursing education. At an early stage nurse researchers attempted to define concepts such as ‘mentor’ and ‘mentorship’ and to clarify the roles and functions of mentors without reaching consensus (Myall et al. 2008). Later, researchers focused on students’ (mentees’) and mentors’ experience of mentoring. Mentor support, preparation and assessment is drawing more attention now (Sawatzky and Enns, 2009; Hyrkäs and Shoemaker, 2007; Kalischuk et al. 2013).

### **1.3 Measurement of mentorship in nursing education**

Due to lack of specific measurement tools, nursing academia and professionals often use tools from business such as Mentoring Functions Scale (Scandura, 1992; Scandura and

Ragins, 1993; Pellegrini and Scandura, 2005; Hu et al. 2011), Mentoring function Scale (Noe, 1988) and Sands' tool (Sands, 1991) to measure mentors' function, behaviour, and relationships. These mentorship tools in different fields may vary in conceptualization and measuring different aspects of mentorship, therefore some researchers in nursing focused on developing their own tools catering for their specific needs (Berk et al, 2005; Chow and Suen, 2001). However the robustness of these instruments is unknown.

#### **1.4 Measurement tools selection and development**

When choosing or developing a measurement tool, several points need to be considered.

##### **1.4.1 Theoretical framework**

To select or develop a measurement, the first thing to determine is what to measure. Usually researchers measure some complicated latent variables which cannot be observed directly, so clarity of the phenomena under study is important. Theoretical frameworks can help to clarify these (Devis, 2003). A proper theory can help to define the boundary, content and structure of a latent variable, which will give clear guidance in the development of a new instrument. This theory can come from a related area or be tentatively constructed based on research on the measurement problem. Users can judge if a tool following a certain theory matches their requirements.

##### **1.4.2 Psychometrics**

To judge a measurement, it is imperative to know its psychometric properties: reliability and validity. Philosophically, to measure something is to explore the true value of an object under measurement (which is never known); or the accuracy of a measurement; the ability to

differentiate subjects with different levels of a trait; consistency and agreement of measurement (Streiner and Norman, 2008).

### **Reliability**

Reliability means to what extent the measurement of a scale is reproducible (Streiner and Norman, 2008). Mathematically and practically, the three aspects of reliability: test-retest reliability; internal consistency; and inter-rater reliability, are commonly explored to demonstrate the quality of a scale, or to be more precise, the interaction of a scale with a certain group of people in a certain context. Test-retest reliability is applied to explore consistency of a measurement over time, in a group of subjects (Streiner and Norman, 2008 p.182). Items or scales showing low test-retest reliability may imply a problem in understanding, which suggests that actions, such as re-wording, are necessary.

Internal consistency reliability measures whether the items in a scale are correlated to the latent trait under evaluation and it is the most frequently used method to express a scale's reliability (Hogan and Cannon, 2003). Items showing low internal consistency reliability in an instrument indicate that they are measuring different concepts, and could be deleted. Since internal consistency is based on a single test, the results should be interpreted with caution (Streiner and Norman, 2008).

Inter-rater agreement or inter-scorer reliability tests different raters' deviation using the same tool to rate the same subject. It considers the effect of different raters' variance and error on measurement accuracy and consistency besides subjects' variance and error (Streiner and Norman, 2008). If inter-rater reliability is low, it may indicate that the scale under investigation is defective or that the raters need to be trained.

Reliability is essential for assessment of a scale's quality, which can have an impact on the validity and decide the maximum of validity (Streiner and Norman, 2008), but, unlike validity, it cannot assure you how true the outcomes are and whether it measures the trait you intend to measure.

### **Validity**

Validity is the extent to which a tool measures the concept that it purports to measure. It allows inference from raw scores of a scale to the trait under measurement. Validity has different categories and the frequently cited 'three C' validities are discussed here: content validity; criterion validity; and construct validity.

Content validity indicates whether a scale contains all the aspect of the concept under study and whether there are any irrelevant items in a scale. It can be achieved through subjects, expert panels and researchers' judgement. But experts' subjective judgement without statistical testing among large samples casts some suspicions on it (Streiner and Norman, 2008), and this implies that more empirical and 'harder' evidences of validity are needed, such as criterion validity and construct validity.

Criterion validity measures the correlation of a new scale with a 'gold standard' tool, which exists to measure the same concept; the higher the correlation is the better the new instrument. The reason for developing a new scale against the old one may be due to considerations of economy, doing less harm or taking less time. If the research is exploring a new area without any instrument or any existing 'gold standard', it is impossible to test the criterion validity of a new tool, but it is feasible to establish its construct validity.

When constructing a new construct (latent variable), people need to demonstrate that this new construct is better than existing constructs. It includes many categories: convergent and



divergent validity, factorial validity, i.e. exploratory factor analysis (EFA) and confirmatory factor analysis (CFA).

Convergent validity is intended to measure the correlation between a new scale and a standard tool assessing a different trait which is assumed to be correlated with the trait under test: for instance, life quality may be associated with social support. Divergent validity is, on the contrary, to test the correlation between a new trait under test and a trait which is assumed not to be correlated with, for example, depression may not be associated with intelligence.

Factorial validity investigates how many factors the observable items can converge to in a latent construct depending on the loading and cross-loading coefficients, which gives a parsimonious understanding of a new construct. To establish factorial validity, usually factor analysis (EFA and/or CFA) is used. EFA purports to explore the structure of a construct based on data through factor extraction and rotation and selection of an appropriate level of 'loading' (essentially correlation) of items on putative factors (Gefen and Straub, 2005).

