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
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

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

© 2017. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/
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 turnover 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 somethings 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 instruments 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; Kavoosi et al. 
1995), having proper instruction on administration and scoring, to see more detail about this 
15
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


18


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


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


© 2017. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/


Table 1 Criteria and search strategies used in literature review

<table>
<thead>
<tr>
<th>Criterion 1: Limiters</th>
<th>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 &amp; older) Academic Search Premier: Language: English Business premier resource : English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criterion 2: Terms / concepts / keywords</td>
<td>Mentor* N3 (behav* or skill? or role? or activit? or function* or relation*) and (scale or tool or instrument or questionnaire or inventory).</td>
</tr>
<tr>
<td>Criterion 3: Content</td>
<td>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</td>
</tr>
<tr>
<td>Criterion 4: Fields of science</td>
<td>Mentorship in business and organization, education and psychology, medicine and allied health fields</td>
</tr>
<tr>
<td>Criterion 5: Scale review</td>
<td>Show proper items and dimensions</td>
</tr>
<tr>
<td>Criterion 6: Accessibility</td>
<td>Likelihood of availability (time and budget constraints)</td>
</tr>
</tbody>
</table>
Figure 1 Flow diagram of the systematic review

Records identified through database searching based on Criterion 1, 2 (n=3153)

Rejected duplications (n =841)

Records after duplication removed (n=2312)

Reject papers not about mentoring based on Criterion 3A (n =1718)

Records based on title (n=594)

Reject records about youth and pupil mentoring based on Criterion 4 (n=203)

Records based on title (n = 391)

Reject paper just measuring mentorship outcome based on Criterion 3B (n = 211)

Records based on abstract (n = 180)

Reject paper of qualitative study or theoretical review or discussion based on Criterion 3C (n = 148)

Records based on full text (n = 32)

Reject paper without proper scale based on Criterion 5 (n = 6)

Records based on full text (n = 26)

Reject paper not accessible with reason based on Criterion 6 (n = 2)

Records accessible (n = 24)

Paper collected from reference lists (n=4)

Records included in the final data analysis (n = 28)
Table 2 General information of the mentors' measurement tool in this review

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Education</th>
<th>Business</th>
<th>Health</th>
<th>Research</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>11</td>
<td>7</td>
<td>3</td>
<td>1</td>
<td>USA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Australia</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Canada</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>China</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>7</td>
<td>9</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Psychometrics</td>
<td>Poor</td>
<td>Fare</td>
<td>Good</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>13</td>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Poor means no statistic test of psychometrics
Fare means reported content validity, EFA OR CFA
Good means reported more EFA OR CFA.
Table 3. Measurement tools identified

