Nurse Education Today 40 (2016) 20–28

Contents lists available at ScienceDirect

Nurse Education Today

journal homepage: www.elsevier.com/nedt

A review of mentorship measurement tools

Yanhua Chen [a,*], Roger Watson [b], Andrea Hilton [b]

[a] The Infectious Disease Department of the First Affiliated Hospital of Lu Zhou Medical College, Luzhou, China
[b] The University of Hull, Kingston upon Hull, UK

ARTICLE INFO

Article history:
Received 3 October 2015
Received in revised form 17 January 2016
Accepted 26 January 2016

Keywords:
Nursing education
Mentor
Behaviour
Measurement tool
Theoretical framework
Psychometrics

ABSTRACT

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 were 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, 3 from health science, and 1 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 specialised 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 conceptualised differently in different fields and is less mature in nursing than in business. Measurement of mentorship is heading to a more specialised 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.

1. Introduction

1.1. Mentorship

Mentorship flourished after the work of Levison et al. (1978) in business and organisation. 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 organisation 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 are 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

* Corresponding author at: 25 Tai Ping Street, Jiangyang District, Luzhou, China.
E-mail addresses: chen_yanhua25@163.com (Y. Chen), r.watson@hull.ac.uk (R. Watson), a.hilton@hull.ac.uk (A. Hilton).

http://dx.doi.org/10.1016/j.nedt.2016.01.020
0260-6917/

Scandura, 2005; Hu et al. 2011), Mentoring Function Scale (Noe, 1988), and Sands' tool (Sands et al., 1991) to measure mentors' function, behaviour, and relationships. These mentorship tools in different fields may vary in conceptualisation 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 (DeVellis, 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).

1.4.3. 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.

1.4.4. 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