While CFA is used to test if the presumed construct can be confirmed by any target sample, therefore, the first step is to specify a construct, then loadings and other model fit indices should be checked and the model can be modified based on the set criteria.

All the above psychometric theory is based on classical test theory. More sophisticated test theory and techniques such as item response theory (IRT), e.g. Mokken scale and Rasch model, have been developed and they are used as a norm by some health rating scales developers (McDowell, 2006).

### **1.5 Samples and utility**

Both reliability and validity are not intrinsic property of a scale, but connected with the scores of the samples being tested; therefore when researchers choose some scales they need to

compare the target samples' characteristics with the sample having been tested or test the scale again with their own samples. Through continuous use, measurement tools can provide more psychometric and suitability evidence in different area, these further information may give users more confidence and reference.

Due to there being no systematic information about existing mentorship tools, this study aims to review mentorship assessment tools systematically and provide comprehensive and objective information when nursing educators need to select measurement scales or develop their new scales.

## **2. Methods**

A literature review informed by PRISMA 2009 guidelines.

### **Search terms and strategies**

The following search terms and strategies were used: mentor\* N3 (behav\* or skill? or role? or activit? or function\* or relation\*) and (scale or tool or instrument or questionnaire or inventory). The time limiter was set from January 1985 to June 2015 as the earliest tool was developed then and mentorship flourished at similar time; language was limited to English; age group limiters as adult, over 18 were applied in different data bases as mentoring in nursing applied among adult groups. Truncation was used and the reference lists were also inspected for a more comprehensive search.

### **Database**

Databases included those from the disciplines of business and organization, health science, psychology and education:

- CINHAL

- Medline
- PsycINFO
- Academic Search Premier
- ERIC
- Business premier resource

The reasons for searching these databases were: mentorship started in business and organization and its mentorship is relative to leadership and management and staff development in nursing education field. In general higher education field, mentorship is applied in varying situations as mentioned before, which is relevant to nursing students and teachers mentoring in nursing school and clinical setting. Other fields such as medicine and other allied health field may also provide useful and relevant measurement tools.

#### **Inclusion criteria**

- Articles about mentoring function/role/behaviour/activities scale development and validation
- Mentoring papers in the fields of business and organization, education, nursing, medicine and allied health

#### **Exclusion criteria**

- Studies not about mentoring or not using concept 'mentor\*'
- Mentoring scales in other fields like youth or pupil mentoring
- Research measuring mentorship outcome such as job satisfaction, career development and other outcomes and predictors

- Papers reporting qualitative research or discussing mentorship
- Studies about scale development showing no proper items or dimensions
- Papers not accessible

### **Data management and selection process**

Data management and selection process following PRISMA 2009 are shown in Figure 1. The criteria referred to can be found in Table 1.

### **Data extraction**

Data extracted were: content of instrument, samples, psychometrics, theoretical frameworks and utility as these were informative for instruments selection and application as discussed above.

### **Data synthesis**

As the heterogeneity among the data was obvious and the nature of this review was not to compare effects, meta-analysis is not suitable, therefore each instrument was presented in an integrative way.

## **3. Results**

Using the search strategies in the six databases, 3153 papers were identified, after removing duplications 2432 were left, then following the inclusion and exclusion criteria 28 papers linked to 22 scales were left as shown in Figure 1.

The majority of the tools were developed in the USA (N=17); the number of tools increased steadily over three decades; they were mainly developed in education (n=11) and business (n=7). Mentorship measurement was pioneered by the business discipline with a universally

accepted theoretical framework, i.e. career function and psychosocial function, and the trend of scale development is developing: from focusing on the positive side of mentorship shifting to negative mentoring experiences and challenges (Eby et al. 2008; Ensher and Murphy, 2011). In education and nursing, measurement has taken to a more specialized focus: researchers in different contexts have developed scales to measure different specific aspects of mentorship. The vast majority of the tools show psychometric evidence of content homogeneity and construct validity (factorial validity), but lack more comprehensive and advanced tests are needed, shown in Table 2 and 3.

#### **4. Discussion**

##### **Theoretical framework/conceptualization**

In the field of business and organization, mentorship is conceptualised as two domains (career development and psychosocial support) and nine key behaviours: sponsorship, role modelling, exposure-and-visibility, acceptance-and-confirmation, coaching, counselling, challenging assignments, friendship and protection (Kram, 1983; Kram and Isabella, 1985), and this structure is supported by five scales (Dreher and Ash, 1990; Pollock, 1995; Ragins and McFarlin, 1990; Noe, 1988; Schockett and Haring-Hidore, 1985) shown in Table 3. Later, the two-function model was split into a three-function structure (career, psychosocial and role modelling function), and was confirmed by Scandura (1992), Scandura and Ragins (1993), Pellegrini and Scandura (2005), and Hu et al (2011). This implies that the conceptualization in the business and organization field has reached consensus and the situations of mentorship application are similar, i.e. staff development.