<table>
<thead>
<tr>
<th>Reference (authors)</th>
<th>Scale name and number of items</th>
<th>Subscales</th>
<th>Psychometric</th>
<th>Participants</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Busch, 1985</td>
<td>Mentoring instrument: mentors’ perception (69)</td>
<td>Three subscales: mutuality, comprehensiveness and career</td>
<td>Describe CV, EFA</td>
<td>463 professor in 67 education departments in 40 states in the USA</td>
<td>Based on the two functions model of mentoring; showing some psychometric evidence; used in business and industry.</td>
</tr>
<tr>
<td>Schockett and Haring-Hidore, 1985</td>
<td>Mentoring function 16</td>
<td>Two factors: career development and psychosocial support</td>
<td>EFA</td>
<td>144 college students</td>
<td>Based on the two functions model of mentoring; showing some psychometric evidence; used in business and industry.</td>
</tr>
<tr>
<td>Noe, 1988</td>
<td>Mentoring Functions Scale (MFS, 29)</td>
<td>Two functions: psychosocial function and career function</td>
<td>EFA CONT, ICR</td>
<td>139 educators(mentee) and 43 mentors from 9 university</td>
<td>Based on the two functions model of mentoring; showing some psychometric evidence; used in business and industry.</td>
</tr>
<tr>
<td>Ragins and McFarlin, 1990</td>
<td>Mentor Role Instrument (MRI, 33)</td>
<td>Two factors: career and psychosocial function</td>
<td>CFA ICR</td>
<td>181 protégés in three organizations</td>
<td>Based on the two functions model of mentoring; showing some psychometric evidence; used in business and industry.</td>
</tr>
<tr>
<td>Drefer &amp; Ash, 1990</td>
<td>Global Measure of Mentoring Practices (GMMP, 18)</td>
<td>Two factors: career and psychosocial function</td>
<td>EFA, content validity, ICR</td>
<td>Business school graduates (147 women and 173 men)</td>
<td>Based on the two functions model of mentoring; showing some psychometric evidence; used in business and industry.</td>
</tr>
<tr>
<td>Wilde and Schau, 1991</td>
<td>Mentoring instrument: mentees perception (65)</td>
<td>Four factors: mutual support, comprehensiveness, mentor profession development, research together.</td>
<td>EFA</td>
<td>177 PhD students</td>
<td>No claim of theoretical framework; showing some psychometric evidence; aiming at measuring mentorship; used by nursing teaching staff.</td>
</tr>
<tr>
<td>Sands and Parson, 1991</td>
<td>Ideal mentoring function (29)</td>
<td>Four factors: friend and support, career guide, information, and intellectual guide</td>
<td>ICR, CRIT, CONT</td>
<td>Assistant Professor (136) Associate Professor (117) Full Professor (94)</td>
<td>Using Ericson’s theory as underpinning; showing some psychometric evidence: aimed at measuring postgraduate students mentoring.</td>
</tr>
<tr>
<td>Scandura, 1992; Scandura and Ragins, 1993; Pellegrini and Scandura, 2005; Hu et al, 2011</td>
<td>Mentoring function Scale (MFS,20-15- 9)</td>
<td>Three subscales: psychosocial function, career function and role modelling</td>
<td>MGCFA, CFA, EFA</td>
<td>244 managers; 377 employed undergraduate and MBA students from 3 universities; 195 employees in USA 309 full-time workers in Taiwan</td>
<td>Based on the two functions model of mentoring; establishing continuance length; showing some psychometric evidence, widely used in the field of education.</td>
</tr>
<tr>
<td>Pollock, 1995</td>
<td>Mentoring functions (19)</td>
<td>Two factors: career and psychosocial function</td>
<td>CFA ICR</td>
<td>383 managers from 50 organizations</td>
<td>Based on the two functions model of mentoring; showing some psychometric evidence; used in business and industry.</td>
</tr>
<tr>
<td>Cohen, 1995</td>
<td>Principles of Adult Mentoring Inventory (PAMI, 55)</td>
<td>Six behavioural functions: relationship emphasis, information emphasis, facilitative</td>
<td>ICR</td>
<td>No report</td>
<td>The principle of adult mentoring inventory was the earliest research. Consistent with times, it was cited widely, with just internal validity evidence; also in nursing staff mentoring, with some psychometric evidence, widely used in business and industry.</td>
</tr>
</tbody>
</table>