In education, however, there are no universally recognised theoretical frameworks for mentoring (Crisp and Cruz, 2009; Jacobi, 1991), although some are mentioned and used. For instance, Anderson and Shannon's (1988) construct of five functions of mentoring: teaching, sponsoring, encouraging, counselling and befriending has often been cited and taken as a theoretical underpinning of educational mentoring scales (Rose, 2003), but this construct was not confirmed by Rose's research (2003). Cohen's six-function theoretical framework (Cohen, 1995) is often cited, but it was not fit for mentoring medical students (Roger et al. 2005), nor mentoring general college students' (Lightfoot, 2000). A new four-factor framework (psychological and emotional support; degree and career support; academic subject knowledge support; and the existence of a role model) (Crisp and Cruz, 2009), and a three-dimensional framework of PhD mentoring (integrity, guidance, and relationship) have emerged (Rose, 2003), but they need more testing. This suggests that, in education, mentorship is conceptualized differently as it is used in varying situations, such as mentoring of teaching staff, mentoring of college students and mentoring of PhD students and that mentorship. This can help educators to develop differing instruments to measure their specific mentorship but will make comparison across areas difficult.

Among the three tools from nursing Berk et al (2005), Jakubik (2008), Chow and Suen (2001) adopted different theoretical framework as they developed instrument to measure mentorship in varying areas: clinical staff, education staff and nursing students' clinical teaching, which is similar to the situation in education. This confirms the specialization direction of mentorship application and assessment and implies that more measurement tools will be developed in future.

### **Psychometrics**

Regarding reliability, the most frequently tested reliability in the 22 tools is internal consistency reliability, but it can just tell how similarly the items in one scale behave and can be inflated by increasing the number of items. No studies included test-retest reliability or inter-rater reliability; this may imply that mentorship assessment is at the stage of construct understanding and exploring, while the precision, agreement and consistency of mentorship measurement has not been so acute as those measurement in medical and psychology, which will influence the results of diagnosis and treatment, but more efforts are needed to get more accuracy assessment of quality and effectiveness of mentorship (Allen et al, 2008).

With regard to validity, content validity is the basic, all the tools reported it. However three tools did not go beyond that in testing of their psychometrics (Cohen, 1995, Chow and Suen, 2001; Berk et al, 2005), which is not sufficient for a measurement tool. Among construct validity, factorial validity is investigated more frequently than others in the review, which is useful to understand the structure of a complicated phenomenon by simplifying multiple items into a few factors; convergent and divergent validity are also explored (Rose, 2003; Eby et al. 2008) in business and education, not in nursing, which implies that mentorship measurement or scale development and validation is relatively new and immature in nursing. Measuring equivalence/invariance is tested using multi-group confirmatory factor analysis by Hu et al (2010), which is a new development in measurement and should be measured before a tool is used in different cultures and sample groups. This implies that mentorship measurement approaches a more scientific direction in a cross-culture comparison when business becomes more and more internationalized.

Criterion validity was used by Eby et al. (2008), suggesting that mentoring measurement is still young compared to other tests, such as IQ test: no gold standard of mentorship measurement exists. No advanced test theory like item response theory is applied. Above all,

to achieve reproducible and accurate assessment and to guide behaviour change in mentorship, mentoring scales need more, and more advanced psychometric evidence, compared to health measurement and other psychometric testing tools, e.g. IQ, personality, suppression.

The sample sizes varied from 43 to 463; some were too small for reliable results (Pamuk and Thompson, 2009); none of these sample sizes were large enough to establish a norm for a certain group of people such as managers and mentees in business or graduates from business schools, PhD students, college students, professors, nurses and so on. The main point is the representativeness of samples, usually local samples were used; therefore, when researchers choose any instrument they need to compare the samples with their population under study or test the suitability of the tools in advance in their own samples.

### **Extent of use**

The Mentoring Functions Scale (Scandura, 1992; Scandura and Ragins, 1993; Pellegrini and Scandura, 2005; Hu et al. 2011) has become increasingly popular in business and other fields, which may be due to its short length and stable three-dimensional structure, while the continuing testing and upgrading of the instrument is another reason. The two-dimensional mentoring scales (Dreher and Ash, 1990; Noe, 1988; Ragins and McFarlin, 1990) are also widely used in different area with up to 1000 citations (Allen et al. 2008). All these scales are used in nursing to assess teaching staff mentoring in nursing school (Altuntas, 2012; Short, 1997; Chung and Kowalski, 2012) and assess clinical nursing staff mentoring in clinical placement (Weng et al. 2010; Salami, 2008). One commercially used scale: The Alleman Mentoring Activities Questionnaire (Alleman, 1987, Alleman and Clarke, 2002) may be the most widely used in nursing (Richard, 1995; Jones, 1997; Aponte, 2007; Kavooosi et al.

1995), having proper instruction on administration and scoring, to see more detail about this



instrument read Gilbreath et al's work (2008). This reflects the fact that mentorship originated from business and is obtaining public acknowledgement across disciplines.

In education, more new scales were developed recently (Harris, 2013; Koc, 2011), but further study is needed, while in health science, medical educators and researchers began to develop their own mentorship scales (Fleming et al. 2013), focusing on research mentoring. Very few of them were used in nursing except one (Sands et al. 1991) which was used to assess faculty mentoring in nursing school (Frandsen, 2003).

Nursing professionals chose some assessment tools from business or education to measure staff nurses' mentorship or nursing teachers' mentoring as stated above, but no study using these scales to measure pre-registered students' mentorship in the field of clinical learning was identified. This implies a conceptualization difference between student mentoring and staff mentoring in nursing. Among the tools from nursing, one nursing student mentoring scale (Chow and Suen, 2001; Suen and Chow, 2001) includes 33 items measuring mentors' behaviour, but suffers from little psychometric evidence (reported face and content validity) and questionable theoretical framework. Its theoretical framework, derived from the five roles of mentors defined by ENB (Chow and Suen, 2001), has been replaced by the new eight roles (Nursing and Midwifery Council, 2008); new themes, such as evidence-based nursing, assessment and accountability, evaluation of learning, have been added; both jeopardize the acceptability of this scale. The content and the outcome of the measure was cited by many nursing researchers (Andrews et al. 2006; Bray and Nettleton, 2007; van Eps et al. 2006; Lambert and Glacken, 2005; Myall et al. 2008), but it has not been applied or tested further. Although Berk's (2005) mentoring scale did not present any psychometric evidence, it was cited widely and used by medical and nursing teachers (Dimitriadis et al. 2012): this may be due to its high face and content validity, or that no other psychometrically sound scale in

medicine and allied health field exists. Jakubik's scale (2008, 2012) was used by herself as it is relatively new.

### **Conclusion**

Mentorship measurement was pioneered by the business discipline with a universally accepted theoretical framework. In education and nursing the measurement is heading to a more specialized direction, as mentorship takes place in different contexts and the conceptualizations vary. The vast majority of the tools show psychometric evidence of content homogeneity and construct validity (factorial validity), but more comprehensive and advanced tests are needed. Mentoring measurement is less mature in nursing, both the psychometric evidence and conceptualization need further study; therefore, scales from the business and education fields are used to measure mentorship of staff nurses or teaching staff, but none have been used to assess the mentoring of nursing students.

Conflict of interests. None declared.

Funding: No special source of funding.

## Reference

Alleman, E. (1987) Alleman Mentoring Scales Questionnaire, Form A. OH: Leadership Development Consultants, Inc.

Alleman, E. J., & Clarke, D.L. (2002) Alleman mentoring activities questionnaire revised. Sharon Center, OH: Silverwood Enterprises

Allen, T. D., Eby, L. T., O'brien, K. E. & Lentz, E. (2008) The State of Mentoring Research: A Qualitative Review of Current Research Methods and Future Research Implications. Journal of Vocational Behavior. Vol. 73 (3) pp. 343-357.

Altuntas, S., (2012) Mentorship relations among academicians nurses in Turkey: An assessment from the viewpoints of mentors and mentees. Nurse Education Today. Vol.32 (6) pp.652-656.

Anderson, E. M., and Shannon, A. L. (1988). Toward a conceptualization of mentoring. Journal of Teacher Education. Vol. 39(1): 38–42.

Andrews, G.J., Brodie, D.A., Andrews, J.P., Hillan, E., Gail Thomas, B., Wong, J. and Rixon, L., (2006) Professional roles and communications in clinical placements: a qualitative study of nursing students' perceptions and some models for practice. International Journal of Nursing Studies. Vol.43 (7) pp.861-874.

Andrews, M. and Wallis, M., (1999) Mentorship in nursing: a literature review. Journal of Advanced Nursing. Vol.29 (1) pp.201-207.

Aponte, M., (2007) Mentoring: career advancement of Hispanic Army nurses [thesis], George Mason University

Berk, R.A., Berg, J., Mortimer, R., Walton-Moss, B. and Yeo, T.P., (2005) Measuring the Effectiveness of Faculty Mentoring Relationships. Academic Medicine. Vol.80 (1) pp.66-71.

Bray, L. and Nettleton, P., (2007) Assessor or mentor? Role confusion in professional education. Nurse Education Today. Vol.27 (8) pp.848-855

Busch, J.W., (1985) Mentoring in Graduate Schools of Education: Mentors' Perceptions. American Educational Research Journal. Vol.22 (2) pp.257-265.

- Chow, F. and Suen, L., (2001) Clinical staff as mentors in pre-registration undergraduate nursing education: students' perceptions of the mentors' roles and responsibilities. Nurse Education Today. Vol.21 (5) pp.350-358.
- Chung, C., E. and Kowalski, S., (2012) Job Stress, Mentoring, Psychological Empowerment, and Job Satisfaction Among Nursing Faculty. Journal of Nursing Education. Vol.51 (7) pp.381-388.
- Cohen, N. H. (1995). The principles of adult mentoring scale. In M. W. Galbraith & H. C. Norman (Eds.), Mentoring: New strategies and challenges (pp. 15–32). New Directions for Adult and Continuing
- Crisp, G. (2009) Conceptualization and initial validation of the College Student Mentoring Scale (CSMS). Journal of College Student Development. Vol. 50 (2) pp. 177-194.
- Crisp, G. & Cruz, I. (2010) Confirmatory factor analysis of a measure of “mentoring” among undergraduate students attending a Hispanic serving institution. Journal of Hispanic Higher Education. Vol. 9 (3) pp. 232-244.
- DeVellis, R. F. (2012). Scale development: Theory and applications (Vol. 26). 3rd ed. London: Sage publications
- Dimitriadis K; von der Borch P; Störmann S; Meinel FG; Moder S; Reincke M; Fischer MR, (2012 ) Characteristics of mentoring relationships formed by medical students and faculty Medical Education Online. Vol. 13(17) pp. 17242
- Dreher, G.F., & Ash, R.F. (1990). A comparative study of mentoring among men and women in managerial, professional, and technical positions. Journal of Applied Psychology. Vol. 75 pp.539–546.
- Dibert, C. & Goldenberg, D. (1995) Preceptors' perceptions of benefits, rewards, supports and commitment to the preceptor role. Journal of Advanced Nursing. Vol. 21 (6) pp. 1144-1151.
- Eby, L.T., Durley, J.R., Evans, S.C. and Ragins, B.R., (2008) Mentors' Perceptions of Negative Mentoring Experiences: Scale Development and Nomological Validation. Journal of Applied Psychology. Vol.93 (2) pp.358-373.