© 2017. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Scale Name</th>
<th>Description</th>
<th>Analysis</th>
<th>Participants</th>
<th>Validation</th>
<th>Theoretical Framework</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suen and Chow, 2001; Chow and Suen, 2001</td>
<td>Scale of students mentoring</td>
<td>Five subscales: befriending, guiding, advising, counselling and assisting</td>
<td>Face and CONT</td>
<td>No report of further validation</td>
<td>Based on ENB's five roles of mentors; showing little psychometric evidence; aiming at measuring mentors' behaviour in clinical nursing education; cited widely, but no further use.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rose, 2003; 2005</td>
<td>Ideal Mentor Scale (IMS, 34)</td>
<td>Three factors: integrity, guidance, and relationship</td>
<td>CONT, CONV, ICR, EFA, CFA</td>
<td>Three samples: 82 PhD students 250 PhD students 380 PhD students</td>
<td>Based on Anderson and Shannon's (1988) five functions of mentors: teaching, sponsoring, encouraging, counselling, and befriending; showing wide range of psychometric evidence; aiming at measuring graduates' ideal mentor; used in general PhD mentoring and nursing field.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Berk et al., 2005</td>
<td>Mentorship Effectiveness Scale (12)</td>
<td>One</td>
<td>No reported psychometric evidence</td>
<td>No participants reported</td>
<td>No claimed theoretical framework; showing no psychometric evidence; aimed at measuring nursing teaching staff mentorship; used by nursing and medical science educators.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fowler and O'Gorman, 2005</td>
<td>Mentoring Functions 39</td>
<td>Eight functions: personal and emotional guidance, coaching, advocacy, career development facilitation, role modelling, strategies and systems advice, learning facilitation and friendship</td>
<td>EFA, CFA</td>
<td>272 mentees and 228 mentors from eight public-sector and five private-sector organizations.</td>
<td>Based on the two-function model of mentoring; showing some psychometric evidence; newly developed.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hudson et al., 2005</td>
<td>Mentoring for effective primary science teaching (MEPST, 45)</td>
<td>Five factors: personal attributes, system requirement, pedagogical knowledge modelling, feedback</td>
<td>CFA</td>
<td>331 final-year preservice teachers</td>
<td>No claimed theoretical underpinning; based on literature review, the new scale was constructed to measure science teachers' mentorship; showing no further use.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eby et al., 2008</td>
<td>Negative Mentoring experience scale (NMES, 36)</td>
<td>Three factors: Performance problem, interpersonal problem, destructive relational patterns</td>
<td>CFA, CONT, CRIT, CONV, DISC</td>
<td>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</td>
<td>Social exchange theory is the theoretical underpinning; showing wide range of psychometric evidence; newly developed and validated; aiming at measuring negative mentoring experience.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jakubik, 2008; 2012</td>
<td>Jakubik's mentoring benefit questionnaire (MBQ, 36)</td>
<td>Four subscales: knowledge, personal growth, protection, and career advancement</td>
<td>Face and CONT, ICR, EFA, CFA</td>
<td>453 paediatric nurses</td>
<td>Based on the mutual benefits theory (Zey, 1991) in the business field; showing some psychometric evidence; aimed at measuring staff nurses' mentoring; newly developed and used by the author herself later.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crisp, 2009; Crisp and Cruz, 2010</td>
<td>College Student Mentoring Scale (CSMS, 25)</td>
<td>Four factors: psychological and emotional support, degree and career</td>
<td>ICR, CFA, MGCFA</td>
<td>351 college students</td>
<td>Based on some previous measures; newly developed.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Authors</td>
<td>Instrument</td>
<td>Factors</td>
<td>Sample Size</td>
<td>Theoretical Framework</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
<td>---------</td>
<td>-------------</td>
<td>-----------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pamuk and Thompson, 2009</td>
<td>Technology mentor benefits instrument (28)</td>
<td>Three factors: technical benefit items, academic benefit items/profession pedagogical benefit</td>
<td>43 graduate students</td>
<td>Bandura's social learning theory was used. It aims to measure the benefits of technology mentoring in education field (a graduate student mentors a faculty for technique development); newly developed.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ensher and Murphy, 2011</td>
<td>Mentoring relationship challenge scale (MRCS, 23)</td>
<td>Three factors: requiring commitment and resilience, measuring up to mentors standards, and career goal and risk orientation</td>
<td>312 managers and professionals in varying industries</td>
<td>Social exchange theory is the theoretical underpinning; showing some psychometric evidence; newly developed and validated; aiming at measuring mentoring relationship challenge in business.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harris, 2013</td>
<td>Perception of Mentoring relationships survey (PMRS, 24)</td>
<td>Three subscales: benefits of mentoring, mentor’s role and mentee’s role</td>
<td>43 university students for CONT 391 university students for EFA</td>
<td>Social learning theory is the theoretical underpinning; with some psychometric evidence; aiming at measuring college students mentoring; newly developed.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fleming et al. 2013</td>
<td>Mentoring Competency Assessment (MCA, 26)</td>
<td>Six competencies of mentors: maintaining effective communication, aligning expectations, assessing understanding, addressing diversity, fostering independence, and promoting professional development</td>
<td>283 mentors (professor) and 283 mentees (associate professor) from 16 universities</td>
<td>No claimed theoretical framework was identified; measures researcher mentors’ competency in medicine; newly developed.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.