Ensher, E.A. and Murphy, S.E., (2011) The Mentoring Relationship Challenges Scale: The impact of mentoring stage, type, and gender. Journal of Vocational Behaviour. Vol.79 (1) pp. 253-266.

Ferro, A., DeWit, D., Wells, S., Speechley, K. N., & Lipman, E. (2013) An evaluation of the measurement properties of the Mentor Self-Efficacy Scale among participants in Big Brothers Big Sisters of Canada Community Mentoring Programs. International Journal of Evidence Based Coaching & Mentoring. Vol. 11(1) pp.146-161.

Fleming, M., House, M. S., Shewakramani, M. V., Yu, L., Garbutt, J., Mcgee, R., Kroenke, K., Abedin, M. Z. & Rubio, D. M. (2013) The mentoring competency assessment: validation of a new instrument to evaluate skills of research mentors. Academic Medicine: Journal of the Association of American Medical Colleges. Vol. 88 (7) pp. 1002.

Fowler, J. L. & O'gorman, J. G. (2005) Mentoring Functions: A Contemporary View of the Perceptions of Mentees and Mentors. British Journal of Management. Vol. 16 (1) pp. 51-57.

Frandsen, G.M., (2003) Mentoring nursing faculty in higher education [thesis], Saint Louis University

Gefen, D. & Straub, D. (2005) A practical guide to factorial validity using PLS-Graph: Tutorial and annotated example. Communications of the Association for Information Systems. Vol. 16 pp. 91-109.

Gilbreath, B., Rose, G. L., and Dietrich, K. E. (2008) Assessing mentoring in organizations: an evaluation of commercial mentoring instruments. Mentoring & Tutoring: Partnership in Learning. Vol. 16 (4) pp. 379–393.

Harris, S.M., (2013) Development of the perceptions of mentoring relationships survey: A mixed methods approach. International Journal of Multiple Research Approaches. Vol.7 (1) pp.83-95.

Hogan, T. P. & Cannon, B. (2003) Psychological testing: A practical introduction. New York: Wiley

Hu, C., Pellegrini, E.K. and Scandura, T.A., (2011) Measurement invariance in mentoring research: A cross-cultural examination across Taiwan and the U.S. Journal of Vocational Behaviour Vol.78 (2) pp.274-282.

Hudson, P., Skamp, K. & Brooks, L. (2005) Development of an instrument: Mentoring for effective primary science teaching. Science Education. Vol. 89 (4) pp. 657-674.

Hyrkas, K. & Shoemaker, M. (2007) Changes in the preceptor role: re-visiting preceptors' perceptions of benefits, rewards, support and commitment to the role. Journal of Advanced Nursing. Vol. 60 (5) pp. 513-524.

Jacobi M. (1991) Mentoring and undergraduate academic success: a literature review. Review of Educational Research. Vol. 61(4) pp.505-532.

Jakubik, L.D., (2012) Development and testing of the Jakubik Mentoring Benefits Questionnaire among paediatric nurses. Journal of Nursing Measurement. Vol.20 (2) pp.113-122.

Jakubik, L.D., (2008) Mentoring beyond the first year: predictors of mentoring benefits for pediatric staff nurse protégés. Journal of Paediatric Nursing. Vol.23 (4) pp.269-281.

Jones, J. M. 1997. A study of mentoring relationships in baccalaureate nursing programs in North Carolina. Unpublished thesis (PhD.), University of North Carolina.

Greensboro.Kalischuk, R. G., Vandenberg, H. & Awosogga, O. (2013) Nursing preceptors speak out: An empirical study. Journal of Professional Nursing. Vol. 29 (1) pp. 30-38

KAVOOSI, M. C., ELMAN, N. S. & MAUCH, J. E. (1995) Faculty mentoring and administrative support in schools of nursing. Journal of Nursing Education. Vol. 34 (9) pp. 419-426.

KOC, E. M. (2011) Development of mentor teacher role inventory. European Journal of Teacher Education. Vol. 34 (2) pp. 193-208.

Kram, K. E. (1983). Phases of the mentoring relationship. *Academy of Management Review*, Vol. 26, 608–625.

Kram, K. E. (1985). Mentoring at work. Glenview, IL: Scott, Foresman and Company.

- Lambert, V. & Glacken M (2005) Clinical education facilitators: a literature review. Journal of Clinical Nursing. 14. pp. 664–673
- Levison, D. J., Darrow, C. M., Klein, E. G., Levinson, M. H. & Mckee, B. (1978) The seasons of a man's life. New York: Ballentine.
- Lightfoot, C. A. (2000) The impact of academic, social and mentoring experiences on the persistence of minority and non-minority college students. (Doctoral dissertation, University of Houston, 2000). Dissertation Abstracts International, 61/07, 2623. (UMI No. AAT 9979230).
- McDowell, I. (2006) Measuring health: a guide to rating scales and questionnaires. Oxford: Oxford University Press.
- Myall, M, Levett-Jones, T & Lathlean J (2008) Mentorship in contemporary practice: the experiences of nursing students and practice mentors. Journal of Clinical Nursing. Vol.17 pp.1834–1842
- Noe, R.A., (1988) An Investigation of the Determinants of Successful Assigned Mentoring Relationships. Personnel Psychology. Vol.41 (3) pp.457-479.
- Pamuk, S. & Thompson, A. D. (2009) Development of a technology mentor survey instrument: Understanding student mentors' benefits. Computers & Education. Vol. 53 (1) pp. 14-23.
- Pellegrini, E. K. & Scandura, T. A. (2005) Construct equivalence across groups: An unexplored issue in mentoring research. Educational and Psychological Measurement. Vol. 65 (2) pp. 323-335.
- Pollock, R. (1995) A test of conceptual models depicting the developmental course of informal mentor-protégé relationships in the work place. Journal of Vocational Behavior. Vol. 46 (2) pp. 144-162.
- Ragins, B.R. and McFarlin, D.B. (1990) Perception of mentor roles in cross-gender mentoring relationship. Journal of Vocational Behaviour. Vol. 37pp.321-339

- Richard, Lee W. (1996) The relationship between organizational culture and the frequency, quality and array of mentoring activities among nurse executive mentors. Worldviews on Evidence-Based Nursing. Vol. 8 (4) pp. 212-23
- Rogers, J. C., Monteiro, F. M., & Nora, A. (2005). Mentoring for faculty development instrument validation study. Paper presented at the 38th STFM Annual Spring Conference, New Orleans, LA.
- Rose, G.L. (2003). Enhancement of mentor selection using the Ideal Mentor Scale. Research in Higher Education. Vol.44 pp. 473–494.
- Rose, G. L. (2005) Group Differences in Graduate Students' Concepts of The Ideal Mentor. Research in Higher Education. Vol. 46 (1) pp. 53-80.
- Salami,S.O. (2008) Psychosocial Factors as Predictors of Mentoring among Nurses in South western Nigeria. Journal of Workplace Learning. Vol. 20 (5) pp. 348-363
- Sands,R.G., Parson,L.A.& Duane(1991) Faculty mentoring faculty in a public university. Journal of Higher Education. Vol. 62 (2) pp. 174-93
- Sawatzky, J.-A. V. & Enns, C. L. (2009) A mentoring needs assessment: Validating mentorship in nursing education. Journal of Professional Nursing. Vol. 25 (3) pp. 145-150.
- Scandura, T.A., (1992) Mentorship and career mobility: An empirical investigation. Journal of Organizational Behaviour. Vol.13 (2) pp.169-174.
- Scandura, T.A. and Ragins, B.R., (1993) The effects of sex and gender role orientation on mentorship in male-dominated occupations. Journal of Vocational Behaviour. Vol.43 (3) pp. 251-265.
- Schockett, M. R. & Haring-Hidore, M. (1985) Factor analytic support for psychosocial and vocational mentoring functions. Psychological Reports. Vol. 57 (2) pp. 627-630.
- Short, J. D. (1997) Profile of administrators of schools of nursing, part II: mentoring relationships and influence activities. Journal of Professional Nursing. Vol. 13 (1) pp. 13-18.



Suen L.K.P. & Chow.L.W. (2001) Students' perceptions of the effectiveness of mentors in an undergraduate nursing programme in Hong Kong. Journal of Advanced Nursing. Vol. 36(4) pp. 505-511

Streiner, D. L. & Norman, G. R. (2008) Health measurement scales: a practical guide to their development and use. 4th ed. Oxford: Oxford university press

van Eps; Cooke M; Creedy DK; Walker R.( 2006) Student evaluations of a year-long mentorship program: a quality improvement initiative. Nurse Education Today. Vol. 26 (6) pp. 519-24.

Weng, R.H., Huang, C.Y., Tsai, W.C., Chang, L.Y., Lin, S.E. and Lee, M.Y., (2010) Exploring the impact of mentoring functions on job satisfaction and organizational commitment of new staff nurses. BMC Health Services Research. Vol.10 pp.240-240.

Wilde, J.B. and Schau, C.G., (1991) Mentoring in graduate schools of education: Mentees' perceptions. Journal of Experimental Education. Vol.59 (2) pp.165-179.

Zey, M.G. (1991) The mentor connection. Transaction Publishers, New Brunswick London.

Table 1 Criteria and search strategies used in literature review

Criterion 1: Limiters	Published Date: 01/01/1985-30/06/2015; Medline: English Language; Human; Age Related: All Adult: 19+ years; Cinhal: Language: English; Human; Eric: Educational Level: Higher Education; Language: English; PsycINFO: English; Age Groups: Adulthood (18 yrs & older) Academic Search Premier: Language: English Business premier resource : English
Criterion 2: Terms / concepts / keywords	Mentor* N3 (behav* or skill? or role? or activit? or function* or relation*) and (scale or tool or instrument or questionnaire or inventory).
Criterion 3: Content	Articles about developing and validating scales of mentoring function / role / behaviour / activities are included, and three kinds of papers listed below are excluded: A. Studies not about mentorship, using concepts as coaching, preceptorship, supervision B. Quantitative studies measuring mentoring outcome such as job satisfaction, commitment and so on C. Qualitative research or theoretical review / discussion
Criterion 4: Fields of science	Mentorship in business and organization, education and psychology, medicine and allied health fields
Criterion 5: Scale review	Show proper items and dimensions
Criterion 6: Accessibility	Likelihood of availability (time and budget constraints)

Figure1 Flow diagram of the systematic review

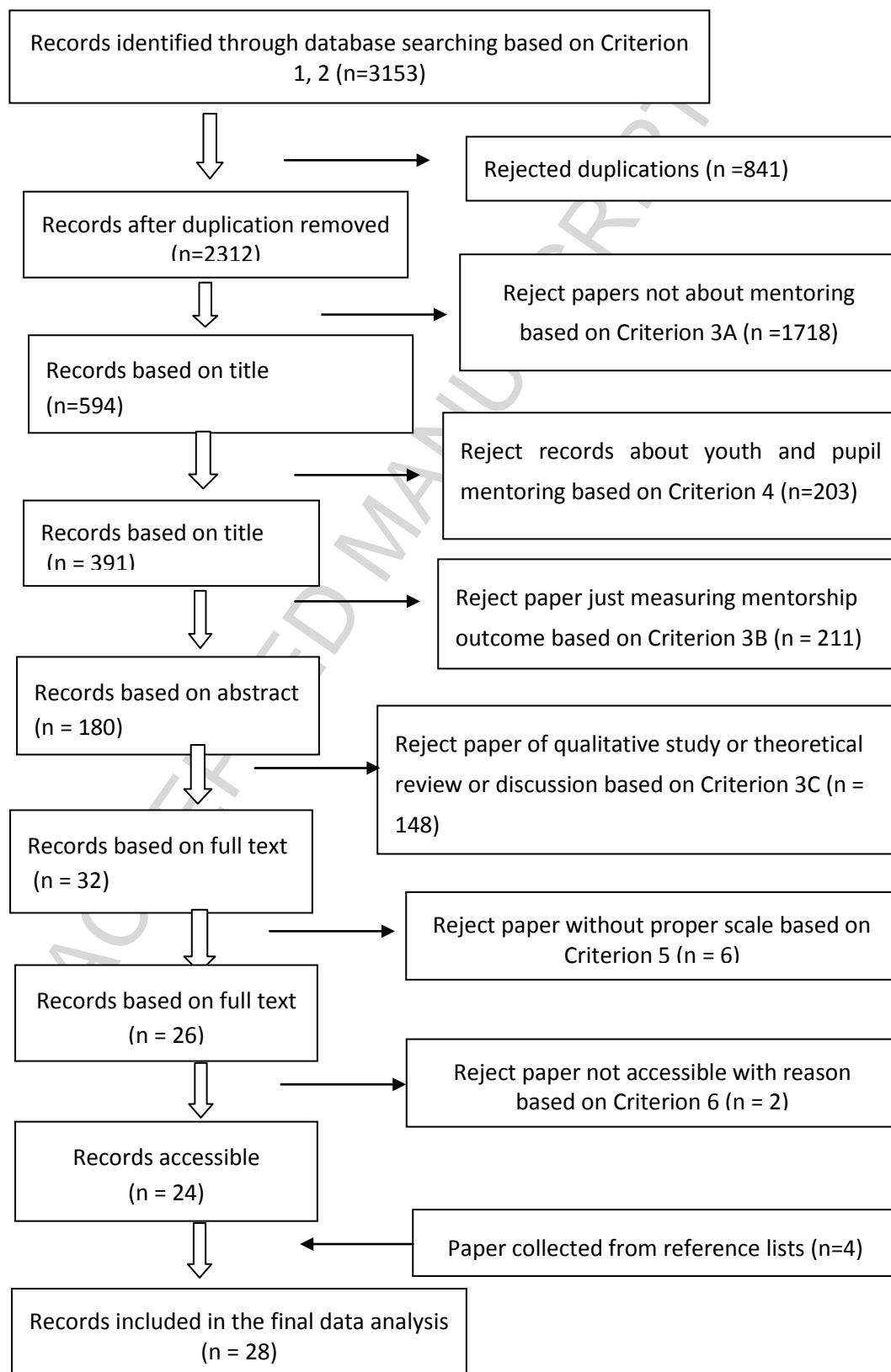


Table 2 General information of the mentors' measurement tool in this review

Discipline	Education	11	Business	7	Health	3	Research	1
Country	USA	18	Australia	2	Canada	1	China	1
Time	<1990	3	1990-2000	7	2000-2010	9	>2010	3
Psychometrics	Poor	4	Fare	13	Good	5		
<p>*Poor means no statistic test of psychometrics</p> <p>Fare means reported content validity, EFA OR CFA</p> <p>Good means reported more EFA OR CFA.</p>								

Table 3. Measurement tools identified

Reference (authors)	Scale name and number of items	Subscales	Psychometric	Participants	Comments
Busch, 1985	Mentoring instrument: mentors' perception (69)	Three subscales: mutuality, comprehensiveness and career	Described CV, EFA	463 professor in 67 education departments in 40 states in the USA	Based on showing measurement mentor
Schockett and Haring-Hidore, 1985	Mentoring function 16	Two factors: career development and psychosocial support	EFA	144 college students	Based on mentor evidence educati
Noe, 1988	Mentoring Functions Scale (MFS, 29)	Two functions: psychosocial function and career function	EFA CONT, ICR	139 educators(mentee) and 43 mentors from 9 university	Based on mentor evidence industr mentor
Ragins and McFarlin, 1990	Mentor Role Instrument (MRI, 33)	Two factors: career and psychosocial function	CFA ICR	181 protégés in three organizations	Based on mentor evidence business staff m
Dreher & Ash, 1990	Global Measure of Mentoring Practices (GMMP, 18)	Two factors: career and psychosocial function	EFA, content validity, ICR	Business school graduates (147 women and 173 men)	Based on mentor evidence industr
Wilde and Schau, 1991	Mentoring instrument: mentees perception (65)	Four factors: mutual support, comprehensiveness, mentee profession development, research together.	EFA	177 PhD students	No claim showing aiming student
Sands and Parson, 1991	Ideal mentoring function (29)	Four factors: friend and support, career guide, information, and intellectual guide	ICR, CRIT, CONT	Assistant Professor (136) Associate Professor (117) Full Professor (94)	Using I theory showing aimed a mentor staff.
Scandura, 1992; Scandura and Ragins, 1993; Pellegrini and Scandura, 2005; Hu et al. 2011	Mentoring function Scale (MFS, 20-15- 9)	Three subscales: psychosocial function, career function and role modelling	MGCFA, CFA, EFA	244 managers; 377 employed undergraduate and MBA students from 3 universities; 195 employees in USA 309 full-time workers in Taiwan	Based on mentor establish continu length; internat academ graduat
Pollock, 1995	Mentoring functions (19)	Two factors: career and psychosocial function	CFA ICR	383 managers from 50 organizations	Based on mentor validity industr
Cohen, 1995	Principles of Adult Mentoring Inventory (PAMI, 55)	Six behavioural functions: relationship emphasis, information emphasis, facilitative	ICR	No report	'The pr cited w consist student

		focus, confrontation focus, mentor model, and student vision			
Suen and Chow, 2001; Chow and Suen, 2001	Scale of students mentoring (33)	Five subscales: befriending, guiding, advising, counselling and assisting	Face and CONT	No report of further validation	Based on showing aiming in clinical widely,
Rose, 2003; 2005	Ideal Mentor Scale (IMS, 34)	Three factors: integrity, guidance, and relationship	CONT, CONV, ICR, EFA, CFA,	Three samples: 82 PhD students 250 PhD students 380 PhD students	Based on (1988) teaching counselor wide range aiming mentor and nur
Berk et al. 2005	Mentorship Effectiveness Scale (12)	One	No reported psychom etric evidence	No participants reported	No claim showing aimed at staff me medica
Fowler and O'Gorman, 2005	Mentoring Functions 39	Eight functions: personal and emotional guidance, coaching, advocacy, career development facilitation, role modelling, strategies and systems advice, learning facilitation and friendship	EFA, CFA	272 mentees and 228 mentors from eight public-sector and five private-sector organizations.	Based on mentor evidence
Hudson et al. 2005	Mentoring for effective primary science teaching (MEPST,45)	Five factors: personal attributes, system requirement, pedagogical knowledge modelling, feedback	CFA	331 final-year preservice teachers	No claim based on was con teacher further
Eby et al. 2008	Negative Mentoring experience scale (NMES, 36)	Three factors: Performance problem, interpersonal problem, destructive relational patterns	CFA, CONT, CRIT, CONV, DISC	420 business students for CONT; 89 mentees and 80 mentors (director or managers working in two universities) for CRIT, CONV and DISC; 132 mentors of graduate students for CFA	Social c underp psychol develop measur experie
Jakubik, 2008; 2012	Jakubik's mentoring benefit questionnaire (MBQ, 36)	Four subscales: knowledge, personal growth, protection, and career advancement	Face and CONT, ICR, EFA, CFA	453 paediatric nurses	Based on (Zey, 1 showing aimed at mentor the autl
Crisp, 2009; Crisp and Cruz, 2010	College Student Mentoring Scale (CSMS, 25)	Four factors: psychological and emotional support, degree and career	ICR, CFA, MGCF	351 college students	Based on some p measur newly c

		support, academic subject knowledge support, and role model			
Pamuk and Thompson, 2009	Technology mentor benefits instrument (28)	Three factors: technical benefit items, academic benefit items/profession pedagogical benefit	EFA, ICR, CONT,	43 graduate students	Bandur used. It technol (a grad techniq develop
Ensher and Murphy, 2011	Mentoring relationship challenge scale (MRCS, 23)	Three factors: requiring commitment and resilience, measuring up to mentors standards, and career goal and risk orientation	CT EFA, ICR	312 managers and professionals in varying industries	Social c underpi psychol develop measur challen
Harris, 2013	Perception of Mentoring relationships survey (PMRS, 24)	Three subscales: benefits of mentoring, mentor's role and mentee's role	CONT, EFA, ICR	43 university students for CONT 391 university students for EFA	Social l underpi evidenc student
Fleming et al. 2013	Mentoring Competency Assessment (MCA, 26)	Six competencies of mentors: maintaining effective communication, aligning expectations, assessing understanding, addressing diversity, fostering independence, and promoting professional development	EFA, CFA, ICR	283 mentors (professor)and 283 mentees (associate professor) from 16 universities	No clai identiifi compet develop

ICR: internal consistency reliability  
 CRI: criterion validity,  
 DIS: discriminant validity,  
 CONC: concurrent validity  
 CONT: content validity  
 EFA: exploratory factor analysis  
 CFA: confirmatory factor analysis  
 MGCFA: multi-group confirmatory factor analysis  
 CONV: convergent validity

## Highlights

1. Mentorship and its research is led by business. Some robust measurement tools from business and education have been used to measure mentorship among nursing educators.
2. More specific tools for mentorship measuring different types of mentorship and different aspects have been developed in different area, which gives nursing educators more choices.
3. Nursing researchers' started to develop their own measurement tools, but more effort needs to be invested in theoretical framework construction and psychometric evidence building.
4. No proper measurement tools to assess nursing students' mentorship in clinical learning has been identified.